Question Group GEN (General Social Simulation Questions)

1.What is the difference between top-down and bottom-up modeling techniques – and which simulation methods are top-down, which bottom-up. Give an example that shows the difference between the two approaches (also draw a sketch showing the layers, agents, interaction and direction of inference; choose a different example than sheep & wolves).

Bottom-up BU	**Top-down TD**
More focus on individual level because heterogeneous development of agents is possible. In a BU simple rules can be sufficient to explain complex phenomena, for example flocks of birds may be very hard to describe if the focus is on the swarm pattern but are a lot simpler described though the rules of movement of individual birds.	Many processes may seem complex from a TD approach, no heterogeneous agents, aren't accurate representations of most socio-economic processes, interaction between individuals don't have a high relevance.
Microsimulation, Cellular Automata, Agent Based Modelling	System Dynamics

2.What is a self-organizing system? What is stigmergy? What is an emergent phenomenon (emergence)? Which modelling techniques can simulate emergent phenomena?

Self-organizing systems don't need central planning or a leader, and there are no predefined directions or patterns that they follow. However, a pattern can still occur which is in this case self-organizing (e.g. Birds forming swarms and observe the behavior of the other birds, there is no clear leader bird but still they follow the direction of the swarm).

Stigmergy is when agents self-organize through their environment. Ants for example do not follow the actions of other ants directly (compared to bird swarm for example) but they leave behind traces of pheromone when they find food on a successful path, other ants start to follow more intense paths and patterns emerge.

A phenomenon is emergent if a new category needs to be formed to describe it, which is not required to describe the behavior of the underlying components of the phenomena.

Examples:

Motion of Atoms ⟶ Temperature

Motion of Birds ⟶ Flocking Behavior

Motion of Cars ⟶ Traffic Jams

Motion of People in stadium ⟶ Mexican Wave

A single Atom does not have a Temperature, but a collection of atoms has a temperature since it is an emergent property of the atom collection's motion.

Modelling techniques: Only Bottom-up techniques are possible. Individuals need to be modeled (emergence is the result of interaction) + Multiple levels need to be possible (emergence effect on 2nd layer)

3.Why is prediction of social processes problematic - compare purely physical processes with social processes? Why did the apple not hit the ground and what does this represent – and is it relevant in the context of social simulations? Which social systems can be better predicted (e.g. with analytical methods or for longer periods) than others and why?

Physical processes follow exact laws of nature (e.g., Gravity) that can be shown with equations that can be solved (e.g. Apple hitting the ground after 1 s after solving equation).

Social processes don't have an exact future state that can be predicted, because not all information is completely known by everyone, and the behavior of many individuals has an influence on the outcome. Making predictions in social systems is therefore much more complicated.

The apple didn't hit the ground because of interference (e.g., someone catching the apple before hitting the ground) by an individual (only this individual knew of the interference before = asymmetric information). This changed the outcome, and therefore the predictions were wrong.

For medium to large homogeneous systems (= systems with huge number of individuals) short to medium term predictions are possible because often individual behavior is evened out completely. Except if there are single individuals with a very high influence on the whole system. Long-term predictions are problematic because the structure will change over time.

4.Which four computational social simulation techniques have been presented in the lecture? Shortly describe them, including the number/meaning of agents and (especially important) also state for which kind of process/situation each technique would be suited well!

1. System Dynamics (Top-down):
System Dynamics is a computational social simulation technique that focuses on understanding the behavior of complex systems using feedback loops and stock-and-flow diagrams. It models the system as a set of interconnected components and captures the aggregate behavior of the system. Agents in System Dynamics models are often represented as aggregate entities or variables, such as population, capital, or inventory. This technique is suited for studying mechanisms, such as population growth, economic cycles, and policy analysis.

2. Microsimulation (Bottom-up):
Microsimulation is a computational social simulation technique that models the behavior of individual entities within a system and simulates their interactions. It focuses on capturing the heterogeneity and individual-level dynamics of agents. Agents in microsimulation models are typically represented as discrete entities with specific attributes and behaviors like people, households, firms. This technique is well-suited for studying complex social phenomena at the individual level, such as labor markets, transportation systems, healthcare systems.

3. Cellular Automata (Bottom-up):
Cellular Automata (CA) is a computational social simulation technique that represents a system as a grid of cells, where each cell can take on different states. The state of a cell is updated based on a set of rules that define its behavior and interactions with neighboring cells. Agents in CA models are typically represented as cells with simple local rules. This technique is well-suited for studying spatially explicit phenomena and emergent patterns, such as urban growth, ecosystem dynamics, traffic flow.

4. Agent-Based Modeling (Bottom-up):
Agent-Based Modeling (ABM) is a computational social simulation technique that models a system as a collection of autonomous agents, each with its own set of attributes, behaviors, and decision-making rules. These agents interact with each other and the environment, leading to the emergence of macro-level phenomena. ABM allows for the representation of heterogeneity, interactions, and adaptive behavior at the individual agent level. It is suited for studying social processes with complex interactions, such as social networks, opinion dynamics, crowd behavior, and the spread of infectious diseases.

Question Group SDM (System Dynamic Modelling Questions)

5. What is a causal loop diagram? What are feedback loops? What is the difference between the two methods? What are they good for and when would they better not be used? Give examples for positive and negative feedback loops (one each)!

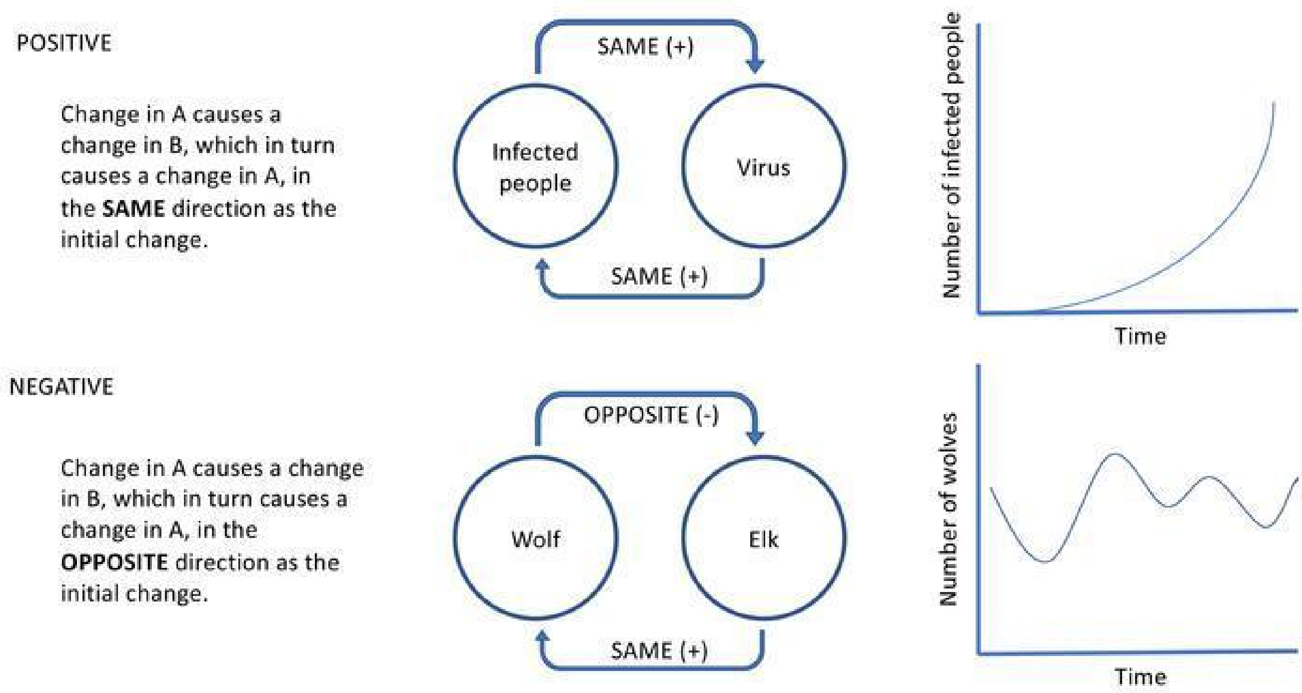
A causal loop diagram is a visual representation of feedback loops that helps to visualize how different variables are interrelated in a system dynamics modeling. It consists of some elements, arrows, and the corresponding sign (positive or negative). The arrows tell the direction of causality for a link, and the sign for a link tells whether the variables at the two ends move in the same (+) or opposite (-) direction.

A causal loop diagram is good for presenting the basic ideas incorporated in a model in a manner that is easily understood, without having to discuss them in detail. They shouldn't be used when it is important to determine the dynamics of a system.

Feedback loops are logical chains of cause and effects in a system. The basic modes of behavior in dynamic systems are identified along with the feedback structures generating them. It is used when an element of a system indirectly influences itself.

Positive feedback loops generate processes like growth, amplify deviations and reinforce change. Negative loops are acting to bring the state of the system in line with a goal (defined by gap) or desired state by counteract any disturbances that move the state of the system away from the goal (balance, equilibrium, stasis)

Difference?
A causal loop diagram is a visual representation that illustrates the relationships between variables in a system. Feedback loops, on the other hand, are the dynamic elements within a causal loop diagram that describe how changes in variables influence each other through reinforcing (positive) or balancing (negative) effects. Feedback loops are the mechanisms that drive the behavior and dynamics of a system.



6. What is a stock and what is a flow? Give three different pairs of examples with one stock and one corresponding flow each including sketches in the notation used in the lecture! Further, give some examples of common misconceptions of stocks and flows!

A stock is a quantity at a certain point in time that is accumulated over time by inflows and/or depleted by outflows, for example: Population, Public debt, Account balance.

A flow variable is measured over an period of time and it is a mechanism that increase (in-flow)/ decrease(out-flow) a stock, for example: Births/Deaths, Year 's budget deficit, Deposit/ Withdrawal

Examples:

- Capital (stock) increase or decline by investment/depreciation (flows)
- Population (stock) increase or dec decline by births/deaths (flows)
- Guests in a hotel (stock) increase or dec decline by arrival/departures (flows)

One common misconception is mistaking stocks for flows or vice versa. Stocks refer to the accumulation or quantity of a variable at a specific point in time, while flows represent the rate of change of that variable over time. For example, considering a bank account balance as a flow (money coming in and going out) rather than a stock (the actual amount of money in the account) can lead to misunderstandings.

7. What is a system dynamics simulation model compared to causal loop and stock flow diagrams? How is such a system dynamics simulation developed (simple words)? What are such simulations used for? Which kind of systems/processes are they well suited to simulate?

A system dynamic simulation model is a computer-based mathematical modeling approach for policy analysis and design that helps to understand the behavior of complex systems, by using stocks, flows and internal feedback loops.

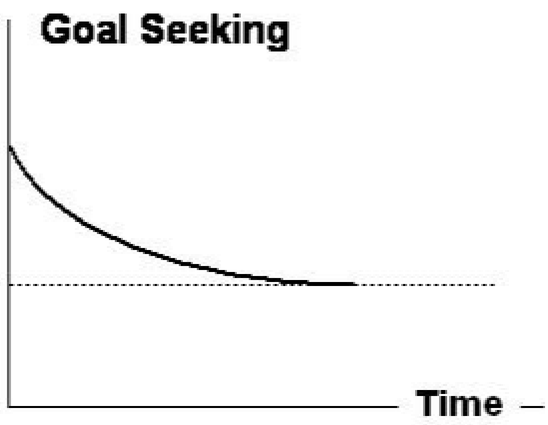
The simulation is developed through these steps:

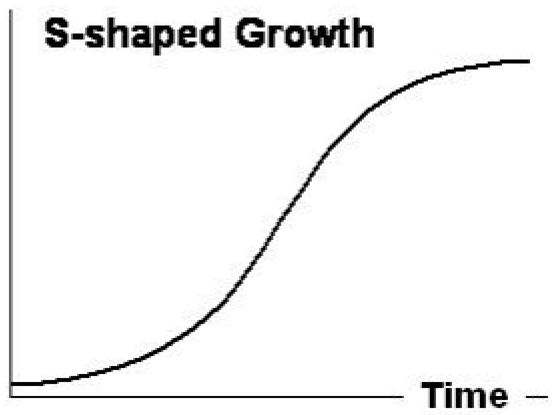
(1) Identifying the most important stocks and flows that change
(2) Identifying the sources of information that impact the flows.
(3) Identifying the main feedback loops.

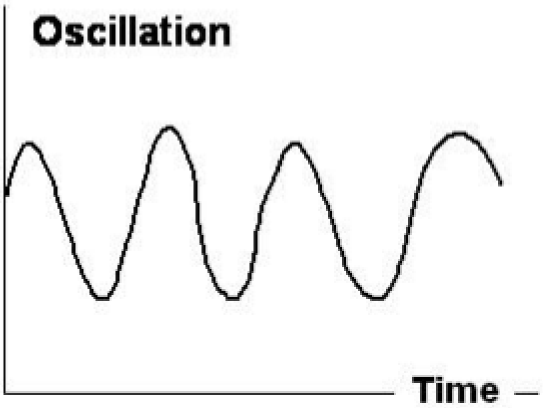
These simulations are used for understanding the sources of policy resistance and designing more effective policies by studying social systems, capturing the structure, and predicting short/ long term effects.

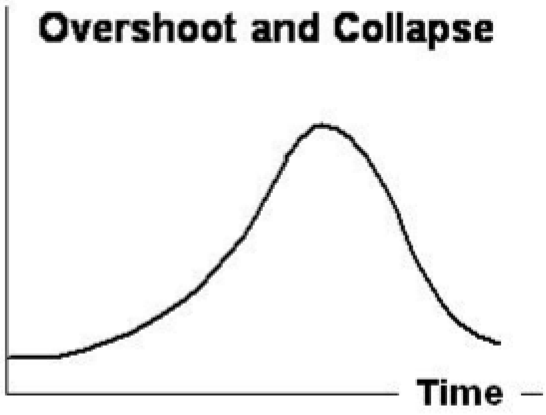
SD works good for a more abstract and simpler simulation of complex systems that have unknown "effects" that we cannot consider. Instead of including every little detail in the model, we can just model the aggregation e.g., we can just model the birthrate of a population rather than trying to model the underlying phenomenon or interactions between individual agents. Example for SD: Predator - Prey model.

8. Sketch the following modes of dynamic behavior: Goal Seeking, S-shaped Growth, Oscillation, Overshoot and Collapse. For each of these, give examples of systems/processes that feature this behavior. Which of these do you think are most likely to be found in natural systems in the long run – give examples of systems and their qualitative behavior!









1. Goal Seeking: Goal seeking refers to a system's ability to adjust its behavior to reach a desired target or goal. An example of goal seeking is a thermostat-controlled heating system. When the temperature drops below a set point, the system activates the heater to bring the temperature back to the desired level. Similarly, an autopilot system in an aircraft continuously adjusts the control surfaces to maintain a stable flight path.
2. S-shaped Growth: S-shaped growth, also known as logistic growth, is commonly observed in population dynamics. One example is the growth of a population of organisms in an ecosystem. We see in the beginning that the growth is exponential. However, it gradually slows down until the system's condition achieves an equilibrium state. This is happening because the population of a system will reach its capacity resulting in the decrease of resources that can sustain its population.
3. Oscillation: Oscillation refers to the repetitive back-and-forth movement or variation of a system. An example is a simple pendulum. When displaced from its resting position, the pendulum swings back and forth, exhibiting oscillatory motion. Another example is an electronic circuit with a feedback loop that can create sustained oscillations, as seen in an oscillator circuit used in radios or clocks.
4. Overshoot and Collapse: Overshoot and collapse occur when a system temporarily exceeds its limits or carrying capacity, leading to a subsequent decline or collapse. One example is the boom-and-bust cycles observed in predator-prey relationships. When the population of prey animals increases, it can lead to a corresponding increase in predator numbers. However, as the predators consume more prey, the prey population declines, eventually causing a decrease in the predator population due to a lack of food.

Oscillation is the most common pattern found in natural systems, especially when thinking of physics there are a lot of examples: a pendulum, waves of different kinds (e.g. sound, ocean, electromagnetic, etc.), string instruments and more. Exponential growth can be found in natural systems as well but is not as common as oscillation, for example the expansion of the universe.

9. From a system-theoretic view: What is the danger of a policy that only regards the short run? Why are complex socioeconomic systems hardly predictable (especially when seen from an SD perspective)? Give an example of a social system where the short run differs from the long run and why this is problematic!

A Policy that only focuses on short term effects can cause long term effects that were not taken into consideration by policymakers. This is caused by the delay of feedback within the systems, and it can take a long time to see the unintended consequences.

Some reasons why socioeconomic systems are hard to predict:

- Human intuition is sometimes wrong about the expected behavior.
- Policy intervention sometimes only focuses on either short term or long-term effects.
- External policy intervention is neutralized or counteracted by the system's internal feedback.
- External shocks can hardly be predicted within the model and therefore the focus should not be on factoring them in but to structure the system to withstand uncertainty.
- In the real-world, there is no equilibrium in complex systems, but continual change.

Example: Road building policy to avoid traffic jams: A policy is put in place to invest and build better road infrastructure because of traffic jams. In the short term, this might have a positive effect with less traffic jams and faster commuting times due to more and bigger roads being available. Over time, due to the convenience of using this better infrastructure, more people rely on commuting to places by car and therefore start to create more traffic. In the long run the city will have more traffic than before with more space used up for road infrastructure but will have the same problem as before or even worse on roads that were not renewed to meet the new demand. This is problematic because it leads to the exact same problem as before, but now in an even bigger dimension due to the higher number of cars.

10. What is (are) the world model(s) (Club of Rome)? How many target variables did the world3 model have? What did the authors try to show with the model? What were their main conclusion from the models results? Was prediction of exact values a goal?

The world model referred to as the "Club of Rome" is a system dynamics model known as World3. It was developed in the early 1970s by a team of researchers at the Massachusetts Institute of Technology (MIT). The model aimed to simulate the interactions between various components of the global system, including population, resources, pollution, and industrial production.

The World3 model had several target variables that were included in its simulations. The main variables included population, industrial output, pollution, food production, non-renewable resource depletion, and quality of life indicators.

The authors of the World3 model sought to explore the long-term behavior of the global system under different scenarios and policies. They aimed to understand the potential consequences of exponential population growth, resource depletion, and pollution on human well-being and sustainability.

The main conclusions drawn from the model's results were that if current trends in population growth, resource consumption, and pollution continued unchecked, the world would face significant challenges. The model projected that the exponential growth of population and industrialization would lead to overshoot and collapse, potentially resulting in resource scarcity, pollution-induced ecological damage, and a decline in quality of life.

It's important to note that the World3 model was not designed to predict exact values or specific events. Instead, it aimed to demonstrate the potential dynamics and trends of the global system under different assumptions and scenarios. The primary goal was to raise awareness about the interconnectedness of various factors and the potential consequences of unsustainable growth patterns.

Question Group MCS (Microsimulation, Cellular Automata and Social Networks Questions)

11. In the lecture we discussed different methodologies/approaches to (econometric) microsimulation. Name and shortly describe these different methods/approaches. Shortly explain / give an example for which kind of analysis/type of question each of the four methods is most relevant. Which of those approaches do you think is nowadays the most relevant?

- **Static microsimulation**: Is mostly applied to short-term predictions such as the changes in tax policy. Static microsimulation models are typically straightforward models that employ static aging techniques, changing specific variables on the original microdata file, which is a file that contains data on the traits of population units, such as individuals, households, or businesses, collected by a survey, to produce a file with the demographic and economic traits anticipated in the following year. A dramatic increase in the tax of Gasoline would make a good example, as people might not use their cars anymore, instead they switch to different transportation methods, thus paying less taxes.

- **Dynamic microsimulation**: Is usually applied in long-term prediction. A more thorough and realistic population approach may be found in dynamic microsimulation models. They use a given probability to determine whether or not each individual in the microdata file will get married, divorced, have a kid, leave school early, find employment, experience unemployment, retire, or pass away as they individually age each individual from year to year. In most cases it is used to inspect the effects in demographic developments and the impact on social security expenditures.

- **Longitudinal microsimulation**: Similar to the dynamic simulation, but with a main difference that this model works on a microdata file that includes only one age group, and then ages the microunits so that an entire life cycle of this cohort is imitated. This approach consists of ignoring the age distribution of a population and concentrating solely on simulating a certain age group, along with their descendants if necessary. Longitudinal microsimulation can be used in social security and retirement policy, where it could determine how policy changes would affect people's lifetime earnings, savings levels, and retirement choices.

Every approach mentioned is extensively used and relevant nowadays, with regard to the research question and the available data.

12. Which kind of elements does a microsimulation consist of? How are the transition probabilities obtained and what kind of tools and data are required for that? How are microsimulation models initialized? Where does that data come from? Btw. What is the difference between projections and forecasts?

Elements:

- Individuals (Households, vehicles, firms, etc.)
  - Each unit is a record containing a unique identifier and a set of associated attributes.
  - E.g. People’s main attributes: age, sex, marital status, employment status, income, tax payments, etc.
- Rules aka transition probabilities
  - Deterministic rules (p = 1, e.g., tax brackets)
  - Stochastic rules (p < 1, e.g., chance of dying)

Transition probabilities depict the probability of moving from one state into another. They are obtained from the analysis of historical, empirical and longitudinal data. Since microsimulations often use available detailed information regarding the state of people and their families, one tool that we would need is a data collection tool, statistical analysis tools. In the 1970s and early 1980s microsimulation models were run on large mainframe machines, which would make it very hard for it to be accessible to many specialists. However, we can mention that nowadays, it is easier to run large microsimulation models, so probably all you would need is a workstation or a personal computer, with user-friendly software.

Initialization based on samples empirical data.

- Take real world sample.
- Calculate distribution.

Data: Empirical Data

- Often only representative samples of data
  - sample of full records
  - micro-census data
- Sometimes countries’ complete fiscal and social security data used.
  - Immense amounts of empirical data – calculation time
  - Data security issues

Projections and forecast: They seem similar at a first glance, but they have different meanings. Projections are usually based on assumptions regarding future trends and do not take into account the changes that might occur in any policy, or even other intrusions. Lets take for example, projecting future population growth, which could assume that there will be no changes of current trends in the future. On the other hand, forecasts take these changes, that might impact the modeling of the system, into account. Taking the same example with population growth, forecasts consider changes such as birth rate statuses

13. What characterizes cellular automata models? What does a cellular automata model consist of (describe)? Give an example of how a simple cellular automata model might work (rules) – think of a simple 1- or 2-dimensional model.

Some elements that characterize cellular automata are the cells and the states they can be in, the grid they "live on", the rules that change their states, etc.

A cellular automaton consists of a regular grid of cells that have a finite number of states. For each cell, a set of cells called its neighborhood is defined relative to the specified cell. An initial state (time t = 0) is selected by assigning a state for each cell. A new generation is created (advancing t by 1), according to some fixed rules (generally, a mathematical function). This determines the new state of each cell in terms of the current state of the cell and the states of the cells in its neighborhood.

One example is Conway's Game of Life, is a zero-player game, that is, the state of the system depends exclusively on its initial state. It takes place on an infinite two-dimensional grid, in which cells can be alive or dead. Each cell interacts with its eight neighbors. The time is simulated discreetly, and cells are updated once per time step. There are some cell's rules:

- Under population: every cell with less than two live neighbors dies.
- Overpopulation: every cell with more than three live neighbors dies.
- Reproduction: every dead cell with three alive neighbors becomes alive
- Any live cell with two or three alive neighbors lives on to the next generation.

14. Are cellular automata useful to model social systems/processes - think of the Rumor Mill and the Voting models? What are CA models good at and is that still relevant? What are the limitations of cellular automata models?

CA can be used to model social processes if we consider the Rumor Mill and the Voting models. The Rumor Mill is a CA that model the spread of a rumor in a "world" and the Voting model shows how cells can change their voting depending on their neighbors and various other rules.

CA are good at investigating the outcomes at the macro scale of millions of simple microscale events, and it can be calculated very fast due to simple structure.

The limitations depend on the CA variant:

- In the more classic variants that have only 1 state, all the cells are identical with passive interaction, the problem is the extreme variance of results with small changes in the initial configuration, rule, or timing.
- But in all variants the spatial arrangement is always a central issue

15. How are networks represented mathematically, which elements does that representation consist of? Describe the different characteristics that edges in a network may have (add sketches) and give examples for networks where the different characteristics matter!

In computational social simulation, networks are commonly represented mathematically using graph theory. A network, also known as a graph, consists of two main elements: nodes and edges.

1. Nodes: Nodes represent the entities or individuals in the network. They can represent people, organizations, or any other relevant units in the simulated social system.
2. Edges: Edges are the connections or relationships between nodes. They represent the interactions, associations, or dependencies etc.

Different characteristics that edges in a network may have:

1. Binary Edges: Binary edges are the simplest form of connections, where an edge exists or doesn't exist between two nodes. They indicate the presence or absence of a relationship. Binary edges can be represented as either solid lines for existing connections or as no lines for non-existent connections.
   For example, in a friendship network, a binary edge may represent a friendship between two individuals.

   Sketch:

   Node A ----- Node B

2. Weighted Edges: Weighted edges assign a numerical value to each connection, indicating the strength, intensity, or weight of the relationship between nodes. These values can represent various attributes, such as the frequency of interaction, trust, or influence.
   For example, in a communication network, a weighted edge may represent the frequency of communication between two individuals.

   Sketch:

   Node A ---(0.7)-- Node B

3. Directed Edges: Directed edges denote a one-way relationship between nodes. They indicate the presence of a connection from one node to another, but not necessarily vice versa. Directed edges are commonly used when the relationship has an inherent directionality.
   For example, in a citation network, a directed edge may represent a citation from one academic paper to another.

   Sketch:

   Node A --> Node B

4. Signed Edges: Signed edges capture the positive or negative nature of relationships. They indicate whether the connection between nodes is favorable or unfavorable, cooperative or antagonistic.
   In a trust network, a signed edge may represent the level of trust (+1 for trust, -1 for distrust) between individuals.

   Sketch:

   Node A ----[+]---- Node B

16. Explain the following network measures and how they are calculated (formulas or concise explanations): Node Degree, Characteristic Path Length, Clustering Coefficient!

<u>Node degree:</u>

It represents the number of edges connected to a given node. The node degree is often used to assess the importance or centrality of a node within the network. There are two types of node degree measures:

1. the degree of a node in an undirected network
2. and the in-degree and out-degree of a node in a directed network.

Degree of a Node in an Undirected Network:

In an undirected network, the degree of a node is simply the count of edges that are directly connected to that node. It indicates the number of immediate neighbors a node has. Mathematically, the degree of a node "v" in an undirected network "G" is calculated as follows:

Degree(v) = Number of edges incident to node v, for example, if a node has three edges connected to it, its degree would be 3.

In-degree and Out-degree of a Node in a Directed Network:

In a directed network, nodes have both incoming edges (in-degree) and outgoing edges (out-degree). The in-degree of a node refers to the number of edges that are directed towards that node, indicating the number of nodes pointing to it. The out-degree of a node represents the number of edges that originate from that node, indicating the number of nodes it points to. Mathematically, the in-degree and out-degree of a node "v" in a directed network "G" are calculated as follows:

In-degree(v) = Number of edges with v as the destination

Out-degree(v) = Number of edges with v as the source

For example, if a node has an in-degree of 2 and an out-degree of 3, it means that two nodes are pointing towards it, and it is pointing towards three other nodes.

<u>Characteristic Path Length L:</u>

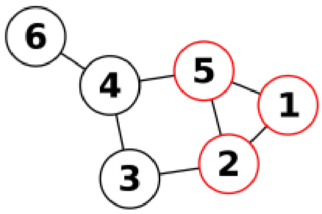
- It calculates the shortest path (number of edges) from each node to every other node.
- L is the average of all these shortest paths

$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j)$$

<u>Clustering Coefficient:</u>

Is a measure of the degree to which nodes in a graph tend to cluster together.

A clique is a subset of nodes such that every two distinct nodes in the clique are connected to each other.



The red nodes are a clique. If a node has kv neighbors, the local clustering coefficient is:

ci = 2 Nv / kv (kv - 1)

where:

- Nv = actual number of links that exist between neighbors of v
- kv (kv - 1) / 2 = number of links that could exist between neighbors of v

and the global clustering coefficient is the average of all ci, this is number of closed triplets/numbers of all triplets.

17. What are strong ties and weak ties in social networks? To whom would a person have strong ties and to whom weak ties? Why can weak ties still be important – explain and give examples! What does this have to do with network structures / small worlds?

Strong ties are close family and close friends, they have a strong influence on decisions, common behavioral norms and tend to move in common circles and have high overlaps of information.

Weak ties refer to acquaintances, remote family, and colleagues, they tend to move in other circles, have access to “new” information and build bridges to other parts of the wider social network (society).

These weak ties are crucial in binding groups of strong ties together. They bring circles of networks into contact with each other, strengthening relationships and forming new bonds between existing relationship circles.

18. What is the small worlds phenomenon (what does it describe)? Describe the initial experiment that investigated the phenomenon! How is it connected to power-laws? Which kinds of networks are likely to show the small worlds phenomenon? How can a small worlds network form?

The small world phenomenon was found out while examining the average path length for social networks of people in the United States. The research was groundbreaking in that it suggested that human society is a small-world-type network characterized by short path-lengths, it is normally associated with the phrase:" six degrees of separation" (idea that all people are six or fewer social connections away from each other).

Milgram's experiment developed out of a desire to learn more about the probability that two randomly selected people would know each other. Imagine the population as a social network and attempt to find the average path length between any two nodes. Milgram's experiment was designed to measure these path lengths by developing a procedure to count the number of ties between any two people.

Social networks are an example of the small world phenomenon. The small world network initially begins with some people that are socially very active, as previous "social success" is an attractor. Eventually, those with many friends get more, and form hubs. They also retain long-term connections.

The distribution of links to nodes in a small-world follows a power-law. The power law means that the vast majority of nodes have very few connections, while a few important nodes (hubs) have a huge number of connections.

Question Group ABM (Agent Based Modelling Questions)

19. What are the main modeling elements of an agent-based model (the rough common structure) and what characterizes an agent-based model? Give examples of these elements by reference to a specific ABM! How does an agent-based model differ from the other simulation models/methods that we discussed?

Agent based modeling (ABM) is a modeling concept, where individual agents interact with each other in some kind of environment.

Main features:

- Agents: represent social actors, have individual attributes like age, and interact with each other. E.g.: The birds in the Flocking ABM.
- Environment: The agents all exist in an environment which has some properties (e.g., grid, geographical map) and can interact with agents (e.g., taxes). E.g.: The sky or space where the birds fly.
- Interactions: The agents have some behavior and interact with each other and the environment. E.g.: birds observe the movements of other birds and form flocks.
- Learning: Agents can learn from experience. E.g.: The patterns that are formed from a flock of birds
- Ontological Correspondence: The agents belong to a certain category, which also exists in the real world. E.g.: a flock

Differences to other approaches:

Cellular Automata (CA) only allows for 1 attribute per cell, have a restrictive environment (e.g., cells can only be on a grid), and the cells adhere to global rules to determine the next state of the model. In ABM the individual agents can have multiple attributes, are able to interact with each other and the environment, and thus can even influence the rules of the environment and behavior of other agents.

Microsimulations offer some heterogeneity between individual actors, similarly to ABM. But the rules are fixed globally. ABM allows for more complex behavior, and the behavior of the agents might even influence the rules and vice versa. ABM also offers the individuals the opportunity to learn from experiences.

In System Dynamics (SD) there are no individual actors, in comparison to ABM. One important aspect of SD is that the model can influence itself through feedback. However, this is limited, and the individual elements cannot adapt to changes. ABM is very similar, but more powerful, since all the components of ABM (actors, environment, …) can influence each other and thus, the rules and behaviors of the simulation can change over time.

20. What are agents in the context of ABM models? Further, describe the five main features that agents often have in an ABM model!

Agents represent social actors, have individual attributes like age, and interact with each other and the environment. E.g.: a person, household, company, country.

<u>The five main feature that agents have:</u>

1. Ontological Correspondence
    a. Meaningful representation of real individuals
2. Autonomy
    a. Decide on their own, based on:
        i. Individual situation: sex, age, wealth, health, ...
        ii. Locality effects: different interaction partners, area, position, ...
        iii. Local rules – behave different already due to above, but:
            1. May learn to choose from rules from a common set (weights)
            2. Or learn / devise individual behavioral rules (uncommon)
3. Social Ability
    a. Can interact with other agents, exchange information, observe, talk, etc.
4. Reactivity
    a. Can react to external stimuli.
        i. On other agents' behavior – imitate, learn, protect, defend…
        ii. On the state of / changes in the environment – time, resources…
5. Proactivity
    a. Can act proactively, i.e., plan in advance.

21. What is part of the environment (the modelling element) in an ABM model? Furthermore, which different spatial topologies can be included in an ABM and for which kind of research question might each of them be relevant?(CHATGPT, questions only)

The environment in an ABM model typically consists of the following components:

1. Space: Space represents the physical or conceptual area in which agents move and interact. It can be two-dimensional, three-dimensional, or even abstract.

Q:

- How does the spatial arrangement of agents affect the emergence of social patterns?
- What impact does the physical distance between agents have on their interaction behavior?
- How does the availability of different spatial locations influence resource distribution among agents?

2. Resources: Resources refer to the entities or elements that agents can acquire, utilize, or compete for within the environment. Resources can be tangible (e.g., food, money, energy) or intangible (e.g., information, reputation, social capital).

Q:

- How do agents compete for limited resources within the environment?
- What strategies do agents employ to maximize their resource acquisition?
- How does the distribution of resources affect the social dynamics and outcomes of the system?

3. Constraints: Constraints represent the limitations or rules that agents must adhere to when navigating and interacting within the environment.

Q:

- How do different constraints imposed on agents shape their decision-making and behavior?
- What are the effects of specific rules or limitations on the emergence of collective behavior?

- How do agents adapt their strategies when facing different types of constraints?

Spatial topologies − 1D, 2D, 3D; Line, Board (Grid), Torus, Sphere, − GIS data, real maps, simplified maps etc.

22. Describe Schelling's Segregation Model, what it tries to explain and how (elements, rules, etc.). What does the model predict/explain? What did Schelling use to simulate it originally?

Schelling's Segregation Model aims to explain how different ethnicities of people often group together in US cities. This grouping phenomenon is also called segregation.

The model consists of a grid, on which people of 2 ethnicities are placed on. These individuals want a certain share of their neighbors to have the same ethnicity as them. The share is fixed and called the tolerance value (or threshold). If the individuals are not satisfied, they move to an empty space on the grid.

The model found that even if individuals only want a very small share of their neighbors to have the same ethnicity, this is enough to lead to segregation.

Originally Shelling simulated the model with a checkerboard as the grid, and with zinc and copper coins, for the two ethnicities of people.

23. Describe the simple bird-flocking model! What does this simple model show (i.e. why is it interesting for our purposes)? What are the behavior rules (micro-level behavior) of the birds and to which behavior do they lead (which macro-level regularities)?

It shows how complex behavior (self-organization) can emerge from simple rules.

3 simple rules of the individual birds:

- Alignment: birds adapt direction to the nearest birds
- Separation: keep minimum distance
- Coherence: birds will move towards other nearby birds

It leads to flocks emerging and complex patterns of the swarms on the macro level.

24. Describe the demographic prisoner's dilemma (DPD) from the second assignment! Describe the rules of the game and the expected dynamics – especially the differences to the regular prisoners' dilemma (non-iterated, non-demographic).

In the demography, there are two types of individuals.

- cooperators who cooperate.
- defectors who act selfishly.

The individuals interact with each other according to the classical prisoner's dilemma. Over time, they can gain or lose wealth according to their interactions (prisoner dilemma matrix). If they gain enough wealth they can reproduce and if they lose too much wealth they can die. Prisoners have a fixed role (defect or cooperate) and an initial position on the map.

In the classical prisoner's dilemma, there are just two individuals who try to find an optimal solution for minimizing their jail time. After the two individuals make their separate choice of either cooperating or defecting, there is an outcome state, and the game is over. The demographic prisoner's dilemma consists of multiple individuals and goes over multiple rounds (iteration), and the goal of the individuals is wealth maximization.

Question Group EXP (Simulation Experiment Questions)

25. What is the difference between conducting a proper simulation experiment and exploring a simulation model? What is an experimental setup in the context of simulations / what does it specify? What are target variables in this context? (CHATGPT)

- Conducting a proper simulation experiment:
    - Involves designing and executing a controlled study.
    - Tests specific hypotheses or research questions.

  - Follows the scientific method.
  - Manipulates variables, collects data, and analyzes results.
  - Aims to gain insights into the behavior of the social system being simulated.
- Exploring a simulation model:
  - Involves interacting with the model and making observations.
  - Aims to gain a deeper understanding of its dynamics and emergent properties.
  - Can be more open-ended and aimed at discovering patterns or phenomena.
  - Often used to generate hypotheses or inform future simulation experiments.

Experimental setup in the context of simulations:

- Refers to the configuration and conditions of the simulation experiment.
- Specifies the:
  - Initial state of the simulation.
  - Values of manipulated variables.
  - Constraints or parameters.
  - Methodology for data collection.
- Ensures consistency and validity of the simulation results.

Target variables:

- The variables that researchers are specifically interested in.
- Chosen based on the research question or hypothesis being investigated.
- Represent the aspects of the social system or phenomenon under study.
- Are the focus of data collection and analysis during the simulation experiment.
- Aim to understand, predict, or explain the behavior of the social system.

26. What are experimental designs with regard to computer simulations? What kind of designs did we mention in the lecture? What is a rerun and what does an experimental design have to do with the number of reruns of an experiment? Which experimental design does a basic Netlogo Behaviorspace experiment use?

Experimental designs refer to the structured approaches used to conduct experiments and investigations using computer-based models.

<u>Experimental Designs:</u>

- Factorial Design:
  - Full factorial design: we take every combination of every meaningful parameter combination – huge number of simulations runs. It is the default approach.
  - Fractional factorial design: we take a specific partition of the multidimensional parameter space
- Random Design:
  - Latin Hypercube design: Random but meaningful combinations of parameters
  - Many other alternative methods

<u>Rerun:</u>

A rerun refers to running a simulation model multiple times using the same experimental conditions and settings. The number of reruns of an experiment is determined by the desired level of statistical accuracy, robustness, and stability of the results. By conducting multiple reruns, researchers can assess the variability in the simulation outputs and obtain more reliable and robust estimates. We should have reruns for every parameter estimation to even out the effects of stochastic components.

<u>Behaviourspace</u> = built-in Netlogo tool for unattended simulation runs

- Uses only full factorial designs (disadvantage: huge number of runs)
- Runs a model multiple times with a different model’s settings and records the results of each test run.

27. What is the difference between model verification and validation regarding agent-based models? Describe both shortly, then explain techniques for validation – which techniques are needed for which type of agent-based model (three types were mentioned with respect to scale)!

Verification & Validation:

Model Verification is used for checking if the model was programmed correctly and does not contain any bugs.

Model Validation is used to check if the "right" model was created. In other words, we want to check if our model even describes the social phenomenon we want to study, or if we created a model that unintentionally describes something different.

Techniques for model validation:

1. Sensitivity Analysis: Is used to determine how good the model and theory fit together. This is done by testing the model with different parameters and analyzing the output. Due to the randomness of ABM the model should be run multiple times with the same settings. Since there often are also many parameters and possible values for these parameters, often a huge number of runs must be done. Using knowledge about the parameter values like a practical minimum value, can be used to reduce the runs. Additionally, it might be beneficial to randomly sample the parameter values to further reduce the amount of runs.
2. Empirical Validation: Another approach is to compare the model with empirical data. This might not work for every model, since sometimes empirical data is just not available, especially when doing predictions.

Which validation fits to which model:

- Abstract Models (the very general bois):
    - check if the general phenomenon makes sense.
- Middle Range Models (the in between bois):
    - empirically check if the overall results of the model like the phenomenon we want to observe (qualitative similarity)
- Facsimile Models (the very detailed bois:
    - empirically check if the values of the model are like the values we expect in the real world. (quantitative similarities)

Types of Models:

- Abstract Models:
    - Are used to demonstrate some general social processes with the goal of getting overall patterns from some simple behavioral rules. The results can then be used to further develop theories and maybe create more specific models.
    - An example would be Schelling's segregation model, which shows how ethnic segregation in US cities could emerge, based on some simple rules. The model uses very simple rules and is not specific to any US city.
- Middle Range Models:
    - Try to describe the characteristics of certain social phenomena and are more specific than Abstract Models. They should still offer some generality, so that the insights can for example be applied to not only 1 industrial district in a certain city, but generally to all industrial districts.
    - An example would be the modeling flow of knowledge in "innovation networks". In this model companies trade with each other, and knowledge can be gained through internal research or by using knowledge obtained from other companies. With better knowledge these companies can make better products. Analyzing the number of partners, a company has in such a model shows that the number follows the power law, a very common empirical phenomenon seen in social networks.
- Facsimile Models:
    - Are very detailed models that aim to describe a quite specific social phenomenon as exactly as possible. Since they are so detailed, they can be used for reproducing historic developments and for predicting future developments. However, the real world still has some element of randomness, so the results should not be seen as an absolute truth.

- An example would be a model of the inventory of a business, with the goal of evaluating the consequences of restocking their items a bit later or earlier.

28. What are the reasons for including stochastic elements included in simulations – especially agent-based models – i.e. which problems shall they counter? On the other hand, which challenges / problems arise from the stochastic influences in the model? What is necessary to alleviate these problems again?

Including stochastic elements in simulations, particularly in agent-based models, serves several important purposes and addresses specific problems:

1. Capturing Uncertainty: Stochastic elements allow for the representation of uncertainty and variability in real-world systems. Many social, economic, and behavioral phenomena exhibit randomness and unpredictability. By incorporating stochastic elements, simulations can reflect the uncertainty in individual behaviors, interactions, and outcomes, resulting in more realistic representations.
2. Emergent Behavior: Stochastic elements contribute to the emergence of complex behavior and patterns in agent-based models. Small variations in agent behavior or interactions can lead to diverse and unpredictable outcomes at the system level.
3. Robustness Testing: Stochastic elements help in testing the robustness and sensitivity of a simulation model.

However, the inclusion of stochastic influences in models also presents challenges:

1. Reproducibility: Stochastic elements introduce randomness, which can lead to non-deterministic outputs.
2. Statistical Significance: Randomness can make it difficult to establish statistical significance and make definitive conclusions.

To alleviate the problems arising from stochastic influences, researchers can employ the following strategies:

1. Multiple Repetitions: Running the simulation multiple times, known as reruns, helps capture the variability and assess the statistical significance of results.
2. Sensitivity Analysis: helps to understand how robust the model is.
3. Statistical Techniques: such as hypothesis testing, can be applied to analyze and interpret stochastic simulation outputs.

29. Shortly describe sensitivity analysis! Why is sensitivity analysis needed? What are the main challenges of conducting a sensitivity analysis? What is the connection between experimental designs and sensitivity analysis?

Sensitivity Analysis is used to determine how good the model and theory fit together and how robust a model is. This is done by testing the model with different parameters and analyzing the output.

It is needed to validate and improve the model. It also helps with focusing on the most important parameters that affect the output.

<u>A big challenge is the number of runs:</u>

1. Due to the randomness of ABM the simulation must be run multiple times with the same settings.
2. This must be done for each parameter setting we want to test => huge number of runs

The connection between experimental designs and sensitivity analysis lies in the fact that experimental designs, such as factorial designs, can be used to conduct sensitivity analysis.

1-10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29