Utility, games, and narratives

Fioretti, Guido

University of Bologna

10 August 2009

Utility, Games, and Narratives

Guido Fioretti
University of Bologna
guido.fioretti@unibo.it

August 10, 2009

1   Introduction

This chapter provides a general overview of theories and tools to model individual and collective decision-making. In particular, stress is laid on the interaction of several decision-makers.

A substantial part of this chapter is devoted to utility maximization and its application to collective decision-making, Game Theory. However, the pitfalls of utility maximization are thoroughly discussed, and the radically alternative approach of viewing decision-making as constructing narratives is presented with its emerging computational tools. In detail, the chapter is structured as follows.

Section (2) presents utility maximization and Game Theory with its Nash equilibria. The most important prototypical games are expounded in this section. Section (3) presents games that are not concerned with Nash equilibria. Section (4) illustrates the main paradoxes of utility maximization, as well as the patches that have been proposed. Section (5) expounds the vision of decision-making as constructing a narrative, supported by a rare empirical case-study. Section (6) aims at providing computational tools for this otherwise literary vision of decision-making. Finally, section (7) concludes by assessing the pros and cons of competing approaches.

This chapter touches so many issues that a complete list of references to the relevant litterature would possibly be longer than the chapter itself. Instead of references, the names of the most important scholars in each field have been made, so the interested reader should be able to reconstruct the relevant bibliography by herself. A few exceptions have been made for very specific works, that have been mentioned in footnotes.

2   Utility and Games

Let $\{a_1, a_2, \ldots a_m\}$ be a set of *alternatives*. Let a_i denote a generic alternative, henceforth called the i-th alternative where $i = 1, 2, \ldots m$.

By selecting an alternative, a decision-maker obtains one out of several possible *consequences*. Let $\{c_{i1}, c_{i2}, \ldots c_{in}\}$ be the set of possible consequences of alternative a_i. Let c_{ij} denote a consequence of a_i, where $i = 1, 2, \ldots m$ and $j = 1, 2, \ldots n_i$.

The *expected utility* of alternative a_i is:

$$u(a_i) = \sum_{j=1}^{n_i} p(c_{ij})u(c_{ij}) \tag{1}$$

where $p(c_{ij})$ is the probability of obtaining consequence c_{ij} and $u(c_{ij})$ is the utility of consequence c_{ij}.

It is suggested that the one alternative should be chosen, that maximizes expected utility. Frank Ramsey, Bruno De Finetti and Leonard Savage demonstrated that this is the only choice coherent with a set of postulates that they presented as self-evident.

Among these postulates, the following ones are mentioned in this chapter:

Transitivity Transitivity of preferences means that if $a_i \succ a_j$ and $a_j \succ a_k$, then $a_i \succ a_k$.

Independence Independence of irrelevant alternatives means that $a_i \succ a_j$ iff $a_i \cup a_k \succ a_j \cup a_k$, $\forall a_k$.

Completeness Completeness means that $\forall (a_i, a_j)$, a preference relation $\succ$ is defined.

Utility maximization is neither concerned with conceiving alternatives, nor with the formation of preferences, which are assumed to be given and subsumed by the utility function. Probabilities may eventually be updated by means of frequency measurement, but at least their initial values are supposed to be given as well. Thus, utility maximization takes as solved many of the problems with which its critics are concerned.

Utility maximization takes a gambler playing dice or roulette as its prototypical setting. In fact, in this setting the set of alternatives is given, utilities coincide with monetary prizes and probabilities can be assessed independently of utilities. For some critics of this decision theory, gambling is not an adequate prototype of most real-life situations.

The interaction of several utility-maximizing decision-makers is covered by *Game Theory*. Game Theory assumes that collective decision-making is the combination of several individual decision processes, where each individual maximizes its utility depending on the alternatives selected by the other individuals. Since selecting an alternative implies considering what alternatives other players may select, alternatives are generally called *strategies* in Game Theory.

Utility takes the name *payoff* in Game Theory. Games in which one player does better at another's expense are called *zero-sum games*. Games may be played once, or they may be repeated.

In general, Game Theory attempts to find out equilibria in games. If each player knows the set of available strategies and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the current choice of strategies and the corresponding payoffs constitute a *Nash equilibrium*. Since this implies stepping in another player's shoes in order to figure out what (s)he would do if one selects a particular strategy, Nash-equilibria are fixed points in self-referential loops of the kind "I think that you think that I think ...".

Note that being at a Nash equilibrium neither implies that each player reaches the highest possible payoff that (s)he can attain, nor that sum of all payoffs of all players

is the highest that can be attained. This is a concern for economics, for it implies that individual interest may not produce the common good.

If a game is repeated a Nash equilibrium may be either realized with *pure strategies*, meaning that players choose consistently one single alternative, or *mixed strategies*, meaning that players select one out of a set of available strategies according to a probability distribution. Accepting the idea of mixed strategies often allows to find Nash equilibria where there would be none if only pure strategies are allowed. However, the realism of random decision-makers choosing strategies according to a probability distribution may be questioned.

Most of the games analysed by Game Theory involve two, or in any case a very limited number of players. On the contrary, *evolutionary games* concern large populations of players playing different strategies, that are subject to an evolutionary dynamics regulated by *replicator equations*. Successful strategies replicate and diffuse, unsuccessful strategies go extinct. Occasionally, new strategies may originate by random mutation.

The equilibrium concept of evolutionary games is that of *evolutionarily stable strategies*. An evolutionary stable strategy is such that, if almost every member of the population follows it, no mutant can successfully invade. Alternatively, evolutionary games may be played in order to observe typical dynamics, in which case they become akin to the influence games that will be handled in § (sec:influenceGames).

The following games propose prototypical modes of human interaction. Games used by experimental economics in order to evince human attitudes do not pertain to this list.

2.1 The Battle of Sexes

Imagine a couple. The husband would most of all like to go to the football game. The wife would like to go to the opera. Both would prefer to go to the same place rather than different ones. If they cannot communicate, where should they go?

The payoff matrix in figure (1) is an example of Battle of the Sexes, where the wife chooses a row and the husband chooses a column. Aside, a generic representation of the game where $L < M$.

This representation does not account for the additional harm that might come from going to different locations and going to the wrong one, i.e., he goes to the opera while she goes to the football game, satisfying neither. Taking account of this effect, this game would bear some similarity to the Game of Chicken of § (2.7).

This game has two pure-strategy Nash-equilibria, one where both go to the opera and another where both go to the football game. Furthermore, there is a Nash equilibrium in mixed strategies, where the players go to their preferred event more often than to the other one.

None of these equilibria is satisfactory. One possible resolution involves a commonly observed randomizing device, e.g., the couple may agree to flip a coin in order to decide where to go.

	Opera	Football			
Opera	3, 2	0, 0		M, L	0, 0
Football	0, 0	2, 3		0, 0	L, M

Figure 1: A payoff matrix for the Battle of the Sexes (left) and its generic representation (right). The left number is the payoff of the row player (wife), the right number is the payoff of the column player (husband). In this generic representation, *L* is the payoff of the least preferred alternative whereas *M* is the payoff of the most preferred alternative.

	Stag	Hare			
Stag	3, 3	0, 1		C, C	S, B
Hare	1, 0	1, 1		B, S	D, D

Figure 2: A payoff matrix for the Stag Hunt (left) and its generic representation (right). The left number is the payoff of the row player, the right number is the payoff of the column player. In this generic representation, *C* is the payoff that accrues to both players if they cooperate, *D* is the payoff that accrues to both players if they defect from their agreement, *S* is the sucker's payoff and *B* is the betrayer's payoff.

2.2 The Stag Hunt

Rousseau described a situation in which two individuals agree to hunt a stag, which none of them would be able to hunt alone. Each hunter may eventually notice a hare and shoot at it. This would destroy the stag hunt, so the other hunter would get nothing.

An example of the payoff matrix for the stag hunt is pictured in figure (2), along with its generic representation. The stag hunt requires that $C > B \geq D > S$.

This game has two pure-strategy Nash-equilibria, one where both hunters hunt the stag, the other one where both hunters hunt a hare. The first equilibrium maximizes payoff, but the second equilibrium minimizes risk. There exists also a mixed-strategy Nash-equilibrium, but no payoff matrix can make the hunters play "stag" with a probability higher than $1/2$.

In addition to the example suggested by Rousseau, Hume provides a series of examples that are stag hunts. One example addresses two individuals who must row a boat. If both choose to row they can successfully move the boat. However if one does not, the other wastes his effort. Hume's second example involves two neighbours wishing to drain a meadow. If they both work to drain it they will be successful, but if either fails to do his part the meadow will not be drained.

Several animal behaviours have been described as stag hunts. For example, the

	Cooperate	Defect
Cooperate	3, 3	0, 5
Defect	5, 0	1, 1

C, C	S, B
B, S	D, D

Figure 3: A payoff matrix for the Prisoner's Dilemma (left) and its generic representation (right). The left number is the payoff of the row player, right number is the payoff of the column player. In this generic representation, C is the payoff if both players cooperate, D is the payoff if both defect from their agreement, S is the sucker's payoff, B is the betrayer's payoff.

coordination of slime molds. In times of stress, individual unicellular protists will aggregate to form one large body. Here if they all act together they can successfully reproduce, however the success depends on the cooperation of many individual protozoa. Also, the hunting practices of orca are an example of a stag hunt. Here orcas cooperatively corral large schools of fish to the surface and stun them by hitting them with their tails. Since this requires that fish do not have ways to escape, it requires the cooperation of many orcas.

2.3 The Prisoner's Dilemma

The Prisoner's Dilemma is a central subject in economics, for it apparently contradicts its basic assumption that common good arises out of self-interested individuals. This difficulty is eventually overcome by repeating the game.

The basic formulation of the Prisoner's Dilemma is as follows. Two suspects, A and B, are arrested by the police. The police has insufficient evidence for a conviction, and, having separated both prisoners, visits each of them offering the same deal: if one testifies for the prosecution against the other and the other remains silent, the betrayer goes free and the silent accomplice receives the full 10-year sentence. If both stay silent, both prisoners are sentenced to only six months in jail for a minor charge. If each betrays the other, each receives a five-year sentence. Each prisoner must make the choice of whether to betray the other or to remain silent. However, neither prisoner knows for sure what choice the other prisoner will make.

The Prisoner's Dilemma describes any situation where individuals have an interest to be selfish, though if everyone cooperates a better state would be attained. Examples may include unionising, paying taxes, not polluting the environment, or else. Figure (3) illustrates a payoff matrix for the Prisoner's Dilemma, as well as its generic representation. The Prisoner's Dilemma requires that $B > C > D > S$.

The Prisoner's Dilemma has only one Nash equilibrium at (D,D). All individual incentives push towards this equilibrium. Nevertheless, this equilibrium is not socially optimal.

Eventually, the difficulty raised by the Prisoner's Dilemma can be overcome if play-

ers can repeat the game (which requires $2C > B + S$). In particular, by playing the Prisoner's Dilemma as an evolutionary game with large numbers of players and strategies it is possible that islands of cooperation sustain themselves in a sea of selfish choices. Robert Axelrod pioneered this line of research in 1984, finding out that a "tit-for-tat" strategy was the most efficient: start with cooperating whenever you meet a new player, but defect if the other does.

2.4 The Traveller's Dilemma

The Traveler's dilemma is a non-zero-sum game in which two players attempt to maximize their own payoff, without any concern for the other player's payoff. The game was formulated by Kaushik Basu and goes as follows.

An airline loses two suitcases belonging to two different travellers. The suitcases contain identical antiques. An airline manager tasked to settle the claims of both travellers explains that the airline is liable for a maximum of 100 $ per suitcase, and in order to determine a honest appraised value of the antiques the manager separates both travellers and asks each of them to write down the amount of their value at no less than 2 $ and no more than 100 $. He also tells them that if both write down the same number, he will treat that number as the true value of both suitcases and reimburse both travellers that amount. However, if one writes down a smaller number than the other, this smaller number will be taken as the true value, and both travellers will receive that amount along with a bonus/malus: 2 $ extra will be paid to the traveller who wrote down the lower value and a 2 $ deduction will be taken from the person who wrote down the higher amount. The challenge is: what strategy should both travellers follow in order to decide what value they should write down?

If this game is actually played, nearly all the times everyone chooses 100 $ and gets it. However, rational players should behave differently.

Rational players should value the antique slightly less than their fellow traveller, in order to get the the bonus of 2 $. For instance, by pricing at 99 $ one would get 101 $, whereas the opponent would get 97 $. However, this triggers an infinite regression such that 2 $ is the only Nash-equilibrium of this game. Thus, being rational does not pay.

The Traveller's Dilemma suggests that in reality people coordinate and collaborate because of their bounded rationality. If they would be smarter than they are, they would obtain less.

2.5 The Dollar Auction

The dollar auction is a non-zero sum sequential game designed by Martin Shubik to illustrate a paradox brought about by rational choice theory. In this game, players with perfect information are compelled to make an ultimately irrational decision based on a sequence of rational choices.

The game involves an auctioneer who offers a dollar bill with the following rule: the dollar goes to the highest bidder, who pays the amount he bids. The second-highest bidder also must pay the highest amount that he bids, but gets nothing in return.

Suppose that the game begins with one of the players bidding 1 cent, hoping to make a 99 cent profit. He will quickly be outbid by another player bidding 2 cents, as a 98 cent profit is still desirable. Similarly, another bidder may bid 3 cents, making a 97 cent profit. Alternatively, the first bidder may attempt to convert their loss of 1 cent into a gain of 97 cents by also bidding 3 cents. In this way, a series of bids is maintained.

However, a problem becomes evident as soon as the bidding reaches 99 cents. Supposing that the other player had bid 98 cents, they now have the choice of losing the 98 cents or bidding a dollar even, which would make their profit zero. After that, the original player has a choice of either losing 99 cents or bidding $1.01, and only losing one cent. After this point the two players continue to bid the value up well beyond the dollar, and neither makes a profit.

2.6 Pure Coordination Games

In 1960 Thomas Schelling introduced *pure coordination games*, which are a sort of a puzzle for Game Theory ever since. Pure coordination games are one-shot games where players face a set of alternatives knowing that a positive payoff will only accrue to them if they coordinate on the same choice. For instance, two subjects may be shown a city map and asked, independently of one another, to select a meeting point. Or, subjects may be asked to select a positive integer. In the first case they obtain a positive payoff if they select the same meeting point; in the second case, if they select the same integer.

The difficulty of pure coordination games derives from the fact that players cannot communicate and that the game is not repeated. The astonishing fact about pure coordination games is that players make an agreement much more often than they would do if they would play randomly.

In general, the explanation is that pure coordination games generally entail cues that single out one choice as more "salient" than others. For instance, subjects asked to select a meeting point generally end up with the railway station, whereas the majority of those asked to name a positive integer select the number 1.

However, this suggests that coordination may eventually be attained because of conventions, habits or values that do not enter the description of decision settings. People may not even be aware of what makes them coordinate with one another.

2.7 The Game of Chicken

The game of Chicken models two drivers, both headed for a single lane bridge from opposite directions. One must swerve, or both may die in the crash. However, if one driver swerves but the other does not, he will be called a "chicken". Figure (4) depicts a typical payoff matrix for the Chicken Game, as well as its generic form.

Chicken is an anti-coordination game with two pure-strategy Nash-equilibria where each player does the opposite of what the other does. Which equilibrium is selected depends very much on the effectiveness in signaling pre-commitment before the game is played. For instance, a driver who disables brakes and steering wheel in front of the other driver may induce him to swerve. One real-world example is a protester who

	Swerve	Straight			
Swerve	0, 0	−1, +1		V/2, V/2	0, V
Straight	+1, −1	−10, −10		V, 0	$\frac{(V-C)}{2}, \frac{(V-C)}{2}$

Figure 4: A payoff matrix for the Game of Chicken (left) and its generic representation (right). The left number is the payoff of the row player, the right number is the payoff of the column player. In this generic representation, V is the value of power, prestige, or of the available resource to be obtained, C is the cost if both players choose "straight".

handcuffs himself to an object, so that no threat can be made which would compel him to move

Betrand Russell remarked that the nuclear stalemate was much like the Game of Chicken: [1]

> As played by irresponsible boys, this game is considered decadent and immoral, though only the lives of the players are risked. But when the game is played by eminent statesmen, who risk not only their own lives but those of many hundreds of millions of human beings, it is thought on both sides that the statesmen on one side are displaying a high degree of wisdom and courage, and only the statesmen on the other side are reprehensible. This, of course, is absurd. Both are to blame for playing such an incredibly dangerous game. The game may be played without misfortune a few times, but sooner or later it will come to be felt that loss of face is more dreadful than nuclear annihilation. The moment will come when neither side can face the derisive cry of 'Chicken!' from the other side. When that moment is come, the statesmen of both sides will plunge the world into destruction.

John Maynard Smith and G. Price re-interpreted the Game of Chicken in the context of animal behaviour. Their Hawk-Dove game has the same payoff matrix as in figure (4), where "swerve" and "straight" correspond to the following strategies, respectively:

Dove Retreat immediately if one's opponent initiates aggressive behaviour;

Hawk Initiate aggressive behaviour, not stopping until injured or until the opponent backs down.

Whilst the original Game of Chicken assumes $C > V$ and cannot be repeated, the Hawk-Dove game lacks this requirement and is generally conceived as an evolutionary game.

The strategy "Dove" is not evolutionary stable, because it can be invaded by a "Hawk" mutant. If $V > C$, then the strategy "Hawk" is evolutionarily stable. If $V < C$

---

[1] Bertrand W. Russell, *Common Sense and Nuclear Warfare*. London, George Allen and Unwin, 1959.

there is no evolutionarily stable strategy if individuals are restricted to following pure strategies, although there exists an evolutionarily stable strategy if players may use mixed strategies.

2.8 The War of Attrition

The war of attrition is a game of aggression conceived by John Maynard Smith in which two contestants compete for a resource of value V by persisting while constantly accumulating costs. Equivalently, this game can be seen as an auction in which the prize goes to the player with the highest bid B_h, and each player pays the loser's low bid B_l.

The war of attrition cannot be properly solved using its payoff matrix. In fact, the players' available resources are the only limit to the maximum value of bids. Since bids can be any number if available resources are ignored, the payoff matrix has infinite size. Nevertheless, its logic can be analysed.

Since players may bid any number, they may even exceed the value V that is contested over. This may appear irrational at first sight. However, each bidder pays B_l. Therefore, it would seem to be in each player's best interest to bid the maximum possible amount rather than an amount equal to or less than the value of the resource V.

However, if both players bid higher than V, the high bidder does not so much win as lose less, in the sense that $-B_l < V - B_h < 0$ — a Pyrrhic victory. In contrast, if each player bids less than V, the player bidding B_l will lose, and the other player will benefit by an amount $V - B_l$.

Since there is no value to bid which is beneficial in all cases, there is no dominant strategy. However, this does not preclude the existence of Nash-equilibria. Any pair of strategies such that one player bids zero and the other player bids any value equal to V or higher, or mixes among any values V or higher, is a Nash-equilibrium.

The War of Attrition is akin to a Chicken or Hawk-Dove game — see § (2.7) — where if both players choose "swerve"/"Dove" they obtain 0 instead of $V/2$ as in figure (4).

The evolutionarily stable strategy when playing it as an evolutionary game is a probability density of random persistence times which cannot be predicted by the opponent in any particular contest. This result has led to the conclusion that, in this game, the optimal strategy is to behave in a completely unpredictable manner.

3 Influence Games

The following games are not concerned with Nash-equilibria. Players are not assumed to figure out which alternatives other players might choose, originating infinite regressions that can only stop at Nash-equilibrium points.

Rather, boundedly rational players are assumed to follow certain rules, that may be quite simple but need not be necessarily so. The game then concerns what collective behaviours can be generated by mutual influence.

Evolutionary games reach this approach when they report simulations of the interaction of given rules rather than predicting evolutionary stable strategies. Such is the

case, for instance, of Axelrod's simulations of the Prisoner's Dilemma — see § (2.3).

This section aims at providing formal tools for this approach. In particular, two prototypical games have been investigated.

The *Ising model* (originally developed in physics, where it is also known as *spin glass* model) is concerned with imitation. The *minority game*, also known as the *El Farol bar problem*, is about the opposite of imitation, i.e., about doing the opposite of what others do.

3.1   The Ising Model

The Ising model was originally developed in physics in order to study the interaction between atoms in a ferromagnetic material. For this reason its agents can only take two states, or opinions — originally, $+$ and $-$ — and are fixed in space. However, although many imitation models are generally more complex than the Ising model, the closed-form solutions of the Ising model may guide the builder of more complex models in the process of understanding their behaviour.

In general, the Ising model is not presented as a game. It is done here in order to stress its symmetry with the minority game.

Let N players be denoted by means of an index $i = 1, 2, \ldots N$. Players must choose between an alternative $A = -1$ and an alternative $A = 1$.

The payoff of a player does not only depend on the alternative that he has chosen, but also on the average of the alternatives chosen by the other players. Let m denote this average.

Since we want to reproduce situations where the individual follows the herd, the effect of m should be the stronger, the more homogeneous the group. Since $A \in \{-1; 1\}$ and consequently $m \in \{-1; 1\}$, we can reach this goal by requiring that payoff depends on a term Am. This term may eventually be multiplied by a coefficient $J > 0$.

A stochastic term ε is necessary in order to understand our game as a system jumping between many equilibria. This term will disappear when expected values will be taken.

In the end, let us choose the following functional form for the payoff of a player:

$$u(A) = v(A) + JAm + \varepsilon \tag{2}$$

where $u(A)$ is the total payoff of a player and $v(A)$ is its individual component.

Furthermore, let us assume that this individual component takes the following form:

$$v(A) = \begin{cases} -h & \text{if } A = -1 \\ h & \text{if } A = 1 \end{cases} \tag{3}$$

where $h \in \mathfrak{R}, h > 0$.

By assuming that the stochastic terms ε_i are Gumbel-distributed, we can apply the logit model. By combining eq. (2) and eq. (3) we derive the following expressions for the probability that a player selects one of the two alternatives:

$$p\{A = -1\} \quad = \quad \frac{e^{\mu(-h - Jm + \varepsilon)}}{e^{\mu(-h - Jm + \varepsilon)} + e^{\mu(h + Jm + \varepsilon)}} \tag{4}$$

10

$$p\{A=1\} \quad = \quad \frac{e^{\mu(h+Jm+\varepsilon)}}{e^{\mu(-h-Jm+\varepsilon)} + e^{\mu(h+Jm+\varepsilon)}} \tag{5}$$

The expected value of the selected alternative is $E\{A\} = -1 \cdot p\{A = -1\} + 1 \cdot p\{A = 1\}$. Since it is also $E\{A\} = m$ we obtain the following expression:

$$m = \tanh(\mu h + \mu J m) \tag{6}$$

where $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$ is the hyperbolic tangent.

Eq. (6) provides an analytic description of a game with herd behaviour on two alternatives described by means of a mean-field approximation. It admits a closed form solution that provides the following findings:

- If $\mu J < 1$ and $h = 0$ there exists one single solution at $m = 0$. Consider that this is a discrete-time system, so its attractors are stable if all eigenvalues of the state transition function are in $(-1, 1)$. Intuitively, $\mu J < 1$ means that this system is globally stable. Furthermore, $h = 0$ means that the individual component of utility is zero so the players have no incentive to choose one of the two alternatives. Consequently, the stochastic term makes $m = 0$ the only solution.

- If $\mu J < 1$ and $h \neq 0$ there exists one single solution with the same sign as h. In fact, as in the previous case the system is globally stable so it admits one single solution. However, since in this case there exists an individual component in their utility function, this component determines what equilibriun arrives to. If players generally prefer $A = -1$ the equilibrium will be $m \approx -1$, if players generally prefer $A = 1$ the equilibrium will be $m \approx 1$.

- If $\mu J \geq 1$ and $h = 0$ there exist three solutions: $m = 0$ and $m = \pm m(\mu J)$. In fact, the system is globally unstable but locally stable equilibria may exist. Since the individual component of utility is zero, the system may either tend towards $m = 0$, or $m \approx -1$, or $m \approx 1$.

- If $\mu J \geq 1$ and $h \neq 0$, the following subcases must be distinguished:

  - If, for any given μ and J, there exists a threshold $H(h) > 0$ such that $\mid h \mid \leq H$, then three solutions exist, one with the same sign as h and the other two with opposite sign. Condition $\mid h \mid < H$ means that the individual component of utility is limited even if not zero. Therefore, results are similar to the previous case.

  - If, for any given μ and J, there exists a threshold $H(h) > 0$ such that $\mid h \mid > H$, then there exists one single solution with the same sign as h. In fact, if the individual component of utility can take any value, then the whole system is force into its direction.

In the Ising model, each player observes the average behaviour of all other players. If each player observes only the behaviour of his neighbours, one obtains Schelling's model of racial segregation.

History	Prediction
0 0	0
0 1	1
1 0	1
1 1	1

Figure 5: An example of a strategy based on the two previous steps of the minority game. The first column lists all possible stories. The second column, depending on past history, makes a prediction.

3.2 The Minority Game

The minority game originates from a consideration inspired to economist Brian Arthur by the *El Farol* bar in Santa Fe, New Mexico (USA). Arthur remarked that people go to the bar in order to meet other people, but they do not want to go when all other people go, because the bar is too crowded on such occasions. Thus, they want to do the opposite of what most people do — go to the bar when most people stay at home, stay at home when most people go to the bar. Evidently, the "El Farol bar problem" cannot have a stable equilibrium. In fact, once the majority observed what the minority did, it wants to imitate it, which turns the minority into majority, and so on endlessly.

Physicists Yi-Cheng Zhang and Damien Challet remarked that this is the essence of the dynamics of the stock market. In fact, in the stock market those traders gain, who buy when stocks are low (because most traders are selling) and sell when stocks are high (because most traders are buying). So all traders want to belong to the minority, which, being impossible, generates instability. Among the physicists, the "El Farol bar problem" became "the minority game".

Let us consider N players who either belong to a group denoted 0 or a group denoted 1. Players belonging to the minority group receive a positive payoff. Players belonging to the majority group have a payoff zero.

Strategies are functions that predict which will be the minority group in the next step given the minority group in the m previous steps. Thus, a strategy is a matrix with 2^m rows (dispositions with repetition of two elements of class m) and two columns. The first column entails all possible series of minority groups in the previous m steps, henceforth *histories*. The second column entails the group suggested for the next step. As an example, figure (5) illustrates a strategy with $m = 2$.

Each player owns s strategies. If $s = 1$, the game is trivial because the time series of the minority group is periodical.

If $s > 1$, players choose the strategy that cumulated the greatest amount of payoffs. Thus, a number of feedbacks may arise between what strategies are chosen and their capability to predict the minority. In this game, players must adapt to an environment that they themselves create.

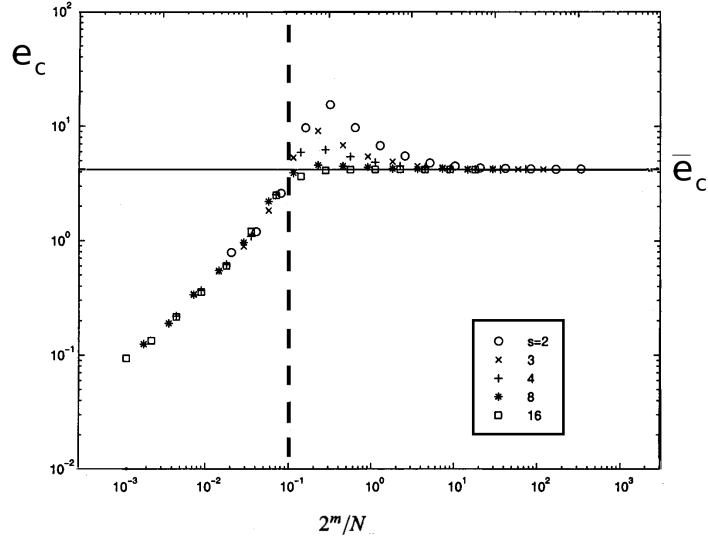An important magnitude in this game is the variance of the time series of the num-

Figure 6: Efficiency of coordination e_c as a function of the number of histories in a strategy $2^m/N$, for different values of the number of available strategies s. The horizontal line at $e_c = \bar{e}_c$ marks the efficiency level when player select a strategy at random. The vertical dashed line marks the point where e_c can be greater than $\bar{e}_c$.

ber of players belonging to group 1 (or, equivalently, group 0). Henceforth, this magnitude will be denoted by σ^2.

The average of the number of players belonging to each group is generally close to $N/2$. If σ^2 is small, then the distribution of the number of players belonging to group 1 is concentrated around $N/2$. This implies that the minority is large, eventually close to its maximum $(N/2 - 1)$. On the contrary, if σ^2 is large the number of players belonging to group 1 tends to be either much smaller or much larger than $N/2$, implying that the minority is often very small.

Let us consider σ^2/N in order to normalize to the number of players. Let us define the *efficiency of coordination* $e_c = N/\sigma^2$ as the reciprocal of the extent to which players behave differently from one another.

Figure (6) depicts numerical simulations of e_c as a function of the number of histories in a strategy $2^m/N$. Graphs are shown for different values of s. The horizontal line marks the value that e_c attains if players would make a random choice among the strategies available to them.

With low m the efficiency of coordination is low. This happens because if memory is short, then players have greater difficulties to adapt to the changing features of the game.

If only few strategies are available $(s = 2, s = 3, s = 4)$, at intermediate values of m many players guess the correct strategy so e_c increases above the level that can

be attained if strategies are chosen randomly. This threshold is marked by the dashed vertical line. However, this effect disappears if many strategies are available ($s = 8, s = 16$). In this case the decision process becomes similar to a random choice so even at intermediate values of m the efficiency of coordination is close to the level attained if strategies are chosen randomly.

Independently of the number of available strategies, with increasing m the value of e_c tends to the level attained when strategies are chosen randomly. In fact, a history of length m occurs again after 2^m steps on average, so a strategy that is successful with a particular history needs 2^m steps in order to be successful again. With very high values of m, no strategy can present itself as particularly successful; therefore, a nearly-random dynamics ensues.

Let us consider what information is available to players. The only information available to them is what group was the minority in previous time steps. Let this information be carried by a variable W_t, where $W_t = 0$ means that at time t the group 0 has been minority, $W_t = 1$ otherwise. The issue is whether this information is used efficiently; if it is not, there may exist exist arbitrage possibilities for players who utilize information more efficiently than their peers.

Let us consider W_t and W_{t+1} as distinct signals. Let us compute their mean mutual information $I(W_t, W_{t+1})$. [2]

Mean mutual information measures whether the information entailed in the outcomes of two steps of the game, taken together, is greater than the sum of the information entailed in the outcomes of the two steps independently of one another. Thus, mean mutual information says whether a player, by observing the time series of the outcome of the game, could do better than his peers. If the minority game is describing a stock market, $I(W_t, W_{t+1}) > 0$ means that a trader could gain from arbitrage.

Let us introduce *information efficiency* $e_i = 1/I(W_t, W_{t+1})$. Being the reciprocal of mean mutual information, information efficiency is high when mean mutual information is low, i.e., when information is efficiently exploited by the player so there is little room for arbitrage.

Figure (7) depicts numerical simulations of e_i as a function of the number of stories in a strategy $2^m/N$. Graphs are shown for different values of s.

One may observe in figure (7) a sudden drop of e_i in the $[0.3, 1]$ interval. This is entailed in the interval $[0.1, 1]$ where e_c was observed to rise above the level corresponding to random choice in figure (6). Thus, we may subsume the behaviour of the minority game as in table (1):

Table (1) shows that the minority game has two large behaviour modes, one inefficient in coordination but efficient in the exploitation of information, the other one efficient in coordination but inefficient in the exploitation of information. In between, a tiny space where the efficiency of coordination and the efficiency of information exploitation may change dramatically depending on s and m.

---

[2]Given a source of binary symbols $\{a_1, a_2, \ldots a_M\}$ issued with probabilities $p_1, p_2, \ldots p_M$, the average information that they convey is defined as $H(A) = \sum_{i=1}^{M} p(a_i) \lg_2 1/p(a_i)$ and it is called *information entropy*. Suppose that there is a second source issuing symbols $\{b_1, b_2, \ldots b_N\}$ with information entropy $H(B)$. Let $H(A, B)$ denote the information entropy of the whole system. *Mean mutual information* $H(A) + H(B) - H(A, B)$ measures to what extent the two sources interact to correlate their messages. Mean mutual information is zero if the two sources are independent of one another.
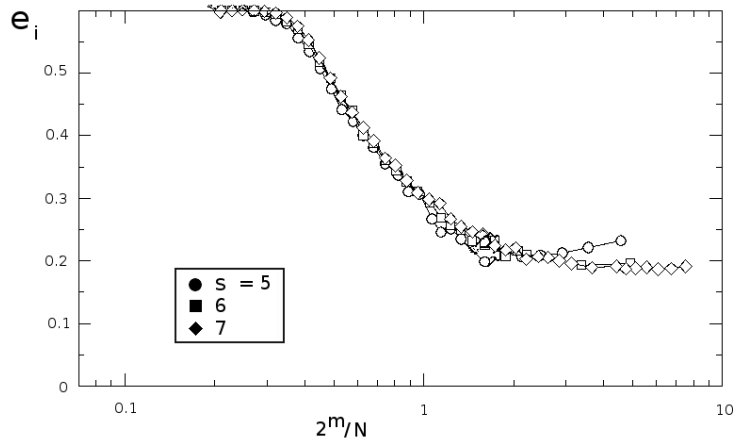
Figure 7: Efficiency of exploitation of information as a function of the number of stories in a strategy, normalized to the number of players.

$2^m/N < 0.1$	$2^m/N > 1$
inefficient coordination low e_c	efficient coordination high e_c
efficient information exploitation high e_i	inefficient information exploitation low e_i

Table 1: Efficiency of coordination and efficiency of information exploitation in the minority game.

15

Since the minority game is a stylised representation of stock markets, we may ask in which region stock markets operate. It is well known that very many traders operate in stock markets, so we may assume that N is very high. Human bounded rationality suggests that traders do not make use of complicated algorithms that take account of events far back in the past, so m should be in the order of a few units. Consequently, $2^m/N$ is likely to be very small.

This suggests that financial markets are characterised by low coordination, which implies irregular oscillations where large majorities and small minorities may appear. At the same time, financial markets are efficient in exploiting information. Thus, the observation of its time series offers few possibilities to extrapolate its future courses.

4 Some Pitfalls of Utility Maximization

Utility maximization strikes its adepts for its elegance, simplicity and beauty. Unfortunately, empirical tests have have shown that in many situations decision-makers do not follow its prescriptions.

The situations where decision-makers typically do not behave as utility maximizers are generally known as *paradoxes*. Some paradoxes can be reduced to utility maximization by means of special additions to the basic theory. In this section the main paradoxes will be discussed, together with their eventual resolution within the utility maximization framework.

4.1 Ellsberg's Paradox and Sub-Additive Probabilities

Suppose that a decision-maker is placed in front of two urns, henceforth denoted A and B. The decision-maker is informed that urn A entails white and black balls in equal proportion, e.g., urn A may contain 10 white balls and 10 black balls. Regarding urn B, the decision-maker knows just that it entails white and black balls. Suppose to ask the decision-maker to evaluate the probability to extract a white ball from urn A and the probability to extract a white ball from urn B.

Since urn A entails white and black balls in equal proportions, the probability to extract a white ball from urn A is 0.5. On the contrary, nothing is known regarding the proportion of white to black balls in urn B. In cases like this, the so-called "principle of insufficient reason" — i.e. the fact that there is no reason to think otherwise — suggests to imagine that also urn B entails white and black balls in equal proportions. Thus, also in this case the probability to extract a white ball is assessed at 0.5. And yet, something is not in order: intuitively, urn B should be characterized by a greater uncertainty than urn A!

Ellsberg's paradox actually deals with the size of the sample on which probabilities are evaluated. In fact, Ellsberg's paradox places two extreme situations aside.

In the case of urn A, since we know that it entails white and black balls in equal proportions we are able to compute probability with infinite precision. It is just like extracting a ball (and replacing it afterwards) infinite times. We are measuring probability on a sample of infinite size.

In the case of urn B, lack of knowledge on the proportion of white to black balls is equivalent to estimating the probability of extracting a white ball prior to any extraction. It means that the probability must be measured on a sample of size zero. We guess its value at 0.5, but the reliability of our estimate is very low.

One possibility for overcoming Ellsberg's paradox is representing uncertainty by means of two magnitudes. The first one is probability whilst the second one is sample size. In general, the size of the sample is expressed by a *precision indicator*.

Another possibility is to resort to the theory of sub-additive probabilities as exemplified by Itzhak Gilboa and David Schmeidler. While according to classical probability theory the sum of the probabilities of an exhaustive set of events must be equal to 1, according to the theory of sub-additive probabilities this holds only if probabilities are measured on a sample of infinite size. In all other cases the probabilities take values such that their sum is smaller than 1.

Let us consider the following example: We are playing dice in a clandestine gambling room. Since we fear that we are playing with an unfair die, we may not assign probability $1/6$ to each face, but rather less, e.g. $1/8$. Thus, the sum of the probabilities of all faces is $6 \times 1/8 = 3/4$, which is smaller than 1. Subsequently, if we have a possibility to throw the die many times — i.e. if we can increase the size of our sample — we may find out that the die is unfair in the sense that, e.g., face "2" comes out with probability $1/3$ while the other faces come out with probability $2/15$. The sum of all these probabilities is $5 \times 2/15 + 1/3 = 2/3 + 1/3 = 1$.

Let us return to Ellsberg's paradox. In the case of urn A, the probability to extract a white ball is 0.5 and the probability to draw a black ball is 0.5. The sum of these probabilities is 1. In the case of urn B, the decision-maker may judge that the probability to extract a white ball is, for instance, 0.4, and that the probability of extracting a black ball is also 0.4. The sum of these probabilities is 0.8, but this does not constitute a problem for the theory.

4.2 Allais' Paradox and Prospect Theory

The following experiment was carried out by Maurice Allais having Leonard Savage as a subject, who was a major advocate of expected utility maximization and nevertheless did not behave according to its prescriptions. Subjects are asked to choose between the alternatives A and B reported on the rows of table (2). It is empirically observed that most people choose alternative (B).

Subsequently, the same subjects are confronted with the alternatives C and D reported on the rows of table (3). It is empirically observed that most people choose alternative (C).

Let us now examine the expected utilities of two pairs of alternatives (A,B) and (C,D). Preferring (B) to (A) means that $u(2,400) > 0.33 \times u(2,500) + 0.66 \times u(2,400)$, which can be written as $0.34 \times u(2,400) > 0.33 \times u(2,500)$. Unfortunately, preferring (C) to (D) implies the opposite, i.e. that $0.33 \times u(2,500) > 0.34 \times u(2,400)$. So it turns out that most people do not behave rationally if they maximise utility.

Allais' paradox is due to the presence of a tiny probability of not obtaining anything in alternative (A). Thus, it is due to aversion to risk.

	Consequence 1	**Consequence 2**	**Consequence 3**
Alternative A	receive $ 2,500 with prob. 0.33	receive $ 2,400 with prob. 0.66	receive nothing with prob. 0.01
Alternative B	receive $ 2,400 with prob. 1.00		

Table 2: The first choice in Allais' experiment.

	Consequence 1	**Consequence 2**
Alternative C	receive $ 2,500 with probability 0.33	receive nothing with probability 0.67
Alternative D	receive $ 2,400 with probability 0.34	receive nothing with probability 0.66

Table 3: The second choice in Allais' experiment.

Daniel Kahneman and Amos Tversky introduced non-linear transformations of utilities and probabilities in order to balance risk aversion. The transformed utilities and probabilities can describe the observed behavior as expected utility maximization. This is called Prospect Theory.

A prospect is a set of pairs $\{(c_1, p_1), (c_2, p_2), \ldots\}$, where c_j is a consequence that will obtain with probability p_j. As a preliminary step, prospects with identical consequences are summed, dominated prospects are eliminated and riskless components are ignored.

Prospects Theory prescribes that the utilities and the probabilities of the above prospects be transformed according to the following rules:

1. Utility is transformed by means of a non-linear function $v = f(u)$ such that $f'(u) > 0$ and $f''(u) < 0$ for $u > 0$, $f'(u) > 0$ and $f''(u) > 0$ for $u < 0$, with $|f''(u)|_{u<0} > |f''(u)|_{u>0}$.

2. Probabilities p are transformed into "weights" w by means of a non-linear function $w = g(p)$ such that $g(0) = 0$ and $g(1) = 1$ but $\exists \overline{p} \in (0, 1)$ such that $\forall p < \overline{p}$ it is $g(p) \geq p$ and $\forall p > \overline{p}$ it is $g(p) \leq p$.

3. Weights w are transformed into coefficients q by means of the following rules:

$$
\begin{array}{llll}
q^-_{-h} & = & w^-(p_{-h}) & \text{for} \quad j = -h \\
q^-_i & = & w^-(p_{-h}+\ldots+p_i) - w^-(p_{-h}+\ldots+p_{i-1}) & \text{for} \quad -h < j \leq 0 \\
q^+_i & = & w^+(p_i+\ldots+p_k) - w^+(p_{i+1}+\ldots+p_k) & \text{for} \quad 0 \leq j < k \\
q^+_k & = & w^+(p_k) & \text{for} \quad j = k
\end{array}
$$

where w^- and q^- refer to prospects with negative utility, denoted with an index $j \in [-h, 0]$, whereas w^+ and q^+ refer to prospects with positive utility, denoted with an index $j \in [0, k]$.

The v and q obtained at the end of this procedure can be used much like utilities and probabilities, respectively. Prospect Theory succeeds to eliminate the inconsistencies highlighted by Allais' paradox, but it does not explain why it works. It should be called a heuristic, rather than a theory.

4.3 Preference Reversal in Slovic's Paradox

Let us consider a series of bets with different characteristics. [3] For instance, a series of bets on different horses, or playing on a series of different slot machines, or a series of unfair dice different from one another. The game consists of choosing to bet on a specific horse, choosing to play on a specific slot machine or selecting a specific die to throw. In other words, the game consists of choosing one bet out of a series of bets.

In order to simplify matters, let us consider series composed by two bets. More specifically, let us consider the four pairs of bets in table (4).

For any pair of bets, subjects are asked to select either bet A or bet B. On average, the number of subjects who prefer A to B is slightly greater than the number of subjects who prefer B to A.

At this point, a different game is played. Subjects are asked to imagine that they own a lottery ticket for each bet, and that they have a possibility to sell it. That is, they can either wait for the outcome of each bet, where they may win or loose with a certain probability, or they can sell the ticket. In order to compare the willingness to play to the willingness to sell the ticket, subjects are asked to fix a minimum selling price for each bet.

In general, it is empirically observed that most people ask a higher price for bets B than for bets A.

However, for each pair of bets, bet A has the same expected (utility) value than bet B. Thus, utility maximizers should be indifferent between A and B. On the contrary, subjects have a slight preference for A if they are asked to play one of the two bets but they definitely prefer B if they are asked to fix a selling price.

The distinguishing feature of bets A is that the first consequence has a much higher probability than the second one. Thus, one assumes that it is the difference of probability values that orientates decision-making.

The distinguishing feature of bets B is that the first consequence concerns a much larger amount of money than the second one. Probabilities, on the contrary, are sometimes very similar, sometimes very different from one another. Thus, one assumes that it is the difference of money values that orientates decision-making.

---

[3] A comprehensive introduction to this topic is: Paul Slovic and Sarah Lichtenstein (eds.), *The Construction of Preference*. Cambridge, Cambridge University Press 2006.

PAIR OF BETS   I

	Consequence 1	Consequence 2
Bet A$_I$	win $ 4.00 with probability 0.99	loose $ 1.00 with probability 0.01
Bet B$_I$	win $ 16.00 with probability 0.33	loose $ 2.00 with probability 0.67

PAIR OF BETS   II

	Consequence 1	Consequence 2
Bet A$_{II}$	win $ 3.00 with probability 0.95	loose $ 2.00 with probability 0.05
Bet B$_{II}$	win $ 6.50 with probability 0.50	loose $ 1.00 with probability 0.50

PAIR OF BETS   III

	Consequence 1	Consequence 2
Bet A$_{III}$	win $ 2.00 with probability 0.80	loose $ 1.00 with probability 0.20
Bet B$_{III}$	win $ 9.00 with probability 0.20	loose $ 0.50 with probability 0.80

PAIR OF BETS   IV

	Consequence 1	Consequence 2
Bet A$_{IV}$	win $ 4.00 with probability 0.80	loose $ 0.50 with probability 0.20
Bet B$_{IV}$	win $ 40.00 with probability 0.10	loose $ 1.00 with probability 0.90

Table 4: Slovic's experiment.

If subjects are asked to play the bets their attention is caught by probabilities, so either they are indifferent or they prefer A. If subjects are asked to sell lottery tickets their attention is caught by money values, so they prefer B.

Slovic's paradox shows that preferences change if the decision-maker focuses on the probability of a consequence or, rather, on its utility (here, money value). This means that human beings are unable to evaluate probabilities and utilities independently of one another.

Slovic's paradox — often known as "preference reversal" — is destructive for utility maximization. In fact, it undermines the assumption that a utility function and a probability function can be defined, independently of one another. Slovic's paradox suggests that uncertain belief cannot be split into utilities and probabilities.

Obviously, several attempts to reconcile preference reversal with the theory of rational choice have been made. Preference reversal can be accommodated with the theory of rational choice if either violations of transitivity, or of independence, or of completeness of preferences are accepted. While the attempts to reconcile preference reversal with the theory of rational decision by relaxing transitivity or independence of preferences did not receive much attention because these properties are essential for our idea of rationality — see § (2), the more recent idea of dropping completeness deserves some discussion. In fact, allowing preferences to be incomplete amounts to accept the idea that a utility function can be defined, at most, for *some* alternatives. Possibly, just the simplest and most repetitive ones.

4.4   Arrow's Paradox

The following paradox of social choice is due to Kenneth Arrow. Let A, B and C denote three alternatives, and let 1, 2 and 3 denote three individuals. Let us assume that:

- Individual 1 prefers alternative A to alternative B and alternative B to alternative C. Thus, he prefers alternative A to alternative C.

- Individual 2 prefers alternative B to alternative C and alternative C to alternative A. Thus, he prefers alternative B to alternative A.

- Individual 3 prefers alternative C to alternative A and alternative A to alternative B. Thus, he prefers alternative C to alternative B.

If these three individuals constitute a democratic community with a majority rule, then this community prefers A to B (individuals 1 and 3) and alternative B to alternative C (individuals 1 and 2). Thus, if the community wants to have transitive preferences, it must prefer A to C. But, the majority of its members (individuals 2 and 3) prefers C to A!

The setting of Arrow's paradox can be seen as a game, where individual utility maximizers are set together. Arrow's paradox shows that there are conditions where the outcome contradicts a basic assumption of utility maximization, even if individuals do not.

Several proposals have been made in order to overcome Arrow's paradox. The most common way out is to allow individuals to have different preferences if all alternatives

are presented to them, or if they are presented with pairs of alternatives. Or, one may limit voters to two alternatives presented in tournaments — Arrow's paradox would disappear, but the final choice is not necessarily the one that would be preferred by the largest possible majority.

5   Logic of Consequence and Logic of Appropriateness

As we have seen in § (4), utility maximization is not a good descriptor of decision processes. Its proponents — notably, Leonard Savage — have objected that utility maximization is not meant to be a faithful description of what people actually do, but rather a prescription of what they should do. It pretends to be a normative theory, although it is not a descriptive theory.

However, the preference reversals highlighted by Slovic point to such a huge distance between theory and reality, that the normativeness of utility maximization might be questioned. If utilities do not exist, it may make little sense to tell decision-makers that they should maximize them. Furthermore, if evolution shaped human reasoning along patterns that are different from utility maximization, we ought to be careful to declare these patterns "illogical", or "irrational". Rather, it may make sense to observe how human beings actually make their decisions, understand the rationale, and eventually revise our theories.

James March traced a distinction between the "logic of consequences" that underlies utility maximization, and a "logic of appropriateness" where human beings behave according to what they deem appropriate depending on past experiences and social pressures to conformity in specific settings. Note that the logic of appropriateness does not separate an individualistic step (Utility) from social interation (Game Theory).

Human minds are viewed as coherence-seeking machines that make use of available information in order to construct a plausible interpretation of reality, be it social roles, scientific theories, or else. By drawing causal relationships and eliminating inconsistencies a decision-maker tells herself a story that explains why certain facts are the way they are and why certain people did what they did. This story, a founding story that suggests a decision-maker what it is appropriate to do, is called a *narrative*.

The construction of a narrative may require that issues that do not fit into the picture are ignored, downplayed or forgotten. It may require that opinions are changed even dramatically, and yet their purporters candidly claim that they have always been coherent throughout their lives, or that they have been coherent in spite of having changed their opinion, if their story is seen from a particular point of view.

Albeit disturbing for our idea of rationality, the extent and easiness with which human beings distort previous experiences is proven by a number of experiments in psychology. Daryl Bem and Michael Ross have shown that it is easy to induce the subjects of experiments to change opinion while they are still convinced to have been coherent throughout the whole experiment. Anthony Greenwald, Michael Ross, Kenneth Gergen and Eugene Winograd have shown that people construct coherent narratives of their past, and that they remember past events to the extent that they fit these narratives. Other experiments by Michael Ross, Anthony Greenwald, Kathy McFarland and Michael Conway have shown how they may eventually change their interpretation of

the past and construct a new narrative if new evidence must be accommodated. The empirical evidence tells us that human beings are ready to lie to themselves in order to build coherent narratives. [4]

This attitude is puzzling, because distorting reality in order to construct a coherent narrative is at odds with our idea of rationality. So either human nature is inherently irrational, or our idea of rationality is incorrect.

According to James March, re-inventing the past is a crucial ability that enables decision-makers to conceive new goals and figure out a strategy in an uncertain future. Later, a similar argument has been made by Karl Weick under the label of "sensemaking". Essentially, these authors suggest that in order to make decisions in the face of an uncertain future it is good to have a narrative that explains the past as if previous decisions had been made along a coherent line. This line guides the decision-maker into the future, providing a rationale for action even if certainties are very few.

So here comes a straightforward argument for normativeness. If seeking coherence has the purpose of constructing a narrative, and if narratives are useful, then a decision theory based on constructing narratives should be regarded as rational, and openly prescribed.

In business, politics and other fields, narratives may constitute the bulk of strategies. David Lane and Robert Maxfield have made a years-long field observation of the elaboration and modification of the narrative of a Silicon Valley firm. [5] This study is worth reporting, because it is very clear in making us understand that narratives are useful precisely because they provide a guidance in the face of an uncertain future, and that their usefulness is not impaired by the fact that their coherence is based on an arbitrary interpretation of reality.

5.1   A Real Story

In 1990, *Echelon*, a Silicon Valley company, launched *LonWorks*, an innovative technology for distributed control. Previously, control was centralized into one main processing unit. With *LonWorks*, each electrotechnical device is endowed with a microprocessor and can communicate with all other devices, so all device control each other. Distributed control is more resilient than centralized control, and easily implements modular architectures to which additional devices can be added.

Distributed control is particularly suited to the automation of office spaces in large buildings, post-Fordist productive plants, as well as any setting where a large number of heterogeneous devices must coordinate their operations while retaining some flexibility. Thus, in its early days *Echelon* focused on partnerships with large producers of the devices to be automated, e.g., a producer in the field of heating and air conditioning was offered a possibility to integrate a microchip in their devices, as well on lifts, doors and windows in order to integrate all controls in a large building, from lighting to heating to theft protection.

---

[4] Detailed references to this litterature can be found in Guido Fioretti, *Either, Or. Exploration of an emerging decision theory*. Working paper available SSRN.

[5] David A. Lane and Robert R. Maxfield, "Ontological Uncertainty and Innovation". *Journal of Evolutionary Economics*, 15 (1) 2005: 3–50.

With some disappointment, *Echelon* had to recognize that the *LonWorks* technology was not exploited in its full potentialities. In fact, each large producer was specialized in one tiny sector so it had neither the power nor the capability to implement *LonWorks* on all devices. For instance, a producer in the field of heating and air conditioning found it difficult to install *LonWorks* on doors, windows, lights and lifts, for the automation of these devices was covered by other firms. Indeed, the difficulty was that *Echelon* was attempting to create a new market — one may call it a market for automation — in a marketplace that was covered by producers of several physical devices at a time.

Echelon was conscious of the enormous difficulties connected with the creation of a new market. Nevertheless, it deemed that long-term relations with a few specialized producers would pay in the long run. *Echelon* had a narrative, saying that large specialized producers would slowly but persistently adopt and impose *LonWorks*. Consequently, it invested all of its resources in these relations.

By 1994, *Echelon* was loosing confidence in this narrative. *Echelon* started to approach large system integrators of ICT, such as *Olivetti* and *Ameritech*. However, the crucial move was that of hiring a person for this job, who did not come from Silicon Valley as all other executives did. Through this employee, *Echelon* approached smaller companies, that integrated devices from different producers. Some people at *Echelon* conceived the idea of embedding *LonWorks* in a box that could be attached to any electrotechnical device, of whatever producer.

Scholars of technological innovation know how difficult it is for visionary employees to convince their boss of the value of their idea. In the case of *Echelon* the CEO embraced enthusiastically the new idea, because it appeared to fit with his previous experience.

Echelon's CEO had been the successful entrepreneur of a small firm that exploited digital technologies to produce PBX (private branch exchange) systems with innovative features. This firm had been able to displace giants such as *AT&T* by providing small independent installers with a superior product. When this CEO met small independent integrators of electromechanical devices, he mapped the new idea onto his previous experience.

In 1996, and within a few months, *Echelon* changed its narrative. *Echelon* presented itself as a provider of an innovative microchip for independent system integrators, a microchip that could be installed on any electrotechnical device, of whatever producer.

Most importantly, *Echelon* told itself that it had always pursued this strategy. Nobody in the firm seemed to be aware that the firm's strategy had changed. According to the narrative that they had developed, they had always done what they were doing.

Moreover, when faced with evidence that the firm had adopted a different strategy, management wished that the final publication by Lane and Maxfield would not stress this aspect (Lane: personal communication). This makes sense, for according to our idea of rationality narratives should reflect "objective information", and decision-makers should stick to it. Thus, management did not want to appear irrational.

However, the case of *Echelon* highlights that constructing a narrative by re-interpreting the past may be good and useful for decision-makers. In fact, the reported case reveals that by re-interpreting its mission *Echelon* was able to direct its investments. If the future is uncertain, as it is often the case, interpreting the past in order to find a direction

for the future is a sensible activity.

Rather, the trouble is with our idea of rationality. Since re-interpreting the past is regarded as irrational, then it must be done in secrecy. In reality, if re-interpreting the past has positive effects, then it should be prescribed.

6 Tools for the Logic of Appropriateness

Although the logic of appropriateness cannot propose itself with a ready-made and ready-to-use formula such as utility maximization, there exist some tools that can be used to reproduce its building blocks. These are essentially classification tools, that form concepts out of information, and coherence tools, that arrange concepts into coherent stories.

In particular, the following tools will be reviewed in this section:

1. Unsupervised neural networks;

2. Evidence Theory;

3. Constraint Satisfaction Networks.

Unsupervised neural networks reproduce the formation of mental categories out of a flow of information. Evidence Theory assumes that an actor receives information on possibilities and arranges them into coherent hypotheses. Constrain Satisfaction Networks arrange concepts into coherent explanations. Although they have not been integrated into one another, they all concern the process of selecting some items from the flow of experiences, arranging them in a coherent narrative, and deciding accordingly.

The logic of consequence makes sense in the restricted realm of games of chance, where it is possible to overview an exhaustive set of possibilities and enlist all of the consequences of any alternative. On the contrary, the logic of appropriateness makes sense precisely because quite often such conditions do not hold. Thus, this review does not cover tools concerned with classification in a *given* set of categories, such as Case-Based Decision Theory and supervised neural networks.

6.1 Unsupervised Neural Networks

Human mental categories are not defined by pre-specified similarity criteria that the objects to be classified should fulfil. Rather, mental categories are continuously constructed and modified according to the similarity of a just-received piece of information to the pieces of information that have already been stored in existing categories. For instance, a child observing house chairs may start with an idea of "chair" as an object having four legs, then observe an office chair with one leg and take away the number of legs from its definition of "chair". The point here is that definitions are made *once* mental categories exist. Mental categories are not constructed around existing definitions.

In some cases, definitions are not even possible, because a mental category entails objects that do not have any common feature. [6] For instance, the mental category expressed by the word "game" refers to children amusing themselves with toys, adults involved in a serious competition on a chess board and as well as a set of wild animals. One may speculate that man transposed the emotions involved in hunting into the more intellectual context of chess, and that the fact that chess was an amusement suggested some similarity to what children were supposed to do. So pairwise intersections of the meanings of the word "game" exist, but this does not imply that all meanings have a common intersection. Therefore, a definition of "game" is not possible.

Unsupervised neural networks are able to reproduce the idea that mental categories arise out of adding examples. In fact, these networks construct categories around the most frequent input patterns, mimicking the idea that a child creates a category "chair" upon observation of many such objects.

Neural networks are composed by a set of *neurons* which produce an output $y \in \Re$ by summing inputs $x_1, x_2, \ldots x_N \in \Re$ by means of coefficients $a_1, a_2, \ldots a_N$:

$$y = \sum_{i=1}^{N} a_i x_i \qquad (7)$$

For any set of coefficients a_i, this simple device is able to classify inputs in a category by yielding the same output y for several input vectors $\mathbf{x}$. In fact, there exist several vectors $\mathbf{x}$ whose weighted sum yields the same y. For instance, if $\forall i$ it is $a_i = 1$, then e.g. $y = 10$ can arise out of $\mathbf{x}' = [9 \; 1]$, $\mathbf{x}'' = [2.5 \; 7.5]$, as well as many other vectors. In this sense, the neuron classifies the input vectors $[9 \; 1]$ and $[2.5 \; 7.5]$ in the same category.

Note that a neuron has no difficulty to classify input vectors that do not perfectly fit its categories. For instance, if there is a category $y = 10$ and a category $y = 11$, an input vector $\mathbf{x}''' = [2.1 \; 8]$ is classified in the category $y = 10$ just as $\mathbf{x}'$ and $\mathbf{x}''$.

The shape of the categories implemented by a neuron depends on the coefficients a_i. For instance, if $a_1 = 0.5$ and $a_2 = 20$ the input vector $\mathbf{x}' = [9 \; 1]$ yields $y = 24.5$ and may not lie in the same category as $\mathbf{x}'' = [2.5 \; 7.5]$, which yields $y = 151.25$.

The coefficients a_i may be chosen by the user of the network during a training phase, in which case we are dealing with a *supervised* network. Alternatively, the coefficients a_i may be initialised at random and subsequently changed by the network itself according to some endogenous mechanism. In this case we have an *unsupervised* network, of which the Kohonen network is the best known instance. [7].

In unsupervised networks, the ability of a neuron to change its categories stems from a feed-back from output y and a feed-forward from input $\mathbf{x}$, towards coefficients a_i:

$$\frac{da_i}{dt} = \phi(\mathbf{a}, y) x_i - \gamma(\mathbf{a}, y) a_i \; \forall i \qquad (8)$$

---

[6] See George Lakoff, *Women, Fire, and Dangerous Things*. Chicago, The University of Chicago Press 1987.

[7] The basic reference is Teuvo Kohonen, *Self-Organization and Associative Memory*. Berlin-Heidelberg, Springer 1989.
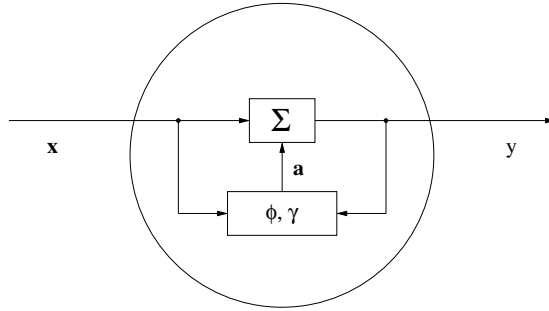
Figure 8: The neuron of an unsupervised network. The feed-backs and -forwards are responsible for the most notable properties of unsupervised networks, including the absence of a training phase. In a sense, the "training phase" of supervised neural networks may be seen as a feed-back and -forward passing through a human operator.

where $\phi(\mathbf{a}, y)$ and $\gamma(\mathbf{a}, y)$ may be linear or non-linear functions.

In equation 8, the term $\phi(\mathbf{a}, y)x_i$ enables the neuron to learn input patterns. It entails both a feed-back (from y) and a feed-forward (from x_i). This *learning term* makes a_i increase when *both* y and x_i take high values, thereby enhancing those coefficients that happened to yield a high y when a particular x_i was high. Thus, the structure of coefficients vector $\mathbf{a}$ ultimately depends on which vectors $\mathbf{x}$ appeared most often as inputs.

The learning term is such that the neuron learns the patterns that it receives most often. This is sufficient to make the network work, but makes it unable to construct different categories if different patterns appear. Furthermore, since the learning term works by multiplying inputs and outputs, it may produce an explosive output. This should be curbed in order to use the network.

For both reasons, a *forgetting term* that makes the coefficients a_i decay towards zero is in order. In equation 8 the forgetting term is $\gamma(\mathbf{a}, y)a_i$. It depends on a feed-back from output y and, most importantly, on coefficient a_i itself.

Figure 8 illustrates the feed-backs and -forwards within a neuron of an unsupervised network.

Simple, but non trivial examples of equation 8 are: $\dot{\mathbf{a}} = \mu y\mathbf{x} - \nu\mathbf{a}$, $\dot{\mathbf{a}} = \mu\mathbf{x} - \nu y\mathbf{a}$, $\dot{\mathbf{a}} = \mu y\mathbf{x} - \nu y\mathbf{a}$, $\dot{\mathbf{a}} = \mu y\mathbf{x} - \nu y^2\mathbf{a}$, where μ and ν are constants. Each functional form corresponds to different strengths of the learning and forgetting terms.

In general, a network of neurons is able to discriminate input information according to much finer categories than a single neuron can do. As a rule, the greater the number of neurons, the finer the categories that a network constructs. However, a neural network is useful precisely because it is able to classify a huge amount of information into a few broad categories. If categories are so fine that they track input information exactly, then a neural net becomes useless. Thus, the number of neurons that a network should possess depends on the variability of the input as well as on user needs.

However, the behaviour of a neural network does not only depend on the number of its neurons, but also on the structure of the connections between them. In fact, just

like the capabilities of neurons depend on feed-backs and -forwards, the capabilities of a neural network depend on linkages that eventually enable information to circulate in loops. If information can circulate within the network, then the whole network acquires a *memory*.

It is a *distributed* memory, fundamentally different in nature from the more usual *localised* memories. Localised memories such as books, disks, tapes etc., store information at a particular point in space. This information can only be retrieved if one knows where its support is (e.g. the position of a book in a library, or the address of a memory cell on a hard disk).

On the contrary, in a neural network each neuron may be part of a number of information circuits where information is "memorised" as long as it does not stop to circulate. Although this *is* a memory, one cannot say that information is stored at any particular place. Hence the name.

For obvious reasons, the information stored in a distributed memory cannot be retrieved by means of an address. However, a piece of information flowing in a particular loop can be retrieved by some other piece of information that is flowing close enough to it. Thus, in a distributed memory information can be retrieved by means of *associations* of concepts, with a procedure that reminds of human "intuition". Indeed, the connectionist idea of a distributed memory is a possible explanation for the existence in both humans and animals of an *associative memory*, i.e. the ability to establish an association between a particular stimulus and specific responses if the stimulus is repeated sufficiently often.

6.2 Evidence Theory

Evidence Theory is a branch of the mathematics of uncertain reasoning that, unlike Probability Theory, does not assume that a decision-maker knows the set of all possible events. [8] Rather than defining a "residual event" for anything that cannot be clearly expressed, Evidence Theory leaves a decision-maker's possibility set open to novelties.

Evidence Theory makes use of a particular class of monotone uncertainty measures, Choquet capacities of infinite order. Furthermore, it assumes that no operation is attached to the possibility set, which frees a decision-maker to define a "residual event" by complementation. Novel possibilities can appear in the possibility set in the course of the calculations, and the possibility set is called *frame of discernment* in order to stress its cognitive nature.

Evidence Theory does not take a gambler as its prototypical subject, but a judge or a detective. The reason is that a gambler playing with dice or throwing a coin knows what possibilities can occur. On the contrary, judges and detectives know that unexpected proves and testimonies may open up unexpected possibilities. Possibly, managers making investments, politicians steering their countries, or just anyone in the important choices of her daily life is more akin to a judge or a detective looking for cues than to a gambler looking for luck.

Let us consider a frame of discernment Θ. Let us suppose that a person receives

---

[8]The basic reference is Glenn Shafer, *A Mathematical Theory of Evidence*. Princeton, Princeton University Press 1976.

testimonies, or *bodies of evidence*, as numbers that to various extents support a set of possibilities $A_1, A_2, \ldots A_N$, where $A_1 \subseteq \Theta$, $A_2 \subseteq \Theta \ldots A_N \subseteq \Theta$ and where the A_is are not necessarily disjoint sets. [9] Let us denote these numbers $\{m(A_1), m(A_2), \ldots m(\Theta)\}$, where $m(A_i)$ measures the amount of empirical evidence that supports the possibility A_i.

Numbers m are exogenous to the person (the judge, the detective) who owns the frame of discernment. They are not subjective measures for this person, though they may be subjective evaluations of those who provide the testimonies. Numbers m are cardinal measures of the amount of empirical evidence supporting each possibility.

Since no operation is defined on the frame of discernment, the number m that has been assigned to Θ does not concern any specific possibility. Rather, it indicates how small the evidence is, that supports the possibilities envisaged in the testimony, or, in other words, how strongly a person fears that the possibilities that she is envisaging are not exhaustive. The greater the ignorance of a person on which possibilities exist, the greater $m(\Theta)$.

Note that $m(\Theta)$ can be smaller than any m of the A_is that it entails. Indeed, this applies to the A_is as well: if $A_i \supset A_j$, this does *not* imply that $m(A_i) > m(A_j)$.

Although it is not strictly essential for Evidence Theory, numbers m are generally normalised by requiring that:

$$\sum_{i=1}^{N} m(A_i) + m(\Theta) = 1 \qquad (9)$$

For instance, if the original format of the testimony is:

$$\{5, 32, 12, 3\}$$

by applying eq. 9 we obtain:

$$\{0.096, 0.615, 0.231, 0.058\}$$

whose numbers sum up to one.

Let us suppose that the decision-maker wants to evaluate to what extent the available empirical evidence supports certain hypotheses that she is entertaining in her mind. Since a hypothesis concerns the truth of a possibility or a set of possibilities, hypotheses are subsets of the frame of discernment just as possibilities are. A body of evidence $\{m(A_1), m(A_2), \ldots m(\Theta)\}$ supports a hypothesis H to the extent that some A_is are included or at least intersect H.

Note that, whilst the possibilities A_i entailed in the testimonies cannot be combined with one another (intersected, complemented, etc.) to form novel possibilities, a hypothesis H represents a free construct of the owner of a frame of discernment (the judge, the detective, etc.). This person is absolutely free to conceive any hypothesis, as well as its opposite. Thus, although $\overline{A_i}$s are forbidden, $\overline{H}$ can be safely considered.

---

[9] For simplicity, the theory is expounded with respect to a finite number of possibilities. No substantial change is needed if an infinite number of possibilities is considered.
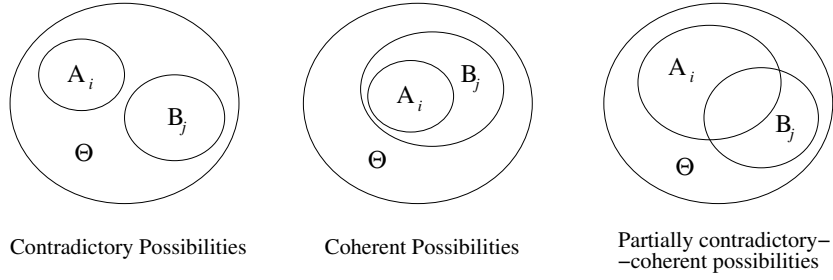
Figure 9: Left, two contradictory possibilities. Centre, two coherent possibilities. Right, two partially coherent, partially contradictory possibilities.

Given a testimony $\{m(A_1), m(A_2), \ldots m(\Theta)\}$, the belief in hypothesis H is expressed by the following *belief function*:

$$Bel(H) = \sum_{A_i \subseteq H} m(A_i) \tag{10}$$

By definition, $Bel(\emptyset) = 0$ and $Bel(\Theta) = 1$. However, this last condition does not imply that any of the possibilities included in the frame of discernment must necessarily realise. It simply means that any possibility must be conceived within the frame of discernment, independently of what possibilities are envisaged at a certain point in time.

The belief function takes account of all evidence included in H. The *plausibility function* takes account of all evidence that intersects H:

$$Pl(H) = \sum_{H \cap A_i \neq \emptyset} m(A_i) \tag{11}$$

It can be shown that belief and plausibility are linked by the relation $Pl(H) = 1 - Bel(\overline{H})$, where $\overline{H}$ denotes a hypothesis opposite to H. If $m(\Theta) > 0$ these two measures are not equivalent, so both of them need to be considered. In general, $Bel(H) \leq Pl(H)$.

Let us suppose that some unexpected facts occur, that are told by a new testimony. The new testimony must be combined with previous knowledge, confirming it to the extent that it is coherent with it. On the contrary, previous beliefs must be weakened if the new evidence disconfirms them.

Let $\{m(B_1), m(B_2), \ldots m(\Theta)\}$ be the new testimony, which must be combined with $\{m(A_1), m(A_2), \ldots m(\Theta)\}$. The new testimony may entail possibilities that are coherent with those of the previous testimony, possibilities that contradict those of the previous testimony, and possibilities that partially support, partially contradict the previous testimony. Figure 9 illustrates contradictory, coherent and partially coherent/contradictory possibilities on the frame of discernment. Contradictory possibilities appear as disjoint sets. A possibility is coherent with another if it is included in it. Finally, two possibilities that are partially coherent, partially contradictory, intersect one another.

Let us suppose that two testimonies

$$\{m(A_1), m(A_2), \ldots m(\Theta)\}$$

30

and

$$\{m(B_1), m(B_2), \dots m(\Theta)\}$$

that satisfy eq. 9, must be combined into a testimony

$$\{m(C_1), m(C_2), \dots m(\Theta)\}$$

that also satisfies eq. 9. Dempster-Shafer's combination rule yields a combined testimony $\{m(C_k)\}$ where the coherent possibilities between $\{m(A_i)\}$ and $\{m(B_j)\}$ have been stressed.

According to Dempster-Shafer combination rule, possibilities $\{C_k\}$ are defined by all intersections of each possibility in $\{A_1, A_2, \dots \Theta\}$ with each possibility in $\{B_1, B_2, \dots \Theta\}$. For any possibility C_k, the amount of empirical evidence is:

$$m(C_k) = \frac{\sum_{A_i \cap B_j = C_k} m(A_i)m(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m(A_i)m(B_j)} \tag{12}$$

The numerator of eq. 12 measures the extent to which both the first and the second testimony support the possibility C_k. In fact, for each possible C_k the sum extends to all pairs of possibilities from the two testimonies that are coherent on C_k (see fig. 9). The more the intersections between the A_is and the B_js that give rise to C_k, and the greater their amounts of evidence, the larger the numerator.

The denominator is the complement to one of those elements of the second testimony that contradict the first one. In fact, the complement to one is made on those A_is and B_js that are disjoint sets (see fig. 9). The denominator represents a measure of the extent to which the two testimonies are coherent, in the sense that all evidence that supports contradictory possibilities is excluded.

Essentially, Dempster-Shafer combination rule says that the evidence supporting possibility C_k is a fraction of the coherent evidence between $\{m(A_1), m(A_2), \dots m(\Theta)\}$ and $\{m(B_1), m(B_2), \dots m(\Theta)\}$. The amount of this fraction depends on the sum of all elements of the testimonies that support C_k.

Dempster-Shafer's rule can be iterated to combine any number of testimonies. The outcome of Dempster-Shafer combination rule is independent of the order in which two testimonies are combined. [10]

The above description made clear that Evidence Theory provides an algorithm for handling an exogenous flow of new, unexpected possibilities. Indeed, the decision-maker of Evidence Theory is not supposed to conceive possibilities. She merely listens to exogenous testimonies that consist of possibilities and degrees of evidence supporting them, and combines these testimonies into a coherent whole by means of Dempster-Shafer theory. She does not conceive novel possibilities out of a creative effort. Rather, novel possibilities — the $\{C_k\}$ — arise out of combination of exogenous inputs.

On the contrary, Probability Theory ascribes its subjects the ability to conceive novel possibilities by applying a set of operations (e.g., union, intersection, complementation) to a given set of elementary possibilities. Under this respect, Probability Theory with its σ-algebras is conceptually akin to classical artificial intelligence, in the

---

[10]A detailed numerical example can be found in Guido Fioretti, "Evidence Theory as a Procedure for Handling Novel Events". *Metroeconomica*, 60 (2) 2009: 283–301.

sense that classical artificial intelligence assumed that intelligent behaviour, expressed by the ability to generate higher-order concepts, results from the application of logical operations on a set elementary concepts. Both lines of reasoning suppose that complex reasoning from the combination of given bricks with given rules.

6.3 Constraint Satisfaction Networks

Parallel constraint satisfaction networks (CSN) arrange concepts into coherent theories. Although they belong to the larger family of neural networks, they do not carry out any classification process.

Constraint satisfaction networks are characterized by:

- Excitatory and inhibitory connections;

- Feedbacks between neurons.

Neurons represent possibilities, or concepts, or propositions. Connections represent inferences: an excitatory connection from neuron A to neuron B means "A implies B", whereas an inhibitory connection from neuron A to neuron B means "A implies ¬ B".

Let a_i denote the activation (the output) of neuron i, with $a_i \in \Re$. Let $w_{ij} \in \Re$ denote the weight by which neuron i multiplies the input arriving from neuron j.

The net excitatory input to neuron i is:

$$enet_i = \sum_j w_{ij}a_j \qquad if \ w_{ij}a_j \geq 0 \tag{13}$$

The net inhibitory input to neuron i is:

$$enet_i = \sum_j w_{ij}a_j \qquad if \ w_{ij}a_j < 0 \tag{14}$$

At each time step, the activation of neuron i is increased by its excitatory inputs and decreased by its inhibitory inputs:

$$\Delta a_i = enet_i(a_{max} - a_i) + inet_i(a_i - a_{min}) \tag{15}$$

where, in general, $a_{max} = 1$ and $a_{min} = -1$.

Feedbacks between neurons make the network maximize consonance:

$$C = \sum_i \sum_j w_{ij}a_ia_j \tag{16}$$

or, equivalently, minimize energy $E = -C$.

Consonance maximization means that those neurons are strengthened, that represent possibilities, concepts or propositions that are coherent with one another. Thus, constraint satisfaction networks can be used to model any cognitive process characterized by a search for coherence. [11] In particular, researchers have emphasized the ability of CSN to construct narratives, much like humans actually do.

---

[11] The basic reference is Paul Thagard, *Coherence in Thought and Action*. Cambridge (MA), The MIT Press 2000.

Notable applications of CSN are the elaboration of scientific theories, which amounts to arrange empirical findings in a network of coherent causal relations, as well as the evaluation of guilt or innocence in a trial, which amounts to fitting testimonies in a coherent frame. CSN have also been used to model post-decisional reduction of dissonance.

Furthermore, CSN can be used to model the process of emphasizing the positive aspects of one alternative and the negative aspects of its competing alternatives until a coherent frame is available and a decision can be made. This oscillation between competing explanations reproduce at least one important aspect of *Gestalt* theories, namely, the idea that the human mind may shift among alternative interpretations of reality, as exemplified by Rubin's vase and other images where at least two interpretations are possible. [12] Many cues suggest that this is the fundamental pattern of decision-making.

A clear limitation of CSN is that they work with given possibilities, concepts, or propositions. In other words, CSN can reproduce the arrangement of possibilities and concepts, not their arousal. In order to include this feature it has been proposed that CSN represent unconscious arrangement of available possibilities and concepts, whereas a conscious process running in parallel would care about the search and elaboration of novel ones.

7 Conclusions

This review presented tools to model decision-making according to two opposing paradigms, namely, the logic of consequence and the logic of appropriateness. The reader may feel unease because scientists do not provide a univocal answer to the demands of the modeller.

However, a pragmatic attitude may suggest that tools should be used depending on conditions. Utility maximization and Game Theory require that all available alternatives and all their possible consequences can be listed. Thus, it may be sensible to make use of these tools when one such exhaustive list is available, eventually releasing the requirement of perfect rationality and the pursuit of Nash-equilibria while assuming some form of bounded rationality as influence games do. Unsupervised neural networks, Evidence Theory and constraint satisfaction networks, on the contrary, may be used when more challenging decision settings must be modelled. The modeller should remember that constructing narratives makes sense because the world is uncertain even in what possibilities may exist, so it is only when modelling such decision settings that these tools make sense.

The trouble, in this last case, is that the tools mentioned above have not been integrated into a unified framework. No simple formula is ready to be used, so the modeller must resort to a higher degree of creativity and intuition. On the other hand, here is an exciting opportunity for modellers to participate to theory development.

---

[12]The simplest picture of this kind is a cube depicted by its edges: it is up to the observer to choose which face stays in the front and which face stays in the rear. Rubin's vase is white and stands against a black background. The observer may see a white vase, or two black profiles in front of one another.