Complex Methodological Individualism[1]

JEAN PETITOT

École des hautes études en sciences sociales (EHESS)
Centre d'analyse et de mathématique sociales
190-198 avenue de France F75244 Paris Cedex 13 France
Bureau 567 – 5e étage
France

Email: petitot@ehess.fr
Web: http://jean.petitot.pagesperso-orange.fr/index.html

Bio-sketch: Jean Petitot, author of 12 books and over 300 papers, is a specialist of mathematical modeling in social and cognitive sciences and is a full member of the International Academy of Philosophy of Science. He has worked on complex methodological individualism and Hayek and co-edited with Philippe Nemo *Histoire du libéralisme en Europe* (Presses Universitaires de France, 2006).

Abstract: The paper begins with a few reminders of the history of complex methodological individualism in the general context of complex systems. It then focuses on Hayek's concepts of complexity, spontaneous order, and cultural evolution. In a second part, it sketches some models of a cooperative "invisible hand" in the theory of evolutionary games, in particular for the iterated and spatialized prisoner's dilemma.

Keywords: complex systems, cultural evolution, evolutionary games, Hayek, methodological individualism, prisoner's dilemma.

I: INTRODUCTION

To understand complex methodological individualism, we need to focus on some preliminary issues related to both emergence and reductionism.

Classically, a reductionist thesis posits that complex high level phenomena, structures and processes can be reduced, as far as their scientific explanation is concerned, to underlying lower level phenomena, structures and processes. The most paradigmatic and best-investigated example is the reduction of macroscopic thermodynamics to microscopic molecular and atomic movements (temperature = mean kinetic energy per degree of freedom, etc.). Let us begin with some conceptual precisions.

1. Reductionism can be a particular scientific thesis concerning a specific scientific theory: it is precisely the case with the reduction of macro thermodynamics to micro statistical mechanics. But it can be also a general metaphysical claim on the ultimate nature of reality. That is the case with different forms of monism. Idealist monism posits the universal reducibility of reality to mind while materialist monism posits the universal causal reducibility of reality to matter and energy. In this paper we will be concerned only with scientific reductionism.

2. Reductionism can concern theories dealing with empirical data and in that case focuses on the problem of causality. But it can also concern purely linguistic and formal theories. For instance lexical definitions or logical axioms consist in trying to reduce complex contents to a small list of primitive contents. In mathematics, many theories deal with the possibility of eliminating higher order concepts, objects or axioms in proofs and deflate rich theories to more restrained ones (theorems of elimination): for instance one can prove that quantifiers can be eliminated in algebraic geometry (Tarsky-Seidenberg) or that a proof using the axiom of choice can be transformed into a proof without the axiom of choice, etc. In this paper, we will be concerned only with theories having an empirical content.

3. Scientific object-oriented reductionism can be *objective* or *methodological*. It is ontological when it concerns explanations in terms of primitive objects (atoms, neurons, etc.) and methodological when it concerns deflationist nominalist explanations (Occam's razor). There are well known examples of eliminative methodological reductionism having eliminated pre-scientific speculative concepts and entities such as "phlogiston" or "vitalist entelechies", etc. To day, a very important debate in cognitive science has to do with the eliminability of "mental" or "conscious" concepts and their reduction to neural concepts (see e.g. the Dennett/Chalmers controversy). In this paper we will be concerned with "objective" reductionism.

4. Now, the main point is that, in our narrow, scientific, empirical, and objective sense, reductionism is by no means eliminativist and it is perfectly compatible with emergence in *complex systems* characterized at least by two levels of reality: a micro underlying level where a great number of elementary units are in interaction and a macro emergent one where *macro self-organized structures* emerge. In such a perspective, reductionism is inseparable from converse concepts such as "emergence", "supervenience" or "functionalism". Functionalism means that macro structures having a functional role can exist only if they are materially implemented in an underlying material substrate, but are at the same time, as functionally meaningful structures, largely *independent* of the fine grained physical properties of the substrate they are implemented in. The paradigmatic example is the opposition software/hardware in computer sciences (see philosophers like Putnam, Fodor, Pylyshyn, etc.) but functionalism also applies in natural sciences where it is an aspect of emergence.

5. There is a general agreement on the fact that in complex systems having different levels of reality at different scales, there exist collective behaviors ruled by laws that are not the laws of the micro underlying level. It is the case for critical phenomena, percolation, self-organized criticality, reaction-diffusion equations, dissipative structures, turbulence, cellular automata, neural networks, ant colonies, swarms, stock markets, etc. According to one's conception of laws, one can develop different conceptions of this empirical fact.

(i) *Eliminativism and epiphenomenalism*: laws being only empirical regularities lacking any objective (and a fortiori ontological) content (Hume's empiricist thesis), emerging structures are purely epiphenomenal and can be scientifically eliminated "salva veritate".

(ii) *Holistic realism* (it is the converse position): laws being real in the ontological sense, the emerging level possesses an ontological reality and cannot therefore be reduced.

(iii) *Causal reductionism and objective emergentism*: laws being objective, that is at the same time empirically grounded and mathematically formalized, the emerging level has no holistic ontological content but is nevertheless much more than a simple empirical regularity. It is *causally* reducible to complex interactions at the micro underlying level but it shares nevertheless some empirical and theoretical *autonomy*.

We will be concerned here with this third type of situation, that is with objective emergentism.

6. The main difficulty that has to be tackled in such a perspective is the relation between causal reduction and theoretical autonomy. *Mathematics* play here the fundamental role. Indeed, the formal equivalent to causal reduction is *mathematical deduction*. But deducibility is a syntactic property and doesn't entail any evident *conceptual* derivation (it is for that very reason that mathematics constitute an authentically "synthetic" knowledge even if their proofs are "analytical"). Therefore the fact that the structures and properties of the macro level can be mathematically deduced from the micro one doesn't mean that the representational content of its conceptual description can be reduced to the representational content of the micro level. A very spectacular example is that found in statistical physics where magnetic critical behaviors can be classified, via the *renormalization group*, in universal classes independent of the specific fine-grained physical structure of the substrate.

The renormalization group is a dynamical method that enables to define these universal behaviors as attractors of a certain dynamics on the space of Hamiltonians. Near the critical temperature ($T = T_c$) the macro variables of the system (magnetization, specific heat, magnetic susceptibility, etc.) follow power laws $(\Delta T)^{\propto}$ (where $\propto$ is called the critical exponent of the variable). Empirical data have shown that there exist universal classes of critical exponents linked by very precise relations. These universal classes depend only on very general abstract dimensional and symmetry properties of the substrate and not on its detailed physical structure. The point is that if you prove mathematically that such a critical phenomenon arises from a symmetry breaking of

an order parameter, this doesn't mean that this macro and abstract symmetry breaking modeled via group theory, has something to do with the concept of a spin of a particle. Causal reduction paralleling mathematical deduction is not a conceptual reduction. Moreover, the universality of critical exponents — that is the existence of invariants — proves that the emerging critical phenomenon under consideration has some measure of autonomy and belongs to an autonomous level of reality.

It is in this framework, that we will present some remarks on *Hayek's catallaxy* and *evolutionary game theory*.

II. COMPLEX METHODOLOGICAL INDIVIDUALISM IN HAYEK

Methodological individualism has to do with the reduction of collective properties to interactions between individual ones. *Complex* methodological individualism advocated by thinkers like Ludwig von Mises, Friedrich von Hayek or Jean-Pierre Dupuy is neither holistic nor eliminativist but *emergentist*.[2] Friedrich von Hayek paradigmatically represents it.

Hayek was one of the first to develop the consequences of the theories of self-organization and spontaneous order in cognitive and social sciences.

1. Hayek and the complexity problem

Hayek always strongly emphasized the specific properties of the complex socio-economic spontaneous order in modern open societies. It is a sophisticated self-organized order where "laissez faire" does not produce anarchy, but an order that is cognitively founded and would be impossible to obtain in another way. Its endogenous complexity is irreducible and, according to Hayek, dooms to failure any rationalist constructivism that would claim to create it artificially. Under the name of "constructivism", Hayek criticizes here a new type of reductionism, namely the possibility to reduce a natural complex order to the application of a system of rational rules. In a certain sense, he claims that there cannot exist an AI expert system for modern societies and markets.

The source of complexity has to be found in the fact that, in an open society, knowledge, competencies and informations are *distributed*, scattered over a great number of cognitively limited and interacting agents. The systemic properties of such systems cannot be conceptually controlled. The political control of social and economic orders rests on a methodological error.

Many consequences derive from this fundamental fact.

(i) Complexity prohibits at the same time a centralized hierarchical organization and a communal link of reciprocity characteristic of small closed communities. In modern open societies the interactions between agents is no longer ensured by consensus on shared values but by exchange of signals such as prices in a market. Market is a way of circulating information in a multi-agent system whose very complexity makes it opaque to its own agents. In a Hayekian "catallaxy" everyone cooperates with everyone else but without any shared ends. The individual aims are incommensurable with each other but mechanisms such as free trade and markets guarantee nevertheless a viable cooperation.

(ii) Complexity is an evolutionary process resulting from a selection of historico-cultural rules of behavior, practices, institutions that are impossible to master conceptually. In that sense, political, juridical, and social constructivism appears to be the dark side of the Enlightenment.

(iii) Rules that govern social exchanges and communication are abstract and formal. Social self-organized complex systems are governed by civil rights guaranteed by public laws.

The main critique raised by Hayek against political constructivism is that it does not understand what a complex order is and is in fact not "progressive" at all but "regressive".

2. Towards a rational justification of Hayekian anti-constructivism

Progressively, these problems have become accessible to scientific inquiry and modeling. In particular, the idea that many common-sense rules have been selected by an evolutionary process and constitute an optimizing collective form of "learning", seems to be essentially right. We will present below an illustration, but let us first recall some aspects of methodological individualism.

2.1. The paradigms of social order

The concept of spontaneous order must be put in historical context. It posits that pluralism and individual freedom are not sources of disorder, anarchy and social struggle but, on the contrary, a factor conducive to higher forms of organization. It stands in sharp contrast with three other paradigms:[3]

1. The paradigm of hierarchical order and absolute power theorized from the Renaissance by Machiavelli (1469-1527), then Bodin (1529-1596) and Hobbes (1588-1679), and put in practice for instance in Spain by Charles V and Philippe II, or in France by Richelieu, Louis XIV and Napoleon. It is in reaction to this form of absolutism that many demands arose for tolerance and human rights, from Grotius (1583-1645), Bayle (1647-1706) and Locke (1632-1704) to Kant (1724-1804), Humboldt (1767-1835) and Benjamin Constant (1767-1830). It was the source of many revolutions: in Netherlands, England, America and France (before the Terror). It was the main origin of modern science, techniques, the industrial revolution, and prosperity.

2. The revolutionary paradigm of rational constructivist order that rejects open society in the name of great ideals of equality and justice and relies on political planning to create a new humanity.

3. The conservative paradigm of natural order, which rejects also open society, but for an opposite reason: it champions a form of organicist holism and accuses modernity for having "atomized" society (individualism) and destroyed "natural communities" (family, corporations, churches, etc.).

The paradigm of spontaneous order posits a new conception of social order as neither natural (permanent and universal) nor artificial (rationally construed), but pluralist and self-organized, non hierarchical and polycentric. Evident examples of such orders are language, law or morals: they are not natural in the strict sense of the term, but they are neither artificial since nobody has ever made them. As the masters of the Scottish Enlightenment David Hume (1711-1776) and Adam Ferguson (1723-1816) emphasized, they are the results of human actions but not of human intentions.

2.2. Methodological individualism[4]

In this context, the problem of methodological individualism (MI) — that is of the reducibility of macro social structures to micro individual interactions of agents — is central. In classical sociology, *holistic realism* is dominant. According to it, social phenomena must be explained in terms of macro-social and supra-individual collective entities prior to individual agents and transcending them: states, churches, parties, classes, nations, markets, etc. Holism aims at explaining how such "real" social entities prescribe norms and values to individual subjects.

(i) For Saint-Simon (1760-1825, *De la physiologie appliquée à l'amélioration des institutions sociales*: 1813) and Auguste Comte (1798-1857, *Système de politique positive*: 1851) holism was a sort of "organicism", a "physiological" conception of the social reality opposing "mechanistic atomism" developed by "social physics".
(ii) With Durkheim (1859-1917, *De la division du travail social*: 1893, *Règles de la méthode sociologique*: 1895, *Les formes élémentaires de la vie religieuse*: 1912) holism is no longer biologically inspired and becomes a true sociological thesis. Social wholes exist and subsist *de re*, and determine the actions of empirical individuals. Of course, there exist "horizontal" interactions between individuals but the true social causality is "vertical" and "top down" and flows from social wholes to individual parts.

(iii) By definition, all variants of socialism and communism are also holistic.

MI considers that holism is a mythology. It rejects any substantial hypostasis of global concepts and develops a modern variant of the fight of nominalism against realist conceptions of universals. For it, as for Occam, social groups are aggregates and not substances. It is called methodological because it concerns explanation and not ontology.

For Karl Popper (*The Poverty of Historicism*, Economica, 1944 and Routledge, 1986), MI is an unassailable thesis according to which all collective phenomena must be reduced to actions, interactions, goals, hopes, thoughts of individual subjects as well as to the traditions they have created and maintained. For Jon Elster (*An Introduction to Karl Marx*, Cambridge University Press, 1986) it is the thesis according to which all social phenomena, their structure and change, can be explained using only individuals with their qualities, beliefs, goals and actions.

The founders of MI are well known:

(i) John Locke (1632-1704). Individuals are the basic social entities but they interact in a contractual society protected by the rule of law.
(ii) Bernard de Mandeville (1670-1733) and his celebrated "*The Grumbling Hive: or, Knaves Turn'd Honest*" (1705) also known as "*The Fable of the Bees; or, Private Vices, Public Benefits*" (1714). He triggered a tremendous

controversy (for instance with Berkeley) because he introduced a principle of *inversion* between individual intentions (micro-level) and non-intentional emerging social properties (macro-level). Individuals are intentionally selfish and governed by their private and local self-interest but their interactions generate, in a *non-intentional* way, a global social order propitious to public interest.

(iii) The Scottish Enlightenment: Hume, Ferguson (see above).

(iv) Adam Smith (1723-1790) and the "*invisible hand*" (*Theory of Moral Sentiments*, 1759, *The Wealth of Nations*, 1776). The essential feature of the invisible hand is that it drives subjects to collective ends that do not proceed from their intentions.

We see that methodological individualism concerns mechanisms of self-organization, which cannot be rationally computed by agents. Social cohesion, cooperation, prosperity are non-intentional effects emerging from an aggregation of selfish interests.

Many variants of MI proceeded from these early works, some more utilitarian and reductionist (John Stuart Mill 1806-1873, Léon Walras, 1834-1910 and Vilfredo Pareto, 1848-1923: *Traité de sociologie générale*, 1916), other more organicist (but not holistic, Herbert Spencer 1820-1903: *The Principles of Sociology*, London, Williams and Norgate, 1882-1898).

Complex MI was founded by the Austrian school: Carl Menger, Ludwig von Mises and Friedrich von Hayek. It rejects of course holist mythology, but it is emergentist; it is neither reductionist nor mechanistic. Carl Menger (1840-1921, *Grundsätze des Volkwirtschaftlehre*, 1871; *Untersuchungen über die Methode des Socialwissenschaften*, 1883) was the first to make explicit this problem of complexity. He was followed by Hayek who gave the best theoretical clarification of self-organized orders (language, religion, law, money, market, state, etc.) that are not the result of a collective intentional will. The parallel with theories in natural sciences (physics, biology, neurosciences) is striking.

(i) Order is a consequence of the coordination of individual agents.

(ii) Emerging collective structures acquire some autonomy even if they are causally reducible to individual interactions.

(iii) They are structurally stable if agents respect rules of law.

(iv) These rules result themselves from a form of cultural evolution.

(v) Emerging structures are non-intentional and unpredictable (no rational planning is possible).

(vi) It is a fundamental error to attribute intentionality to them. That error is one of the main sources of totalitarianism.

(vii) As we will see, cultural evolution is Darwinian in a specific, non biological, sense.

3. Cultural evolution and emerging ethical maxims

At the cognitive level, be it individual or social, according to Hayek, the origin of the rules governing perception and action, as well as that of conventions and norms, is evolutionary. These patterns result from a cultural selection—a collective learning—which is a competitive/cooperative process having favored the individuals and groups that applied them. They are like cultural short-cuts enabling people to behave rapidly and adaptively without having to recapitulate every time all the experiences and beliefs necessary to action. For Hayek, *common-sense* is a library of tacit knowledge routines and practical schemes patterning our experience after generic default schemes. It is necessary to act without being overwhelmed by the overflow of irrelevant informations coming from the environment. For Hayek (as for Mandeville, Hume or Ferguson), common sense norms are not repressive constraints but, on the contrary, cognitive achievements deeply adapted to the contingencies of life. Traditions express an "embodied knowledge" which is "phylogenetic" in the sense of cultural evolution, and it is therefore rational to comply with them "ontogenetically".

In much the same way as in evolutionary biology, phylogenetic a posteriori operate as ontogenetic a priori, common sense rules operate for the subjects as a priori frames. In this sense, we find in Hayek an evolutionary theory of the *self-transcendence* of behavioral rules. Like linguistic rules, they proceed from symbolic institutions whose origin is neither a rational omniscient intelligence nor a deliberative social contract.

We see how Hayek articulates cognitive psychology (the "sensory order") with the sociology of complex spontaneous orders ("catallaxy").

We know that the very concept of cultural evolution is quite problematic. For Hayek, as for Popper, cultural evolution selects groups and not individuals, subjects having to comply with rules that maximize the collective performances of their group. However, for the subject themselves, it is

impossible to understand in what operational sense these norms are socially fruitful because they encode a "phylogenetic" historical evolution. That's why they interpret them as *duties* and *values*. We must emphasize the originality of this conception:

1 As individuals cannot understand the pragmatic efficacy of norms, they accept them for *deontic* reasons. We recognize here a thesis that belongs in Kantian ethic.

2 However, norms being socially useful we recognize also a *utilitarian* conception of ethics (Jeremy Bentham, John Stuart Mill). The main difference is that the "computation" of moral maxims and actions is cognitively inaccessible for individuals.

Therefore, according to Hayek, cultural evolution implies that maxims of action can act for individuals as transcendent "categorical" imperatives[5] while they are at the same time immanent "hypothetical" (pragmatic) imperatives for cultures.[6] For cultures, maxims are caused by the viability of a social order from which individuals gain a lot. As was emphasized by John Gray, Hayekian utilitarianism is *indirect* and exemplifies the general evolutionary principle (Haeckel's law) according to which phylogenetic a posteriori operate ontogenetically as a priori. Hayek was able to reconcile, from within methodological individualism, reductionism with holism: social entities prescribe norms, rules and maxims to individuals.

It is interesting to highlight how Hayek succeeded in renewing the notion of categorical imperative as a deontological (non consequentialist) conception of actions. According to deontological theses, actions must be evaluated in a principled way independently of their consequences, while according to consequentialist theses they must be evaluated on the basis of a computation of the costs and benefits of their consequences. But as that kind of computation is impossible for a finite and limited rational mind, it is performed by cultural evolution. As was emphasized by Jean-Pierre Dupuy, cultural evolution is "utilitarian" but bears on "deontological" maxims that can be interpreted in accordance with a test of "categoricity".

III. THE EXAMPLE OF EVOLUTIONARY GAMES

Let us now consider an example of how social modeling upholds some Hayek's theses. We shall take the problem of selection of social rules and shows how it can be modeled in terms of evolutionary game theory.

Ever since the pioneering work of Robert Axelrod (see, e.g., Axelrod *et al.*, 1998), many models have been dedicated to underlying causal mechanisms of complex adaptive social systems (see e.g. Binmore 1994). The simplest and best-known example is that of the *Iterated Prisoner's Dilemma* (IPD).

3.1. The prisoner's dilemma

Let us first recall the classical Prisoner's Dilemma (Poundstone 1993). There are two players A and B and each player can choose one of two behaviors (strategies): d = defection and c = cooperation. In order to compute gains and losses (profits and deficits), we use a matrix of payoffs with columns $A(c)$ (A plays c) and $A(d)$, and lines $B(c)$ and $B(d)$. Each entry corresponds therefore to a one-shot game and we introduce the players's payoffs: the column player *A's* in the upper right corner and the line player *B's* in the lower left corner. The payoff matrix involves 4 terms:

$T = (d, c)$ = Temptation,
$S = (c, d)$ = Sucker,
$R = (c, c)$ = Reward,
$P = (d, d)$ = Punishment.

For the game to be interesting (there must exist a "dilemma") payoffs must satisfy the set of inequalities:

$T > R > P > S$.

Here is a typical example:

	A(c)	A(d)
B(c)	R = 3 R = 3	T = 5 S = 0
B(d)	S = 0 T = 5	P = 1 P = 1

Behaviors:
d = defection, c = cooperation

Payoffs:
T = (d, c) = 5, S = (c, d) = 0
R = (c, c) = 3, P = (d, d) = 1

Conditions:
T = 5 > R = 3 > P = 1 > S = 0
(T + S)/2 = 5/2 < R = 3

This extremely simple game is not trivial since it represents a situation where *individual* rationality is at odds with *collective* rationality. Indeed:

(i) If column player A plays c, then line player B gets R if he plays c and T if he plays d. As $T = 5 > R = 3$, it's in the interest of B to play d.

(ii) Now, if column player A plays d, then line player B gets S if he plays c and P if he plays d. As $P = 1 > S = 0$, it's still in the interest of B to play d.
(iii) Therefore, if B is *rational* in the individualist sense, he must play d whatever A's behavior. It is said that strategy d strictly *dominates* strategy c : d is better than c whatever the other player's behavior.
(iv) The same holds for A by symmetry.
(v) The rational outcome of the game is therefore the non-cooperative behavior (d, d), which leads to the very bad collective payoff ($P = 1$, $P = 1$).
(vi) But clearly, the cooperative behavior (c, c) would have led to a largely better collective payoff ($R = 3$, $R = 3$).
(vii) So individual rationality selects a poor (lose, lose) strategy (d, d) while a collective rationality would have selected a good (win, win) strategy (c, c).

The dilemma comes from the fact that for the above payoff matrix the double strategy (d, d) is the only *Nash equilibrium* (NE), that is the only strategy having the property that each player would do worse if they changed unilaterally their strategy.

One can generalize this basic example in multiple ways, introducing asymmetries, non-strict inequalities, neutral behaviors (a player can refuse to play), multiple players, probabilistic strategies (a player plays c with probability p and d with probability $1 - p$, etc.). The main result is that the dilemma is *robust*. How can we therefore explain the emergence of *cooperative* collective behaviors through an evolutionary selective process? It is clearly a fundamental problem.

3.2. The iterated prisoner's dilemma (IPD)

The situation changes completely when one *iterates* the game, because defection can then be punished and cooperation rewarded. We can in that case introduce genuine strategies. We must suppose that the number of moves is indeterminate to avoid *backward induction* (the possibility of defining a strategy by going backwards from the desired result to the initial move) that has the property of leading us back to the non-cooperative behavior (d, d) (double defection). We can test strategies such as G = "good" (sucker) = play always c (unconditional cooperation); M = "bad " (meany) = play always d (unconditional defection); *TFT* = "tit for tat" = start with c (initial cooperation), then play what the other player played at the previous move; V = "vindictive" = start with c and play d for ever as soon as the other player plays d (that is defection is punished as an irreversible betrayal), etc. One pits these strategies against one another over a great number of plays (for instance 1000) and one compares their scores. The notion of a Nash equilibrium (NE) must be revised since the strategy (*Id, Id*) that iterates the one-shot NE (d, d) remains a NE. But many other strategies yield the same result as *Id* when playing against *Id*, and there exist too many NE's. Hence the concept of "subgame perfect equilibrium" which is a NE for every sub-game of the game.

For pools of strategies that are not too complex, one finds that the strategy tit-for-tat (*TFT*) has a striking superiority. *TFT* does not win every time but it always gets a very good score. More generally, computer simulations show that the best strategies are nicely cooperative, rapidly reacting to defections ("retaliatory"), rapidly forgiving, and simple ("clear", without wiles). "Good" (sucker) and "bad" (meany) strategies are catastrophic.

3.3. Evolutionary games

Evolutionary game theory considers polymorphic populations of individuals using different strategies and defines new generations using the scores in a generalized competition: strategies with good scores increase their number of representatives while those with bad scores progressively vanish. Evolutionary theory is more realist than the classical one based on individual rationality. It substitutes a collective selective scheme to an impossible variational calculus. Moreover, it enables us to understand the dynamics that drive agents towards global equilibria (see e.g. Hofbauer & Sigmund 1988; Livet 1998; Samuelson 1997; Weibull 1996; Kirman 1998).

Let $\{s_i\}$ be the set of strategies and $\{p_i\}$ their respective probabilities (i.e. the proportions of the population playing them). We suppose that the size N of the population remains constant. Consider the case where there are only two strategies: c with probability = p and d with probability = $1 - p$). It is easy to compute the expectation of gains (utilities $U_c(p)$ and $U_d(p)$) for each strategy as a function of the parameter p. Recall that $T = (d, c)$, $S = (c, d)$, $R = (c, c)$, and $P = (d, d)$. If an agent plays c, the probability that he will play against another agent playing c is p and he will gain $(c, c) = R$, while the probability that he will play against another agent playing d is $1 - p$ and he will gain $(c, d) = S$. If the agent plays d, the probability that he will play against another agent playing c is p and he will gain $(d, c) = T$, while the probability that he will play against another agent playing d is $1 - p$ and he will gain $(d, d) = P$. We get therefore:

$$\begin{cases} U_c(p) = pR + (1-p)S \\ U_d(p) = pT + (1-p)P \end{cases}$$

The mean gain of the population is therefore given by the quadratic expression:

$$U(p) = pU_c(p) + (1-p)U_d(p) = p^2R + p(1-p)S + (1-p)pT + (1-p)^2P$$

that is:

$$U(p) = p^2R + p(1-p)(S+T) + (1-p)^2P$$

The evolution of the probability p is given by the *replication dynamic*

$$p' = p\big(U_c(p) - U(p)\big)$$

3.4. The "tit for tat" strategy: from common sense to dynamical models

In these models, agents are considered as "phenotypes" expressing "genotypes" identified with strategies, and "micro" strategies influence "macro" population dynamics. Simulations (which specialists of complex systems call "computational synthesis") provide extremely interesting results. Axelrod has shown that:

(i) Anti-cooperative strategies are eliminated, cooperation wins and becomes stable.
(ii) *TFT* dominates, but is *fragile* with respect to mutations; indeed sucker mutants *Ic* exhibit exactly the same behavior as *TFT* in a *TFT* environment; they can therefore substitute themselves progressively and "silently" for *TFT*, without any observable effect; but then "bad" mutants *Id* (meanies) can destabilize, invade and destroy the system.
(iii) For a strategy, to react to defections (to be retaliatory) is a condition for being *collectively stable*, that is to resist destabilizations by "bad" mutants.
(iv) If one introduces complex strategies, many subtle phenomena can occur. For instance, a non-cooperative strategy can use another one to eliminate cooperative strategies and eliminate its allies in a second step; social disorders enable some non-cooperative strategies to survive and even win the game, etc.
(v) Simulations show that there exist sophisticated refinements of *TFT*, which improve slightly its results in more complex contexts. But we can say that *TFT* is the most efficacious simple strategy.

Now, it is also an empirical anthropological and cultural fact that since ancient times *TFT* has been selected by common sense.

We meet here a typical *model of common-sense*:

(i) Simulations corroborate an old common sense rule proceeding from collective political and social knowledge.
(ii) But at the same time, they enable us to overcome naive common sense and to develop an *experimental* framework for virtual (modeled) cultural evolutions.

3.5. Sigmund's and Novak's generalizations

Many authors have studied factors that facilitate cooperation in the *IPD* when one changes the space of strategies, the interaction process, the adaptive responses, etc. We will say a few words on the introduction of "topological" relations of neighborhood between agents, each agent becoming able to imitate the one of his neighbors that makes the best score.

Consider for instance extremely simple strategies (i, p, q) where:

i = initial probability of cooperation,
p = probability of cooperation after a cooperative move by the other player,
q = probability of cooperation after a defective move by the other player.

We have trivially $Ic = (1,1,1)$, $Id = (0,0,0)$, $TFT = (1,1,0)$, $c_p = (p, p, p)$ (always c with probability p). Other interesting strategies are $GTFT = \left(1, 1, Min\left(1 - \frac{T-R}{R-S}, \frac{R-P}{T-P}\right)\right)$ and Kraines' "Pavlovian" strategy that resists *Id* well: play c after R or T, and d after P or S.

Sigmund and Nowak have shown that "bad" agents *Id* ("meanies") can win at the beginning of the game. However, *TFT* agents resist. Once the "good" *Ic* ("suckers") have been decimated, the exploiters can no longer abuse them and cooperative strategies of *TFT* type emerge. But after this emergence of cooperation, the *TFT* strategies are themselves overtaken by *GTFT*. However, the *GTFT* strategy is fragile and allows for the return of "bad" *Id*.

3.6. Spatialized IPD

In spatial *IPD*s, there exists a "topology", each agent having a few (fixed) neighbors with whom he interacts (on spatialized cooperation relations, see e.g. Dupuy & Torre 1999). Then defective strategies can no longer invade the system. Once again *TFT* strategies dominate because if two *TFT* agents appear by mutation and meet, they are immediately imitated and their strategy propagates until it has invaded the system. For instance, a sucker *S* with three *TFT* neighbors and a meanie neighbor *M* generated by a mutation is eliminated by this *M* at a first stage, and *M* wins. However, at a second stage the *M* agents must interact and have only *TFT* neighbors. Then *TFT* agents win, and the meanies *M* convert to *TFT*. In other words, fluctuations generating *M* agents are *recessive*. This mechanism explains the strong *stability* properties of evolutionary stable strategies such as *TFT* that cannot be destabilized by mutating invaders.

Let us give Nowak and May's example of systems defined on a square network with 8 neighbors by the payoff matrix:

	A(c)	A(d)
B(c)	R = 1 R = 1	T = b S = 0
B(d)	S = 0 T = b	P = 0 P = 0

Behaviors:
d = defection, c = cooperation

Payoffs:
T = (d, c) = Temptation,
S = (c, d) = Sucker,
R = (c, c) = Reward,
P = (d, d) = Punishment

Conditions:
T = b > R = 1 > P = 0 = S = 0

b is the parameter of the system. Take for instance a random initial configuration with 50% *c* and 50% *d*. One compares the scores (the score of each site being the sum of its gain and of the gains of its 8 neighbors) and each site adopts the strategy of its neighbor (including itself) that gets the best result. The conclusion (Zhen Cao and Rudolph Hwa [1999]) is extremely interesting, and a priori unexpected if one is not familiar with *critical phenomena* in physics and, more generally, with *bifurcation* processes. One gets:

(i) for $b < 1.8$, *c* dominates;
(ii) for $b > 2$, *d* dominates;

for *b* belonging to the interval $B_c = [1.8, 2]$ — called the *critical interval* — there exists a critical transition *c* ® *d*, with multi-scale nested clusters of *c* and *d*.

Here is a *Mathematica*™ implementation I computed using an algorithm due to Richard Gaylord and Kazume Nishidate. We code moves with colors: *c* then *c* = blue; *d* then *d* = red; *c* then *d* = yellow; *d* then *c* = green.

For $b = 1.5$ (under the critical interval) and an initial configuration "InitConfig" 50%-50%, we see (figure 1) that there is an initial catastrophe (in two steps) leading to an overwhelming domination of meanies (red). Then cooperation (*c*, *c*) (blue) restores and dominates progressively, through the extension of residual scattered nuclei having survived the catastrophic initial phase of decimation. Domination of cooperation is by the way non complete and leaves fracture lines of oscillating non-cooperation (*d*, *d*).

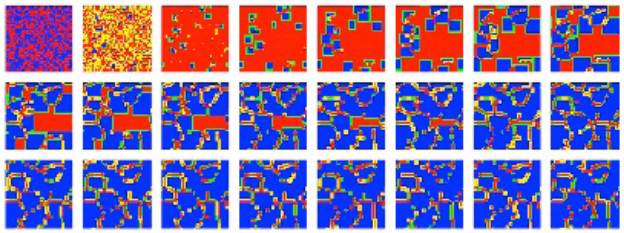

Figure 1. Nowak and May's example of a spatialized iterated prisonner dilemma defined on a square network with 8 neighbors and depending upon a parameter b. Moves are coded by colors: c then c = blue (cooperators); d then d = red (defectors); c then d = yellow; d then c = green. Value of b = 1.5 and the initial configuration "InitConfig" is 50% blue-50% red.

If we represent the temporal evolution of the sub-populations (*c*, *c*) and (*d*, *d*) we see very distinctly the initial decimation followed by a reconquest presenting small oscillating fluctuations. (See figure 2).

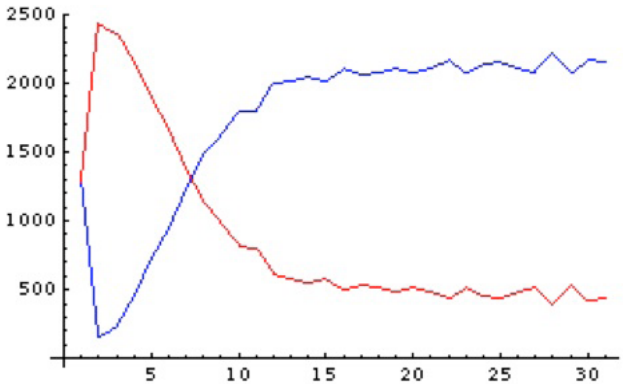

Figure 2. Temporal evolution of the sub-populations (c, c) and (d, d) in figure 1.

For $b = 2.1$ (above the critical interval) and a 50%-50% InitConfig, we see that d dominates immediately and totally (totalitarianism). (See figure 3).

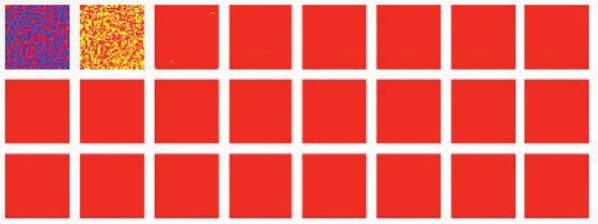

Figure 3. The system of figure 1 for the value of b = 2.1 and a 50%-50% InitConfig.

The curves of evolution are evident (no comment, figure 4):

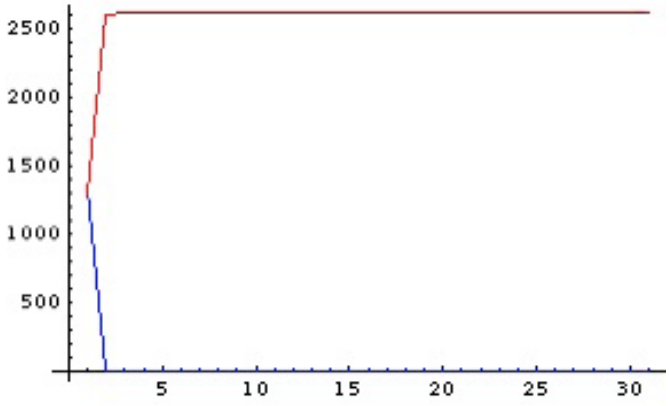

Figure 4. Temporal evolution of the sub-populations (c, c) and (d, d) in figure 3.

But if the value of the parameter $b = 1.85$ belongs to the critical interval and if we take a 50%-50% InitConfig, we see that (d, d) begins to dominate, next that (c, c) begins to reconquer ground by expanding from nuclei having resisted the initial extermination, but that, contrary to the first example $b = 1.5$, multi-scale nested clusters of c and d appear and expand: inside a blue island expanding in a red sea appears an expanding red lagoon, inside which emerges a smaller blue island, etc. (See figure 5).

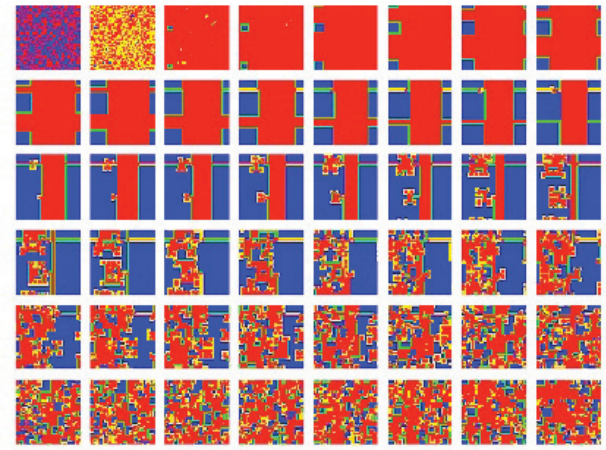

Figure 5. The system of figure 1 for the critical value of b = 1.85 and a 50%-50% InitConfig.

This critical dynamics is very apparent on the evolution curves where the curves (c, c) and (d, d) present, besides small oscillating fluctuations, large scale oscillations. (See figure 6).

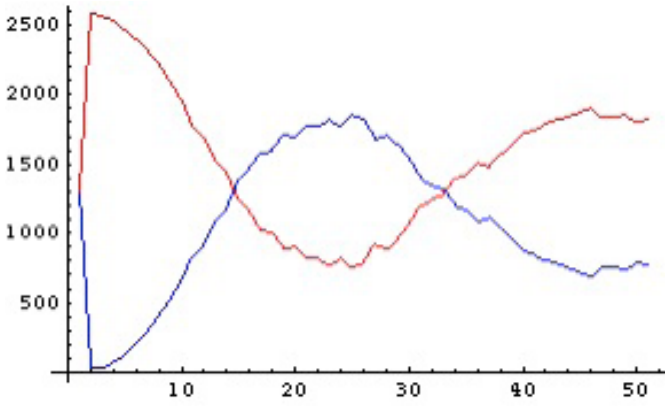

Figure 6. Temporal evolution of the sub-populations (c, c) and (d, d) in figure 5.

IV: CONCLUSION

We have seen on simple examples of evolutionary games how cooperation can spontaneously emerge as collective behavior in a population of individuals. We have retrieved very well-known common sense rules such as the tit-for-tat evolutionary stable strategy and modeled the evolutionary stability explaining how such a rule can have been selected by cultural evolution.

There exists to-day an extremely impressive amount of such computational syntheses that provide us with an experimental method for what we could call "Hayekian science", that is science based on complex methodological individualism, investigating the natural selection of behavior rules, and explaining how collective structures can emerge through self-organizing dynamics.

NOTES

1 This article draws directly from my articles: (i) "Modèles formels de la 'main invisible': de Hayek à la théorie des jeux évolutionniste". In Ph. Nemo and J. Petitot eds. *Histoire du libéralisme en Europe.* Paris: Presses Universitaires de France, pp. 1095-1114, 2006; and (ii) "Reduction and Emergence in Complex Systems". In Richard E. Lee ed. (Foreword by Immanuel Wallerstein), *Session III of Questioning Nineteenth-Century Assumptions about Knowledge, II Reductionism, SUNY Series,* Albany, NY: Fernand Braudel Center Studies in Historical Social Science, 2010. It also draws from a talk I gave at Centro di Metodologia delle Scienze Sociali, LUISS University on April the 28th 2004 at the invitation of Dario Antiseri and Enzo Di Nuoscio and from two talks presentented at the following two conferences organized by Francesco Di Iorio in Paris: (i) *L'évolution culturelle* (CREA, Ecole Polytechnique, 11-12 June 2007) and (ii) *Actualité de l'individualisme méthodologique* (EHESS, CREA-Ecole Polytechnique and Paris ISorbonne University, 22-23 June 2009).

2 On Hayek, see e.g. Hayek (1952), (1982), (1988), Nemo (1988), Nadeau (1998), Petitot (2000). On Dupuy, see e.g. Dupuy (1992), (1999).

3 Philippe Nemo: *Histoire des idées politiques aux Temps modernes et contemporains*, Puf 2002.

4 See Laurent (1994).

5 For Kant a normative judgement is "categorical" when it is independent of any end. Categorical prescriptions are purely "procedural".

6 For Kant a normative judgement is "hypothetical" when it is conditionned by an end and prescribes means to achieve the end (consequentialism).

REFERENCES

Axelrod R., Cohen M., and Riolo, R. (1998). The Emergence of Social Organization in the Prisoner's Dilemma: How Context Preservation and other Factors Promote Cooperation. *Santa Fe Institute Working Paper*, 99-01-002.

Binmore, K. (1994). *Playing Fair*. Cambridge, MA: MIT Press.

Cao, Z. and Hwa, R. (1999). Phase transition in evolutionary games. *Intern. Jour. of Modern Physics A*, 14, 10: 1551-1559.

Dupuy J-P. (1992). *Le sacrifice et l'envie*. Paris: Calmann-Lévy.

Dupuy J-P. (1999). Rationalité et irrationalité des choix individuels. *Les Mathématiques sociales, Pour la Science,* 68-73.

Dupuy, C. and Torre, A. (1999). The morphogenesis of spatialized cooperation relations. *European Journal of Economic and Social Systems*, 13, 1: 59-70.

Hayek, F. (1952). *The Sensory Order: An inquiry into the Foundations of Theoretical Psychology.* Chicago: University of Chicago Press.

Hayek, F. (1982). *Law, Legislation and Liberty*, London: Routledge & Kegan Paul.

Hayek, F. (1988). *The Fatal Conceit*, London and New York: Routledge.

Hofbauer J. and Sigmund K. (1988). *The Theory of Evolution and Dynamical Systems*. Cambridge: Cambridge University Press.

Kirman A. (1998). La pensée évolutionniste dans la théorie économique néoclassique. *Philosophiques*, XXV, 2 : 219-237.

Laurent, A. (1994). *L'individualisme méthodologique*, Que sais-je ? n° 2906. Paris: Puf.

Livet P. (1998). Jeux évolutionnaires et paradoxe de l'induction rétrograde. *Philosophiques*, XXV, 2: 181-201.

Nadeau, R. (1998). L'évolutionnisme économique de Friedrich Hayek. *Philosophiques*, XXV, 2: 257-279.

Nemo, P. (1998). *La société de droit selon F.A. Hayek*. Paris: Puf.

Nemo, P. (2002). *Histoire des idées politiques aux Temps modernes et contemporains*. Paris: Puf.

Petitot, J. (2002). Vers des Lumières hayekiennes: de la critique du rationalisme constructiviste à un nouveau rationalisme critique. *Friedrich Hayek et la philosophie économique. Revue de Philosophie économique*, 2: 9-46.

Petitot, J., (2006). Modèles formels de la 'main invisible': de Hayek à la théorie des jeux évolutionniste, *Histoire du libéralisme en Europe.* (P. Nemo et J. Petitot eds.). Paris: Presses Universitaires de France, pp. 1095-1114.

Petitot, J. (2010). *Reduction and Emergence in Complex Systems*, Session III of *Questioning Nineteenth-Century Assumptions about Knowledge, II Reductionism*. Richard E. Lee (ed.), Fernand Braudel Center Studies in Historical Social Science. Albany: State Iniversity of New York Press.

Poundstone, W. (1993). *Prisoner's Dilemma*. Oxford: Oxford University Press.

Samuelson, L. (1997). *Evolutionnary Games and Equilibrium Selection*. Cambridge, MA: MIT Press.

Weibull, J. (1996). *Evolutionary Game Theory*. Cambridge, MA: MIT Press.