FROM MODELS to DECISION MAKING

9.1 Equation-based versus agent-based model

9.1.1 Motivations

Agent-based modeling (ABM) has become a fashionable tool to simulate complex systems. Of course we know that dynamical systems theory offers a mathematical framework to model and simulate spatiotemporal phenomena. While there is not a strong difference between equation-based and model-based description, certainly not at the implementational level, agent-based model offers a somewhat different modeling philosophy and practice.

An agent is considered an autonomous computational unit, with the ability to bring decisions, or to put it another way, to make selections among possible strategies. Of course, if the definition is loose, we may interpret even a simple McCulloch-Pitts neuron, as an agent, who decides to fire, for a certain set of inputs, and remain at rest for others. Agents have time-dependent internal states, and their states change due to the interaction of other agents. Specific rules determine the interactions among agents. Agents may interact locally with their neighbors, or globally with all other agents. Interacting agents may be also the nodes of say, small world network, so an agent may interact both locally and partially globally.

Agents are supposed to show some elements of intelligence, or at least they are capable of showing some degree of autonomy, adaptation to environmental changes, communication with other agents, and occasionally goal-directed learning.
ABM has multiple intellectual roots. One is cellular automaton, which is a discrete time discrete state space dynamical system, where players are located on the grid points of the arena having a specified geometry. There is a rule for updating the dynamics of the state of each grid points, and the rule is the same for each point. Artificial life and artificial societies have been constructed based on these principles, as it will be reviewed soon.

Another intellectual source of agent-based modeling emerged as a reaction to the neoclassical theory of economics, which dominates the science of economics. As we learned from Brian Arthur (recently reinforced by the popular book [54]), the complex systems approach is not in accordance with the assumptions of the neoclassical theory. The comparison of the two approaches can be seen here:

- **Neoclassical economics**
  - Behavioral model for people:
    * Fully-informed
    * Rational
    * People interact only indirectly with one another (through markets)
    * Focus on equilibrium outcomes

- **Complexity approach**
  * People are adaptive
  * They interact directly with one another
  * Focus on dynamics
  * Methodology: equation vs. agent-based modeling?

Economy is a dynamic process. It is the emergent result of collective interaction of very different players (well, agents), from individual investors, to firms, consumers, banks etc, and ABM seems to be a very good method to simulate, how local interactions among players may lead to macroscopic behavioral patterns.

### 9.1.2 Artificial life

*From cellular automata to the Alife movement*

Cellular automata (CA) are also considered as intellectual roots of agent-based modeling. It is well-known that having been motivated by von Neumann model, John Conway **constructed** (this is the key word: “constructed”: the algorithm does not pretend to implement a real mechanism, it is skilfully constructed to generate interesting phenomena) a set of rules, which was able
to generate complex spatiotemporal patterns. The **game of life**, as it is called, also motivated the birth of the artificial life (Alife) movement. Alife models (written mostly by people belonging to the computer science community) grasp the “possible”, while biologists are interested mostly in the “actual”. I think, Alife models give insight to understanding real biological phenomena, and also “life, as it could be”. However, the myth of simulation, which says “we understand what we can generate by computational algorithms” should not be taken literally. Alife certainly came from the tradition of cybernetics, and helped to popularize the notion that constructive bottom up models also contribute to understand biological dynamic organizations.

**Artificial chemistry**

The Alife movement has strong branches. Artificial chemistry is a formal system. Its building blocks are the set of molecules, the set of reactions, and a rule, which describes the temporal evolution of the state. Formal chemical models may describe chemical events themselves, but may serve as a metalanguage as it was mentioned in 4.2.2. Artificial chemistry (just as formal chemical kinetics) can be used to model prebiotic evolution but also higher-level, even social systems [131].

**Digital evolution**

Michael Conrad (1941-2000) was a pioneer of simulating evolution around 1970 studied a model of the evolution of an artificial open-ended ecosystem [108]. Individuals with given genotypes were defined, and these genotypes were interpreted as instructions, leading to phenotypes. The organisms competed in a simple one-dimensional world for the possession of resources which they use for self-repair and reproduction. Interactions between organisms for providing co-evolutionary selection pressure for increased complexity has been emphasized. Twenty years later his evolutionary computer simulations became more realistic [109].

Somewhat in the spirit of Conrad (and a few others), an artificial ecology and artificial evolution was developed by the ecologist Thomas S. Ray by implementing the basic evolutionary and ecological processes in a computer. The system established is called **Tierra**. It was an abstract, not too biologically detailed model, but programs were able to mutate, replicate and recombine (also called crossover). Evolution was embedded in a computer, and programs evolved by the rules of evolution. Artificial evolution has two goals, first, to help in understanding real biological evolution, second, to provide evolutionary computational algorithms to solve hard optimization problem.
Tierra has been motivated by a few facts and requirements: (i) Life on Earth is a single example of evolution (sample size of one); it is possible (ii) to create and observe a new and independent instance of evolution, (iii) and broaden our perspective of evolution by increasing the sample size (to more than one) (iii) and to observe the properties of evolution in a (silicon-based) digital medium. According to this approach, real carbon-based biology is not the essence of an artificial evolutionary systems.

Of course, digital medium is nothing to do with our physics, Digital evolution is not constrained by the laws of our physics. Cyberspace is nothing to do with our 3D Euclidean space. The laws of conventional physics and chemistry of our organic medium is replaced by the rules of the program languages and operating systems used in the digital medium.

The idea behind the digital evolution is that we don’t have any idea what might be the emergent result of it. Ray suggested a thought experiment. Let’s assume we are all robots made of metal and our brains of silicon chips. We have no experience or knowledge of carbon based life. A robot brings several chemical substances, such as methane, ammonia, hydrogen, water, and some minerals to a scientific meeting. The robot asks: “Do you suppose we could build a computer from this stuff?” The answer is that some engineer-minded robot might see molecular computers, but neurons (not speaking about neural networks, and interacting brain regions) could be never imagined. As in the organic medium evolution established the unpredictable structures, digital evolution also should generate complex information processes.

Tierra motivated the construction of other platforms for the evolution of computer programs, e.g. Avida. It is an extension of Tierra at least in two respects. First, each program is stored in a different part of the memory, second the speed of the program running may be different for the individual programs, According to spirit of Alife, Avida does not simulate evolutionary processes, but “digital organisms in Avida evolve to survive in a complex computational environment and will adapt to perform entirely new traits in ways never expected by the researchers, some of which seem highly creative.” (from the Avida website http://devolab.csc.msu.edu/, Nov 26th 2006). Programs able to evolve behave similarly in many respects that biological evolution. In the paper [310] it was shown how a complex function such as “equal” function evolved by simpler functions:

“... Of course, digital organisms differ from organic life in their genetic constitution, metabolic activities and physical environments. However, digital organisms undergo the same processes of reproduction, mutation, inheritance and competition that allow evolution and adaptation by natural selection in organic forms.”
9.1 Equation-based versus agent-based model

Genetic algorithm

Probably the best known concept related to digital evolution is the genetic algorithm (GA) formulated by John Holland in Ann Arbor, at the University of Michigan [239]. GA is one of the successful winner of the evolutionary computational algorithms (an excellent book about evolutionary computation is [178]).

GA is generally used to solve optimization or search problems. There is a population of candidate solutions (“chromosomes”) and the goal is to progress towards better solution by some algorithm. In the case of GA, this come from biological motivation. The solutions are encoded as strings, traditionally as binary strings of zeros and ones. A fitness function, which measures the quality of a solution, can be assigned the set of solutions.

A GA starts with the generation of a set of candidate solutions, chosen either randomly, or if there is some preliminary information, then around some estimated values. The next step is selection. A subset of the candidate solutions are selected controlled by the combination of (i) a fitness function based deterministic selection and (ii) some random effects. Now comes the next step: reproduction. The selected solutions are modified by crossover and mutation. Crossover is a procedure to establish “children” strings from the “parents” mainly by cutting and glue. The size of the population is fixed, so the worst solutions will be eliminated. A new generation of solutions (having the same size as the initial one) has generally a better fitness, than the previous. The procedure is iterated, and will stop when the fitness functions does not show an improvement.

GA and its different versions, and extensions are used now in problem solving. While the procedure is convergent, often it leads to local and not global optimum. A larger mutation rate (i.e. more randomness) helps to kick off the system from a local minimum.

Genetic programming

GP is an extension of the genetic algorithm. While Holland adopted (typically binary) strings with fixed length, genetic programming uses trees. Hierarchical organizations better characterize the structure of computer programs, and proved to be much more flexible than string-based representations, as John Koza ¹ one of the champions of genetic programming explains [292]. Of course,

¹ John Koza is also involved in the initiation to change the US electoral system and suggested a state-based plan for electing the president by national popular
genetic operators, (mutation, crossover) should be defined for this geometry. An example, how crossover generates new programs is shown at figure 9.1.

Fig. 9.1. Sub-tree crossover creating new programs from existing programs. Sub-trees are selected randomly from two existing parse trees and are then swapped to produce child parse trees. http://www-users.york.ac.uk/mal503/common/thesis/jpegimages/subtree_c.jpg

GP is a method to detect an uncover “complexity”. It helps to design and optimize computer programs. It also offers a technique for interpreting certain data sets. To be more precise, GP can solve certain “inverse problem” of (bio)chemical kinetics, i.e. starting from measured trajectories gen-

vote. There is a downloadable book about this issue; http://www.every-vote-equal.com/tableofcontents.htm; November 27th, 2006
erates the network of chemical reactions. Similarly, it is a method for the automated synthesis of genetic networks, and also of analog electrical circuits. As it is stated, GP was able automatically produce computer programs that is competitive with human performance. The URL http://www.genetic-programming.com/humancompetitive.html (November 17th, 2006) gives a list of problems, which were solved by programs evolved by GP, where many of the solved problems were previously patented. GP produced human-competitive results, and seems to be an excellent method for novelty generation. For more details see [292, 293, 294, 295].

9.1.3 Artificial societies

The spirit of the Alife movement was extended to artificial societies. Joshua Epstein and Robert Axtell published a book on “growing artificial societies”, which still is a paradigmatic illustration of agent-based models of social phenomena. (Epstein was also the author of a book, as it as mentioned earlier on applying equation-based models of social dynamics ...). It could be considered as an outgrowth of the Schelling model.

At the land of Sugarscape

The model is called Sugarscape [155], and it has several versions, with less difficult and more difficult agents – and thus society. Sugarscape is a simulated artificial world populated with artificial agents. The world consists of a square lattice with some sugar distribution. Sugar is the only source of food and is needed to stay alive. Specifically, the sugar concentration had peaks in the northeast and southwest quadrants of the grid. The agents walk around on the lattice looking for sugars. The model is deterministic, only the initial conditions – i.e. the initial position of the agents – are random.

Agent-based models are theoretically completely equivalent to equation-based models, only the usual description is different: for the latter one generally it is a set of equations, for the former one it is generally a list of algorithms. While for Sugarscape the agent-based approach is obviously better suited let us here show an equivalent equation-based description.

This will be a discrete time discrete space deterministic model. Let \( N \) be the size of the \((N \times N)\) grid and \( M \) the total number of agents (dead or alive), these are parameters of the model. Other parameters are the maximum amount of sugar in the grid positions: \( S^* = [s^*_ij] \), this is an \( N \times N \) matrix of integer values between zero and four; the distance giving how far the agents
can see on the lattice: $e_i, 1 \leq i \leq M$, and the number of sugar units the agent digest in a time step: $d_i, 1 \leq i \leq M$; these two parameters define the DNA of the heterogeneous agents.

The variables of the model are the actual amount of sugar at each grid position: the $S = [s_{ij}]$ matrix, vectors giving the actual coordinates of the agents: $(x_i, y_i)$, and a vector giving the amount of reserves an agent has: $w_i$, basically her wealth in this model.

The variables are updated according to the following formulae:

$$s_{ij}(t + 1) = \min \{ s_{ij}(t) + 1 - s_{ij}(t) \sum_k \delta(x_k(t), i) \delta(y_k(t), j), s^*_{ij} \} \quad (9.1)$$

$$w_i(t + 1) = \begin{cases} w_i(t) - d_i(t) + s_{x_i(t), y_i(t)}(t) & \text{if } w_i(t) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (9.2)$$

$$(x_i(t + 1), y_i(t + 1)) = p_i(t + 1) \quad (9.3)$$

where $p_i(t + 1)$ is recursively defined as

$$p_1(t + 1) = \arg\max \{ S_{ij}(t + 1) | (i, j) \in P_1(t + 1) \} \quad (9.4)$$

$$p_k(t + 1) = \arg\max \{ S_{ij}(t + 1) | (i, j) \in P_k(t + 1) \setminus \bigcup_{l=1}^{k-1} \{ p_l(t + 1) \} \} \quad (9.5)$$

and the set of possible moves are defined as

$$P_k(t + 1) = \{ (x_n, y_k(t)) \in N \times N | |x_n - x_k(t)| \leq e_k \} \cup \{ (x_k(t), y_n) \in N \times N | |y_n - y_k(t)| \leq e_k \} \quad (9.6)$$

Now compare this description to the normal algorithmic description of the Sugarscape world, as follows. A $N \times N$ grid is given, there is at most one agent in each grid point. In each time step first the agents eat up the sugar at their current position, then digest some amount of sugar given by a parameter which is different for different agents. If an agent has no sugar at all in her body after this, she starves to death and it is removed from the simulation.

Then the amount of sugar increases by one at each grid position if it less than its maximum level. Then – still in the very same time step – the agents choose where to step next by looking around themselves. Each agent has a vision parameter, the number of steps it can see along each direction (but not diagonally). One by one in predefined order, the agents check the available
cells within their sight and choose the one with the maximum sugar. Then the next time step follows.

While theoretically it is possible to analyze this system in the dynamical systems framework, this is obviously not done. Instead, some average measures are calculated from the state variables or their distribution is plotted and its time evolution is observed.

Fig. 9.2. One of the many implementations of the Sugarscape world. Various variables and parameter of the simulation can be seen on the right panel.

The Sugarscape simulations were traditionally run on a landscape with two sugar mountains, one in the northeast and one in the southwest with some deserts in the two other corners. The agents are placed randomly on the landscape and then one clicks on the ‘Run’ button and waits for what would happen. See Fig. 9.2 for an example landscape.

This very basic model is able to explain how the skewed wealth distribution can be generated from totally symmetric and ‘just’ rules. (In fact it would be nice to see whether the skewed wealth distribution could be obtained by homogeneous agents as well.) The little differences in the agents skills and their starting position is enough to make a huge difference in the long run.

In the basic model agents only die if they starve to death and thus there is no need for reproduction. In the first extension of the model each agent is randomly tagged as female or male and in a more later version they also have age. Agents die at a given age. If an agent is old enough and has some sugar savings then it is considered fertile. If two fertile agents meet (i.e. they are on neighboring cells) then a new agent is created on a neighboring cell. The new agent will inherit also some of the wealth of the parents and also some random variation of their DNA (i.e. vision and digestion parameters).
Introducing reproduction has a number of effects on the Sugaracape society. First, the average vision and metabolic efficiency starts to increase over time. Second, big oscillations started to occur in the size of Sugarscape population, unlike in the basic model. Third, the gap between rich and poor widened ever further.

The second extension added is trading. For trading another type of essential food is introduced: spice; agents needed both sugar and spice for survival. Also, the deserts are replaced by spice mountains in the landscape. From now on if two agents met they are able to trade spice for sugar and sugar for spice and the price is determined by the two interacting agents after some bargaining. The immediate result from the extended model was that trade is good. The average wealth of the society increased compared to the basic version. Another result was that the difference between the poor and the rich grew ever further. Finally, an important result was that even if the constructed supply and demand curves looked like examples from a microeconomics book, the Sugarscape society was never in equilibrium as predicted by traditional neoclassical economics. This is probably not as much as a surprise since in the Sugarscape world a number of phenomena, like horizontal inequality and incomplete information exist which are not allowed in traditional economics.

The last extension we mention here allowed Sugarscape agents to lend and borrow goods. This little addition had a huge impact: some agents started to borrow goods from their neighbors and lend it to others: banks appeared.

For example, in the agent-based model Sugarscape (Epstein and Axtell 1996), each individual agent has simple local rules governing movement, sexual reproduction, trading behavior, combat, interaction with the environment, and the transmission of cultural attributes and diseases. These rules can all be “active” at once. When an initial population of such agents is released into an artificial environment in which, and with which, they interact, the resulting artificial society unavoidably links demography, economics, cultural adaptation, genetic evolution, combat, environmental effects, and epidemiology. Because the individual is multidimensional, so is the society.

From: [154]: Epstein J (2006): Generative Social Science: Studies in Agent-Based Computational Modeling
9.1 Equation-based versus agent-based model

How did the Anasazi native Americans live or could have lived?

Epstein now published Generative Social Science [154], a series of papers about the use of bottom-up social models. His intention is to show that complex social phenomena can be understood by local interactions among agents. A series of papers, very ambitiously, instead of studying virtual worlds, deals with the rise and fall of a real society, actually of the Anasazi native Americans. The historical fact is that the Anasazi abandoned Long House Valley, in about A.D. 1350 after they lived there for about 3000 years.

The Anasazi are a particularly good candidate for a simulation study since there are data available from various sources about the climate, crop production, and the actual place of the settlements in the valley. This makes it possible to compare the results quantitatively to the real events. The scientific question proposed whether it is possible that the Anasazi abandoned the place as a result of their normal, “internal” dynamics or there was some external force which made them chose this alternative. Note that as climate is part of the model, it is counted as internal here.

Let us briefly introduce the model. The unit of the modeling, the agent is one household, the smallest unit in archaeological records. Some attributes of the household are the model variables, these describe the state of the model: (1) the place of its residence, (2) its agricultural location, (3) the “age” of the household, everybody in a household assumed to have the same age, (4) the amount of corn the household owns. Most of the model parameters are set based on real data, these include: nutrition need per individual, the size of the household, the maximum length of grain storage, the parameters of the environment, the fertility of the household, etc. The model contained eight adjustable parameters: minimum and maximum death age, minimum and maximum age of the end of fertility, minimum and maximum fission probability, the average harvest and the harvest variance. (The actual death age, end of fertility and fission probability were chosen uniformly from the interval given with the parameters.) By the careful choice of these eight parameters the model was able to replicate the real world data very closely, except for one thing: the simulated Anasazi never abandoned the valley. This implies that as the model assumptions were not sufficient for explaining the phenomenon (i.e. the abandonment), there must have been some external cultural or environmental reason for it.

It is interesting to compare the Anasazi model to the Sugarscape world described earlier. In the Sugarscape model the (very minimal) assumptions were able to reproduce the real phenomena, so it shows that the assumptions are sufficient (but not necessarily required). Whereas in the Anasazi model the assumptions fail to reproduce one single aspect of the real events (and
succeed to reproduce every thing else), this means that the assumptions are not sufficient for explaining this aspect.

ABM seems to be an appropriate methodological tool to test ideas about social mechanisms which may establish certain macroscopic patterns. It may help to test hypotheses about the spatiotemporal dynamics of these macroscopic patterns emerging due to interaction of agents characterized by their possible strategies and actual decisions.

### 9.1.4 Agent-based computational economics

Agent-based computational economics (ACE) uses models of “virtual economic world” where different (“heterogeneous”) agents participate. Interactions of agents specify the temporal development of an economy from a given initial condition. The agents are computational units, or, if you wish, basically softwares. These software modules implement rules, which might be a simple if-then instruction, but also could be a long set of instructions.

ACE certainly preserves the basic property of dynamical systems description, namely causality. It helps to understand the mechanism of pattern generation (empirical understanding), and to design “good economics” (normative understanding).

Decentralized market economies might be understood, much better as complex adaptive systems (occasionally labeled as CAS) than optimizing rational calculators. A CAS consists of large numbers of adaptive agents involved in local interactions. These microscopic interactions imply macroscopic patterns, which often show regularities, as shared market protocol or behavioral norms, which feedback to the microscopic interactions, as Tesfation [510] states. He also classified the problems of ACE (of course we know, that every classification contains arbitrary elements): (i) Learning and the embodied mind; (ii) evolution of behavioral norms; (iii) bottom-up modelling of market processes; (iv) formation of economic networks (v) modelling of organizations, (vi) design of computational agents for automated markets; (vi) parallel experiments with real and computational agents (vii) building ACE computational laboratories.

A big enterprise of ACE was the establishment of the Santa-Fe Artificial Stock Market model.
The Santa-Fe Artificial Stock Market model

The goal was to set up an artificial stock market where players are inductive agents with bounded rationality, and to show that their interaction leads to real(istic) market dynamics [307].

There are \( n \) stock market traders, and two types of assets. The first is risk-free, pays a constant (obviously low) rate of return \( r \).

First, the market dynamics should be specified. The state of the stock is characterized by a price \( p_t \) and yields a dividend \( d_t \). The traders may act in each discrete time step. Actually either they trade (buy or sell), or don’t, and the price is consequently updated by the equation

\[
p_{t+1} = p_t + \beta(B_t - S_t). \tag{9.7}
\]

Here \( B_t \) and \( S_t \) are the number of buyers and sellers in the actual time interval, and \( \beta \) is a parameter, which controls the “velocity” of the price change. The dynamics of the dividend is specified by a random walk (more precisely by an Ornstein-Uhlenbeck process, (studied in subsection 6.1.3), \( d_{t+1} = d_t + \text{“Gaussian noise term”} \)). The fundamental value is a constructed quantity \( f_{vt} \), \( f_{vt} := d_t/r \). The agent knows the history of \( p_t \), \( d_t \), and \( f_{vt} \), and the goal is to maximize the money owned at the end of the simulation.

Agents are supposed to have inductive rationality, and evolutionary algorithms are adopted. The actions of the agents depend on rules that specify what to do for particular market conditions. The rules have variable strength, which are subject of change by some adaptive learning rule. In addition, new rules are formed by genetic algorithms, and the weakest rules are eliminated and are substituted by strongest rules (modified by random mutations).

The original model had two qualitatively different outcomes [395], one is equilibrium and homogeneous, the other is a nonequilibrium, heterogeneous. More specifically, under simple structural conditions (few agents, few rules per agent etc.) the system’s state is in or around equilibrium. Prices are close to the fundamental values, so fluctuations are small. The agents behave by and large uniformly, based on the logic of the equilibrium theory (buy when the price is smaller than the fundamental value).

Nonequilibrium behavior has many faces. Occasionally there are large deviations from the fundamental values (identified as bubbles or crashes) due to collective interaction among agents. Agents will become heterogeneous, they may adopt highly different rules. Rules are subject to evolutionary changes. Wealth distribution shows a time arrow; initially the distribution is narrow, i.e. more or less everybody has the same wealth. However, the finding of some
more advantageous rule by chance ("luck") may imply the emergence of much more wealth to some agents.

The Santa Fe Artificial Stock Market model was certainly a very influential attempt to make predictions on the market behavior in an environment when the agents have adaptive strategies. For the evaluation of the initial model and further artificial stock markets see [306].

9.2 Game theory: where we are now?

9.2.1 Classical game theory

As it was mentioned in 5.5, game theory was formulated basically by von Neumann and Morgenstern. It emerged as a mathematical theory of economics, and also proved to be an important tool for treating the problem of necessary cooperation to avoid (nuclear and other) catastrophes.

While the establishment of game theory was motivated by real games, the modern applications are related bringing decisions to social situations. Players have to choose among different strategies, and they take into account the possible strategies of all the other players. In cooperative games formations of coalitions are permitted, so finally the game is among different coalitions, and not among individual players. In the von Neumann and Morgenstern book [545] the term “coalition” appears in Chapter V, where zero-sum three-persons game were analyzed. A game is zero sum, when the total resource (award, penalty etc.) is fixed, and does not depend on the chosen strategies. In a non-zero sum game players interests are not necessarily in direct conflict.

Games, however, maybe non-cooperatives, too. In this case coalition formation is forbidden, as John Nash invented. The only non-cooperative game that von Neumann - Morgenstern treated was the zero-sum two persons game, which is necessarily non-cooperative. Nash excluded the possibility of coalition formation, and for these non-cooperative games he developed an equilibrium concept, which became later the Nash equilibrium. There is a situation, when each player choose a strategy, and nobody can better off by changing unilaterally her strategy. Nash showed that such equilibrium should exist for under very general conditions.

Let’s see now how two or more players divide a cake among themselves!
Cake division

The “cake division” problem is well-known. It is actually not a single problem, but a quite big set of different problems associated with the “proper” division of a set of goods (the cake) among \( n \) “players”. In the simplest version of the problem the cake is homogeneous, in other versions it is heterogeneous, i.e. each player has her own preferences regarding to different pieces of the cake.

A division is called proportional if every player receives at least \( 1/n \) of the cake in her own measure. A division is called envy-free if no player would change its piece with another player. Note that if the cake is homogeneous then proportional and envy-free divisions are exactly the same and also note that an envy-free division is always proportional.

A proportional division however is not necessarily envy-free if the cake is heterogeneous for at least three players. It is possible that all three players consider their own piece to be at least \( 1/n \) of the whole but (say) one of them considers another piece to be even bigger.

Note that there is possible to come up with other definitions for the “fair” division, like Pareto-optimality. Also note that the “preferences” of the players are usually expressed in terms of a probability measure on the cake and one needs to restrict the measure to ask meaningful questions, e.g. for \( k \) players we might need to assume that every subset of the cake can be divided into \( k \) equal parts according to the measures of any player, etc, see the cited references for details.

As Brams and Taylor [70] notes, there are in general four sets of problems associated with cake division. The first set includes existence theorems, i.e. determining the required and sufficient conditions for the existence of proportional or envy-free division. These theorems are usually not constructive, they only prove that such a division exists but don’t help to find it [491, 492, 569].

The second set of problems is to give a procedure to obtain a desired, i.e. proportional or envy-free solution. As it happens many procedures given are not algorithms but so-called moving-knife solutions when typically a player or a referee continuously moves a knife over the cake and another player calls stop whenever she thinks the cake should be cut at the current position of the knife.

A two-player moving-knife solution proposed by Austin in 1982 [33] works as follows. First imagine that the cake is one-dimensional. (Some kind ab-

\[ \text{Pareto optimality means that no better solution exists for any player without making at least one player worse off.} \]
Abstraction always necessary if want to formalize problems in the language of mathematics.) A referee moves a knife from the left edge of the cake (line) to the right edge until one player calls “stop” because she thinks the knife is at half-way point according to her measure. (Let this player be Player 1.) Now Player 1 places a second knife at the left edge of the cake and moves both knives together towards the right edge ensuring all the time that the piece between the two knives is 1/2 in her measure. Player 2 calls “stop” whenever she thinks the piece between the two knives is 1/2. As in the final situation, with the right knife at the right edge (and thus the left knife at the same place where Player 1 called stop) Player 2’s piece is larger than 1/2 in her measure there will be a point when it is exactly 1/2.

The third set is actually a subset of the second, but now we only allow strict algorithms (programmable on a Turing-machine if you like) and no moving-knife solutions. Some moving-knife solutions, like the previous example can be easily rephrased as an algorithms, others are not that easy.

The fourth set is the most interesting and relevant to us, so we will focus on this. This problem is to define a game in game theoretical sense in which each player has a non-losing strategy. In other words, we need to come up with a set of rules and a set of strategies for the players which ensure that they will receive at least 1/n of the cake in their own measure. (Or some other definition of “good” division for them.) The set of rules and the strategies are often called protocols in the literature [70]. Players in these games are rational and risk-averse, they don’t choose a strategy which can result less than 1/n of the cake, even if this event is very unlikely.

The classic envy-free protocol for two players and heterogeneous cakes is the “cut-and-choose” protocol: the first player cuts the cake into two halves the second player chooses one of them. The strategy of the first player is to cut the cake into exactly two pieces according to her measure, this she will get at least 1/2. The strategy of the second player is to choose the half which is at least as big as the other (according to her measure of course). While this protocol is envy-free, it is clear that the second player has some “advantage” as she can get a piece bigger than 1/2, but this is impossible for the first player.

Let us first give a simple protocol k players and homogeneous cakes, i.e. here each player has the same preferences. The protocol is given by mathematical induction.

1. If $k \leq 2$ then the problem is solved by the cut-and-choose protocol.

2. Divide the cake into $k - 1$ pieces according to the protocol being described.
3. Each player divides her share into \( k \) pieces.

4. Player \( k \) chooses one from every other players’ little pieces.

The strategy for dividing the cake is to divide equally and the strategy to choose is to choose the largest piece. Here is a short informal proof that the protocol is proportional. It is clearly proportional for two players. If the \( k - 1 \) players divided the cake into pieces \( a_1, a_2, \ldots, a_{k-1} \), \( \sum a_i = 1 \) then if player \( k \) follows her strategy she gets at least \( a_1/k + a_2/k + \cdots + a_{k-1}/k = 1/k \).

If a player divides equally after receiving her initial piece then she will keep holding an equal share.

For envy-free protocol for more players and heterogeneous cakes either, the inquiring Reader is advised to consult [70, 69] and references within.

We also didn’t deal with the cases where the cake is not continuous but discrete, i.e. a set of items are distributed among players with heterogeneous preferences, please see [71] and [433].

Prisoner Dilemma

The most famous non-zero sum game is the Prisoner Dilemma, which analyzes the costs and benefits of two possible strategies, namely cooperation or defection. Here is basic example:

“...In this game the two players are partners in a crime who have been captured by the police. Each suspect is placed in a separate cell, and offered the opportunity to confess to the crime. The rows of the matrix correspond to strategies of the first player. The columns are strategies of the second player. The numbers in the matrix are the penalties: the first number is the penalty of the first player, the second is the penalty of the second player. Notice that the total penalty of both players is smaller if neither confesses so each receives 1. However, game theory predicts that this will not be the outcome of the game (therefore we have a dilemma). Each player having the intention to minimize her penalty, reasons as follows: if the other player does not confess, it is best for me to confess (0 instead of 1). If the other player does confess, it is also best for me to confess (2 instead of 3). So no matter what the other player will do, confession implies one year off the sentence.

The payoff matrix of the game describes the possible outcomes:
Game theory predicts, that each player following her own self-interest will result in confessions (defection) by both players.

**Trust** in the other would minimize the total penalty. This is the basis of the dilemma. But can trust evolve? While the question is beyond the scope of classical game theory, it was investigated by evolutionary game theory. Specifically, first, a more sophisticated (i.e. actually more dynamic, or adaptive) version of the game, the iterated prisoner’s dilemma was introduced by Robert Axelrod, a political scientist, in his book “The Evolution of Cooperation” [34]. The main point is that emergence of cooperation and trust based on reciprocity. This problem leads us to evolutionary game theory.

### 9.2.2 Evolutionary game theory

Evolutionary game theory elaborated by John Maynard Smith (1920-2004) and George Price (1922-1975). Price gave an interpretation of Fisher’s fundamental theorem, basically by using an argument in 3.7.2. They introduced the concept of **evolutionary stable strategies** (ESS). If all members of a population adopts EES, there is not any mutant strategy, which can invade.

In subsection 4.5 the equations behind evolutionary dynamics were discussed. These types of equations are the basis of evolutionary game dynamics. The interpretation of $x_i$ in equation 4.43 is the frequency of strategy $i$, and the equations describe the evolution of strategies. As it was shown [393], these equations are basically equivalent to the Price equations.

### Evolutionary game dynamics

In evolutionary game theory the payoff matrix is identified with fitness. If there are two players, A and B, there are three possibilities: (i) A meets A, both get a; (ii) B meets B, both get d; (iii) A meets B; A gets b and B gets c. Considering a population of A and B players, if $x_a$ is the frequency of A and $x_b$ is the frequency of B than the payoffs for A and B are as

<table>
<thead>
<tr>
<th></th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>not confess</td>
<td>1,1, 3.0</td>
</tr>
<tr>
<td>confess</td>
<td>0.3, 2.2</td>
</tr>
</tbody>
</table>

- **Table:**

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>not confess</td>
<td>1,1, 3.0</td>
</tr>
<tr>
<td>confess</td>
<td>0.3, 2.2</td>
</tr>
</tbody>
</table>
Applying the frequency dependent selection, and using \( x = x_A \) notation, we have

\[
x_i(t) = x(1-x)[a - b - c + d]
\] (9.9)

Depending on the numerical values of the elements five qualitative outcomes are possible (the first two are symmetric) [387, 385]:

- \( a > c \) and \( b > d \) \( \rightarrow \) A dominates B. The whole population tends to consist of A players.
- \( a < c \) and \( b < d \) \( \rightarrow \) B dominates A. The whole population tends to consist of B players.
- \( a > c \) and \( b < d \) \( \rightarrow \) A and B are bistable. The equilibrium point is unstable, and the initial condition determines whether the system converges to “pure A” or “pure B” state.
- \( a < c \) and \( b > d \) \( \rightarrow \) coexistence of A and B. The stable equilibrium point is given as
  \[
x^* = \frac{d - b}{a - b - c + d}
\] (9.10)
- \( a = c \) and \( b = d \) \( \rightarrow \) A and B are neutral. Selection does not influence the composition of the population.

Examples for all cases can be found in the repeated prisoner’s dilemma, where an interaction between two players consists of many rounds. “...Tit-for-tat (TFT) is a strategy which cooperates in the first round and then repeats whatever the other player did in the previous round. "Always defect" (AllD) is bistable with TFT if the average number of rounds is sufficiently high. "Always cooperate" (AllC) and TFT are neutral if there is no noise and can coexist in the presence of noise. AllC is dominated by AllD. ... [387].

In case of three strategies the situation is more complicated. Since rock smashes scissors, scissors cut paper, and paper covers rock, there is a non-transitive, circular relationship among the three concepts. Recently it was found that this strategy exists in the biological realm. One version of lizards
live in three different forms (i.e. with orange, blue or yellow throats) with mating strategies which implies results characterized by circular relationship: “orange beats blue, blue beats yellow, and yellow beats orange” [12]. In a game with three possible strategies different long-term behaviors may emerge. It is possible that the three strategies coexist, or after increasing oscillation two strategies are subject of extinction.

Evolution of Cooperation

Is natural selection, i.e. a spontaneous mechanism sufficient to develop moral rules of cooperation from the interaction of self-interested players? Political scientist Rob Axelrod investigated this problem for many years. The starting points of his argument are that (i) biological evolution proved to be useful by adopting altruism; (ii) genetic algorithm used evolutionary principles successfully. Consequently, if the prisoner dilemma is iterated iteratively, strategies containing altruistic elements should perform better than purely selfish ones. The optimistic perspective was that evolutionary mechanisms may establish these altruistic phenotypes even from selfish ones.

It is well known that the “Tit for Tat” (“equivalent retaliation”) 3 algorithm suggested by Anatol Rapoport (one of the founders of systems theory movement, as it was mentioned in subsection 2.2.1) proved to be very efficient. The two steps of the algorithm are:
1. Cooperate!
2. Do what your opponents did in the previous step!

Long-term cooperation has evolutionary advantage to myopic selfishness. The more general rules behind evolution of cooperation is under investigation [325].

Evolution of fairness: the problem:

The Ultimatum Game is quickly catching up with the Prisoner’s Dilemma as a prime showpiece of apparently irrational behavior. In the past two decades, it has inspired dozens of theoretical and experimental investigations. The rules of the game are surprisingly simple. Two players have to agree on how to split a sum of money. The proposer

3 The Hungarian language is not so economic, and it says that “Szemet szemért, fogat fogért”, i.e. eye for eye, tooth for tooth) based on a principle taken from the Old Testament (Exodus 21).
makes an offer. If the responder accepts, the deal goes ahead. If the responder rejects, neither player gets anything. In both cases, the game is over. Obviously, rational responders should accept even the smallest positive offer, since the alternative is getting nothing. Proposers, therefore, should be able to claim almost the entire sum. In a large number of human studies, however, conducted with different incentives in different countries, the majority of proposers offer 40 to 50% of the total sum, and about half of all responders reject offers below 30%.

The irrational human emphasis on a fair division suggests that players have preferences which do not depend solely on their own payoff, and that responders are ready to punish proposers offering only a small share by rejecting the deal (which costs less to themselves than to the proposers). But how do these preferences come about?...


Kin selection

This algorithm is certainly motivated by biological principles. Cooperative hunting in group of animals required altruistic elements. Kin selection was suggested by William D. Hamilton (1936-2000) by using genetic arguments. Hamilton’s rule says that:

\[ r > \frac{C}{B}. \]  

(9.11)

Here, C is the fitness cost to the altruist and B is the fitness benefit to the individual helped, and r is a measure of genetic relatedness. Hamilton’s kin selection principle motivated Edwards Wilson to explain altruism, aggression, and other social behaviors in terms of biological evolution. His book on what he called sociobiology [564] dealt mostly with social animals (such as ants), and a single chapter with humans, provoked sharp debates. The opponents of sociobiology were headed by leading (admittedly leftist) evolutionary biologists, Richard Lewontin and Stephen Jay Gould (1941-2002) who attacked sociobiology for supporting biological determinism. Biological determinism may have as they argued, serious negative social consequences. Sociobiology has been replaced by evolutionary psychology, a less direct, more neutral theory to explain the evolution of human behavior and culture by mechanisms of natural selection [46].
Direct reciprocity

Iterated prisoner’s dilemma is a typical example of direct reciprocity. It is obvious that cooperation may emerge not only among relatives.

Cooperation, however, may exist also on more general form of mutual “reciprocity”, and it might occur in the animal world between different species. An animal helps another and expects help back in the future ([517]): altruism might be mutual. The mechanism suggests that one player’s cooperation increases the probability of the cooperation of the other later. Also, it turned out that while TFT proved to be twice the champion, it has some weakness as well; it is somewhat rigid. Sometimes defections are not intentional, and therefore a certain degree of forgiveness proved to be more advantageous than strict retaliation. A modified strategy, Generous tit-for-tat showed improved performance. An other algorithm Win-stay, lose-shift quickly corrects accidental mistakes and it exploits the other player if she chronically cooperates. It was suggested [325] that there is a simple relationship to characterize the situation when direct reciprocity can lead to evolution of cooperation: the probability $w$ of the encounter between the same individuals should be larger than the the cost-benefit ratio of the altruistic act:

$$w > C/B.$$  \hfill (9.12)

Reciprocal altruism is most likely very rare (or probably non-existent) in animal societies, cooperation is the result of selfishness or kin selection.

Indirect reciprocity

Martin Nowak (roots in Vienna) and Karl Sigmund (Vienna) [386] offered a mathematical model to show that cooperation can emerge even if recipients have no chance to return the help to their helper. This is because helping improves reputation, which in turn makes one more likely to be helped. The indirect reciprocity is modeled as an asymmetric interaction between two randomly chosen players. The interaction is asymmetric, since one of them is the “donor”, who can decide whether or not cooperate, and the other is a passive recipient. However, the result of the decision is not localized, it is observed by a subset of the population, who might propagate the information. Consequently, the decision to cooperate might increase one’s reputation. Those people who are considered more helpful, have a better chance to receive help. The calculation of indirect reciprocity is certainly not easy. A cooperative donor would like to cooperate with a player, who is most likely a cooperator,
and would not cooperate with a defector. The probability, $q$, of knowing someone’s reputation should be larger than the cost-benefit ratio of the altruistic act:

$$q > C/B. \quad (9.13)$$

Evolutionary game theory suggest that indirect reciprocity might be a mechanism for evolution of social norms.

**Network reciprocity**

While in a well-mixed population pure natural selection would favor to defectors, interaction in real populations can better characterized by spatial structures or social networks, where the probability matrix of the binary interaction between individuals has some structure. Recently evolutionary graph theory was offered [315] to study the effect of the network structure on evolutionary dynamics. It was found, that for many network structures (including small-world and scale-free networks) a simple rule describe the condition of favoring cooperativity. The cost-benefit ratio should exceed the average number of neighbors, $k$:

$$B/C > k. \quad (9.14)$$

**Group selection**

The concept of multilevel selection offers that selection acts not only on individuals but on groups, too.

A model for group selection

A population is subdivided into groups. Cooperators help others in their own group. Defectors do not help. Individuals reproduce proportional to their payoff. Offspring are added to the same group. If a group reaches a certain size, it can split into two. In this case, another group becomes extinct in order to constrain the total population size. Note that only individuals reproduce, but selection emerges on two levels. There is competition between groups because some groups grow faster and split more often. In particular, pure cooperator groups grow faster than pure defector groups, whereas in any mixed group, defectors reproduce faster than cooperators. Therefore, selection on the lower level
(within groups) favors defectors, whereas selection on the higher level
(between groups) favors cooperators. This model is based on “group
fecundity selection,” which means that groups of cooperators have a
higher rate of splitting in two. We can also imagine a model based on
“group viability selection,” where groups of cooperators are less likely
to go extinct.

(from [325].)

Under certain conditions, denoting by \( n \) the maximum group size, and by
\( m \) the number of groups, group selection supports evolution of cooperation if

\[
B/C > 1 + (n/m) \tag{9.15}
\]

**Constructive evolution?**

Evolutionary game theory suggests that selection and mutation, the funda-
mental principles of evolution can be supplemented with possible mechanisms
for the evolution of cooperation. In each mechanism the benefit-to-cost ra-
tio of the altruistic act should exceed some critical value. The appropriate
ratio between competition and cooperation may be the the driving force of
constructive evolution.

### 9.3 Widening the Limits to Predictions: Earthquake,
Eruptions Epileptics Seizures, and Stock Market Crashes

While the study of the eruptions of earthquakes, the onset of epileptic seizures,
and the crashes of stock market traditionally are investigated by very differ-
ent disciplines, and in departments, which differ very much in their scientific
culture, complex system approach emphasizes the similarities and offers some
common methods to predict the behavior of these systems, and/or understand
the inherent limits of their predictability.
9.3.1 Scope and limits of predictability

Layman believe that extreme events both in nature and society, such as earthquakes, landslides, wildfires, stock market crashes, destruction of very tall tower buildings, engineering failures, outbreak of epidemics etc. are surprising phenomena, and their occurrence does not follow any rule. Of course, such kinds of extreme events (birth, death, marriages, career steps, etc) are rare, but they influence our everyday lives dramatically. Can we understand, assess, predict and control these events? (An extreme event should not be necessarily negative.) For a very recent excellent, edited book about extreme natural and social events is [7].

Complex systems theory offers different approaches to model these, generally large-scale (global) macroscopic events, which are generated from small-scale (local) microscopic interactions. There are two classes of theoretical approaches, which give quite different answers for the question of eventual predictability.

One possibility is to say, that big earthquakes are nothing else but small earthquakes, which don’t stop. The consequence is that these critical events would inherently be unpredictable, since they don’t have any precursors. This approach is called as self-organized criticality (SOC) and was championed by Per Bak [38].

Didier Sornette, originally a geophysicist, based on the analogy between natural and financial crisis, wrote an exciting, and therefore controversial best-seller [472] about the reasons, and the eventual predictability of the stock market crashes. According to his arguments, catastrophic events (or at least a class of them) result from accumulating amplifying cascades. Based on the

---

Footnote: Per Bak (1947-2003) formulated the problem this way: "How can the universe start with a few types of elementary particles at the big bang, and end up with life, history, economics, and literature? The question is screaming out to be answered but it is seldom even asked. Why did the big bang not form a simple gas of particles, or condense into on big crystal? We see complex phenomena around us often that we take them for granted without looking for further explanation. In fact, until recently very little scientific effort was devoted to understanding why nature is complex. I will argue that complex behavior in nature reflects the tendency of large systems with many components to evolve into a poised, “critical state”, way out of balance, where minor disturbances may lead to events, called avalanches, of all sizes. Most of the changes take place through catastrophic events rather than by following a smooth gradual path. The evolution to this very delicate state occurs without design from any outside agent. The state is established solely because of the dynamical interaction among individual elements of the system: the critical state is self-organized. Self-organized criticality is so far the only known general mechanism to generate complexity."
hypothesis of this theory of intermittent criticality, many stock market crashes are generated by a slow building up of "subterranean forces", and their precursors may be detected. Were this hypothesis true, the predictability of these events may be imaginable.

First, a short comparative analysis of these phenomena is given.

9.3.2 Phenomenology

Earthquake eruption

*Richter scale, Gutenberg-Richter law and others*

We know from the media that the strength of the earthquakes are characterized by using the Richter magnitude scale. Accordingly, to the size of an earthquake a single number can be assigned. The magnitude (M) of an earthquake is proportional to the logarithm of the maximum amplitude of the earth’s motion. For example, an earthquake with magnitude 8, moves the ground 1000x more than a magnitude 5 earthquake.

There are many earthquakes, but only a few (well, still more than enough) make the headline news, as Figure 9.3 shows.

![Day vs. Maximum Amplitude vs. Frequency vs. Magnitude](http://simscience.org/crackling/Advanced/Earthquakes/GutenbergRichter.html)

*Fig. 9.3.* a. magnitude of earthquakes on different days of 1995 in Southern California; b. distribution of earthquakes with different magnitudes.

The number of earthquakes of magnitude M is proportional to $10^{-M}$:
$\log N(M) = -bM, \quad b \sim 1. \tag{9.16}$

$N(M)$ is the number of earthquakes of magnitude greater than $M$.

This statistical finding is called as the Gutenberg-Richter law. Earthquakes are not isolated events, after a mainshock in the same region there are a number of aftershocks with smaller magnitude. The temporal decay of these aftershocks is described with the equation

$$n(t) = \frac{K}{(c + t)^p}, \tag{9.17}$$

where $n(t)$ is the number of earthquakes $n$ measured in a certain time $t$, $K$ is the decay rate, $c$ is the “time offset” parameter; and the parameter $p$ typically falls in the range $0.75 - 1.5$. For $p = 1$, it is called the Omori’s law, stated by the Japanese seismologist Omori in 1894.

Fig. 9.4. Corrected Omori’s law: The decay rate of the magnitude of earthquakes shows hyperbolic decrease.

There is another empirical law (Bath’s law), which states that difference in the magnitude between a mainshock and its largest aftershock is 1.2, independent of the mainshock magnitude. The distribution of the magnitude of the aftershocks also follows the Gutenberg-Richter law.

Epileptic seizures

Epilepsy, as a neurological disease, is characterized by abrupt, unprovoked recurrent seizures, exhibiting pathological electrical activity. There are several types of seizures in epilepsy. They can roughly be classified into generalized seizures and partial seizures. Generalized seizures arise from abnormal electrochemical activity affecting the whole brain at the same time; partial seizures
arise from abnormal electrochemical activity affecting one part of the brain, but which may spread to other parts of the brain.

It was generally believed that the seizure is very really abrupt, and it begins just seconds earlier before its clinical onset. There are some indications, although the clinical relevance is still debated, that significant changes in the electrical activity of the brain could be detected by EEG (electroencephalogram), hours in advance of the onset of seizures. If we had a reliable precursor detection technique ([362, 248, 138]), antiepileptic drugs could be administered by some implantable devices to intervene in time to suppress the development of epileptic seizures.

To be able to make a real prediction, the EEG must show at least two different stages ([458]). First, there has to be a time period between the prediction and the earliest occurrence of the seizure, called seizure prediction horizon (SPH). This period is necessary for example, for drug administration. There should be another time interval, during which the predicted seizure occurs, called the seizure occurrence period (SOP). For obvious reasons, prediction methods should avoid false alarms: false predictions would lead to impairment due to possible side-effects of interventions.

Stock market crash: some comparative analysis

Stock market crash is an abrupt fall of stock prices in a significant region of the stock market. Such kinds of crashes emerge due to the interaction of inherent economic factors and human actions.

There were a number of big crashes in the last few centuries. Kenneth Galbraith (1908-2006), the very influential economist wrote an easily readable small book with the title “A short history of financial euphoria” [191]. The book explains why and how the interplay of irrational expectations and specific economic factors generate boom and bust financial dynamics in each several decades.

Now historical crashes will be briefly reviewed. For more details see e.g. “Stock Market Crash!Net” http://www.stock-market-crash.net/.

The Tulip Bulb Mania

The Tulip Bulb Mania emerged in Holland around 1635. Tulip bulbs first were bought for their aesthetic value, but as their prices increased, they became
Fig. 9.5. Basic operation of a prediction method during an interictal and a preictal period. Seizure onset is marked by vertical lines. a Examples of EEG recordings and b exemplary time course of a feature extracted by a seizure prediction algorithm. The solid, horizontal line indicates the threshold for raising alarms. Alarm events and two consecutive time intervals characterizing a prediction, the seizure prediction horizon SPH and seizure occurrence period SOP are illustrated in c. Note the different time scales for the EEG data and the feature time series. Adapted from [458]

subject of buying and selling. People bought them to make a (large) profit. There was a month when its value increased twenty-fold. At a certain point the Dutch government attempted to control the mania. After its regulatory actions some informed speculator realized that the price could not become more inflated and started to sell bulbs. Other people soon noticed that the demand for tulips could not be maintained. Their attitude propagated very rapidly among people interested in the business, and soon panic, a social collective phenomenon, emerged. During six weeks there was a 90% reduction in its price.

The South Sea Bubble

A very bad financial bubble occurred in England between 1711 and 1722.

---

5 The term “bubble” was adopted in a poem of Jonathan Swift (1667-1745): The Nation too too late will find, Computing all their Cost and Trouble,
The British government offered a deal to South Sea Company to finance a big state debt emerged during the War of the Spanish Succession. The South Sea Company traded with South America (excluding Brazil, as it was a Portuguese territory). After a rumor that South Sea Company had been granted full use of Latin American ports, “It became extremely fashionable to own South Sea Company shares”. It turned out at a certain point for the leaders of the company that the actual commerce did not produce profit, and money was generated mostly from issuing stocks, so their shares became strongly overvalued. Soon after the owners started to sell shares, there was a panic among shareholders and the market crashed. Factors, such as speculations, unrealistic expectations and corruption contributed to the emergence of the bubble. Newton first saw the bubble, but later lost a lot. 6 Jonathan Swift also lost a large amount of money, so he was motivated to write a satire about the British society, which we know as Gulliver’s Travels.

6 “Sir Isaac Newton, scientist, master of the mint, and a certifiably rational man, fared less well. He sold his 7,000 shares of stock in April for a profit of 100 percent. But something induced him to reenter the market at the top, and he lost 20,000. “I can calculate the movement of the stars, but NOT the madness of men.”

In a recent reanalysis of this bubble it was argued by Temin \(^7\) and Voth \(^{509}\) that “Hoare’s Bank, a fledgling West End London banker, knew that a bubble was in progress and that it invested knowingly in the bubble; it was profitable to ride the bubble.”

**Stock Market Crash of 1929**

Since the stock market crash of 1929 is an emblematic event, it was extensively analyzed. Many investors adopted a high-gain or high-loss strategy, called **leverage**, but the stock seemed to be very safe. These, so called **margin investors** knew that there was a bull market \(^8\) from 1921, so believed that stock market always went up. When the bear market started after some intervention of the Federal Reserve (the central bank of the US, founded in 1913) panic emerged on Thursday, October 24th, 1929. Margin investors became bankrupt. Furthermore, since banks also invested their deposit in stock market, even the depositors’ money was also lost.

---

\(^{7}\) This is the second citation for “Temin”. The economist Peter Temin is the younger brother of the geneticist Howard Temin, who discovered the reverse transcriptase discussed in 2.1.2.

\(^{8}\) The terms **bull market** and **bear market** are adopted for longer time periods when the prices are rising or falling, respectively.
Stock Market Crash of 1987

The Black Monday of October 19, 1987 is remembered as the largest (or second largest) one day stock market crash. While the bull market started in 1982, it accelerated in 1986. The growth of some companies by buying others was one driving force in the stock market. The appearance of personal computers was also considered as an excellent opportunity to increase profit. As it happened in other times, euphoria became the predominant attitude, and many new investors entered the market. Strong economic growth implied inflation, and the Federal Reserve’s intervention by raising short term interest rates initiated instability. The big trading firms had portfolio insurance, which all of them wanted to use simultaneously. While the clients asked their brokers to sell, they were not able to act due to the large number of their clients. Together with the fall of the Dow Jones index, the national stock markets also collapsed. The central banks, however, managed to control the crisis. The reduction of short term interest rate by the Fed helped, and firms bought back their, now undervalued stocks.

The Nikkei Bubble

The bull market in Japan was driven first by the natural recovery from the effects of World War II. The ratio of people working in industry had strongly

Fig. 9.10. Price dynamics of rise and fall of the Nikkei Bubble. Adapted from [http://www.stockmarket-crash.net/nikkei.htm](http://www.stockmarket-crash.net/nikkei.htm) ©Elliot Wave International.
been increased, large corporations, who offered life-time employment, and received loyalty, had been formed. Japanese firms gained fame for “copying and improving Western products and selling them for much cheaper”. In the next stage, the oil crisis combined with the development of high-tech Japanese car industry (as opposed to the low-tech US automotive technology) implied the permanent increase of the Japanese stock market. Finally, with the emergence of electronic giants (such as Hitachi and Sony) it was believed that Japan would dominate the whole microelectronics industry. Speculation started (also in the real estates market: the price of a home in the Tokyo area was much higher than even in Manhattan). To control the inflated economy interest rate was increased. The Nikkei index crashed from 40,000 to 10,000 during several months. As it happens, corruption came to light. The recovery period was extremely long: the Japanese stock market showed a fourteen year long bear market. It is probably over; “Tokyo’s Nikkei share average ended 2006 with a 6.9 percent gain for the year, marking its fourth straight year of growth and its longest bull run in nearly two decades”, as the Reuter reports on December 29th, 2006, (see http://in.news.yahoo.com/061229/137/6anyi.html).

*The Nasdaq Bubble*

![Fig. 9.11. Price dynamics of rise and fall of the Nasdaq Bubble. Adapted from http://www.stock−market−crash.net/nasdaq.htm ©Elliot Wave International.](image)

There was a rapid recovery from the 1987 crisis. The most effective driving force of this process was the propagation of personal computers to the
everyday life, from all type of business applications to electronic entertain-
ment. Hardwares, produced in different geographical regions, were very simi-
lar. What made the difference was the software. Stock prices of software houses
showed a strong increase. Many small companies appeared in the stock mar-
et. Initial shareholders (often even employees of these companies), became
rich very soon due to the increase of the values of these technological stocks.
The Internet catalyzed the super-exponential increase of the stocks. The Nas-
daq Composite Index increased during four years with from 600 to 5000 with
a peak on March 2000 driven by the irrational expectations. Investors realized
at certain point, that the abrupt enrichment of the “dotcom” companies was
a bubble. The Nasdaq crashed to 800 for 2002. Today (February 2nd, 2007)
its value is 2,475.88 and showed 0.30 percentage increase for the day. (So,
probably we are not close to the next bubble).

*History (even financial history) repeats itself*

In all the six cases a very sharp (often, but not always super-exponential:
please note that the vertical axes are in logarithmic scale) growth is followed
by an abrupt fall. The early stage of the rise is economically justified, and it is
related to the appearance of new products, generally related to technological
changes. The initial rise is amplified, (more accurately) over-amplified by in-
creased investment due to expectations for rapid and big profit. *Self-organized
cooperativity* appears when many people invest synchronously. However, the
velocity of the price increase due to a positive feedback mechanism cannot be
sustained. The process becomes unstable. In an unstable situation small, local
perturbations might have dramatic consequences, and say, a relatively small
increase of the interest rate might initiate a crash.

Why there are relatively small number of crashes? In general cases people
remember that price should have a peak sooner or later, and this expectation
has a deterring effect for the potential buyers. In regular cases this feedback
effect can stabilize the price. Stock market crash is the consequence of the
impairment of the positive-negative regulatory loop.

### 9.3.3 Statistical analysis of extreme events

It is not surprising to hear such kinds of questions:

- What is the probability of having big earthquake in California within a
  year?
How large might
- a possible stock market crash be tomorrow?
- the lowest daily return (the minimum) be?
- the highest daily return (the maximum) be over a given period.

Standard statistical procedures neglect data deviating “very much” from the others. (These data are called outliers). Extreme value analysis uses statistical methods to analyze the rarely occurring events. Very characteristically, extreme events occur in the tails of probability distributions as a function of the “size” of the events (such as energy, duration etc . . .).

Gumbel distribution and other extreme value distributions

“It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.’ (Emil Gumbel)

Emil Gumbel (1891-1966) a famous pacifist, contributed very much to the establishment of statistical methods to describe extreme deviations from an “average” behavior. Extreme value analysis, a branch of mathematical statistics, estimates the probability of extreme floods, large insurance losses, market risk, freak waves, tsunamis, etc.

The Gumbel distribution is defined by the probability density function (PDF):

\[ f(x; \mu, \sigma) = \exp\left(-\left(\exp\left(-\frac{x - \mu}{\sigma}\right)\right)\right) \]

for

\[-\infty < x < \infty\]

It has a location (\(\mu\)), and a scale (\(\sigma\)) parameter, but the Gumbel PDF has no shape parameter. The location is the value with the greatest observed frequency, while the scale characterizes the practical minimal and maximal values. This means that the Gumbel pdf has only one shape, which does not change. Fig. 9.12 visualizes the parameter-dependence of the Gumbel distribution.
Fig. 9.12. The shape of the PDF is skewed to the left. By increasing $\mu$ the pdf is shifted to the right. As $\sigma$ increases the PDF spreads out and becomes shallower.

While the Gumbel distribution shows a light-tail (exponential decay), other classes of the “extreme value distributions” behave different. The Fréchet distribution has heavy tail, and the Weibull distribution is characterized by a bounded tail. Though the Gumbel distribution is not symmetric, still its two parameters could be well estimated.

The Gumbel distribution proved to be acceptable for describing the majority of Asian stock markets [119], but the Fréchet distribution fits modeling extreme rainfall better [291]. While the Gumbel distribution proved to be a very good method to estimate the probabilities of extreme events, such as floods, forest fires, financial losses, there are many new results (e.g. for its derivation random variables were assumed and so restricted to be independent and identically distributed), see [8].

Distributions with power-law tails (like the distributions of earthquakes and avalanches) have extreme value statistics described by Fréchet distributions. Distributions that have a strict upper or lower bound have extreme value distributions that are described by Weibull statistics.

Once again, in light of the recent re-emerged interest in power law distributions, what they are and how theory are used in analyzing extreme events?

**Power law distributions**

Long tail distributions have already been discussed in section 6.3. Power law distribution is characterized by the distribution function

$$P(\xi > x) = ax^{-k}$$  \hspace{1cm} (9.20)

where $k$ is the shape parameter.

In estimating the possible occurrence of extreme events it is important to know the probability that a particular sample will be larger than $x$:
\[ P(x) = \left(\frac{x}{x_{\text{min}}}\right)^{-\alpha+1} \]  

(9.21) 

while \( \alpha > 1 \).

The spatiotemporal distribution of the number of Californian earthquakes was analyzed [450], and showed power law distribution 9.13.

![Empirical probability density functions of the number of earthquakes in the space–time](image)

**Fig. 9.13.** Empirical probability density functions of the number \( r \) of earthquakes in the space–time \([5 \times 5 \text{ km}^2 \times 1 \text{ day}]\), from California SCEC catalog. Adapted from [450]

According to the analysis of Mantegna and Stanley (presented in a highly cited paper on econophysics) [327] the S&P500 index initially shows lognormal distribution followed by a heavy tail power law distribution, as Fig. 9.14 shows.

Power laws appear to describe different financial fluctuations, such as fluctuations in stock price, trading volume, and the number of trades.
9.3 Widening the Limits to Predictions

9.3.4 Towards predicting seizures

The analysis of EEG curves based on dynamical approach offered methods to separate the different qualitative regions before and during a seizure. The challenge is to define good measures. 9

Measures based on the so-called recurrence points were offered by [314]. Three different measures were derived. One of them (called recurrence rate) measures the number of events, when a trajectory visits a predefined region of a state. While by visual inspection only two stages (inter-ictal and ictal periods) can be discriminated, the analysis helped to show the existence of a third stage, the pre-ictal period, which separated the two stages previously mentioned, as Fig.9.15 shows.

9 For the analysis of complex dynamics by recurrence plot analysis, see http://www.recurrence-plot.tk/ February 16th, 2007).
Fig. 9.15. The upper figure shows an EEG curve, while the bottom visualizes the temporal change of the recurrence rate. Seizure initiation is accompanied by lower signal complexity due to the higher synchrony.

In our own lab in Budapest, Zoltán Somogyvári analyzed the data came from Magda Szente physiological lab in Szeged [470]. The analysis of the slow dynamics of epileptic seizure led to the following results:

- epileptic seizures may begin much earlier than their clinical onset
- relative self-excitation is a control parameter to induce/suppress epilepsy.

He also found that complexity reduction happens during the transition to the seizure, as Fig. 9.16 shows.

Fig. 9.16. There is a remarkable complexity reduction. Upper panel: postsynaptic potential in time. Lower panel: phase plane at the pre-epileptic and the seizure region. The first plot is shows some chaotic character, while the second is much more ordered.
We might have some sense about the general, formal mechanisms of the onset of critical states.

Common message to explain the emergence of epilepsy and emergence of earthquake:

Transition from pre-epileptic and pre-seizmic states toward critical states:

- gradually increased spatial and temporal correlations
- more dramatic change in energy release

better supports the “intermittent criticality” than the “self-organized criticality” hypothesis.

based on the argument of [314].

9.3.5 Towards predicting market crashes: analysis of price peaks

Irrational expectations often initiate a super-exponential growth, which leads to instability and a crash. Are there stock-market behaviors controlled by general laws? Bertrand Roehner, an econophysicist from Paris, hopes to find the dynamic equations (i.e. causal explanation) of price dynamics. He first starts with “modest goals”, such as finding regular patterns behind speculations [435]. Specifically, the shape of price peaks is characterized by two symmetry parameters. An idealized curve is shown in Fig. 9.17. Three time points, i.e. the initial time, the time of the peak and the time of the stop of the fall, denoted by \( t_1, t_2 \) and \( t_3 \) respectively, should be determined. To each time point a price value, \( p_1, p_2 \) and \( p_3 \) is assigned. The duration of the bull market and bear market is measured by \( T_{\text{up}} := t_2 - t_1 \), and \( T_{\text{down}} := t_3 - t_2 \).

The rising phase of curve leading to peaks is rarely slower than exponential, and the real important difference whether it is exponential (so it tends to infinite during infinite time), or super-exponential leading to finite-size singularities. The peak and bottom amplitude is defined as \( A := p_2/p_1 \), and
\( B := \frac{p_3}{p_1} \). Roehner suggests an empirical relationship to resilience: there seems to be correlation between \( A \) and \( B \):

\[
B = aA + b,
\]

where the value of \( a \) is about 0.4 – 0.5. From the positive sign of \( a \) implies that for larger peak amplitude, the bottom amplitude is also higher. With other words, large peak amplitude defends the system to fall to low.

As we have known since Newton, the best predictions can be given by dynamical models. So, let’s discuss the different versions of dynamical models to catch the most important aspects of extreme events.

### 9.3.6 Dynamical models of extreme events

**Model of self-organized criticality: Sandpile model**

Self-organized criticality suggests that the same effect may lead to small, but also to very large avalanches, so the outcome is not really predictable. A famous toy model was offered Bak, Tang, and Wiesenfeld 1987 [39].

**Self-organized criticality: a generating algorithm**

Each point \( z(x, y) \) in the grid has a number \( n \) associated with it (say, average slope of the sand pile at that point).
For a randomly selected point:
\[ z(x, y) \rightarrow z(x, y) + 1 \]
if \( z(x, y) > \) threshold, then
\[ z(x, y) = z(x, y) - 1 \]
\[ z(x \pm 1, y) \rightarrow z(x \pm 1, y) + 1 \]
\[ z(x, y \pm 1) \rightarrow z(x, y \pm 1) + 1 \]

If this redistribution results in \( n \) to be too big for any nearby grid points
than start next iteration; else: stop and take randomly another point
\( z(x, y) \)

Figure 9.18 illustrates the actual and simulated avalanche distributions. Avalanches show power law distribution, i.e. perturbation of a single point
may result avalanche with very different size.

Fig. 9.18. Avalanches: self-organized criticality. In the lefthandside figure the num-
ber of avalanches and number of grains involved in an avalanche are plotted in
log-log space. The middle figure shows a region with avalanches While their sizes
are different, each of them were triggered by the addition of a single grain. The
righthandside figure shows the frequency distribution of avalanches with differ-
ent size. From Introduction to Self-Organized Criticality & Earthquake; Winslow
http://www.geo.lsa.umich.edu/ruff/Geo105.W97/SOC/SOCeq.html

The model of self-organized criticality suggests that one mechanism to-
wards catastrophic events are the occurrence of small events which don’t stop,
and by this way they are leading to unpredictability.
‘Explosions’: Finite time singularities and intermittent criticality

Finite time singularity roughly speaking means that a dynamical variable gets an infinite value during finite time. This phenomenon is qualitatively different from the exponential growth, when infinite value can be attained during infinite time only. Figure 1.3 showed both exponential and super-exponential growth. The simplest representative dynamical evolution equation leading to a finite-time singularity is:

\[
\frac{dx}{dt} = x^m, \text{ its solution is: } x(t) = x(0) \left( \frac{t_c - t}{t_c} \right)^{-\frac{1}{m-1}}, \quad (9.22)
\]

with \( m > 1 \), while \( m = 0 \) leads to linear, and \( m = 1 \) to exponential growth, as the figure 9.19 shows.

Such kinds of equation with \( m > 1 \) implements large, “higher-than-linear” (HTL) positive feedback, which seems to be a general mechanism behind finite time singularities. It ensures that the “instantaneous doubling time” tends to zero after a finite period.

General mechanism for finite time singularities

As cybernetics envisioned, the lack of the stabilizing effects of negative feedback mechanisms may lead to catastrophic consequences. If there are no mechanisms to compensate for the effects of HTL positive feedback, the processes lead to finite time singularities.
Positive feedback seems to be a general mechanism behind eruption of earthquakes, financial crisis (stock market crashes, hyperinflation), and epileptic seizures, as we review here briefly.

**Positive feedback phenomena leading to earthquakes**

The analysis of the role of positive feedback mechanisms in amplifying seismic activities led to the conclusion [453] that great earthquakes are critical singular events, and even simple models help to explain the generation of such kinds of singularities. Their statement is that, opposed to conclusions of the theory of self-organized criticality, the precursors of the dynamic process leading to singularities can be found.

What kinds of mechanisms may lead to accelerated seismic activity?

Models that have been proposed to describe accelerating seismicity before large earthquakes fall into two general classes: (i) related to mechanical feedback; (ii) based on the decay of the “stress shadow” from a prior event.

(i) The power law acceleration of regional seismicity is due to positive feedback in the failure process at constant large-scale stress. This feedback can be the result of stress transfer from damaged to intact regions, or it can result from the effect of damage in lowering the local elastic stiffness. Sammis and Sornette [453] showed that for both cases simple models lead to finite-time singularities.

(ii) The phenomenon of “stress shadow” inhibits seismic activity after a big earthquake. After the 1906 earthquake in San Francisco, the seismic activity dramatically reduced for about seventy years. The recovery of a stress shadow might be followed by accelerated seismicity resulted from the increasing stress as the shadow is eroded by tectonic loading. Finally this acceleration may lead to finite time singularities.

**Positive feedback phenomena leading to financial crisis**

The onset of two types of financial crisis, stock market crash and hyperinflation is generated by the positive feedback between the actual and the expected growth rate.

Large stock market crashes, are the social analogues of big earthquakes (as all analogies, this is also has scope and limits). Sornette’s view is that the stock market crash is not induced by single local events (such as a raise
in the interest values, or other governmental regulations), but due to the unsustainable velocity of price increase. This speculative increase will take the system or more and more unstable situation. Finally, the market collapses due to any small disturbance.

While the conventional economic theory, which is based on the equilibrium between demand and supply, the complex systems approach suggests that partially due to our susceptibility to imitate each others behavior, there is a period when both demand and supply increase which finally lead to singularities. Equilibrium theory works well when the negative feedback effects expresses its stabilizing effect to the increase due to positive feedback. While in “normal situations” the activities of “buyers” and “sellers” neutralize each other, in “critical situations” there is a cooperative effects due to the imitative behavior (“everybody wants to buy since everybody else has already bought”). So, the positive feedback is HTL, and the increase is unsustainable. For the comparison of the “equilibrium” and the “intermittent criticality” approaches see figure 9.20.

---

**Fig. 9.20.** Comparison of the equilibrium and the ”intermittent criticality” approaches. The equilibrium between demand and supply can be stabilized due to the interplay between positive and negative feedback (upper panel). In case of the dominance of HTL positive feedback the process tends towards crisis, since the growth cannot be sustained.

*Positive feedback phenomena leading to epilepsy*

The stable dynamic operation of the brain is based on the balance of excitatory and inhibitory interactions. The impairment of the inhibitory synaptic transmission implies the onset of epileptic seizures. Epileptic activity occurs in a population of neurons when the membrane potentials of the neurons are ”abnormally” synchronized. Both experiments and theoretical studies suggest the
existence of a general synchronization mechanism in the hippocampal CA3 region. Synaptic inhibition regulates the spread of firing of pyramidal neurons. In experimental situations inhibition may be reduced by applying drugs to block (mostly) GABA-A receptors. If inhibition falls below a critical level, the degree of synchrony exceeds the threshold of normal patterns, and the system dynamics switches to epileptic pattern. Collective phenomena occurring in neural networks, such in case of disinhibiton-induced epilepsy have been studied successfully studied by combined physiological computational studies by Roger Traub and Richard Miles. Their book [516] is one of my favorite, and influenced me very much.

![Fig. 9.21. An excitatory-inhibitory networks supplemented with self-excitatory and self-inhibitory connections](image)

**A key to the prediction?: log-periodic corrections**

*Log-periodic correction of the dynamics towards crisis*

On a refined scale it was found for different systems and variables (from acoustic energy to stock market price) that the “power law dynamics” is modulated with log-periodicity, as Figures 9.22 and 9.23 show.

To describe also the log-periodic correction of the “power law”, dynamics leading to finite-time singularities has been corrected.

In case of stock market crash, the simple price equation

$$\log[p(t)] = A + B(t_c - t)\beta$$  \hspace{1cm} (9.23)

has been refined by Sornette to as

$$\log[p(t)] = A+B(t_c-t)^\beta \rightarrow \log[p(t)] = A+B(t_c-t)^\beta [1+C \cos(\omega \log((t_c-t)/T))]$$  \hspace{1cm} (9.24)
Fig. 9.22. Energy release rate of approaching rupture. Based on Sornette 2002

Fig. 9.23. Examples of log-periodicity: The S&P 500 index on Wall Street prior to the October 1987 Crash. The US dollar against the Deutschmark and Swiss Franc prior and around the collapse. The continuous line is a regression curve with a power law superimposed log-periodic oscillation. Based on [471].

From a formal point of view, log-periodicity reflects the fact that critical exponent or more generally dimensions can be “complex”, (of course, here “complex” means such kinds of numbers which when squared give negative values).

Fractal dimensions have an important role, as we know, to characterize strange attractors related to chaotic dynamics. The notion of the “complex fractal” means a further generalization of the notion. The measured variable shows “power law” dynamics related to the real value of the exponents, and a log-periodic corrections due to the imaginary part of the exponent.
What did we learn?

The theory of complex systems suggests:

- extreme events:
  - may be predicted
  - their precursors can be detected

- there are methodological similarities to analyze and model different “critical events” occurring in physical, life and social phenomena

- unbalanced higher-than-linear positive feedback is the source of crisis

- there are initial results and many open problems.