Cognitive Science 1: Midterm 1 Review

## Lecture 1: Introduction

Cognitive Science: interdisciplinary science of the mind, inherently recursive, easy human actions difficult for computers, intangible, addresses classical questions:

- where does knowledge come from?

- what is the nature of thought?

- are there uniquely human aspects of cognition?

**Rationalist:** knowledge comes partly from reason, some aspects of which may be innate

- Plato, Leibniz, Boole, Frege, Turing, Chomsky

**Empiricist:** neither principles nor ideas are innate. Knowledge comes from experience

- Aristotle, Locke, Skinner

**Computational Representational Understanding of the Mind (CRUM):** thinking is performed by computations operating on representations

- Basically, formalized information processing on symbols that have meaning (key hypothesis for how cog. sic is performed)

## Lecture 2: Innateness

**Rationalist** (innate knowledge/knowledge from reasoning): Plato

**Empiricist** (knowledge based on experience): Aristotle (taught by Plato)

**Aristotle:**

- Blank Slate/Tabula Rasa: human mind functions as a blank slate ==> we learn from our experiences

**Plato:**

- We have a sense of innate knowledge ==> soul must understand things since the dawn of man (has innate knowledge)
- Meno: boy knew how to do the math all along since that knowledge was innate ==> need the correct stimulus to recall this innate knowledge
- Learning is remembering
- Poverty of the stimulus (limited experience: "Plato's Problem"): how can humans know as much as we know when our contacts with the world are brief, personal, and limited
- Objects in the world are shadows of their ideal "pure" forms, but are similar enough to remind us of these forms

**John Locke:**

- No innate knowledge ==> neither ideas nor principles are innate
- Same idea as Aristotle: "the blank slate"  but now called "the white paper"==> empiricist idea
- Mind gains all abilities to reason and use knowledge from experience
- Two sources of knowledge:
    - Sensation: yellow, hot, cold, etc. (feelings)

    - Reflection: (operations of our mind) ==> perception, thought belief, reason, etc.
- Counter example to rationalism: children's gradual acquisition of knowledge

**Leibniz:**

- John Locke is wrong: experience is necessary, but not sufficient to account for knowledge
- "Seeds of eternity": the something beyond experience that allows us to gain knowledge ==> basically, flashes of insight that allow us to use some sort of reasoning (why Leibniz isn't purely empirical)
- Why we're different from beasts == seeds of eternity (animals base their "knowledge"/actions solely on empiricism)
- **"Veins of Marble":** if John Locke was just making Hercules out of marble, he would just see the Hercules itself in the blank marble (metaphor for mind: isn't predisposed to any given ideas and can just take on whatever empirical form). Leibniz argues that because the marble is veined, the veins are in such a way to encourage the formation of Hercules ==> the mind is predisposed to certain ideas and information, but it requires experience and reason to uncover/refine them

**Induction:** generalizing beyond the data given to form a conclusion (not just looking at the facts, but rationalizing and interpreting them)

- Same thing as veins in the marble: if we have a good hunch about how to solve something, we can solve it better

**Deduction:** going from known truth to derive new truths with absolutely certain conclusions

**Lecture 3: Alphabet of Human Thought**

**Theory of thought:** comprehensive theory that predicts the thoughts one has while explaining the thoughts that people already have, theory connects systems to the world, modern cognitive science

**Induction:** start at particular experiences, end at uncertain, universal truths

**Deduction:** start at known truths, end at other known truths with absolute certainty

**Aristotle:** wanted to classify every valid syllogisms (basically, he wanted to classify every example of the transitive property) . . . syllogisms = conclusions from propositions (define space of possibility, determine validity)

- Ex. All A's are B's, All B's are C's ==> All A's are C's
- Four types of sentences
    - Universal Affirmation: All As are Bs
    - Universal Denial: All As are not Bs
    - Particular Affirmation: Some As are Bs
    - Particular Denial: Some As are not Bs
- Four something to be valid, the premises must only have 1 term in common, and the conclusion has 2 terms not shared by the premises . . . 2 premises and 1 conclusion
- Aristotle's approach: attack a general question by looking at a specific special case (doesn't really work well)

**Leibniz (**rationalist, veins in marble)

- World is optimistic: neither accidental nor undetermined ==> we live in best possible combo of world
- People are fundamentally good, but we limit our goodness by language ("imperfect mirror of intelligible thoughts that makes reasoning obscure")
- Solution: ideal language that perfectly represents relationships between our thoughts (THE DREAM)
- How to do this:

    - Create compendium of all human knowledge
    - Find the key underlying notions and make a symbol for them
    - Come up with rules to manipulate them
- Big catchphrase was "Let us calculate!" ==> believed that language, emotion, and thought could be calculated out, AI ahead of it's time, new ideas deducted from the language for clarity of thought, based on notation, frees mind for more creative thought, higher education

**George Boole:**

- Propositional logic, laws of thought: Represent thoughts as classes/sets of objects (ex. All plants are alive ==> 1 class is plants, other class is living things)
- Key ideas:
    - Let x = things that are black, y = things that are sheep ==> xy = things that are black sheep
    - Let x = things that are sheep ==> xx = things that are sheep (note: not a double sheep): formula is xx = x
    - x + y = all things that are in x or y (union)
    - x - y: things that are in x but not y (difference)
    - x + (1- x) = 1

**Lecture 4: Logic**

Boole's Logic is limited: What if an affirmation is both universal and particular? (Ex. Everybody loves somebody) . . . "multiple generality"

**Frege:**

- **Frege** wrote Begriffsschrift ("concept script"): formal language that tries to capture the relationship between our thoughts
    - X ^ Y = X and Y
    - X v Y = X or Y
    - X ==> Y: If X then Y
    - ~X: not X
- Frege's Problem (proposed by Russell): a set can't be a member of itself ==> Ex. Set of things that are not sparrows ==> the set itself is member of the set
- Frege's Problem #2: If you can't tell that a conclusion C can be drawn from a premise P, how can you check if it really can't be drawn or you didn't just screw up the logic involved? (No way to really check) complicated and time-consuming
- Quantifiers:
  **Extraordinary:** set is extraordinary if it contains itself
  **Ordinary:** set that doesn't contain itself

First-Order Predicate Logic:

- First-order Predicate logic: Universal Quantifier (Upside down A: "For all"), Existential Quantifier (Backwards E: "for some"/"there exists")
    - Universal: true for every object
    - Existential: true for at least one logic

- Limitations:
    - Can't say For every property P, there is some object that has that property
    - Can't say necessarily/possibly, can't deal with time, can't say anything that isn't definitely true/def. false

Inference Rules:

- Modus Ponens: Basically, it's what has to be true in order for the whole statement to be true (given a set of statements above the line, what has to be true to make the whole statement true goes below the line) . . . Affirming the Antecedent (if P then Q, P, therefore Q)
    - X ==> Y
    - X
    - _____
    - Y
- Modus Tollens: Denying the Consequent (If P then Q, Not Q, therefore not P)
    - X ^ Y
    - _____
    - X
    - Y
    - or
    - X
    - _____
    - X v Y
    - or
    - X==>Y
    - ~Y
    - _____
    - ~X
- Invalid Inferences:
    - Affirming the Consequent: (If P then Q, Q, therefore P)
    - Denying the Antecedent: (If P then Q, Not P, therefore not Q)

Tautology: formula which is always true in all possible worlds
Contradiction: formula which is always false in all possible worlds
Syntax: how things are written (properties of formulas)
Semantics: what things mean (statements formulas make about possible worlds)

An “Algebra that yields valid inferences” for Leibniz’s dream? . . .

Russell’s letter to Frege, problem of a set being a member of itself (ex: set O is all sets that do not contain themselves . . . is O ordinary or extraordinary?) . . . flawed.

**Lecture 5: John Campbell's (Get main details)**

Marr’s Levels: Computation, Algorithm, Implementation (Humiliation, Serotonin, Depression)

Schizophrenia - tickling - efference copy (control over your own motor skills)

Chinese room - computer has syntax but no semantics

Problem of consciousness - (Mary the vision scientist) how to explain imagination

Eliminativist: doing away with the mental altogether

Limits of CogSci

**Lecture 6: Brief Detour into Infinity/Beyond**

**Infinity:** Rationalists challenge empiricists by proposing infinity (although our experiences are finite, we are capable of experiencing things that are infinite)

- Infinity often has religious contexts: associated with large, positive concepts (e.g. God, Absolute, etc.)
- Countably Infinite vs Infinite:
    - Ex. How many Natural numbers are there? Infinite; How many even natural numbers are there? Infinite
    - Seems like the latter is a smaller subset of the former, yet both are infinite

**George Cantor:** proposes countably infinite (like natural numbers) and uncountably infinite

- Also proposed the Diagonal method to show there is > 1 size of infinity
- Countably Infinite number of Natural Numbers, uncountably infinite number of subsets of natural numbers
    - Proof: assume there are a countably infinite subsets of N ==> Use diagonal method to create a unique subset of N (i.e. 1 greater than number of countable ones). Conclusion: more than countably infinite subsets of N == uncountably infinite

**David Hilbert:** "Entscheidungsproblem": we need a way to take a problem P and conclusion C as inputs, and determine whether C follows from P

- This procedure would fulfill Leibniz's dream of a universal logic language

**Turing:** uses diagonal method to show that no such procedure that Hilbert proposed exists

- Turing Machine: formal model of computation (internal state, infinite tape, finite set of rules)
    - Ex. Q1:0-->Q
    - Start at State Q, Read 1, write 0 to the tape, move right one on the tape, stay in state Q
    - "Rationalist" Turing Machine: converts all 0s to 1s, then stops ==> output doesn't just depend on input
    - "Empiricist" Turing Machine: prints out exactly what you see ==> output solely depends on input

**Lecture 7:** Thought, Computation, and the World

**Turing Machines cont:** Every Turing Machine can be specified as a number

- Finite Turing machine, infinite tape ==> any number can specific its finite components (see Davis reading for details)

**Halting Set:** set of input numbers that causes a turing machine to halt

- Label: Turing machine number
- Package: machine's halting set
- How to find the set of numbers that isn't a halting set for any Turing machine
    - Set up grid with columns being the numbers the Halting Set contains, and rows being the Machine's code number
    - Use diagonal method to create a set D that isn't the halting set of any given turing machine
- Main point: Some claims cannot be decided computationally ==> Leibniz's hope for a language that can solve all problems won't ever work

**Watson** (computer): computer that was able to beat the top two Jeopardy players

- Is Watson actually thinking? No

**Symbols/Symbol Grounding Problem**

- Logical formulas contain propositions that correspond to things in our world ==> when our propositions are true, the corresponding things in the world are true
- Propositions themselves are symbols/representatives of things in the world they correspond to

- How can we prove if the real things are true? Our senses tell us this
    - Inference derives new facts from knowledge
    - Inferences let us take actions that tell us more about things
    - Problem: what do we know for certain if it is only based on our senses?

**Rene Descartes:** only thing we cannot doubt is the existence of our mind (by thinking itself, we establish its existence)

- We can doubt the existence of our body since it is only based on our senses
- Proposes pineal gland is link to solution for mind-body problem

**Dualism:** mind and body are different kinds of entities

- We can't explain minds in physical terms/vice verse

**Monism:** mind and matter are the same thing/ **Materialism:** belief that mind can be reduced to matter (brain)

**Mind-Body Problem:** How does information get to the mind from the senses/info get from mind to acting via senses?

Answering different questions of mind/matter:

- We can ask questions of our minds at a physical level (how neurons fire, etc.) by asking questions of physics/bio
- We can ask questions of how our minds function at an abstract level by looking at them in terms of computational terms

## Lecture 8: Dav's Lecture: Embodiment and Emotions

Computation: what Turing machines do (input, operation, output), symbols grounded in real world

- but! do we really need so many symbols? info already structured in the world

Complex Behavior: symbolic representations of mental states, ethics, representing agents with goals or just following rules? rules vs. innate?

James/Lange: emotions are essentially bodily states (smiling makes you happier)

## Lecture 9: Behavior and the Mind

**Question:** Why wasn't scientific method applied to the mind?

- Boole says: The mind can only be studied in specific instances, and cannot be generalized ==> will greatly differ from person to person, and we will be unable to delineate the differences without getting some of the study's bias in it
- Boole is focused on deduction, rather than induction
- Thoughts have no spatial dimension, is it possible to reduce it to objective simple laws?

**Apperception:** process of making an experience clear in consciousness ==> based on "complication apparatus"

- "Complication Apparatus" was a pendulum that swung with a click happening at time x
- People perceived the click happening after the pendulum hit that spot, despite the fact that the two happened simultaneously
- Conclusion: apperception takes time to happen

**Wundt:** "Father of Experimental Psych" relied on people's resorts of their experience (introspection)

- All judgments based on experiments were subjective, introspection, brass-instrument psych
- Was pioneer in experimental psych (apperception, first psych lab)

**Ebbinghaus:** wasn't a fan of introspection

- Came up with spacing effect: a list of items is better remembered when studied at spaced intervals, rather than all at once
- Relied on objective measures for all of his experiments
- Exponential Forgetting curve: people forget at an exponential rate (downward slide)

**Behaviorism:** focus on behavior rather than mental states, emphasizing the environment's role in behavior (empiricism round 2)

- Recognize no dividing line between man and brute

**John Watson:** leader on behaviorism front (1913)

- Rejected all forms of introspection since the answers to experiments should be found in the environment
- Focuses on animals since they essentially behaved the same as humans ==> led to theories being developed with animals
- "Little Albert Experiment" classically conditioned an infant to form a phobia via classical conditioning with a white rat and a loud noise

**Classical Conditioning:** learning that one cue (conditioned stimulus) is associated with another that natural elicits a reaction (unconditioned stimulus) ==> ringing a bell (conditioned) results in dog drooling (unconditioned)

- **Pavlov's dog**

**Operant Conditioning:** learning that performing an action results in reward (reinforcement) or punishment

**Thorndike:** used operant conditioning when he put cats in boxes with levers ==> they learned quicker to get out since the escape itself was meant as a reward

Operant conditioning and classical conditioning tell us that people are shaped by their environment: EMPIRICISM

**Skinner:** radical behaviorist

- Drop all mental talks ==> theories try to explain behavior, and it is done in terms of the environment
- No evidence that we ever actually acquired knowledge, but rather we just acquire behavior that reflects a new awareness of what happened

1950's Psychology: behaviorism reigns supreme ==> described as general purpose learning mechanisms applied to the environment ==> blank slates are written to

**Lecture 10: Cognitive Revolution**

**Tolman:** “insight” in rats with latent learning, challenged behaviorism idea of gradual reinforced learning

- Cognitive Maps: mental representations (!) easier to learn than rote repetition

1950: behaviorism prominent, formal theory of thought, computer science break throughs . . .

- Universal Turing Machine: can be programmed to simulate any other Turing machine, different hardware same software (mind as a symbolic system)

1956: Birth of Cognitive Science, Symposium on Information Theory at MIT

- **Newell & Simon:** *Logic Theorist*: first AI system, mathematical proofs with inference rules, heuristics inspired by human problem-solving strategies (Whitehead and Russell "responded with delight")
    - ---> Computers can solve problems characteristic of human thought, human thought can guide us in how computer programs should work
- **Miller:** *The Magical Number Seven*: limit on capacity for short term processing of information, 7 "chunks" of information in memory, mental representations are hierarchically structured
- **Chomsky:** *Three Models of Language*: language also governed by hierarchically structured mental representations, not just sequence of words

Behaviorism --> not sufficient or necessary

**Lecture 11: Language**

Whitehead (mathematician, logician, mentor of Russell) vs. Skinner (radical behaviorist) 1934:

- Verbal Behavior: mental state resulting from language encoding an absence uses no environmental stimulus
- Language is . . .

1. Special (hierarchically structured)
2. Uniquely human
3. Better for hypothetical scenarios than fine-grain motor control
4. Infinite (rule-driven)
5. Reliant on mental representations

- Skinner's Response: 1957 *Verbal Behavior*
- Chomsky's Response to Skinner 1959: language reflects knowledge, some innate, relies on mental representations, classic rationalist
    - sound, syntax, meaning: we have developed an understanding of rules for what is linguistically acceptable for English language, while mostly unaware (rule-based thought, infinitely generative)
    - Plato's problem: poverty of stimulus, logical problem of proper language acquisition
    - Gold' Theorem 1967: successful language learning requires either negative evidence or innate structure, thus children have innate understanding

Universal Grammar:

1. Why language is uniquely human
2. Why language is learnable
3. Why there are "language universals" (i.e. linguistic patterns, color naming)