CARTESIAN DUALISM:

- Mind and matter are different kinds of substances.
- **Matter:** Essentially spatially extended.
- **Mind:** Essentially thinks.
- We can think of our mind as being separate from lots of things.
- Thinking is very fundamental to mind and personality.
- Given the essences of mind and matter, how could the mind and body causally interact?
    - **Answer:** They can't
    - Mind being an addition to essential thinking **does not have a physical substance matter.**
- So the mind is a separate, non-physical substance.

CONCEIVABILITY ARGUMENT:

- **P1:**
    - I can conceive that I, a thinking thing, exist without my extended body existing.
    - The mind can live without a physical body.
- **P2:**
    - Anything I can conceive of is logically possible.
    - If you can conceive that his mind could exist without a body, it should at least be logically possible for your mind to exist without a body.
- **P3:**
    - If it is logically possible for X to exist without Y, then X is not identical to Y.
    - If it's possible for one thing to exist without the other thing then 2 things aren't identical to each other.
- **C1:**
    - I, a thinking thing, am not identical with my extended body.
    - It's true that our mental states aren't identical to our physical states.
    - The mind does not have matter and it exists outside of a physical body and it's just a thinking thing

BEHAVIORISM:

- **Methodological Behaviorism:**
    - What science should do.
    - Science should focus on observable, measures entities and behaviors (Watson, Hull, Tulman)
    - The most objective way back then in science was to focus totally on measurable entities and behaviors.
    - A thesis of what science should be oriented towards and what sort of explanations it should utilize.
- **Analytic Behaviorism:**
    - Mental concepts/states refer to behavioral tendencies. (Ryle)
    - The idea is that mental states are just behaviors. (Ryle)
    - **Example**
        - Sally wants ice cream which is going to be a desire of her just having the behavior of seeking ice cream.
        - Behaviorism just says what it is to desire ice cream is going about ice cream seeking behavior
        - Mental states are just in terms of behavioral profiles and don't make any reference to people like internal pictures in their head, beliefs, desires, and thoughts.
        - The only way to measure mental states for other people is just going about mapping out certain patterns of behaviors.
- **Radical Behaviorism:**
    - Internal mental states exist but they are unimportant and not very explanatory. (Skinner)
    - Science should focus on demonstrating relationships between behavior and environment. (Skinner)
    - Internal mental states exist but they are unimportant and not explanatory
    - Science should just focus on the relationship between behavior and the environment.

- Internal mental states are not very important for explanatory scenarios.

- Science should focus on demonstrating between behavior and the environment.

- **Example:**
    - You put a mouse in a skinner box, which is a box with no light or external sounds.

    - Scientists try to teach a mouse how to do a particular thing like a mouse being able to click a particular lever.

    - The idea is that you block out all the stimuli and you see how quickly it takes for the mouse to figure out what they want them to do by reinforcing the behavior with rewards.

    - If a mouse hits the lever and food comes down, they will keep doing it.

BEHAVIORISM TAKEAWAY:
- Mental states are just clusters of behavior

- If the brain feels pain, behaviors will be associated with being in pain.

COUNTEREXAMPLE TO BEHAVIORISM:
- If the behavior defines mental states, which are just clusters of behavior.

- What it is to have mental state pain is just to exhibit behaviors of pain like saying ‘ow’ and recoiling from the object of pain if you touched a stove for example.

- If that’s true then how do Spartans not display pain or have any outward signs of being in pain.

- The idea is that it’s possible that a person can have no reaction to pain, but is experiencing it.

- Or people can be experiencing emotional pain but a person does not exhibit the behaviors we associate with that state.

- If all it is to be in pain is to exhibit pain behaviors and you’re not exhibiting pain behaviors, then you’re not in pain according to the behaviorist.

- The mental state is just the behavior bypassing all of the references to internal representational states since we use traffic science.

- Being able to traffic mental states and predict other people's behavior in terms of these internal mental states is very important.

- It's very difficult to imagine if we weren't able to utilize internal representations like imagistic and propositional representations to understand people.

IDENTITY THEORY:

- Mental states are identical to brain states. (Place, Feigl, and Smart)

- It's saying behaviorism is wrong and we need to define them in some other way.

- It can't just be where mental states are identical to brain states because we think that multiple realizability is possible.

FUNCTIONALISM REVISITED:

- Mental states are defined in terms of their functional profiles (inputs, computations, and output), not in terms of their underlying physical state.

- We can understand profiles kind of computationally in terms of input computations and output.

- In Contra behaviorism, mental states can still be described as internal states, and in contra identity theory, mental states are multiply realized.

- You don't have to look at mental states in terms of patterns of behavior if you're a functionalist

- Functionalism is going to define the mental state at the 2nd algorithmic Marr level of selecting kinds of representation and algorithm.

- Example: Mental states are defined at the second algorithmic Marrian level.

COMPUTATIONAL THEORY OF MIND (CTM):

- **CTM:** The mind is an information-processing computer. (Putnam, Fodor, Rescorla)

- It is a specification from functionalism.

- Functionalism says mental states are defined in terms of their functional profiles.

- In other words, the mind is the software to the brain (Gross 'Meat Machine' analogy)

- The mind is not "like" a computer

- The mind is a computer

- “It is that neural tissues, synapses, cell assemblies, and all the rest are just nature’s rather wet and sticky way of building a hunk of honest-to-God computing machinery” (Andy Clark, 8).

- Mental properties are instantiated by properties but are not identical to physical properties.

- This is a functionalist theory.

- The mind can have computational functional profiles because the mind is a computer and it processes information.

- The mind is like the software to the brain.

- The mind is like a classical computer.

- However, others say the mind is a connectionist computer.

WHAT IS COMPUTATION:

Example - Formal Logic:

- Modus Ponens
  p
  p → q
  ----
  q

- p is some proposition

- If p then q, then q is the conclusion.

- It hangs together and has to be true.

- Forma logic preserves truth regardless of the meaning of the input.

- If p and q are true, then q is true.

Validity and Soundness:

- An argument is valid if and only if it takes the form that makes it impossible for all the premises to be true and the conclusion to be false.

- Regardless of where the premises are true.

        - **Example:**
            - If my cat is made out of popcorn (p)

            - If my cat is made out of popcorn then q = the cat would be a delicious snack (pretty messed up).

            - If it is true that the cat is made out of popcorn and if it is true it would make a delicious snack because it's popcorn, then it couldn't be false that the cat could be a delicious snack.

            - This makes this argument valid because the premise is true.

            - However, it is not sound and sound means you have validity because not all statements are true

            - The cat is not made out of popcorn, so this argument is not sound, but the argument would be valid.

FORMAL SYSTEMS:

- **Example:**
    - You can have different themes of chess boards such as Star Trek, different materials, but all chess boards will have the same rigid rules.

    - That is what makes games formal systems.

SYNTAX VS. SEMANTICS:

- Syntax: refers to the formal structure

- Semantics: Refers to meaning (and reference)

- **Examples:**
    - Natural language
    - "Dog" and "Chien": They refer to the same thing but have different formal structure.

    - **Programming languages:**
        - var x,y,z
        - x = 5; y = 6;
        - z = x + y;

TURING MACHINE:

- David Hilbert had the idea of automatic machines to answer questions.

- He wanted to make a machine that can answer a true or false question.

- At the time calculations were known, but not thinking machines.

- Alan Turing was fascinated with the human mind.

- In 1928 he published a paper that addressed Hilbert's challenges in his machine.

- **The Universal Machine:** Imagine it as a human in a box with a pencil and a stack of paper. They have a book of instructions to follow. You must alter instructions in the book to change the directions.

- Wanted to define the behavior with symbols rather than mechanisms.

- **Steps to the machine as a human:**
    - The person is observing precisely where their pencil is in the first symbol of input known as awareness.

    - They are only aware of one symbol at a time.

    - They then look at the instructions and compare the symbol in their memory to a list of possibilities in the 1st column to find a match.

    - Once a match is found, they move on to the second step of the conditional and operation step.

    - Turing observers that no matter what the instructions are or language, they are only ever going to be writing down symbols on paper and moving their attention to some new location.

    - The last column in the table is called the next state list, which lists the next page to jump to.

    - Each page in the book is a new step, which is called the state of the machine.

    - The process of computation is like turning pages in a book and on each page, you make a decision about what to do next based on where the pencil is pointing on a separate page.

    - This process will eventually lead to a final state and the machine will halt and produce its final result.

- Turing states that anything his universal machine can write down is what we call computable.

- A computable problem is where a turing machine could solve given enough time and space.

- All future machines and computers would be based on how you design the instruction.

- table instead of needing any complex mechanism inside the machine itself.

What a turing machine does:

- A hypothetical machine that seems relatively simple.

- Reads its input from paper tape that is infinitely long.

- The tape has a bunch of squares and each of the squares contains a symbol.

- You could have 2 symbols such as a space or a 1.

- There is a head that can move back and forth along the tape and can look at only one symbol at a time.

- After looking at the symbol, it consults a table of instructions and the instructions tell it what to do based on the state it's in and what symbol it scans.

  Example:

  - If a machine was in State A, and the head is pointed to a 1, the instructions say to write a 1 if it were empty since there is already a 1.

  - The instructions then tell the head to move to the right and change to state C.

  - It will then look at the table for any instructions that have State C and read the symbol, perform the operation, and then switch states.

  - This process will repeat forever until the machine is halted and produces the output.

- This universal machine can actually simulate itself or any other machine given the right instructions.

- The Turing machines make a bold claim that it's possible to invent a single machine that can be used to compute any computable sequence.

- **What does it mean to be computable:**
    - All arguments that can be given are bound to be fundamental appeals to intuition and for this reason rather unsatisfactory mathematically.

    - Alanzo Church came up with something called **Lambda Calculus:**

        - **The Rules of Lambda Calculus:**
            - **Variables:** x, y, z

            - **Function Definition:** 2 x M - takes variable x and returns some other expression.

            - **Function Application:** M N - the ability to apply a function and give it some parameters.

            - From this, Church argues you build on this to calculate anything calculable.

            - This is an abstract way of calculating things, so you must define everything if you want to compute something.

TURING TEST:

- Can an AI fool a human into thinking it's human?

- A benchmark for AI if an AI can fool a human it's a human.

- A tester would be communicating with a machine (AI)

- The idea is could you program a machine where the tester thought it was human.

- Or another test where a tester would have to pick between an AI and a human and see which one is actually a human.

- **Famous Examples of Tests:**
    - ELIZA (Weizenbaum 1966)

    - PARRY (Colby 1972)

    - Chatterbots

    - They fooled doctors into thinking they were paranoid schizophrenic, so that doctors that had experience dealing with patients with paranoid schizophrenia were allies.

    - They were able to display the sort of limited characterics of a paranoid schizophrenic.

- To this day no AI has gotten a gold medal for the turing test because it is very difficult to fool a real human.

- It's very hard to build a machine of general intelligence and to access all the subject matter.

- Intelligence and humanity seem to be separable.

- You can get intelligent but in human behavior and unintelligent behavior.

- There is a complicated balance in human behavior and intelligent behavior.

- **How do we know if a Turing test is a genuine intelligence and just sophisticated mimicry?**
    - We have never had a machine past the silver level, so it is hard to say how intelligent a machine can be.

THE FRAME PROBLEM:

- How to deal with uncertainty and relevance?

- **From a Dennet Fodor's paper who talks about the frame problem:**
    - There was a robot called R1.

    - It's only a task to fend for itself.

    - One day it's designer arranged for it to learn that its spare battery is locked in a room with a bomb set to go off.

    - R1 located the room and formulated a plan to get his battery.

    - R1 saw the battery on a wagon in the room

    - R1 hypothesized that a certain action which was called "pull out wagon room", which resulted in the battery being removed from the room because R1 pushed it out.

    - It acted and succeeded in getting the battery out of the room before the bomb went off.

    - However, the bomb was on the wagon.

- R1 did not know the bomb was on the wagon.

- R1 missed the obvious implication of its planned action.

- **What Went Wrong:**
    - R1 knew it was low on energy and it needed a battery.

    - It knew that the battery was in the room with a bomb.

    - He knew the battery was in the wagon in the room and knew to pull out.
    - R1 was not able to connect, making the implication that if he pulled the wagon out, the bomb would blow up outside the room.

    - The designers said their next robot must recognize not just the intended implication of its act, but also the implications of their side effects.

- They created another bot called R1D1 and put the bot in the same predicament as R1.

- It also had the idea of pulling the wagon out of the room.

- This time it considered the implications for such a course of action.

- The problem is that there are too many implications for its action and it would be a complex task in thinking about all possible implications in a scenario.

- **The real question is, how do we deal with when we're programming AI uncertainty and relevance?**
    - You would have to program everything the robot could interact with

    - You would also have to make the robot scan and somehow program it to make any implication given the information around it

    - It's very hard to program due to the amount of scenarios that can happen.

    - When we are dealing with uncertainty, we are not brute computing.

    - Humans cannot do efficient computing like computers who can do very complex algorithms very fast.

    - However since our brain is like a computer, we are able to figure out relevant set of possibilities and act depending on a scenario.

Success of AI: AlphaGo Zero:

- This AI is programmed only to know the rules of Go.
- It was able to beat 18 time world champion Lee Se-dol 100- 0 in a game of Go.
- There are so many moves you can make in go and AlphaGoZero was able to make intelligent turns.
- There was no brute computing besides giving it the rules of Go.
- This is where AI excels in giving a set of rules for something.
- It was able to play 4.9 million games against itself using a lot of brute computing.
- Our minds are not capable of doing such brute computing.
- Our brains are good at figuring out what to do quickly given a scenario and compute over relevance very quickly and that is something our most advanced AI models are lacking in.
- AI is trafficking in brute computing, which isn't the type of thing the mind is doing.

THE CHINESE ROOM (SEARLE):

- It is a thought experiment.

  The Structure:

  - There is a box with someone inside who follows instructions that is given from an input hole from somebody outside to their left.
  - The person inside then outputs the instructions it receives to another hole outside to their right.
- There is somebody that thinks they are having a conversation in Chinese (this can be in any language).
- The person in the room who does not speak or read Chinese has a huge book, which has similar instructions like a turing machine.
- The book is specifying a set of rules that is very turing like.
- It will specify a very long list of rules you do until you get to the desired output.
- **The question is, is this a good model for how our mind works?**

- This is a computational theory of the mind

- Our mind is like the guy in the room with a huge book of instructions.

- We get input from the world and manipulate it in a set of turing set rules

- Then we produce an output in some way, which is essentially thinking.

- However, this is essentially brute computing and in our brains, we understand how things work rather than just brute computing it.
- Serele condlues that this computation in the Turing method where you just have a big book and you're looking things up.

- It's not the right thing to get at consciousness, intentionality, and mentality.

- There is something missing here and it's not the right kind of thing.

- This is a good example of mimicry, but it's not understand anything due to brute computing.

- Our minds seem to be different from turing machines because we can adapt to things and learn.

- The thing that is missing is the engagement from the real world rather than just a robot body, which this box represents.

TWO KINDS ON INTELLIGENCE:
- **Classifying:** Recognizing patterns in the data.

    - We are good at building things that can recognize and see patterns better than humans.

    - We can build AI to be much more useful than humans when it comes to analyzing data due to its brute computation powers humans can never have.

- **Understanding:** Explaining and building a model of the world.

    - Explaining and understanding models of the world is a lot more difficult because AI would need genuine intelligence and an understanding of consciousness.

    - Humans are so limited in computation, but we are able to figure out what is relevant in a scenario very quickly.

- We can converse information through language and we have consciousness unlike AI.

- Our brains are still able to compute though without brute force, but just not as effectively as computers.