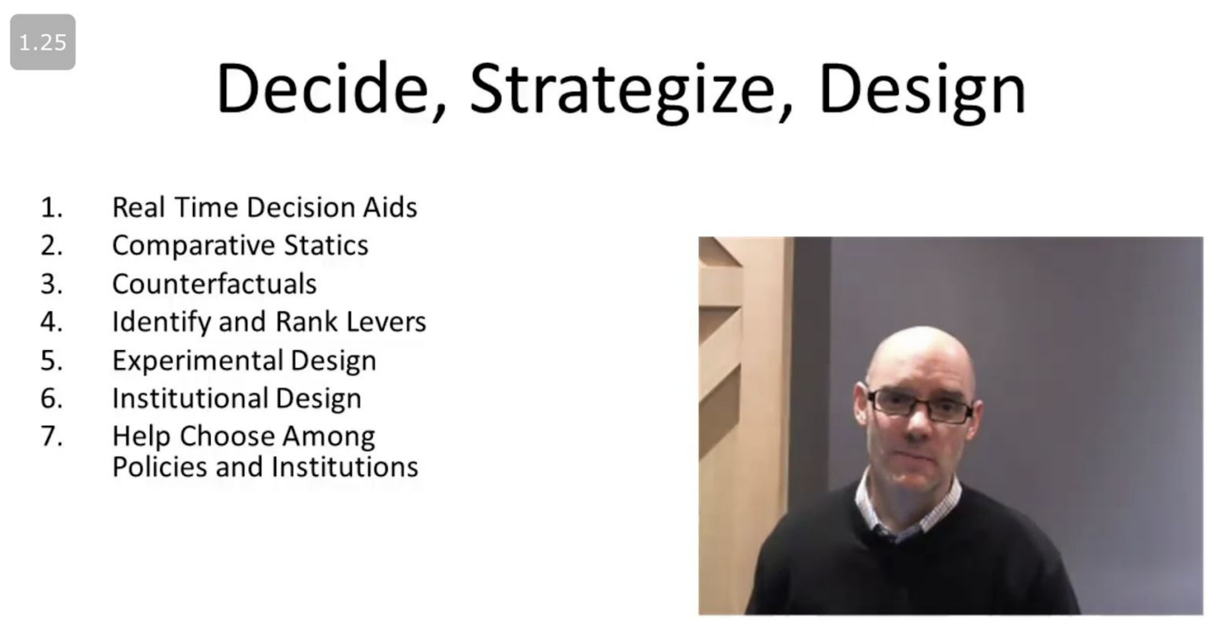
Model Thinking

Week 1

Why model?

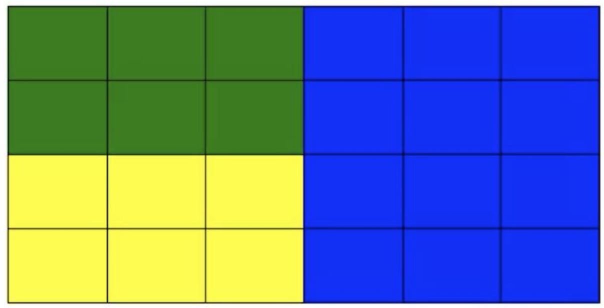
The professor makes a motivation speech about why models are better than we are, what you can gain from them and how we can use them to make better choices.

Models make us think logically creating heuristics or simply enabling things to be measured.



Segregation and peer effects

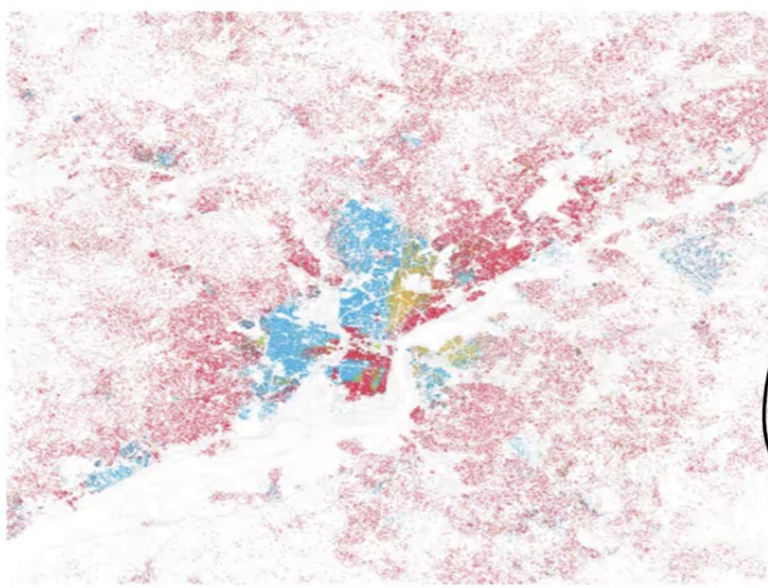
Schelling's segregation model tackles the propensity of people living in a neighbourhood based on who is currently living in the neighborhood, who just left the neighbourhood and so on.



72/90 .8

$$12 \times (1/15) + 6 \times (1/9) + 6 \times (1/45) \neq 72/45$$

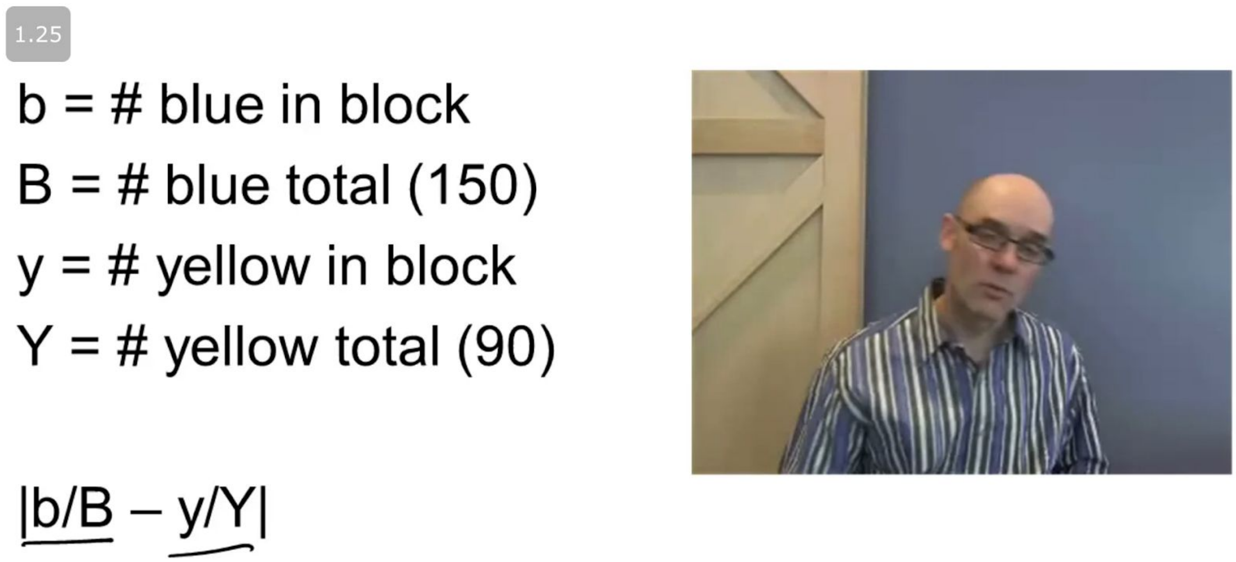
1.25



Phili

0.8

The professor introduces a formula (shown below) to measure segregation in a city made by blocks. If there are different types of blocks (blue, yellow, green), the measure is the weighted average of all these groups of blocks.



Peer effects can be described as how one person might influence the next, and creates a model designed to use thresholds. For example, the Granovetter model:

- if my threshold to buy a hat is 0, then I will buy the hat
- if my threshold to buy a hat is 1, then I will buy the hat if I see one person buying a hat
- if my threshold to buy a hat is 2, then I will buy the hat if I see at least two people buying a hat

Standing ovation models rely on a calculation over the quality and threshold of people to stand after a show is finished.

One formula can be given by: S (signal) = Q (quality) + E (error) > T (threshold)

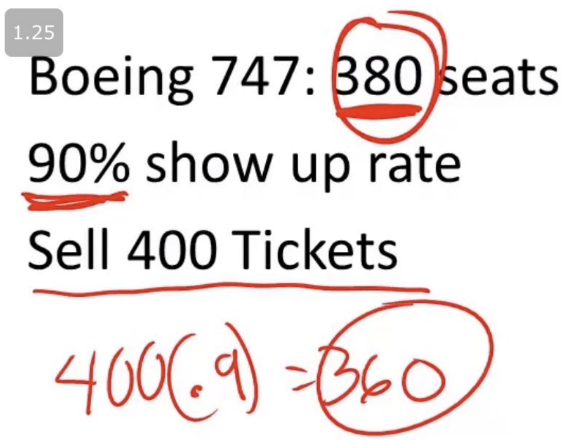
Peer effects vs sorting, are people changing or are people moving?

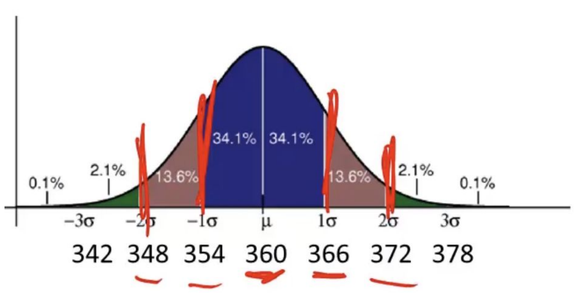
Week 2

Aggregation

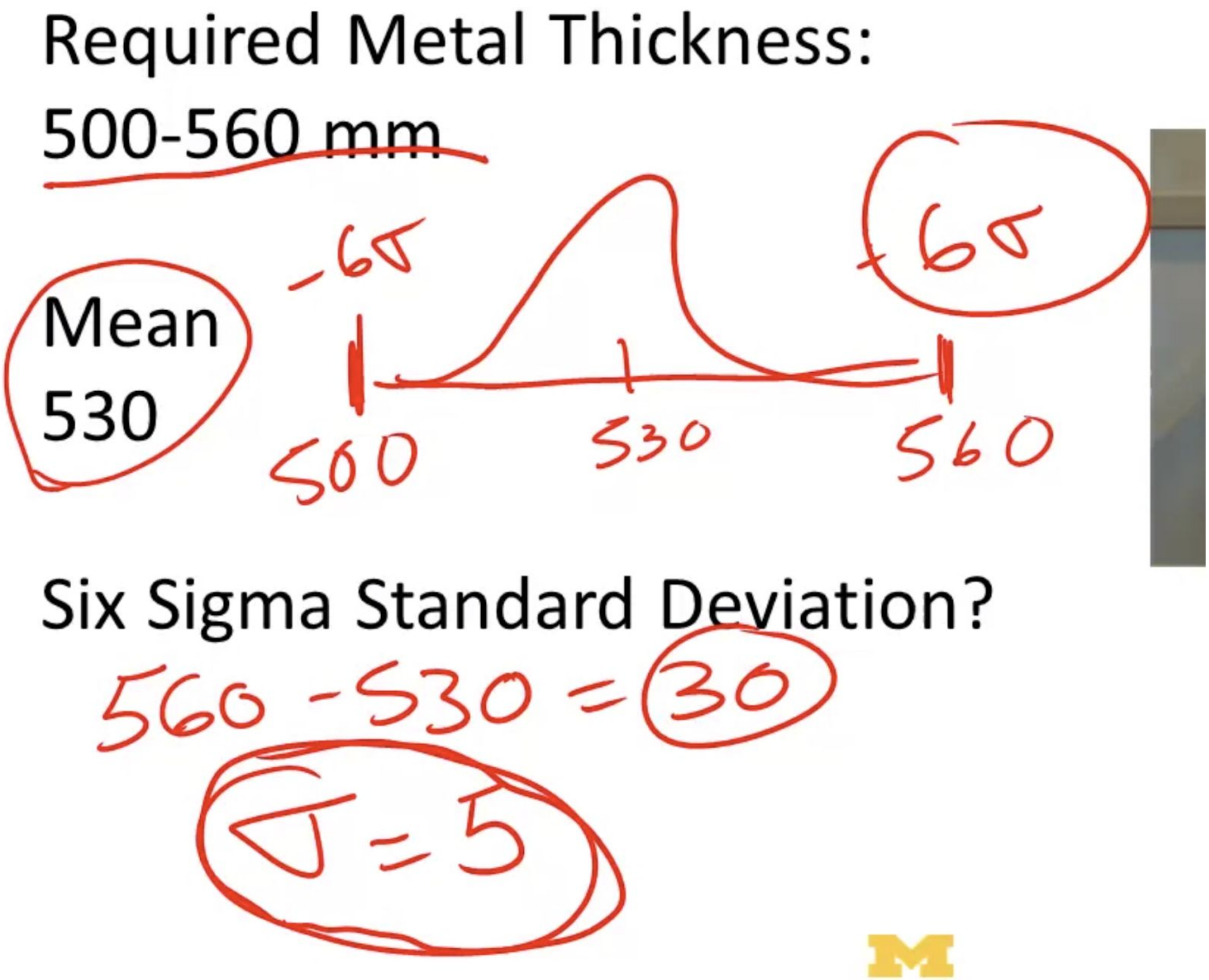
Central limit theorem states that if you add random variables such as they are:
- Independent
- Finite variance

Their sum presents a normal distribution.

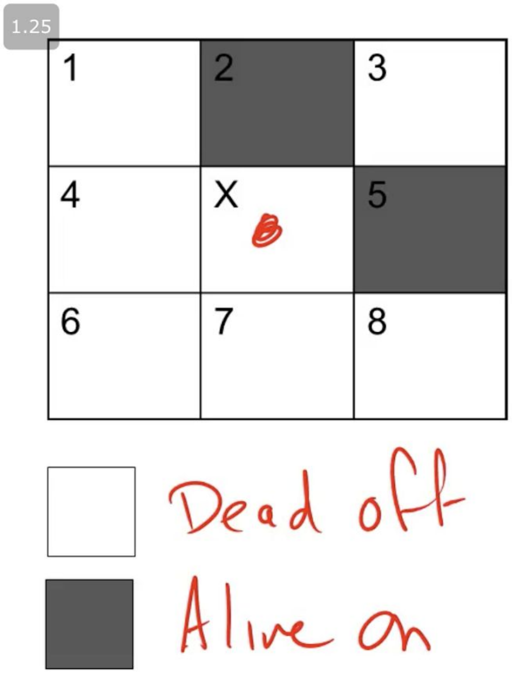


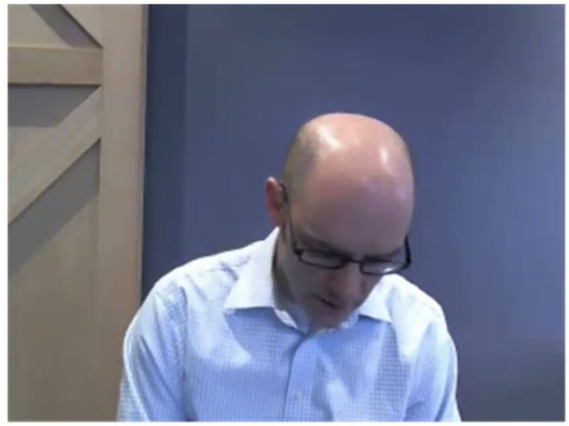


Six sigma is a standard industry method that defines if there are problems with events or products, they are within an interval of +/- 6 standard deviation. Example:



Game of life defines a set of rules that define how a system works generating living patterns.





1.25

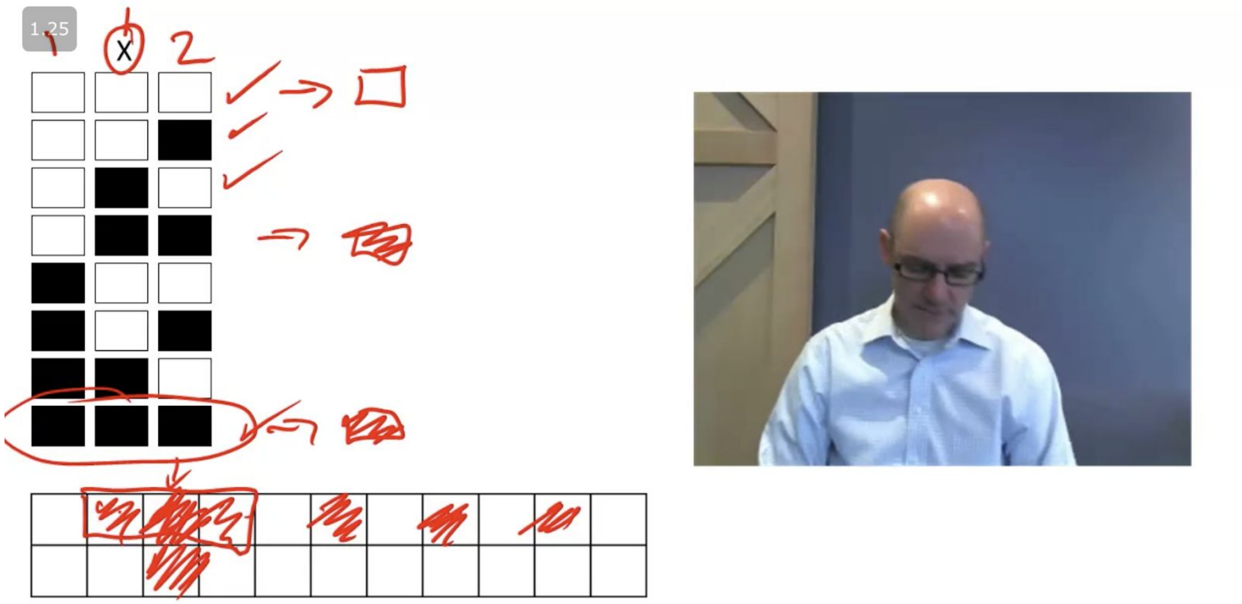
Rules

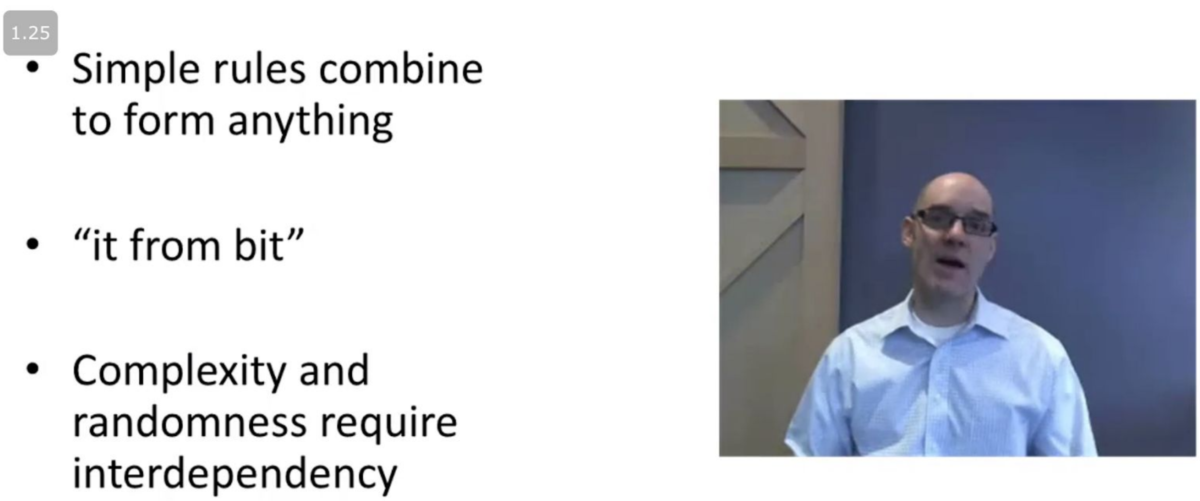
If **off** you turn **on** if
3 neighbors on

If **on** stay **on** if 2 or
3 neighbors on

Cellular automata is a model of computation and can be applied to physics, theoretical biology, and microstructure.

Given certain patterns, there is an initial set and it propagates from top to the bottom over the grid.

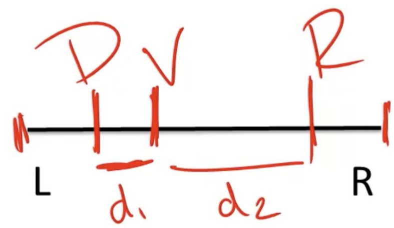




Decision models

Multi-criterion decision making is about taking decisions based on characteristics of the product/proposal. It can be based solely on the linear available parameters or it can take into account weights for certain parameters.

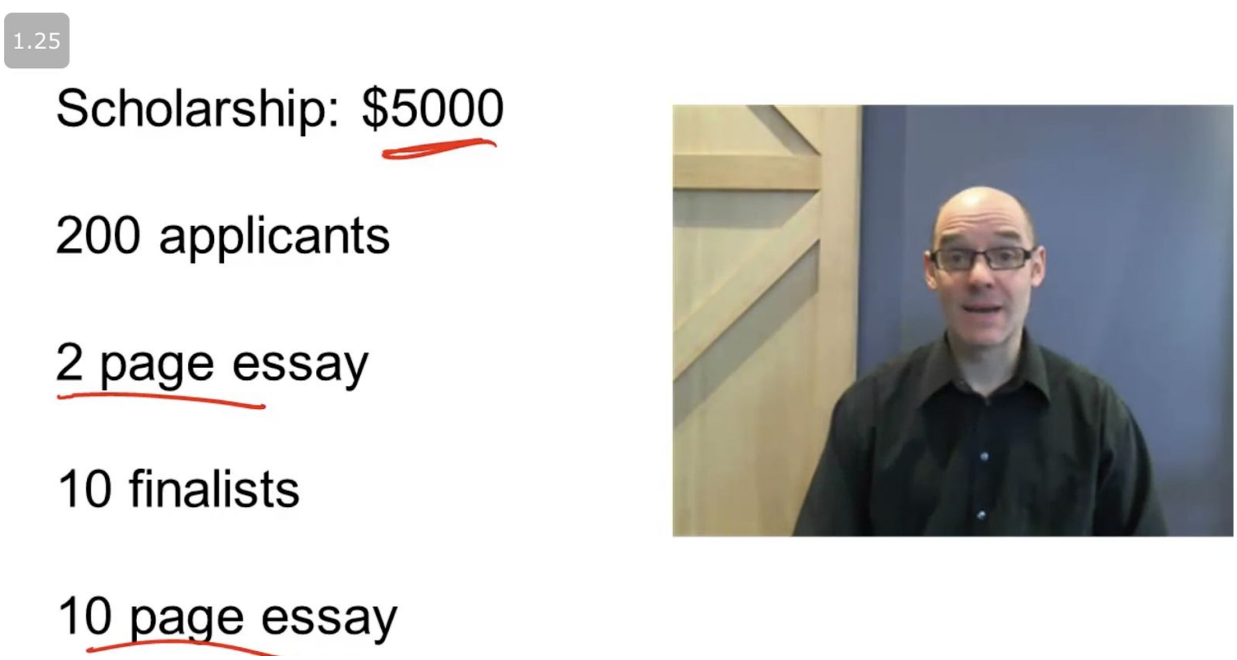
Spatial choice models assume that for certain characteristics (or variables) the weights are going to be different since people are different and can value more certain things. One good example is to calculate the distances of your candidates from "Liberals" and "Republicans".

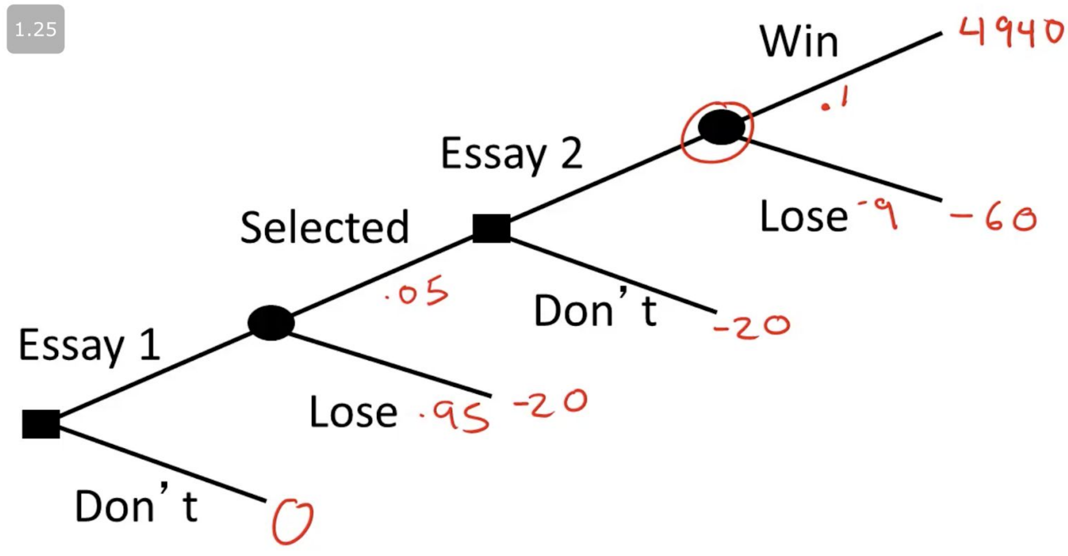


Probability uses three axioms as base rules:

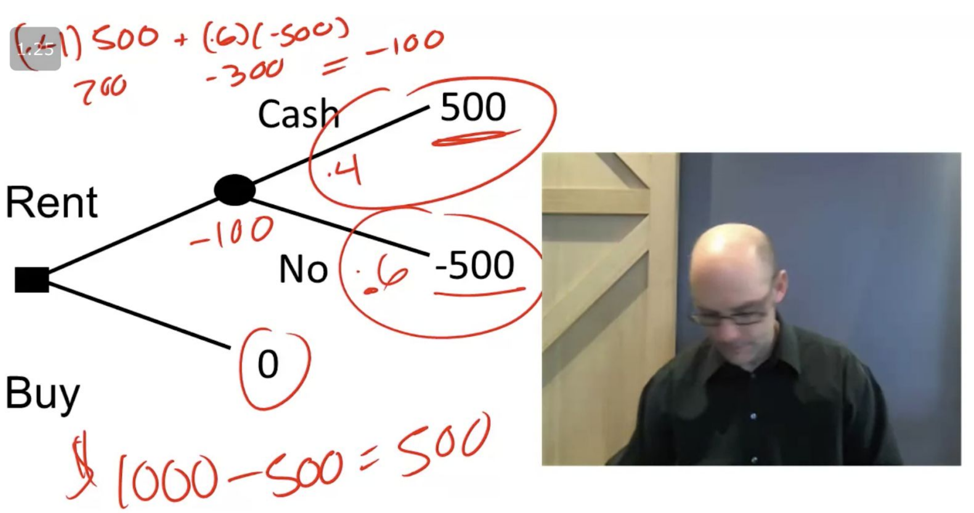
- Axiom 1: a probability of an outcome is always in the interval of [0, 1]
- Axiom 2: the sum of all possible outcomes should be 1
- Axiom 3: if A is subset of B, then P(A) <= P(B)

Decision trees can be used to support when making decisions based on probabilities and final values.





Value of information introduces the concept of calculation of what is the expected outcome of a certain group of rules.



Week 3

Thinking electrons: modeling people

Rational actor models take the assumptions that there is an objective and then the actor is going to optimize this objective function (in a rational way).

- Firms will maximize profits
- Individuals will maximize utility
- Candidates will maximize votes

Decision vs Game objectives vary

- Decision: objective depends only on own action
- Game: objective depends on actions of others

Behavior models take the assumptions that people might deviate from rational actions.

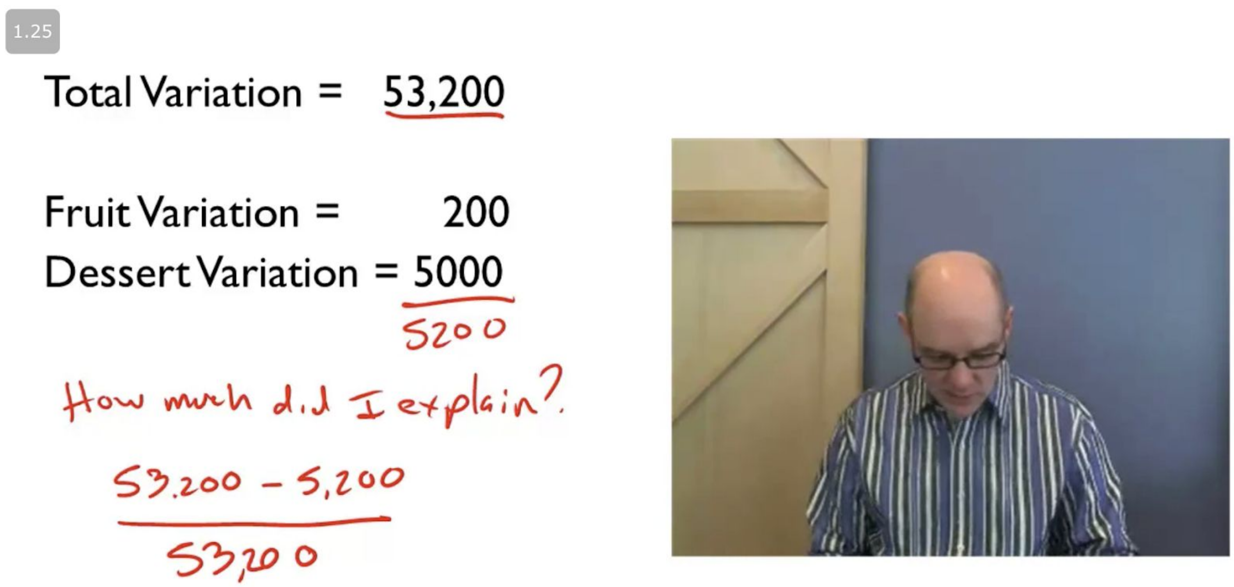
Examples:

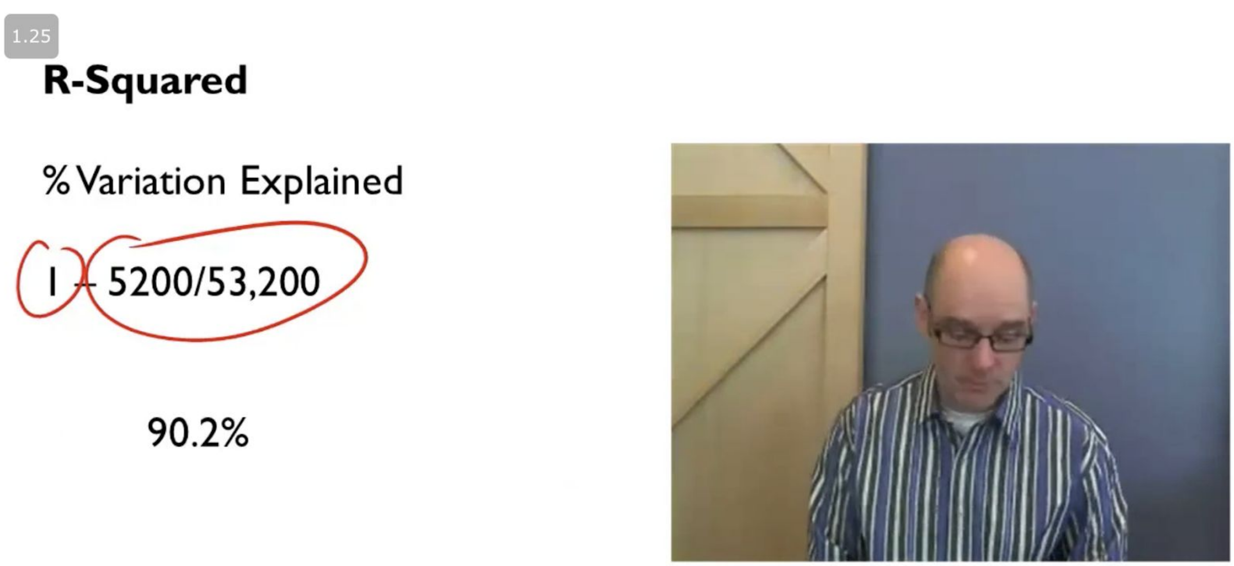
- Prospect theory: people are risk averse, so they might gamble in case of a certain loss
- Hyperbolic discounting: people tend to accept less gains/money when they are closer to the date of the event of lending
- Status quo bias: people tend to let things as they are (status quo) if there isn't confrontation (donate organs checkbox example)
- Base rate bias: when there isn't any baseline, people can use numbers/information uncorrelated

Categorical and linear models

Categorical models categorize the data in order to reduce the variation of a certain feature. Thus, the R-squared is calculated by the following formula:

R-squared = 1 - (model variation / total variation)





- R-squared equals to one means that the model explains a lot
- R-squared equals to zero means that the model explains little

Linear models assume that there is a linear relation between the independent variable (X) and the dependent variable (Y). The models are used to understand the data and to predict.

Example:

$$Cost = 15 * Length + 100$$

The **total variation** is calculated using the observed Y values, given by the formula:

$$(y1 - mean)^2 + (y2 - mean)^2 + \ldots$$

The **model variation** is calculated using the observed and the predicted Y values, given by the formula:

$$(y1 - y1\_hat)^2 + (y2 - y2\_hat)^2 + \ldots$$

The **R-squared** will be given by the formula:

$$1 - (model\ variation\ /\ total\ variation)$$

To try to find the **best fit**, you can solve the equation below.

$$Y = mX + b$$

$$(y1\_hat - y1)^2 + (y2\_hat - y2)^2 + \ldots$$

$$((m * x1 + b) - y1)^2 + ((m * x2 + b) - y2)^2 + \ldots$$

This will result in an equation, thus, it can be solved to find the m and b, getting the best fit for the model ($Y = mX + b$).

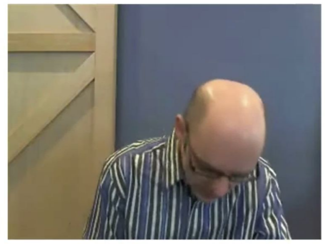
The **regression output** is shown below.

R-Squared represents how much the model can explain about the data
Standard Error shows the standard deviation observed in the data
Observations is the quantity of observations used for this model
Coeff is the coefficients for each independent variable and the intercept (constant)
SE is the standard error of each coefficient (plus and minus over the Coeff value)
P-value determines if a certain estimative can be trusted or not (if p-value is less than 0.05, we reject the null hypothesis that there's no difference between the means and conclude that a

significant difference does exist for the coefficient in question, else, we accept the null hypothesis)

R-Squared: 0.72
Standard Error: 24.21
Observations: 50

	Coeff	SE	P-value
Intercept	25	2	0.000
X1	20	1	0.000
X2	10	4	0.014



From linear to nonlinear models, there are several techniques, such as using splines that split the interval piecewise (for each piece there is an equation), or the use of non linear terms in the models (square root of x, x squared, log of x, and so on).

The big coefficient vs The new reality is the paradox to decide if there is a need for a total new approach instead of tweaking the coefficients of the problem. Examples:

Problem: traffic
Big coefficient: increase number of lanes
New reality: create a rail system

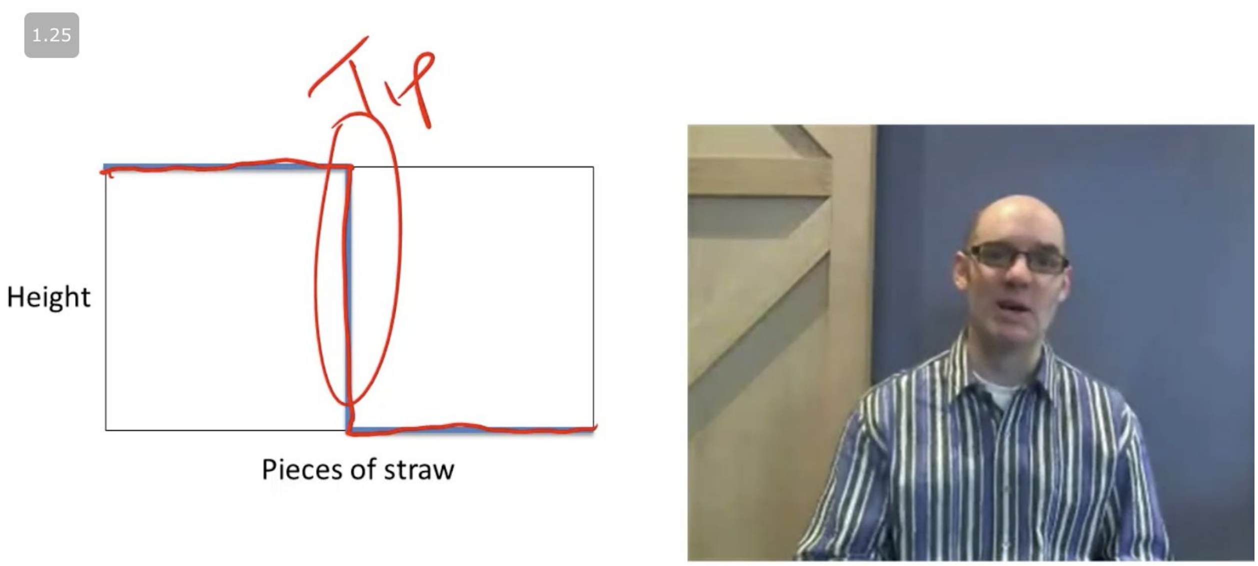
Problem: lung cancer
Big coefficient: tax cigarettes
New reality: universal health care

Week 4

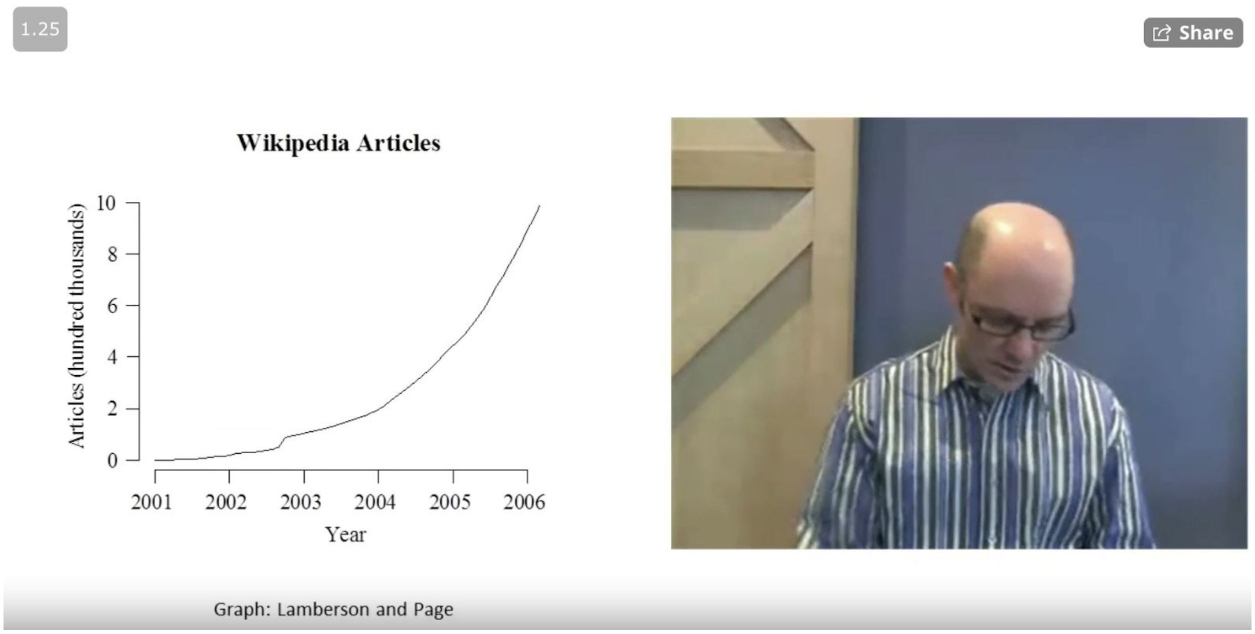
Tipping points

Tipping points are basically a small change that has a huge effect.

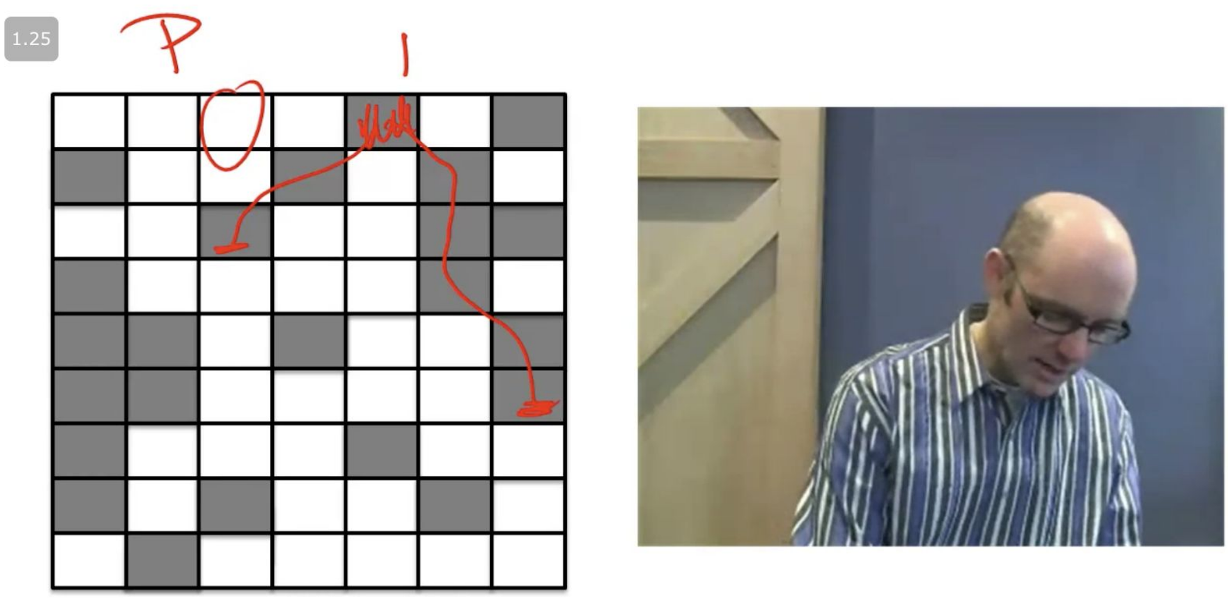
As an example (shown below), if you keep adding straw to a camel's back, at a certain point the camel will fall and that will be your tipping point. The height will fall to zero.



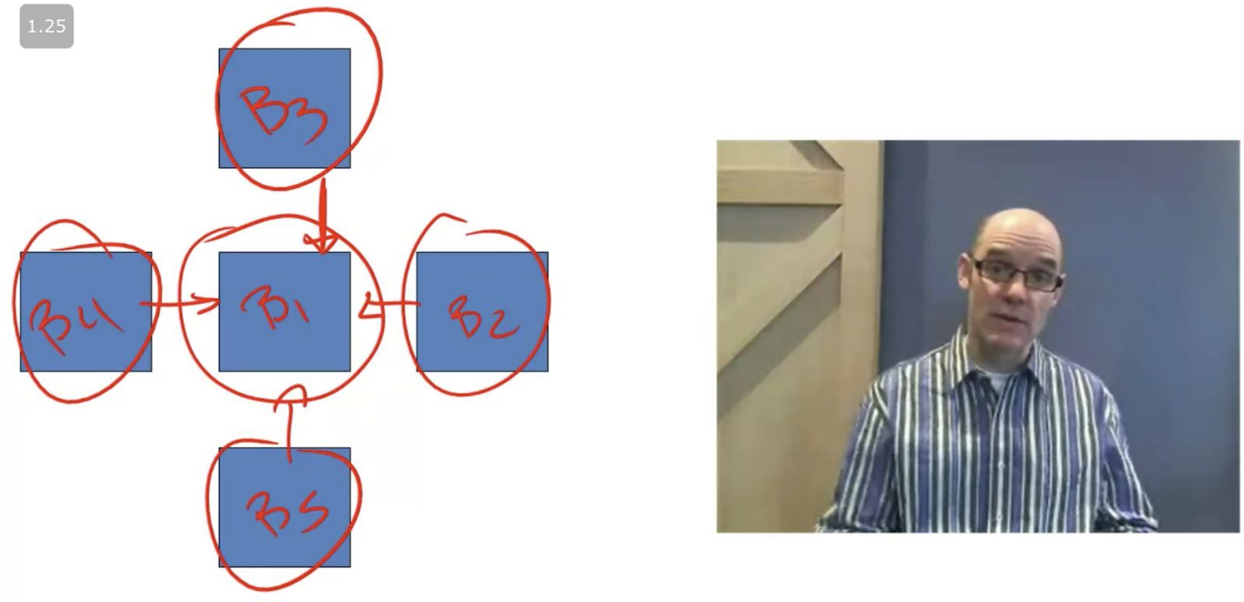
But it's important to notice that tipping point is classified as small changes leading to huge effects. So do not confuse that with **exponential growth** like the number of articles in Wikipedia (shown below).



The Percolation model comes from the physics field. It models if a certain state will percolate or not (shown below). Sometimes it can be that a system will not percolate until it reaches a certain threshold (the tipping point).

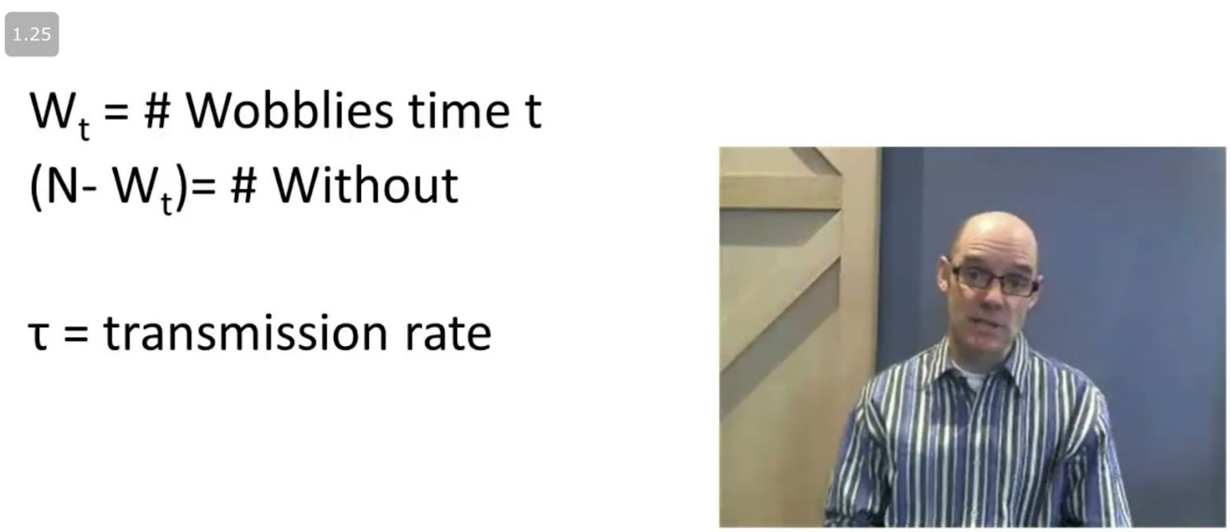


One example can be if banks lend money to one another, they are percolating the bankruptcy in case several banks lended money to a certain bank and this bank goes bankrupt (shown below).



Percolation also explains why suddenly several events of the same subject occur at the same time, or at a very similar period of time. Usually there are missing spots to solve a certain thing and when these spots are covered/discovered, the path is done and the percolation occurs.

Contagion models 1: diffusion models are given by the formula shown below.



1.25

Two people meet:

Probability one gives wobblies to other?

$$\tau\left(\frac{W}{N}\right)\left(\frac{N-W}{N}\right)$$

If c is the contagion rate, then we have the number of contaminants by Wobblies in t+1 as:

1.25

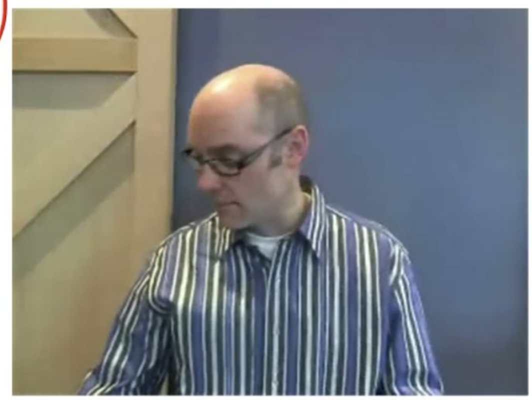
$$W_{t+1} = W_t + \underbrace{(Nc)\tau\frac{W_t}{N}\left(\frac{N-W_t}{N}\right)}_{\text{new}}$$

Contagion models 2: The SIS model comes from the epidemiology field. It's very similar to the diffusion models but it takes an extra parcel into account which is the recovery of people by the disease (formula shown below).

$$W_{t+1} = W_t + (Nc)\tau \frac{W_t}{N}\left(\frac{N-W_t}{N}\right) - aW_t$$

new

cured



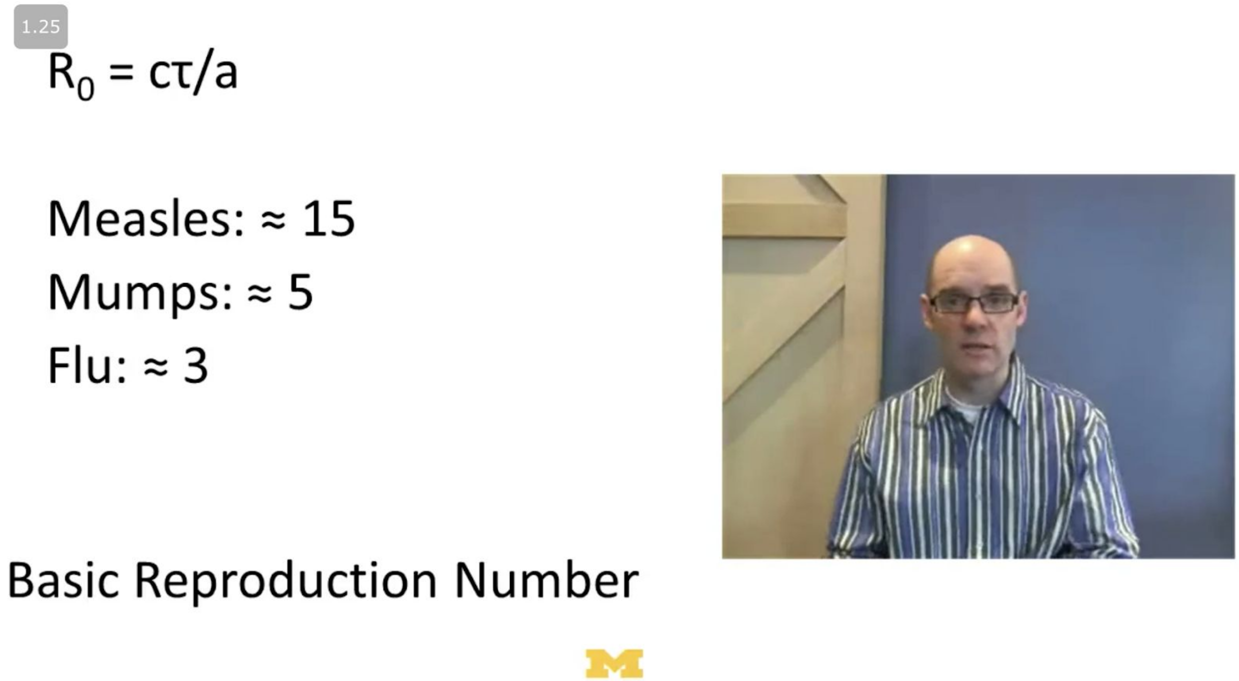
Where **a** is the recovery rate. Then it's possible to calculate if the disease is going to spread based on contagion rate (c), transmission rate (tau) and recovery rate (a).

$R_0 = c\tau/a$

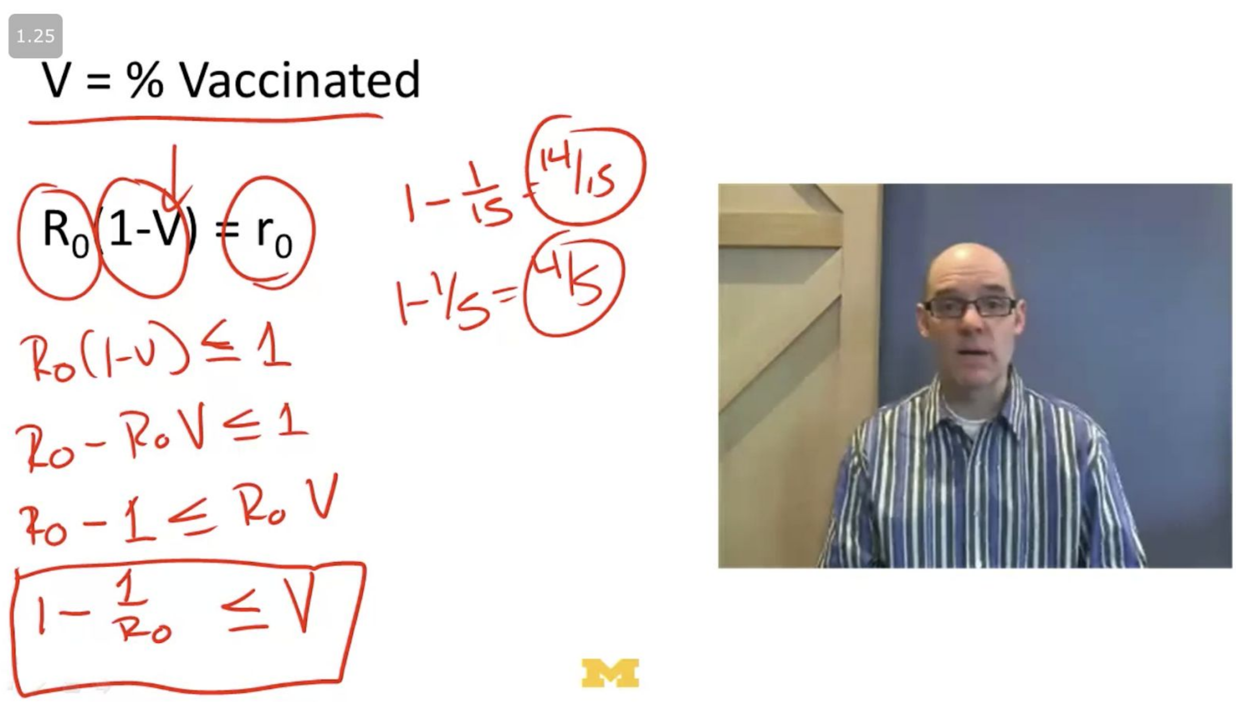
$R_0 > 1$ spreads

$w_{t+1} = w_t +$ Pos

Basic Reproduction Number



Finally, there is also the percentage of people that need to be vaccinated in order to stop the spread.

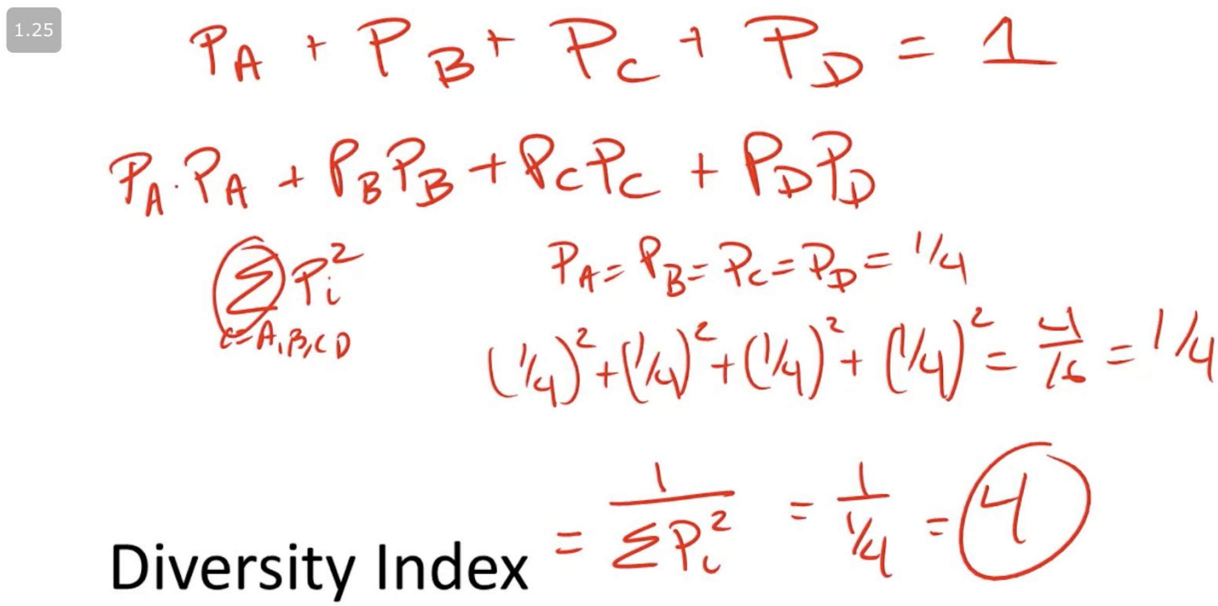


Direct tips is a situation where a particular action tips that same variable.

Contextual tips is when something changes in the environment in such a way that causes the system to move from one state to another (enable it to percolate, for example).

There are some ways to measure tips, the ones presented in this class are:

Diversity index: measures the diversity in the outcomes. Imagine 4 different outcomes (A, B, C and D), and each has a probability of ¼, then the formula and result of the diversity index is shown below.



Entropy: measures the degree of uncertainty. Using the same example as shown above, the formula and result of the entropy calculation is shown below.

1.25

$$-\sum P_i \log_2(P_i)$$

$$\log_2(2^x) = x$$

$$\log_2(\tfrac{1}{4}) = \log_2(2^{-2}) = -2$$

$$-\left[\tfrac{1}{4}\log_2(\tfrac{1}{4}) + 1/4 \log_2(1/4) + 1/4 \log_2(\tfrac{1}{4}) + 1/4 \log_2(1/4)\right]$$

$$-\left[\tfrac{1}{4}(-2) + 1/4(-2) + 1/4(-2) + 1/4(-2)\right]$$

$$-[-2] = 2$$

Entropy

What entropy is showing is the number of bits of information one needs to know in order to identify what the outcome will be.
Thus, in the example above, there are 4 possible outcomes and with only 2 questions it's possible to identify what the outcome will be. Example: " Is it (A or B) or (C or D)?", let's say it gets the first, then, "is it A?", if "yes", then it's A, if "no" then it's B.

Economic growth

Compounding is the process whereby interest is credited to an existing principal amount as well as to interest already paid.

Exponential growth is the growth applied to GDP (gross domestic product) based on the growth of the country per se. Example below with country A (2% growth) and B (6% growth).

Year	2%	6%
0	1000	1000
1	1020	1060
10	1219	1791
35	2000	7686
100	7245	339,302

The **rule of 72** is basically a formula that states if you divide 72 by the percentage of growth of a certain country, it will give you - approximately - the number of years it will take for that country's GDP to double. For example, for country A (2% growth), it will be 72/2 = 36 years (you can see it matches with the table above).

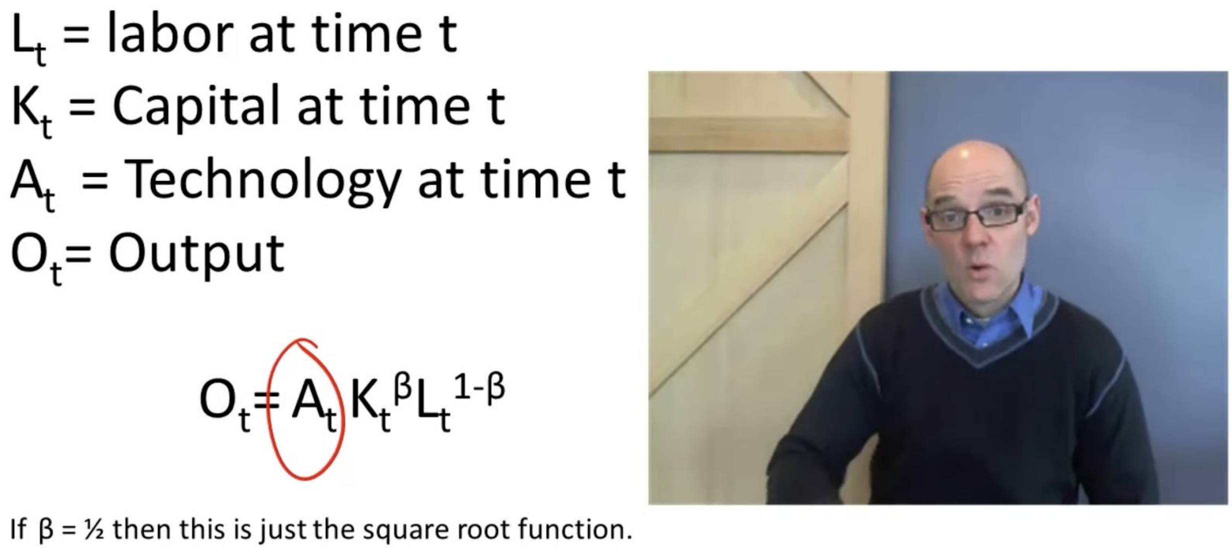
The **basic growth model** states that you have the following:

- Output = (a formula using M as the # of machines)
- Investment = s (saving rate) * output
- Depreciation = d (depreciation rate) * # of machines

The **long run equilibrium** is the condition that states the following:

Investment = Depreciation

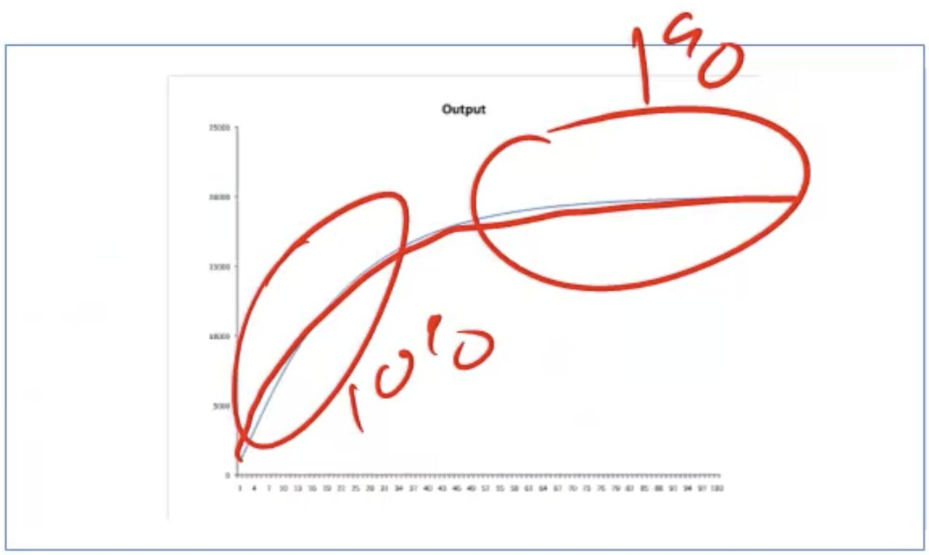
The **Solow Growth Model** is very similar to the basic growth model but adding one parameter (A = technology). The rationale behind it is that using the basic growth model the country's growth will cease without innovation, whereas the solow growth model takes technology into account. The formula is shown below.



The output growth is concave (shown below), thus, you have higher growth in the beginning of the process and it falls over time for a plato.

Economic growth can happen in two ways:

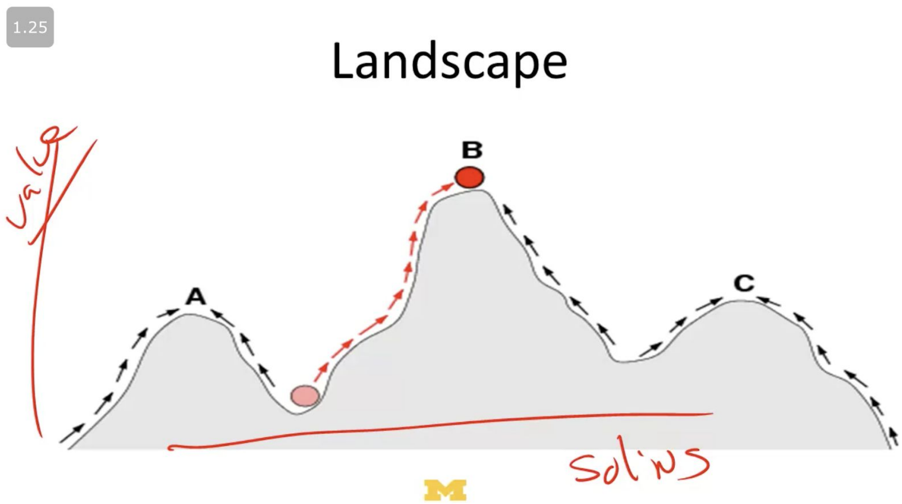
- Due to capital accumulation
- Due to technology advances



Week 5

Diversity and Innovation

Perspectives and innovation, the professor elaborates around how problem solving occurs and how to maximize the value for a certain problem exploring all the possible solutions you have available. One example - shown below - is to imagine a landscape where the x-axis is solutions and the y-axis is value, what one aim is to maximize the value given the possible solutions.



Perspective depends on the problem you have to solve, if you want to define a function that represents a straight line in a cartesian plane, it's easy to do so with a cartesian methodology. In case you want to represent a curve in the cartesian plane, it's better to use a polar coordinate system.

* Better perspectives have fewer local optimas

Savant existence theorem: for any problem there is a perspective that creates a mt Fuji landscape, that is, a solution that is by far the best one in terms of aggregate value in comparison to the other solutions available.

Heuristics are useful for guiding you on how to go to find solutions for your problems once you have defined a perspective. There are several strategies for heuristics, such as:

- **Do the opposite**: try the opposite of is currently propose in the system, for example, arrive to a hotel and say how much you are willing to pay for a daily
- **Big rocks first**: you first solve how to fill in big rocks first and then you fit the smaller rocks

- Diverse heuristics: combine different heuristics available might bring more value to your problem solving

However, there is the **No Free Lunch theory** that says that a heuristic can't be good for all types of problems, thus, if one heuristic is great for a problem, most likely is terrible for another.

Teams and problem solving is interesting for the diversity you can get, that is, different perspectives and solutions, which means that the team can end up better.

Recombination can be super useful when it comes to innovation, for example:

- The **steam engine** uses a combination of pistons, gauges, heat, etc.
- The **bicycle** combines gears, wheels, fractions, etc.
- **Smart phones** combine phones with gps, a camera, and many other features.
- The **Swiss Army knife** features a variety of useful tools.

Each of these inventions was, at the time, a "recombination" of existing ideas that created a new, innovative machine.

Markov Processes

Markov processes are really simple, they're basically formed by **states** and **transition probabilities**. And a markov model holds:

- Finite number of states
- Ability to get from any state to any other
- Fixed transition probabilities

A Markov process matrix of probabilities can be defined as follows.

Markov Transition Matrix

t

	A(t)	B(t)
A(t+1)	0.8	0.25
B(t+1)	0.2	0.75

t+1

$$\begin{bmatrix} 0.8 & 0.25 \\ 0.2 & 0.75 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} (0.8)\cdot 1 + (0.25)\cdot 0 \\ (0.2)\cdot 1 + (0.75)\cdot 0 \end{bmatrix}$$

The vector (1, 0) can be interpreted as the current state (100 percent allocated in status A, and 0 percent allocated in status B), and the result sum is the new distribution for each state (0.8*1 + 0.25*0, and so on).

The markov process can be in equilibrium when we get this situation:

$$\begin{bmatrix} 0.8 & 0.25 \\ 0.2 & 0.75 \end{bmatrix} \times \begin{bmatrix} p \\ (1-p) \end{bmatrix} = \begin{bmatrix} p \\ (1-p) \end{bmatrix}$$

$$.8p + .25(1-p) = p$$

Thus, with simple algebra the value for p that sets the system in equilibrium can be found.

The **Markov Convergence Theorem** states that for a Markov model to converge, the process needs to fill the four following assumptions, and the equilibrium will be unique.

A1: Finite **states**

A2: Fixed **transition probabilities**

A3: Can eventually get from any one state to any other

A4: Not a **simple cycle**

* not a simple cycle means that there is not only two states and they will not fall into 100% into state A and after 100% into state B

Important notes:

- Initial state doesn't matter
- History doesn't matter (what happened during the process of convergence)
- Intervening to change the state doesn't matter

Week 6

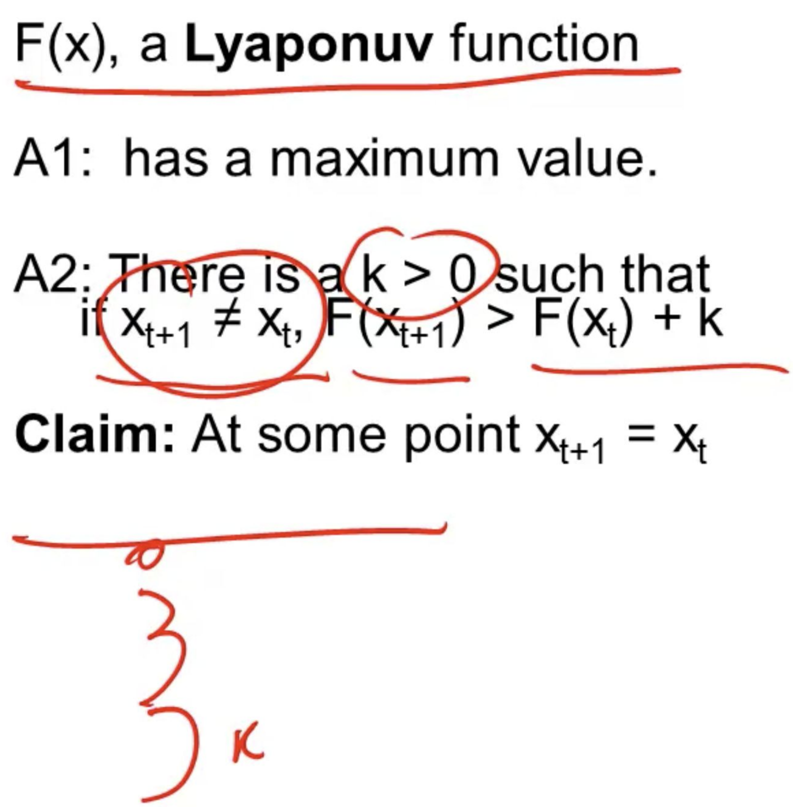
Lesson learned: other than equilibrium, models can have only three outcomes (shown below).

- Complex
- Periodic (cycle)
- Chaotic (random)

Week 7

Lyapunov Functions

Lyapunov functions is a tool that can translate a model into a function (mapping models into outcomes), and in doing so, it's assured that this model can go to equilibrium.



There is also the contrary regarding the minimum value.

F(x), a **Lyaponuv** function

A1: has a minimum value.

A2: There is a k > 0 such that
if $x_{t+1} \neq x_t$, $F(x_{t+1}) < F(x_t) - k$

Claim: At some point $x_{t+1} = x_t$

However, Lyapunov functions shouldn't be applied to cases where there is the presence of externalities, such as the example below (political coalitions, mergers, alliances, dating).

- North Korea trades Nuclear Weapons for Oil
- Iraq trades Oil for Nuclear Weapons
- USA is not involved but it's unhappy with the trade

1.25

How Long Until Equilibrium?

$F(x_1) = 100$

$k = 2$

$max = 200$

$\#periods \leq 50$

$\frac{100}{2} \leq 50$

Collatz Problem is a case that shows that even simple problems can take a long time to converge to an equilibrium.

HOTPO (Half Or Three Plus One):
1. If even, divide by two
2. If odd, times three and plus one
3. Stop if reach one

Examples:

5:

16, 8, 4, 2, 1

7:

22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1

Summarising:

- If you can construct a Lyapunov function, it means that the system goes to equilibrium. Thus, you can compute the time to equilibrium as well
- Externalities are a reason why systems don't go to equilibrium
- The equilibrium need not be unique or efficient

Coordination and Culture

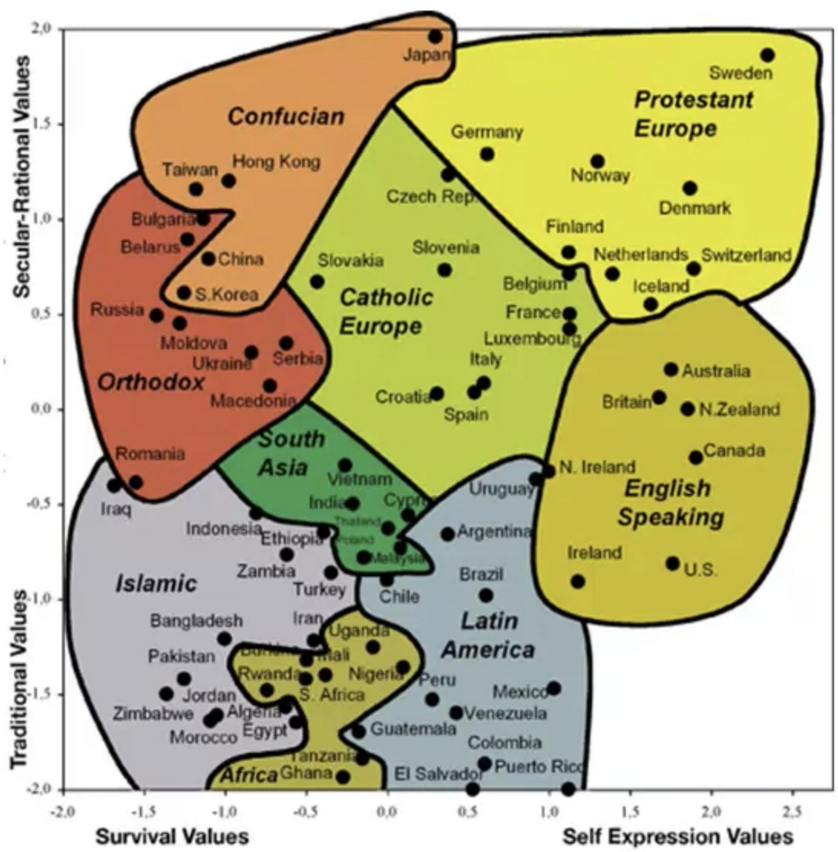
According to Tylor (1871), the definition of **culture** is a complex whole which includes:
- Knowledge
- Belief
- Art
- Law
- Morals
- Customs

According to Boaz (1911), the definition of **culture** is the totality of mental and physical reactions and activities that characterize behavioral responses to environment, others, and to himself.

According to Trilling (1955), the definition of **culture** is:

Trilling (1955): When we look at a people in the degree of abstraction which the idea of culture implies, we cannot but be touched and impressed by what we see, we cannot help but be awed by something mysterious at work, some creative power which seems to transcend any particular act or habit or quality that may be observed. To make a coherent life, to confront the terrors of the outer and the inner world, to establish the ritual and art, the pieties and duties which make possible the life of the group and the individual – these are culture, and to contemplate these various enterprises which constitute a culture is inevitably moving.

In summary, there is the need for similarities within interests.



www.worldvaluessurvey.org

Coordination is connected with cultures and societies, since people tend to move themselves towards coordination naturally, that's coordinating between themselves in order to align behaviours.

Week 8

Path Dependence

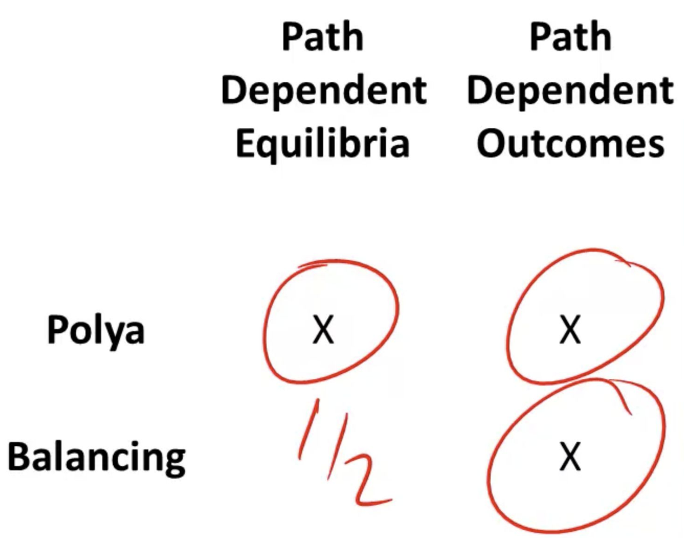
Path dependence means that the outcome probabilities depend on the sequence of past outcomes. And path dependence equilibrium states that the percentage of red balls in the long run depends on the path.

"Phat" dependence means that the outcome probabilities depend upon past outcomes but not their order.

Polya process is defined as adding a new ball that is the same color as the ball selected. Interesting note:

- Result 1: any probability of red balls is an equilibrium and equally likely
- Result 2: any history of blue and red balls is equally likely

Balancing process is defined as adding a new ball that is the opposite color as the ball selected.



Chaos is determined solely by the starting point of the process, thus, is not path dependent.

Networks

The **structure of networks** can be divided by three things:

- Logic: how it forms
- Structure: measures (degree, path length, connectedness, clustering coefficient)
- Function: what it does

The networks are formed by objects (nodes) and these objects are connected by lines (edges), you can have both directed and undirected networks.

Examples:

- The nodes are the US states, the edges are whether these states share a common border (undirected networks)
- The nodes are students, the edges are whether people look at each other for fashion design (directed networks)

Some definitions of the structure:

- **Degree**: number of edges attached to a node (node), and average degree of all nodes (network)
- **Path length**: the minimum distance (number of edges) to go from one node to another
- **Connectedness**: average path length between all pairs of nodes in a network
- **Clustering coefficient**: number of possible triangles (triangle connection of nodes) overall in the network

Week 9

Randomness and Random Walk Models

The **paradox of skill** states when you have the very best competing, the differences in their skill levels may be close, so the winner will be determined by luck.

The **skill luck model** can be given by the formula below, and it can support us assess outcomes, show we can anticipate reversion to the mean, give us good feedback, and help fairly allocate resources.

$$\text{Outcome} = a * \text{Luck} + (1 - a) * \text{Skill}, \; a \text{ in } [0, 1]$$

The **binary random walk model** states that both possible outcomes can happen with equal probability.

- Result 1: after N outcomes, you should expect the mean to be zero (in case one outcome is +1 and the other -1)
- Result 2: for any number K, a random walk will pass both -K and +K an infinite number of times
- Result 3: for any number K, a random walk will have a streak of K heads (and tails) an infinite number of times

When it comes to random walks and wall street, the best book that defines this is A Random Walk Down Wall Street (quote below).

> The **efficient market hypothesis** is associated with the idea of a "random walk," which is a term loosely used in the finance literature to characterize a price series where all subsequent price changes represent random departures from previous prices. The logic of the random walk idea is that if the flow of information is unimpeded and information is immediately reflected in stock prices, then tomorrow's price change will reflect only tomorrow's news and will be independent of the price changes today. But news is by definition unpredictable and, thus, resulting price changes must be unpredictable and random.

Burton Malkiel
Journal of Economic Perpectives 2003

Colonel Blotto

The **Colonel Blotto model** is about strategic mismatch and allocation of resources. The rationale behind the model can be shown with the following example.

2 players each with T troops

N fronts (T >> N)

Actions: allocations of troops across fronts

Payoffs: # fronts won

A **zero sum game** is where one player/team wins, and one player/team loses, no other option.

Week 10

Prisoner's Dilemma and Collective Action

The **prisoner's dilemma game** is about two players making choices of cooperating or defecting, and based on each choice there is the score's outcome for both players.

There are two special situations important to be noted:

- **Pareto efficient**: when players cooperate or not, the outcome for each player will be T or F (the better off option for the players, when an economy has its resources and goods allocated to the maximum level of efficiency, and no change can be made without making someone worse off)
- **Nash equilibrium**: when both players defect, the outcome for each player will be R (each player will do what is best in the individual perspective)

		P2: C	P2: D
P1	C	T,T	0,F
P1	D	F,0	R,R

$T > R$

$F > T$

$2T > F$

There are seven ways to cooperation models (shown below), each one relies in different types of incentives for people to cooperate and can be applied to different situations.

1. Repeated
2. Reputation
3. Networks
4. Group
5. Kin
6. **Laws**
7. **Incentives**

Collective action problem is when the problem itself contains several different people involved and there is an action for each person j, the payoff for person j is given by the formula below (x_j is the action of person j, or can be seen as the cost).

Let x_j be the action of person j.

$$\text{Payoff to j} = -x_j + \beta \sum_{i=1}^{N} x_i$$

β in (0,1)

Mechanism Design

Mechanism design is about setting the possible actions people can take and what the payoffs are going to be for each action.

Within mechanism design there is the **Clarke-Groves-Vickery pivot mechanism**, and the formula each person involved in the mechanism is going to pay is shown below.

Clarke-Groves-Vickery Pivot Mechanism:

Each person claims value: V1, V2, V3

If V1+V2+V3 > Cost

Do the project

Person 1 pays

Max {Cost – V2 –V3, 0}

Week 11

Learning models: replicator dynamics

Replicator dynamics is a way to define the proportion of actions taken by agents given a certain proportion of actions and the payoff (formula below) in previous times.

$$\mathrm{Pr}_{t+1}(i) = \frac{\mathrm{Pr}_t(i)\pi(i)}{\sum_{j=1}^{N} \mathrm{Pr}_t(j)\pi(j)}$$

Fisher's theorem states that a situation can be described by the union of three different models: There is no cardinal; Rugged landscape; replicator dynamics.

The idea behind Fisher's Theorem is that higher variation increases the rate of adaptation.

Variation or six sigma, it all depends on the assumptions to be taken. Such as:

- For **fixed landscapes**, use six sigma
- For **dancing landscapes**, use Fisher's Theorem

Prediction and the many model thinker

The **Diversity Prediction theorem** states that the crowd's errors are equal to the average error minus the diversity (shown below).

Diversity Prediction Theorem

Crowd's Error = Average Error - Diversity

$$(c-\theta)^2 = \frac{1}{n}\sum_{i=1}^{n}(s_i-\theta)^2 - \frac{1}{n}\sum_{i=1}^{n}(s_i-c)^2$$

θ = True value

S_i = Individual i's prediction

And, the definitions are as follows:

Crowd's error = final prediction (crowd's mean) minus the observed value squared
Average error = sum of individual's prediction minus the observed value squared
Diversity = sum of each individual prediction minus the final prediction squared