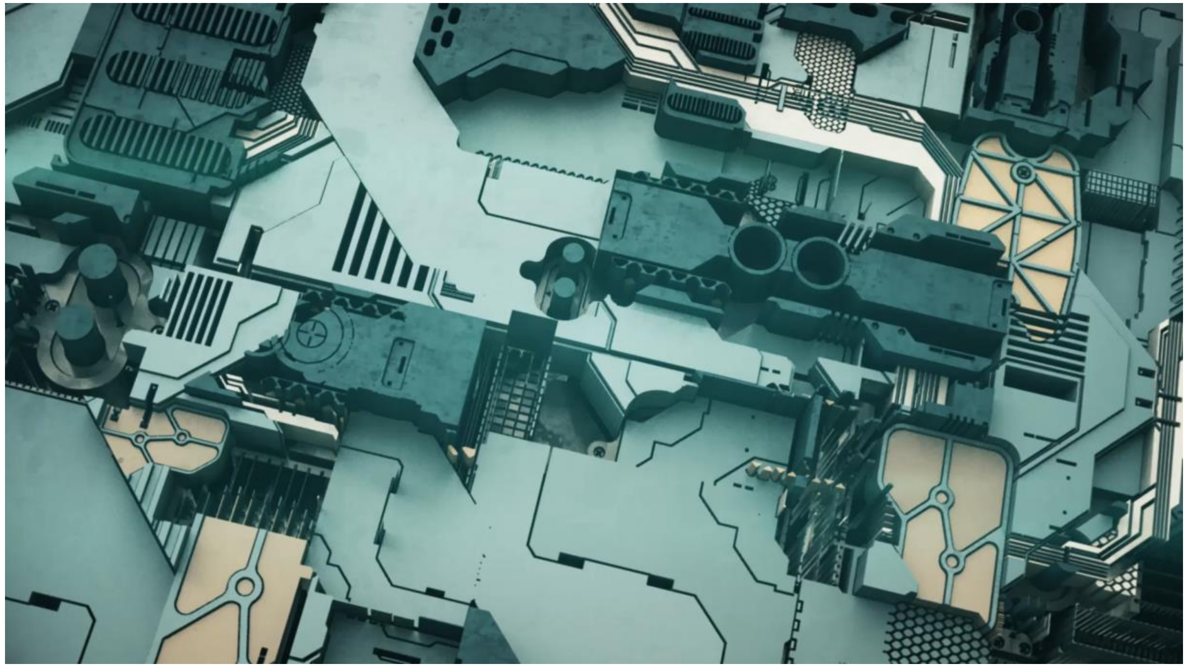
Data modeling and simulation

Prepared By: Dr. Ahmad S. Tarawneh

Updated By: Dr. Saed Alqaraleh



Outline

- Definitions
- Why simulation
- Modeling
- Pros and cones
- Area of applications

Definitions

- The title is "Data Modeling and Simulation", and consists of several terms:
- Data: refers to qualitative or quantitative attributes of a variable or set of variables.
  - It's collected through observations, measurements, or analysis and can be represented in various formats, such as numbers, text, symbols, images, etc.
- Modeling: refers to the process of representing a system, using specific tools, to study its behavior
  - model is a representation of a real-life system
- Simulation refers to replicating real-life scenarios using appropriate data and tools

Cont.

- The result of a simulation or analysis, can be predictions of what the system is expected to do or what outcomes can be anticipated, given data
- For example, given the population in an area of a city, car types, and number of cars, we can:
  - Predict the number of employees to employ in a gas station
  - Number of gas pumps to install
  - Estimated profits each month, given weather conditions and tourism seasons
- All of these can help in the decision-making process
  - Which all data science is about

Goals

- A model (or a simulation model) can be used to investigate a wide variety of "what if" questions about real-life systems
  - Potential changes to the system can be simulated, and predict their impact on the system
  - Find the adequate parameters before the actual implementation
- For example,
  - What If We Added One New Pump To The Station
  - What If The Weather Was Snowy 3 Days In A Month
  - What If A New Gas Station Opened In The Same Area
  - Etc.

Validation

- A simulation system can be validated by comparing its results with real-life data

Simulation: This involves running the model to predict how the system will behave under different conditions.

Prediction: The results of the simulation, which represent potential future outcomes or behaviors.

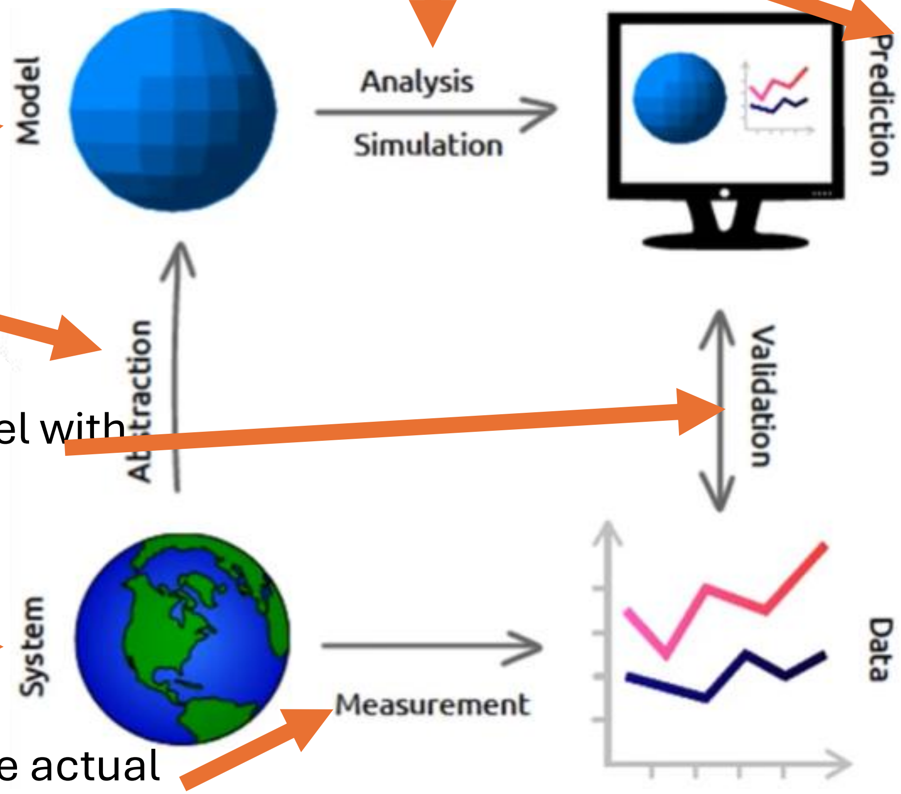
Model: This is a simplified representation of the system, often expressed mathematically or computationally.

Abstraction: This process involves simplifying the complex system into a model that focuses on the most relevant aspects.

Validation: Comparing the predictions from the model with the real-world measurements to assess the model's accuracy.

System: This represents the real-world phenomenon or process you're trying to simulate.

Measurement: Collecting real-world data from the actual system.



Advantages of S&M

- Hypothesis about how and why certain phenomena occur can be tested for feasibility
  - Why the number of registered students in a major is low
  - Why interaction between chemical elements is so slow
- We can control the time of the simulation
  - Study the spread of a virus over a year
  - How this spread might affect the medical sector in a country
- Study the interaction of different variables
  - How the social distancing and hygiene rate can affect the spread of a virus

Cont.

- Finding the importance of the variables to the system performance
  - determining whether the number of pumps or the number of employees more significantly affects the profits of a gas station

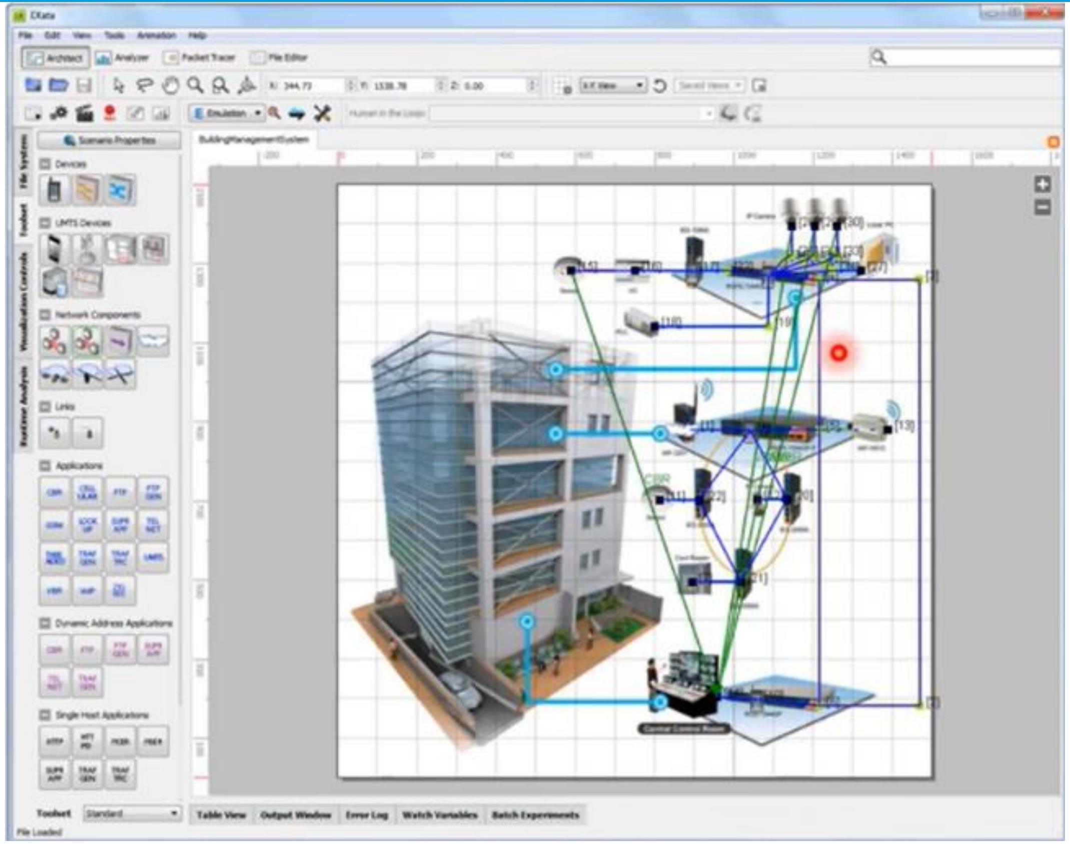
- What if questions can be answered easily
  - helps in creating or improving an existing system
    1. can adding new tables in a library enhance the customer's satisfaction
    2. is opening a new bookshop in an area will be beneficial

disadvantages

- Simulation might be expensive and time-consuming
    - For example, the simulation phase for a mission like the Mars Rover involves not just designing the rover itself but also simulating the entire journey from launch to landing and operation on its surface
    - This might consume a large portion of the budget
- Simulation results might be hard to interpret
  - Need specialists in the domain
- Designing simulations needs special training
  - How to include and remove factors, extract results and visualizations, etc.
- Needs good knowledge in the domain of discourse
  - Reduce The Time Needed To Check The Best Factors To Include, To Remove
  - Useful When There Are Numerous Factors To Manipulate

Applications

- Networks
- Simulation of a network for organization
  - Study the coverage area
  - cameras distribution
- Without simulation such networks might be expensive to update



Health care

- How much a hospital can handle cases before it breaks
- How many beds are needed for a hospital in a city

Road Traffic

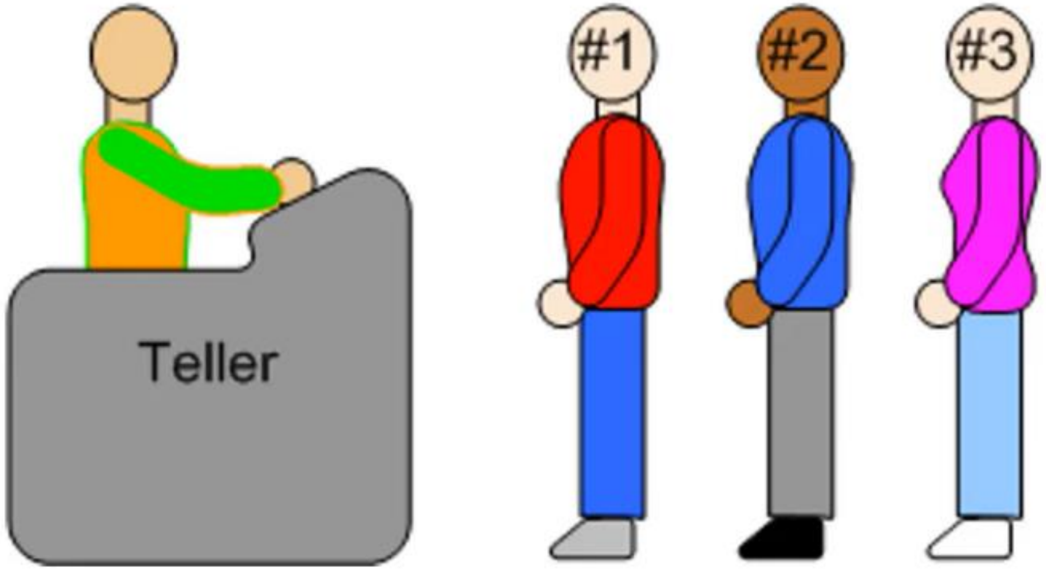
- What is the best width of a road in a given area with properties (population size, type of trucks
- How many traffic lights needed to minimize the traffic jams

Gas Stations

- How many pumps are needed, the width of the paths to prevent crowds. Also, it helps people get fast service.

Banking

- Do we need more employees to serve
- Prevent people from waiting too much to be served



Homework#1

- **Identify and describe three real-world examples of data modeling and simulation from your personal and local environment.**
- **For each example:**
  - Clearly state the problem or question being addressed.
  - Explain how data is collected and represented.
  - Discuss the potential benefits or limitations of **modeling this example**.
  - The systems classification(i.e., Discrete, Continuous, Hybrid).

Complete this assignment using your own knowledge and observations. In other words, don’t use the internet, ChatGPT, or similar apps for this question. Your answer must come from your own environment.

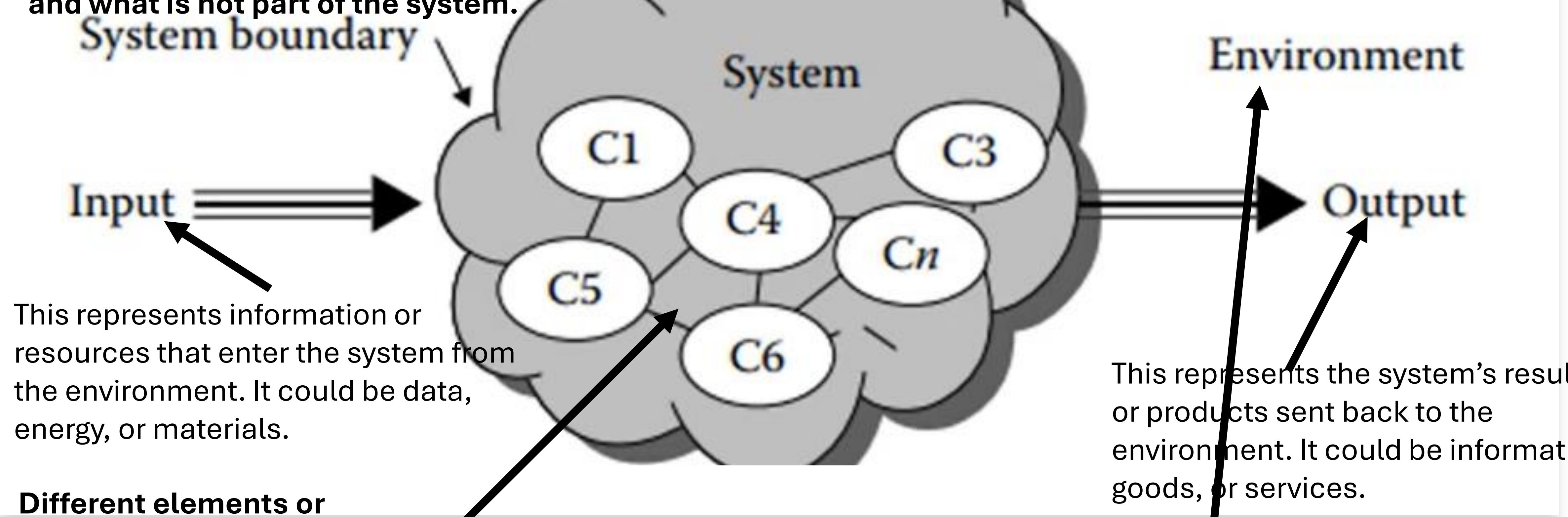
Simulation system

- A system is a set of components that interact with each other to achieve an objective (Goal)
  - **Or can be defined as a group of objects joined in regular interaction or interdependence toward accomplishing some purpose.**
  - A system can consist of several subsystems
- You might study a system that, in its turn is a part of a more complex system
  - The emergency room is apart from a hospital
  - Hospital is a part of the health care system, etc.

Cont.

- For example, if one wants to study a market shopping system to determine the number of cashiers needed to efficiently serve customers purchasing items, the system will include the components involving customers and cashiers.
  - If the role of a store supervisor is also of interest, the system will expand to incorporate this element accordingly.
  - More inclusive system
- Any subsystem that wants to be studied is a system on its own
  - Abstracted from all other objects

System Boundary: Separates the system from its environment. **It helps to define the scope of the system and identify what is and what is not part of the system.**



This represents information or resources that enter the system from the environment. It could be data, energy, or materials.

This represents the system's results or products sent back to the environment. It could be information, goods, or services.

Different elements or subsystems within the larger system

System attributes

- A system can be characterized by the following attributes
  - System boundary
  - Components
  - Environment

This is the external context in which the system operates. It includes the factors and influences that can affect the system's behavior, performance, and interactions.

Cont.

- The system gets its input from the environment
  - A customer enters the grocery store.
- This input affects the components inside the system
  - The number of waiting customers in line increases by 1.
  - A shopping cart is taken.
- Output is the result of the interaction between the components and the input
  - Checkout time: a customer took 5 minutes to complete their purchase.
  - The store's revenue increased by $50.
    - The customer bought groceries and signed up for a store loyalty program.

System components

- Entity: any object of interest in a system
  - Machines in a factory
  - Customers in a store
  - Pumps in a gas station
  - Employee in a library
- Attribute: a property of an entity
  - Capacity
  - Speed of service
  - Balance of a customer
- Activity: action done over a period of time
  - Making deposit
  - Being infected with a virus
  - time to hair cut

Cont.

- State: a collection of variables that are used to describe the state of the system at a given time
  - Individual (healthy, infected, dead, recovered)
  - Machine (idle, busy, down)
  - For example, in a grocery store system, the state variables might be: the number of busy cashiers, the number of customers waiting in line, or the arrival time of the next customer.
- Event: is an immediate occurrence that affects (change) the state of the system
  - Can be **endogenous** (within the system): a customer finished his service
    - this will change the cashier's state from busy to idle.
  - or **exogenous**: event occurred outside the system: a new customer arrived
    - This changes the state of the queue from empty to 1 (or cashier from idle to busy)

Concept	Duration	State Change	Example
Activity	**Yes**	**May or may not**	**Customer being served**
Event	**No**	**Always**	**Customer arriving**

System Name	Entities	Attributes	Activities	Events	State Variables
Traffic Intersection	Vehicles, Pedestrians	Speed, Direction, Type (car, truck, etc.), Age, Gender	Driving, Walking, Crossing	Vehicle arrival, Vehicle departure, Traffic light change, Pedestrian crossing	Number of vehicles waiting in each lane, Traffic light state (red, yellow, green), Number of pedestrians waiting to cross
Hospital Emergency Room	Patients, Doctors, Nurses	Severity of illness, Age, Gender, Insurance status	Treatment, Diagnosis, Surgery	Patient arrival, Patient discharge, Doctor availability, Nurse availability	Number of patients waiting, Number of patients being treated, Number of available doctors, Number of available nurses
Restaurant	Customers, Waiters, Cooks	Hunger level, Order preferences, Table number	Ordering food, Serving food, Preparing food	Customer arrival, Customer departure, Order placement, Order completion	Number of customers waiting, Number of tables occupied, Number of orders in the kitchen
Production Line	Machines, Products, Workers	Capacity, Speed, Quality	Assembly, Inspection, Packaging	Machine breakdown, Product completion, Worker availability	Number of machines operating, Number of products in queue, Number of workers on duty
Call Center	Customers, Agents	Call duration, Problem type, Agent experience	Handling calls, Resolving issues	Call arrival, Call completion, Agent availability	Number of calls waiting, Number of calls being handled, Number of agents available

Examples

Concept	Duration	State Change
Activity	Yes	May or may not
Event	No	Always

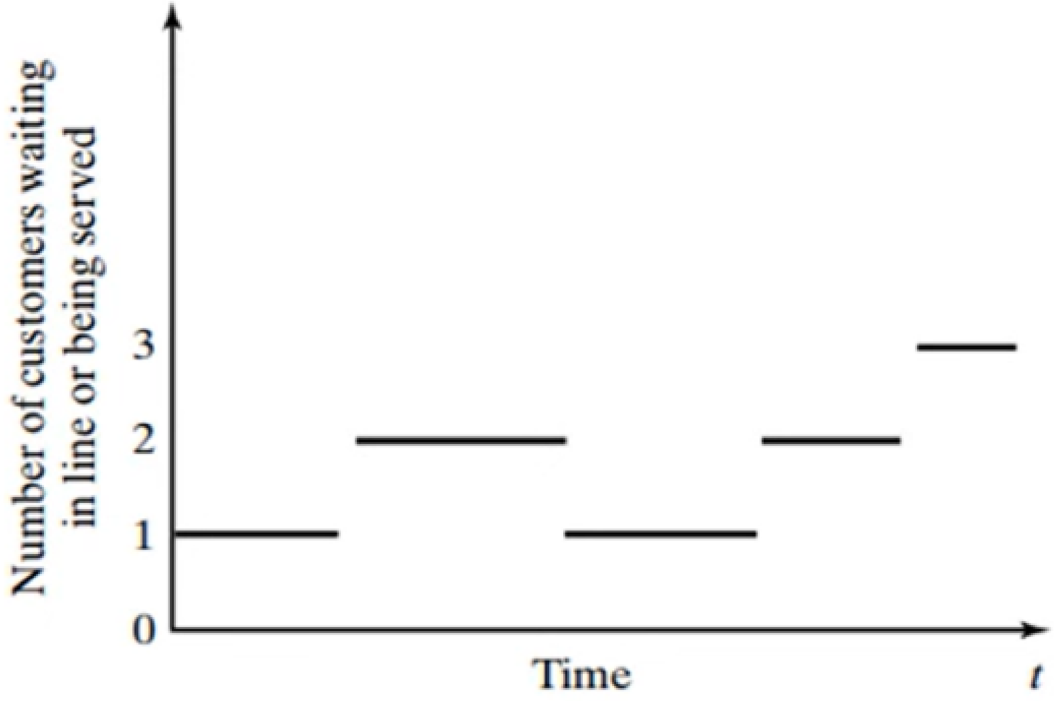
System	*Entities*	*Attributes*	*Activities*	*Events*	*State Variables*
Banking	Customers	Checking-account balance	Making deposits	Arrival; departure	Number of busy tellers; number of customers waiting
Production	Machines	Speed; capacity; breakdown rate	Welding; stamping	Breakdown	Status of machines (busy, idle, or down)
Communications	Messages	Length; destination	Transmitting	Arrival at destination	Number waiting to be transmitted
Inventory	Warehouse	Capacity	Withdrawing	Demand	Levels of inventory; backlogged demands

Systems classification

- The simulation systems can be classified based on time frames, type of measurements taken, type of interaction, type of components, etc.
- According to time frame, the system can be:
  - Discrete
  - Continuous
  - Hybrid

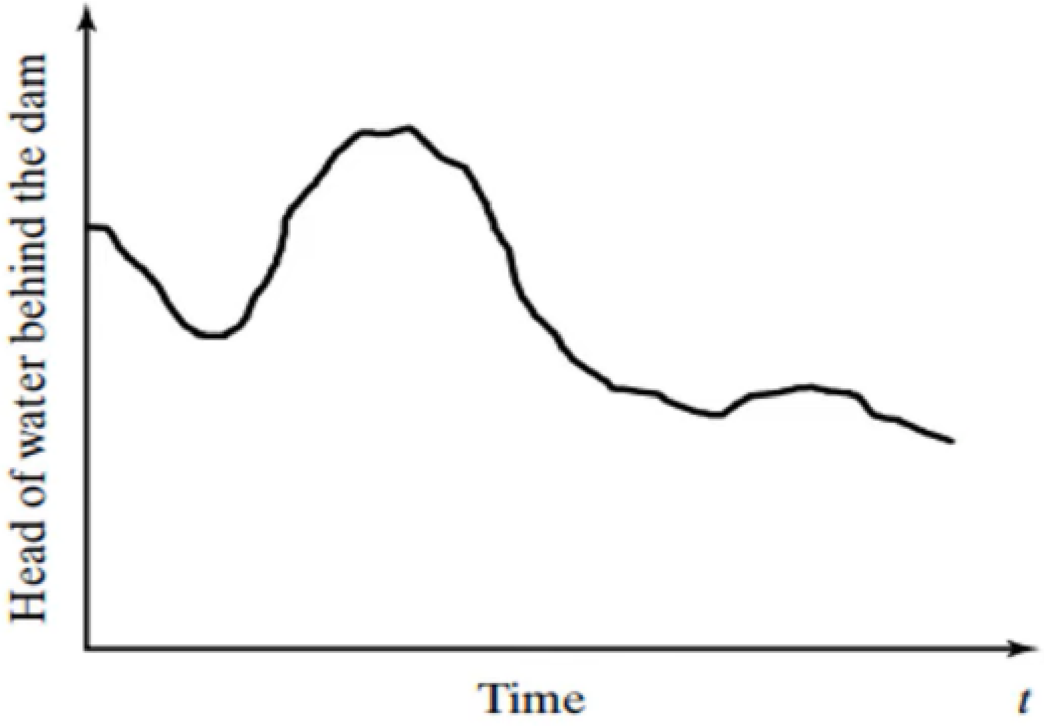
Discrete system

- The state variables change at discrete points in time. **In other words, These systems are characterized by events that occur at specific, distinct points in time**
- For example, banking system, gas stations (queuing systems in general), A traffic light system. The state changes (red, yellow, green) occur at specific intervals, not continuously.



Continues systems

- The state variables change continuously
  - Head of water behind a dam, battery charging, wind speed



Hybrid Systems

- These systems combine both discrete and continuous elements.

A chemical reactor. The chemical reactions within it may be modeled continuously, while the control system that regulates the reactor might operate in a discrete manner (e.g., turning valves on or off).

Systems classification Examples

System Type	Example	Explanation
Discrete	Traffic light system	State changes occur at specific times (red, yellow, green).
Discrete	Inventory management system	Inventory levels change when items are added or removed.
Continuous	A car's speed	Speed can vary smoothly over time.
Continuous	Temperature of a room	Temperature can change gradually over time.
Hybrid	A chemical reactor	Chemical reactions occur continuously, while control systems may operate discretely.
Hybrid	A traffic flow model	Vehicle movements can be modeled continuously, while traffic signals operate discretely.

Components

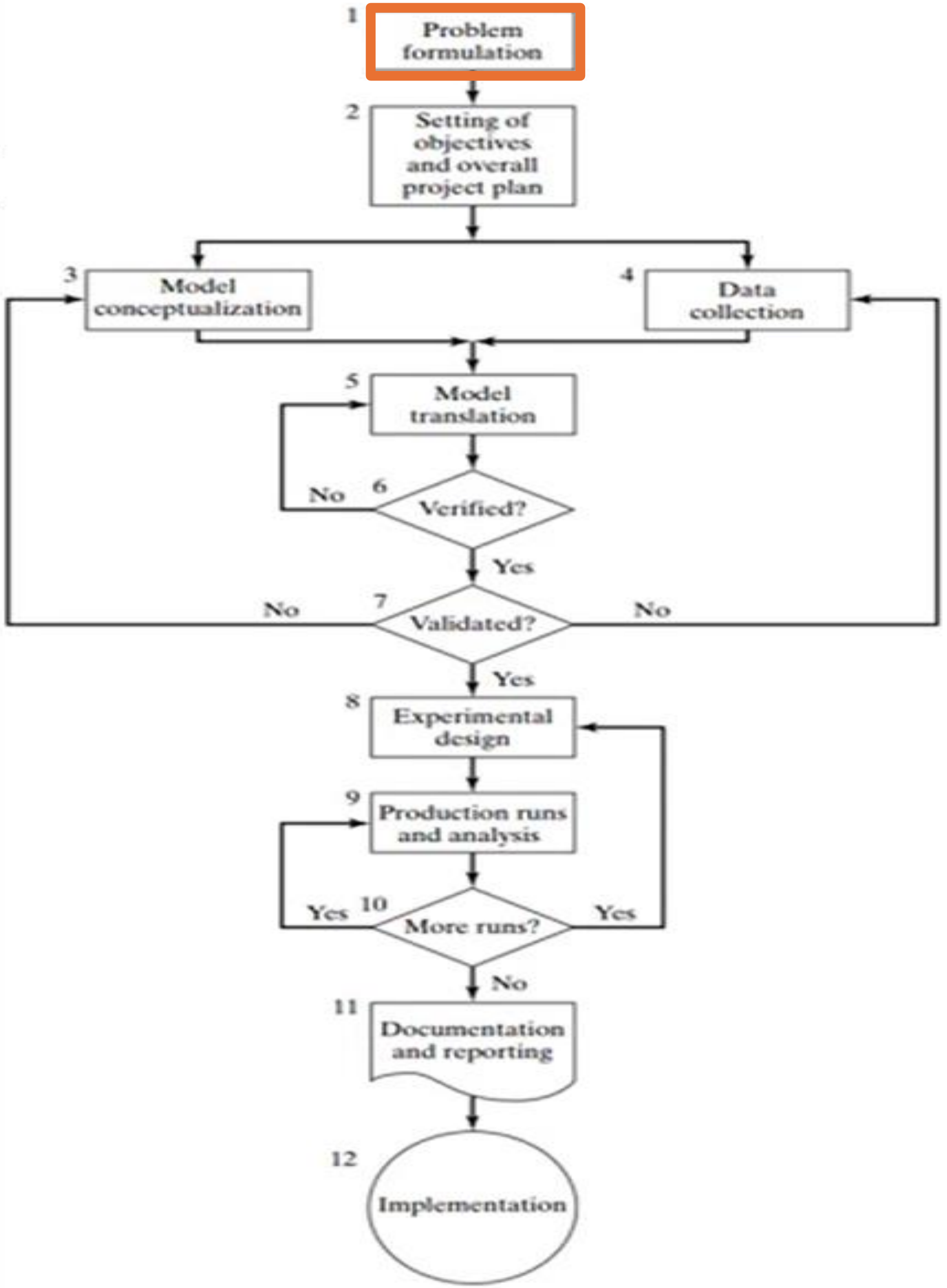
- Also, The system can be classified based on components to:
- Static systems: typically refer to systems that do not change over time.
  - Present a system at a particular point in time
  - or without considering the element of time
  - e.g., Simulating the value of π (pi), Network Design
- Dynamic systems: Systems that changes over time.
  - simulation of a bank from 9 AM to 4 PM

Deterministic and stochastic

- Deterministic: In deterministic simulations, the outcome is strictly determined by the initial conditions, parameters, and fixed rules without any randomness involved
  - Predicting the movement of planets in a solar system based on gravitational laws
  - Linear regression using the least squares method => same results each run
- Stochastic: Stochastic simulations involve elements of randomness
  - Linear regression using gradient decent (random weights) => different results each run
  - Queuing systems

Steps to simulation

- Every study should begin with a statement of the problem.
  If the statement is provided by the policymakers or those that have the problem, the analyst must ensure that the problem being described is clearly understood.
- If a problem statement is being developed by the analyst, it is important that the policymakers understand and agree with the formulation.



Steps to simulation

- The objectives indicate the questions to be answered by simulation.
- At this point, a determination should be made concerning whether simulation is the appropriate methodology for the problem as formulated and the objectives as stated.
- E.G.:
    1. **Objective 1:** To accurately model the current customer service process at the bank's branches, including customer arrival rates, service times, and queue management strategies.
    2. **Objective 2:** To identify the key factors contributing to long wait times during peak hours by analyzing the model's output.
    3. etc.

Steps to simulation

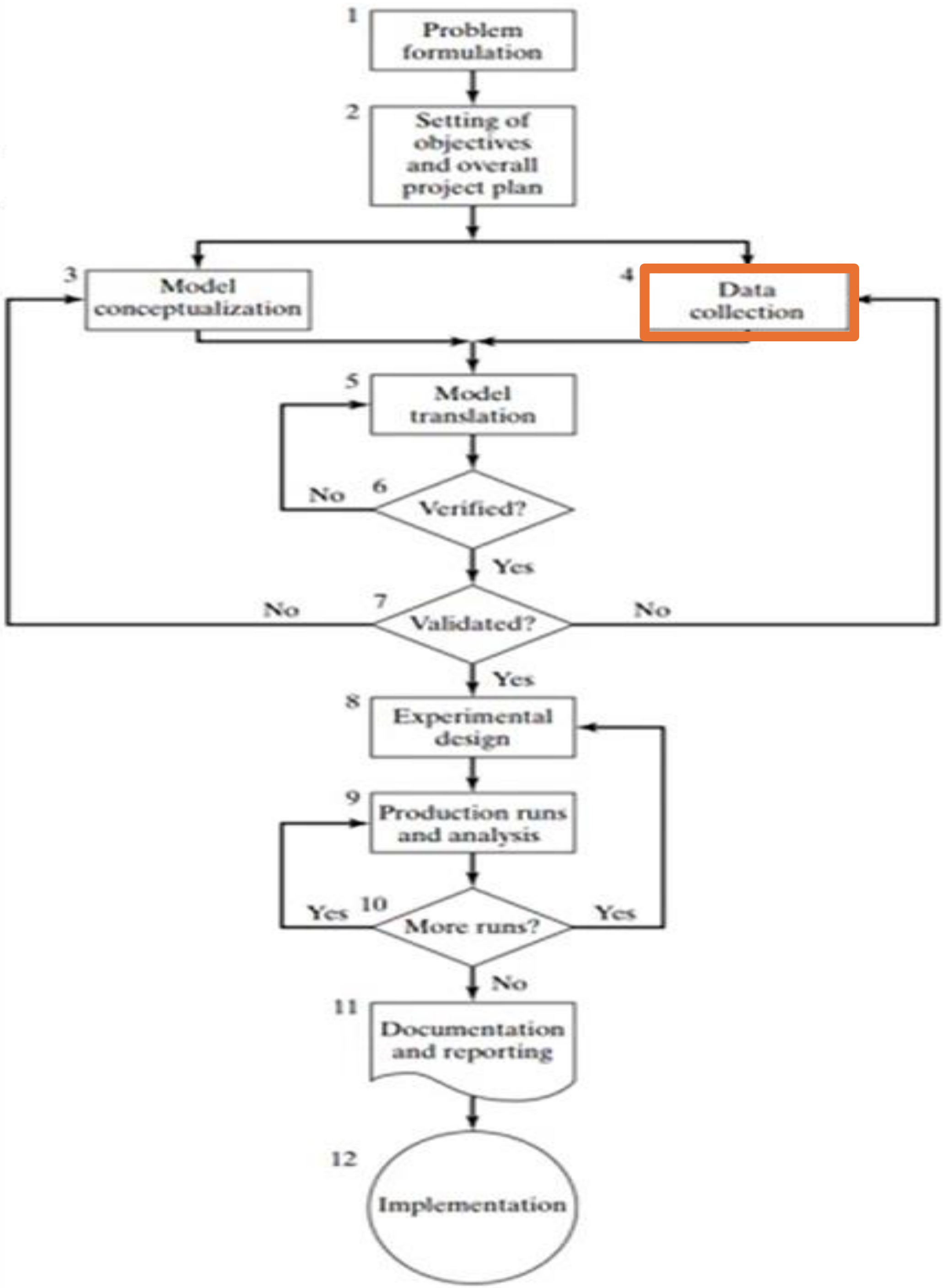
- It is best to start with a simple model and build toward greater complexity.
- However, the model complexity need not exceed that required to accomplish the purposes for which the model is intended.

Steps to simulation

- Start collecting the data from the real life domain to understand the patterns.
- For example, Patient Arrival Times data in the context of a hospital emergency department (ED),

Patient ID	Arrival Date	Arrival Time	Day of Week	Acuity Level	Reason for Visit
001	2024-02-01	08:45	Friday	3	Chest Pain
002	2024-02-01	09:10	Friday	2	Broken Arm
003	2024-02-01	09:30	Friday	4	High Fever
004	2024-02-01	10:05	Friday	1	Minor Laceration

we can add more data, like the departure time



Steps to simulation

- Start modeling, this might be using simulation language (Matlab, Python) or special purpose software

- Check with the decision makers if the simulation is programmed as intended (did we build the model right?)

Steps to simulation

- Did we build the right model
- Does the model give results as expected
- This is done on a prototype simulation
- Compare its results with the actual system

Steps to simulation

- Designing a large-scale experiments
- run the simulation several times or more experiments should be conducted

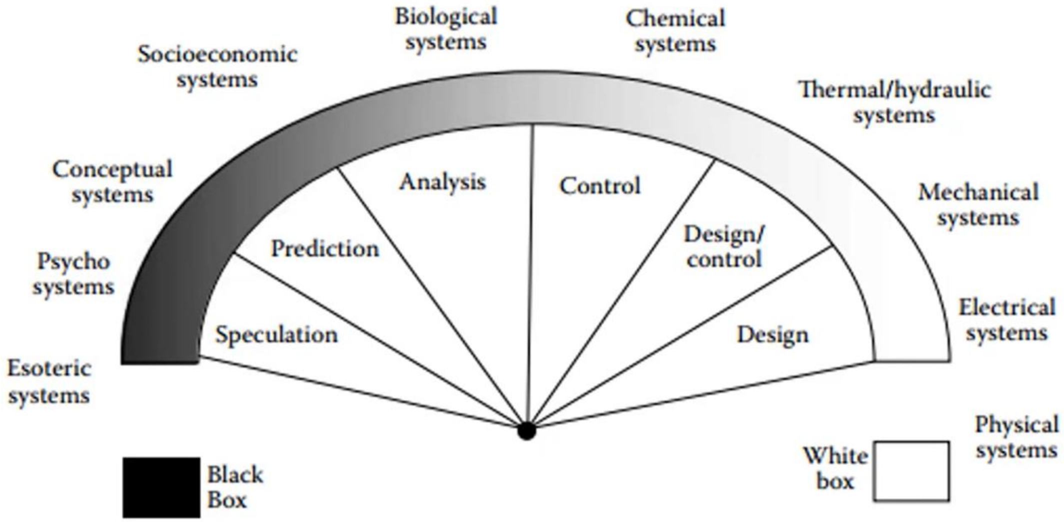
- Thes more experiments might be needed to get the best replication of the real-life scenario

Steps to simulation

- Document the results of the simulation and report the outputs
  - Report the whole history of the simulation
- Make decisions based on the results of the simulation
  - Add a new gas pump
  - Install a new ATM at a street or mole
  - Buy new beds for ER at a hospital
  - etc.

Simulation complexity

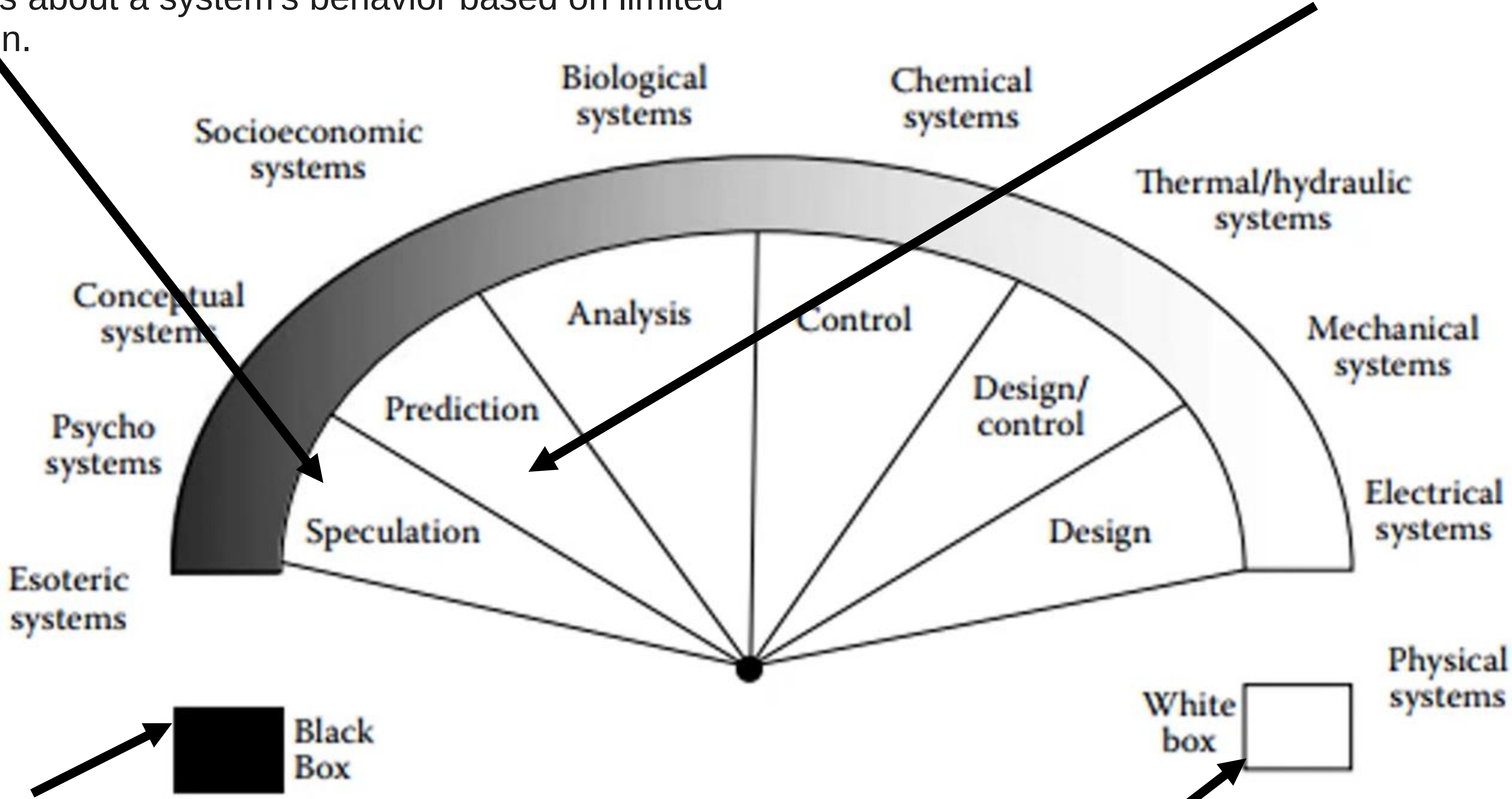
- The systems can be classified based on complexity



Simulation complexity

This technique involves making educated guesses or predictions about a system's behavior based on limited information.

This involves using models and data to forecast the futur behavior of a system.



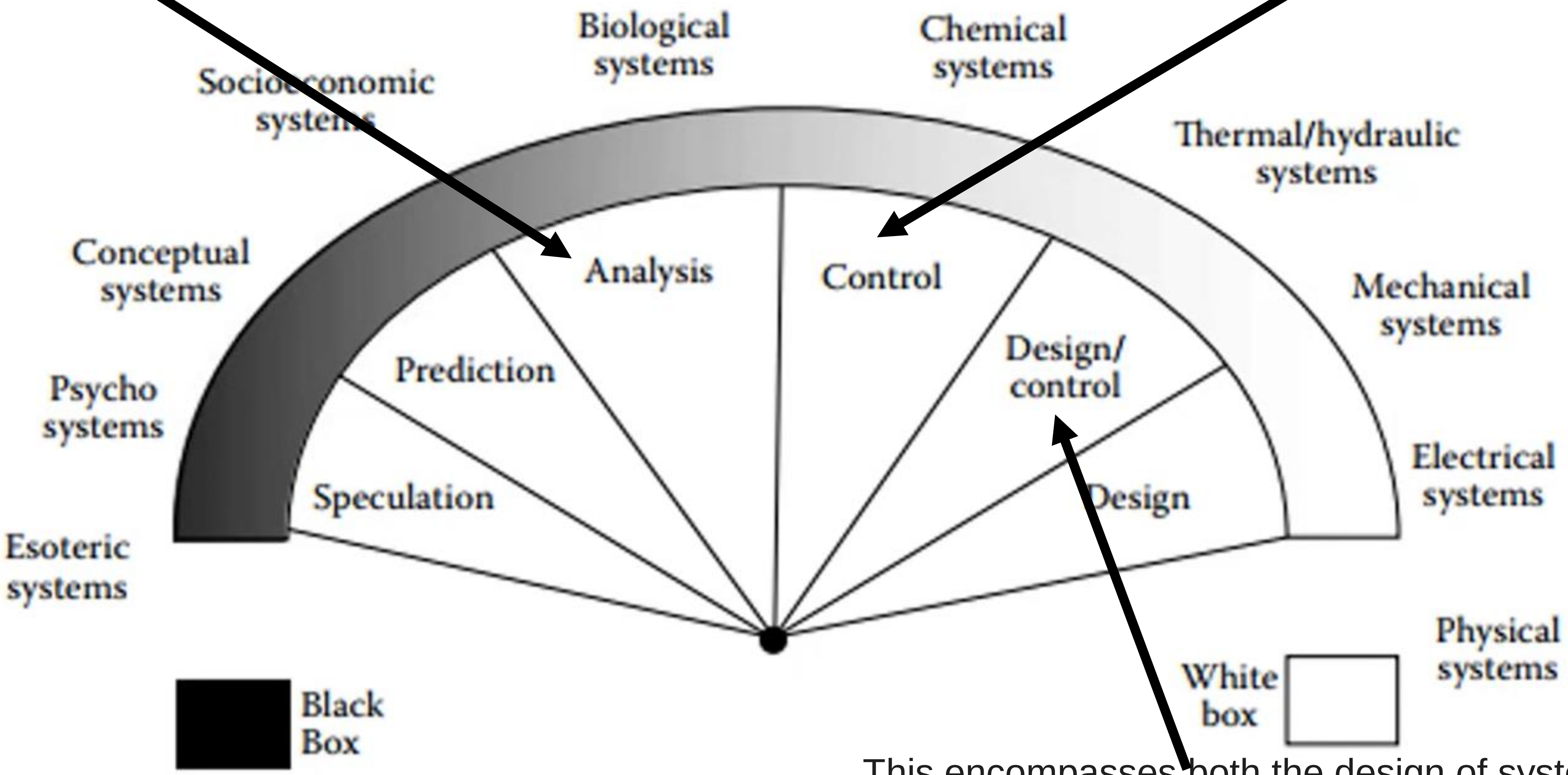
This approach treats a system as an unknown entity, focusing solely on its inputs and outputs without considering its internal structure or mechanisms.

This approach involves a detailed analysis of a system's internal components and their interactions.

Simulation complexity

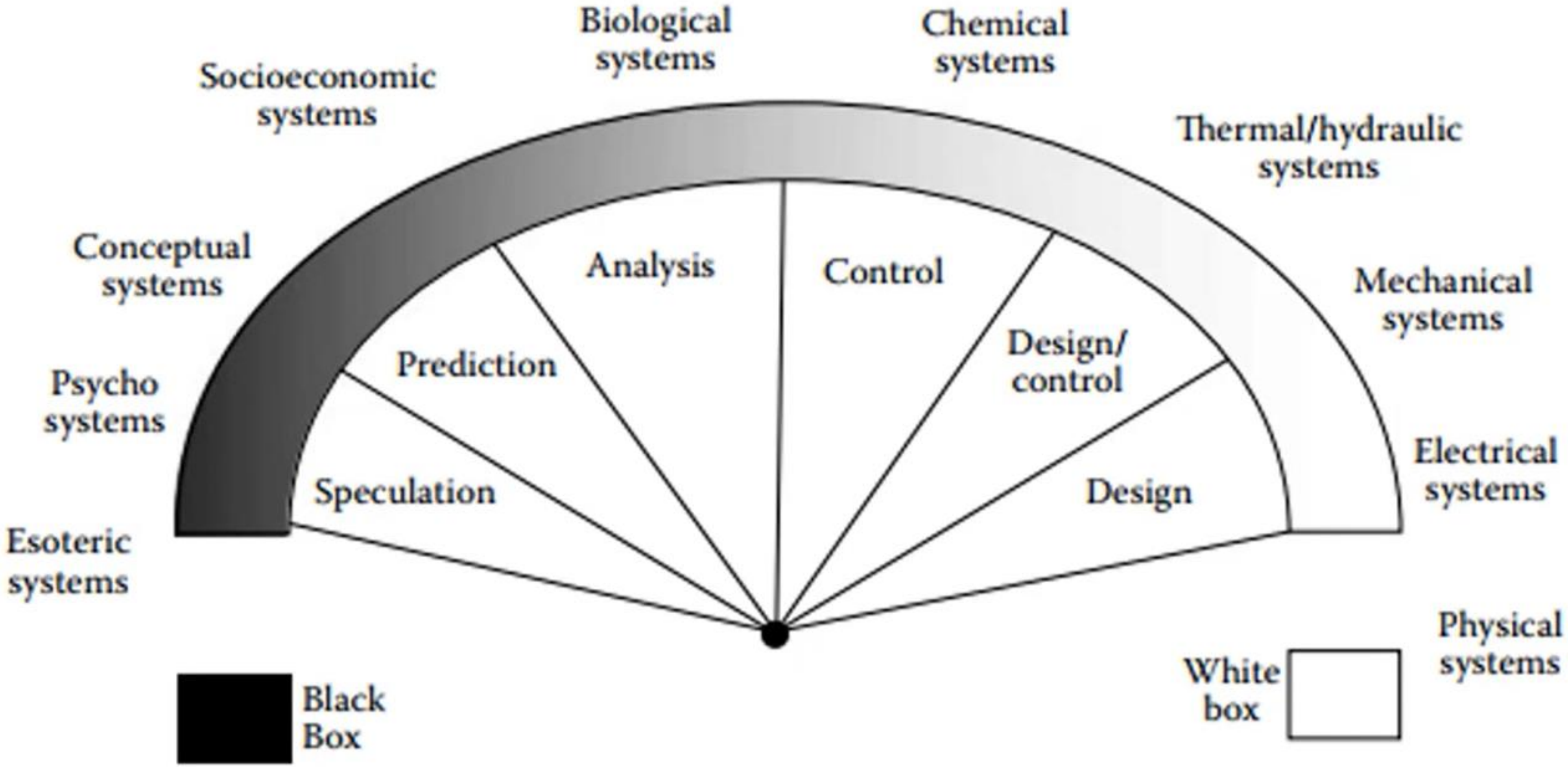
This involves breaking down a system into its constituent parts and studying their relationships.

This involves designing and implementing strategies to influence or regulate a system's behavior.



This encompasses both the design of systems and the development of control mechanisms to ensure desired performance.

Simulation complexity



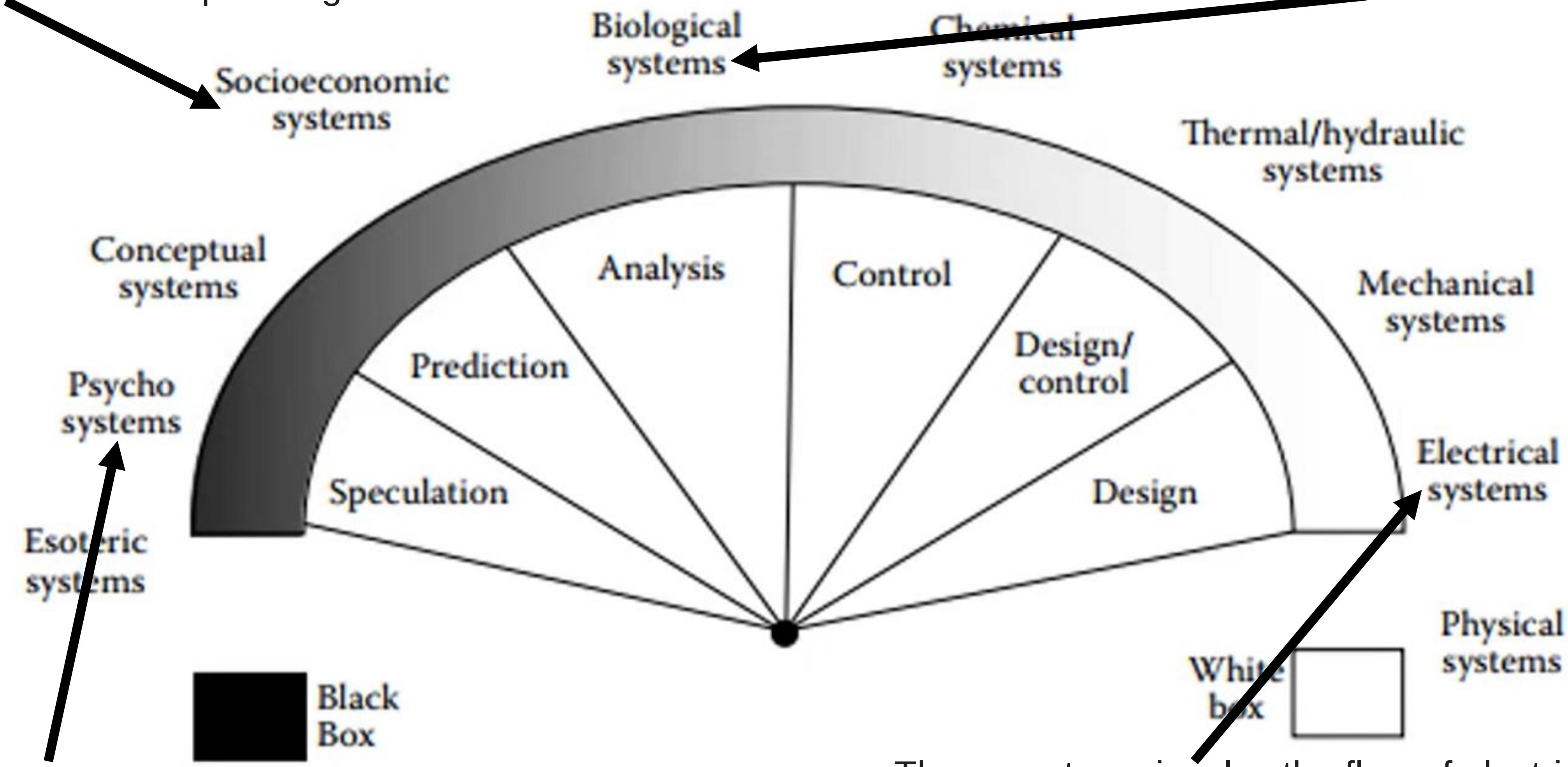
Right ← **Left**

The complexity of systems increases as we move from left to right and from bottom to top. This implies that systems in the upper right quadrant, such as socioeconomic and biological systems, are generally more complex to model and analyze than physical systems in the lower left quadrant.

Simulation complexity

These systems involve human interactions, economics, and society. Examples include financial markets, transportation networks, and urban planning.

These systems deal with living organisms and their interactions. Examples include ecosystems, human physiology, and genetic engineering.



These systems focus on human psychology and behavior. Examples include cognitive science, social psychology, and human-computer interaction.

These systems involve the flow of electrical current and its applications. Examples include power grids, electronic circuits, and communication systems.