CAHIERS D'ÉPISTÉMOLOGIE

Publication du *Groupe de Recherche en Épistémologie Comparée*
Directeur: Robert Nadeau
Département de philosophie, Université du Québec à Montréal

The two meanings of "critical mass": Probing the new frontiers of economics in search of social emancipation

[The use of knowledge about society, Part III]

Christian Arnsperger

Cahier nº 2004-06 **318e numéro**

UQÀM

http://www.philo.uqam.ca

Cette publication, la trois cent dix-huitième de la série, a été rendue possible grâce à la contribution financière du FQRSC (*Fonds québécois de recherche sur la société et la culture*).

Dépôt légal – 3e trimestre 2004
Bibliothèque Nationale du Québec
Bibliothèque Nationale du Canada
ISSN 0228-7080
ISBN: 2-89449-118-2

The use of knowledge about society
Part III

The two meanings of "critical mass":
Probing the new frontiers of economics in search of social emancipation

Christian Arnsperger
FNRS (Brussels) & Université catholique de Louvain (Belgium)
arnsperger@etes.ucl.ac.be

September 2004

Texte du troisième séminaire offert dans le cadre des travaux de l'Équipe de recherche en histoire de la pensée et méthodologie économiques, Département de philosophie, UQÀM, jeudi 30 septrembre 2004.

1. Why normative economics needs deeper epistemic foundations

1.1. Reopening the road from Frankfurt to Vienna, building a bridge to Santa Fe

Would you like it if when opening the newspaper you happened upon a list of axioms which, unbeknownst to you and without any public consultation that you were aware of, a group of economists had put together and submitted to your government, which consulted these economists so that they could provide it with a "solution concept" for the distribution of income? How would you feel if these axioms were affirmed once and for all and there was no visible room for you or your fellow citizens to influence their content in the future?…

If your first reaction to these questions is, "Well, I wouldn't particularly mind, this is the way things are happening anyway"—then my advice to you is: *think again*; reflect on that reaction, probe it to understand *why* it is your reaction, and ask yourself if as a first reaction it can be stabilized into a firm *conviction* about how the society you live ought to work; during your reflection, talk to others, read, write up arguments and tear them up to write up new ones, and so on. You might end up having your own list of axioms as well as a pretty elaborate theory about the kind of society these axioms ought to be translated into, and about how the required changes could be pushed for without hurting or killing those people who don't immediately agree with you. If, on the other hand, you find yourself shrugging your shoulders and saying, "Oh come on, that's all just too much work and brain-wringing, I have things to do with my life"—well, my advice again is: please tell me *why*; how can the "things you want to do with your life" be completely disconnected from the kind of society in which you do them? Is it because in the end you really think there's *nothing wrong* with society, so that you can pursue your goals in it unreflectively—or is it because some customs, habits or sanctions evolved and established by the society keep you from reflecting? Now, if you say to me, "Look, dude, I'm *telling* you, I don't *want* to reflect, I want to *live*," then I'll leave you alone and look around for someone else to reflect with. And if absolutely everyone reacts in the same way as you, I'll consider becoming a Taoist or a Stoic practitioner, trying to live out my own axioms of a better society without anyone else cooperating, accepting the market economy but trying to improve the quality of my participation in it, accepting commercial culture but trying to refine my own tastes in it if I can, and so on. Who knows, maybe I'll find out that you were right after all, or maybe after having

accumulated your own experiences of unreflectively "doing things" you'll remember our aborted conversation and call me up again…

This kind of interactive dynamics is what this paper is about. The broad research program it seeks to promote can be briefly described in terms of a tentative geography of the history of science: reopening the road from Frankfurt to Vienna, and then building a bridge to Santa Fe. Underlying this apparently fanciful geographical metaphor is a tenacious epistemological conviction: it is high time that normative economics entered into contact with "complexity science," and the most fruitful way to establish this contact is by hooking up the social-complexity approach with a specific, methodologically individualistic version of the Frankfurt School's "critical theory" approach. To be more precise, I want to argue that one important way in which normative economics can be made more fruitful is by breaking away from models in which the axioms used to judge the social context are supplied from the top down by an external theorist, and by attempting to model the interacting agents themselves (one of whom might be the theorist, now no longer external) as harboring reflectively critical knowledge about the society in which they are interacting and as using that knowledge to act towards furthering a better society.

1.2. The implicit politology of mainstream normative economics

It has become a truism of contemporary social theory that agents' normative aspirations and representations are no obstacle to a positive social science. Under methodological individualism, three conditions are necessary—though often not sufficient—to ensure the positive character of any theory:

(1) The normative conceptions of agents, their temporal genesis and their various modes of interactive transformation (preference formation, cognitive learning, etc.) are viewed as hard facts.
(2) One of the things about which individuals form normative conceptions are the interaction structures within which they are, and/or could like to be, embedded.
(3) Among these interaction structures are the modes of interactive transformation of the agents' normative conceptions.

If these three conditions are satisfied it is possible, either by conceiving of the axiom-producing theorists as a specific class of agents to be modeled, or by making "normative production" an activity of all or part of the agents, to reduce any normative social theory—among which, normative economics—to a positive theory of *normatively guided inter-individual interactions*. In that perspective, Marc Fleurbaey's discussion of axiomatic theories of economic justice highlights something important; he goes rather against the mainstream by writing that the work of normative economists, namely "logical analysis (study of the mutual compatibility of the axioms) and/or economic analysis (study of the existence of allocations, or of the non-emptiness of allocation rules)" is work that "also fits into the realm of the positive, not the normative, even though it still does not, or does not essentially, concern itself with the causal explanation of observed phenomena" (Fleurbaey [1995: 4]). What he is saying, in essence, is that the axioms being normative expressions of desirable allocation rules does not prevent the theory using them from being itself a positive theory.

The way Fleurbaey justifies this view, however, reveals the deep underlying *politology* of much of mainstream normative economics:

> Thus, the normative dimension is located upstream from economic theories of justice, at the level of the choice of moral values (or, possibly, of the choice of philosophical theories) that inspire the criteria one studies. This choice is the true domain of the political sphere (in the broad sense). In reality, in a society which is sufficiently homogeneous culturally and ideologically, there is frequently a broad consensus about the values that ought to inspire the elaboration of criteria of socioeconomic justice. Thus, the normative dimension of the debates frequently amounts to very little and leaves most of the space to the analysis of how to derive axioms from the values, and how to derive criteria from the axioms. (Fleurbaey [1995: 5])

This passage starts very well and then suddenly (with "In reality…") sweeps crucial problems under the carpet. Restricting the domain of mainstream normative economics to culturally and ideologically homogenous societies appears as a pretty strong retreat from the claim that normative economics is really positive—in fact, conflating the two means literally that mainstream normative economics studies the technical implementation of normative criteria made positive by ideological homogenization. In short, and to put it in perhaps too crude terms, by its implicit structure normative economics is really (independently of the modeler's deeper intentions) the technocratic handmaiden of a conventional social order: in a communist society where everyone is a communist, how to express and implement the underlying shared values; in

an industrial market society where everyone believes in the virtues of competition, in flexibility, and so on, with perhaps a bit of egalitarianism thrown in for social-democratic good measure, how to express and implement the underlying shared values? This is indeed positive theorizing, and it falls under the irony of what Abba Lerner said of all of mainstream economics: "Economics has gained the title Queen of the Social Sciences by choosing solved political problems as its domain" (quoted in Bowles [2004: 1]).

There are really just two ways to ground the axiomatic approach of current normative economics:

- *The paternalistic view*: The economist is a standing partner of the "ruling class" (sovereign, government, parliament, etc.) in the effort to channel myopic or ignorant individuals into "socially optimal" situations—from the ruler's viewpoint, informed by the economist as "impartial spectator." (According to Muller [1993: 15-60], Adam Smith harbored this vision of being co-opted into the ruling spheres so as to be able to "design the decent society" by channeling blind passions into peaceful and efficient market outcomes.)
- *The unanimistic-technocratic view*: The axioms formalized by the economist are simply a recapitulation and clarification of the society's already established norms and values, and the economist's task is to tell "us" how "we" could optimally implement them (where "optimality" itself is part of the values axiomatized).

The second view is defended by Fleurbaey, not the first; but concretely they are rather difficult to keep apart and are perhaps better viewed as two facets of a *single* view of the social world and its political dimensions. The reason is that, in both instances, the agents apart from the theorist and possibly the "rulers" or "administrators" play *no reflective role whatsoever* in the axioms' evolutionary genesis (how did the underlying values arise, and how did they become what they are?) and in their current argumentative establishment (what compels the populace to subscribe to these values rather than others?). Thus, it turns out that the single overarching principle that unites the paternalistic and the unanimistic-technocratic view is point (1) above, combined with the assumption of cultural and ideological homogeneity. The factuality of the agents' normative representations is affirmed because it is required for a positive theory; points (2) and (3), however, are evacuated—agents don't anchor the axioms they supposedly all subscribe to in a

reflective deliberation of the society they want, and above all they don't reflect on how they would like their normative representations formed.

1.3. "Hayekian Critical Theory," or how to bring critical instrumental rationality into complexity science

What this paper wants to suggest is that we should take Fleurbaey's suggestion all the way to its ultimate implications—which, I will claim, will imply a rather different epistemic grounding for normative economics. Instead of getting caught up in strong assumptions about *ex ante* homogeneity and instead of assuming a deep *ex ante* divide between the economist, the philosopher, the politician and "the mass of people," how about an approach that seeks to replace *normative* economics as practiced now by a *positive* theory of *normatively motivated* interactions between economic agents? To do so requires us to bring together two strands of thought not often brought together in economics—not even, I will claim, in the disciplines now forming at the frontier of economic research: on the one hand, the heritage of *Frankfurt School social criticism*, which I want to adapt to a framework of methodological individualism by using concepts from cognitive science; on the other hand, the heritage of *Viennese complexity-oriented social thought*, which goes back to Hayek and has reached new heights recently with the advent of "complexity economics." (For more thorough historical and conceptual discussions of this merger, see Arnsperger [2004a, 2004b, 2004c, 2004d].)

From the Frankfurt School I draw a strong emphasis on *individual agents' capacity, if adequately socialized, to reflect deeply on the society in which they live and on how to improve it*. In other words, I view individuals as endowed with *critical instrumental rationality:* one crucial factor influencing the individuals' rational choices of action is their reflective aspiration to social emancipation. This capacity may become blunted an near-imperceptible in certain oppressive or otherwise alienating social contexts, but that is no reason to assume it away in our models unless we want to add to Fleurbaey's homogenization assumption—which reifies social homogeneity into "how society is"—a non-reflectiveness assumption—which reifies ignorance or indifference into "how people are."

From the Viennese School I draw an equally strong emphasis on *a society's capacity, if it is to be a free society, to single out and isolate phenomena which escape any conscious or purposive attempts at top-down control, and which instead "emerge" (in the complex-systems sense) from*

the interactions of the individual agents. In other words, I view a free society as endowed with what I call a set of *characteristic emergents* which characterize the particular way(s) in which it is a "free" society. In Hayek's non-equilibrium, free-market approach, the main characteristic emergent is market price, which both (*a*) emerges unchecked from all inter-individual interactions (search, matching, etc.) and (*b*) is fed back into individuals' interactive-decision problems; there is no reason why this should be the unique possibility, however. Other characteristic emergents and other emergence contexts can—and must—be envisaged.

Bringing these two strands of thought together will allow us to deepen the epistemic foundations of normative economics to the point where it will have to be called instead *emancipatory social economics* or *critical-process economics*—a specific sub-discipline of economics that deals with the question of how to model a "critical mass" of agents interacting while each pursuing as much as emergent constraints allow his own critical conception of a better society. The objective of this paper is to show how this epistemically deepened normative economics might be constructed. It is therefore a programmatic paper, but I believe it offers a very important research program. After an early start on the puzzle of system dynamics in Keynesian macroeconomics (Arnsperger [1990]), I ventured into normative economics (Arnsperger [1994, 1997]) and policy-oriented analysis (Arnsperger and de la Croix [1996]) before exploring certain radical departures from standard instrumental rationality (Arnsperger [2000], Arnsperger and Varoufakis [2003]). Gradually, *the importance of not assuming away the agents' critical capacity and "emancipatory interest"* asserted itself (Arnsperger [2004a], [2004b]). The present investigation is an attempt to suggest how economic theory might be taken in this direction, while heeding early intuitions about economic life as an organic-evolutionary process rather than a static configuration or even merely multi-period dynamic succession of states.

Section 2 investigates the structure of "complexity science" and shows how the economy is portrayed there as a self-discovering interactive system populated with boundedly rational agents. The learning process of these agents is discussed, and some salient conclusions are drawn as to the exact nature of "emergence" in this approach, in which "critical mass" is a statistical notion that largely dispenses with the agents' own perceptions of social reality. In section 3, this shortcoming is used in order to introduce an alternative avenue for complexity science: that of a self-criticizing interactive system populated with agents who are instrumentally rational while

harboring critical views about what is wrong about their society and how it could be improved. A simple formal model is introduced in which agents interact with a view to shaping their critical views about society, and a new sort of emergent phenomenon called the society's *esprit critique* is shown to be a crucial component of an economy geared towards social emancipation. This phenomenon is used in order to show how a "mass of critics" can emerge within a complex adaptive economic system. Finally, in section 4 the problem of how to deepen the epistemic foundations of normative economics is taken up in light of the preceding discussion, and some simple proposals are made to promote a renewed normative economics designated as *critical-process economics*.

2. The internal structure of Santa Fe economics : "Critical Mass" I

2.1. The economy as a self-discovering interactive system: Collective exploration out of bounded instrumental rationality

Neoclassical equilibrium analysis, of which general equilibrium theory is a subset, implicitly relies on interactions between agents but operates what Auyang (1998: 115-121) calls an "independent-individual approximation": through the feedback via prices, individuals' actions are modeled *as if* these individuals had relinquished disequilibrium-price trading and other non-equilibrium actions—whereas, in fact, all the theory is really able to describe is the mutual compatibility of plans effected by trading at *equilibrium* prices, or more generally (in game theory) the mutual compatibility of strategies *at a Nash equilibrium*. In other words, interactions are "in there," they are not denied by they are made implicit and assumed away. The *as if* narrative accompanying equilibrium models, if taken literally, usually make for fanciful and self-contradictory institutional Towers of Babel, such as when the Walrasian model of decentralized trade is given a centralized-direction, "auctioneer" narrative. Such a focus on equilibrium theory is one of the reasons why mainstream economics has long privileged the economy as an *allocation device* to another figure, that of the economy as a *collective exploration device*. The Vienna School's enduring contribution is to have kept this other figure alive; its enduring failure, as I have argued elsewhere (see Arnsperger [2004c]) is to have used this figure almost exclusively for the purposes of narrow, pro-capitalist apologetics.

Now, if capitalism, or rather the industrial market economy, is to be viewed as a "superior" system in some sense, this alleged superiority has itself to be accounted for through a dynamic process—and not just any process: a *critical process* by which market organization and industrial production arise, albeit in a partly unintended way, from the interactive exploration by *consciously normatively oriented agents* who do not merely pursue momentary necessities. The mainstream work done in the direction of market emergence (see e.g. North, [1981, 1990], Platteau [2000: 241-280]) does not respect this condition; it models institutional dynamics as the unintended result of largely *unreflective* interactions between agents who are following *everyday-life rules* designed to ensure their own and their families' fitness or survival. While this is respectable work in the sense that it appropriately treats the economy as a complex adaptive system, it falls short—as I will argue in section 4—of the *full potential* inherent in the approach.

This approach has today become intimately connected with the name of the Santa Fe Institute and its extraordinary exploration of complexity in all its dimensions (see e.g. Waldrop [1992], Arthur, Durlauf and Lane [1997], Brock [2000]). Viewing the economy as a complex adaptive system means something even less easy to carry out than to write down. Assuming that a certain directed graph G can be taken to possess operational closure so as to be called a system, the agents $i, j, \ldots$ making up the vertices of G interact through strong connections (which is one of the conditions for operational closure) by individually using various elements drawn from three sets of rules—a set $R1$ of interaction rules which are particular ways of realizing, or activating, or disactivating, the various available connections between vertices; a set $R2$ of credit-attribution rules which allow the agent to evaluate the relative success or failure of her actions; and a set $R3$ of revision rules by which the agent modifies her interaction rules in the hope of obtaining higher credit in the next round of interactions. Such a system is complex if the interactions between agents generate so-called emergent properties that are not the sum of all individual-level properties—in other words, individual actions aggregate in nonlinear fashion through interaction. The system is adaptive in a two fold sense: (*a*) the emergent properties of interaction generate certain credit measures (profit, fitness, etc.) which allow the agents to learn from mistakes and successes, and (*b*) perceived lack or loss of credit triggers adaptation *per se*, i.e., a change in interaction rules.

Perhaps the key element in such a system as expounded in its dominant interpretation is that nontrivial collective behavior—or *systemic behavior*—can be generated by very trivial rules in

*R*1, *R*2 and/or *R*3. In other words, agents endowed with bounded rationality (myopic horizon, small calculative abilities, use of simple routines or rules of thumb, etc.) and basing their behavior on expectations which may not at all be consistent with past observations, may nevertheless interact to produce very elaborate-looking patterns in certain aggregate variables (share prices, inflation rate, attendance of a nightclub, etc.). The illusion that to generate such intricate patterns as can be viewed in nature (Ball [1999]) or in social settings (Ball [2004]), there either has to be an omniscient designer or else incredibly sophisticated individuals is dispelled—and in a certain sense, one of the reasons for humanistic pride which had come along with the Enlightenment seems to vanish in thin air: to "explain" the emergence of aggregate sophistication none of man's hard-earned capacities for reasoning, criticism, abstraction, calculation, and so on, are *really necessary*—the mind can be viewed as a machine made up of exceedingly simple subroutines, and sophistication nevertheless comes out in aggregate-scale "subjectless processes" (Dupuy [2002]).

If first- and second-generation cybernetics had not reduced metaphysics to a form of surface noise in the subjectless system itself, one could almost say that what the Santa Fe vision conveys is a "mystique of unintended collective discovery," i.e., a view of the economy inherited from Schumpeter and Hayek as a *self-discovering system*—innovations and novelty flowing from the mere attempts of agents (sometimes successful, often failed) to survive the adaptive tasks bestowed upon each of them by… the aggregate of them. At each step, each of them can marvel at what they had not even dreamed some of them would discover or invent due to systemic pressure, but the next step is as radically uncertain as the previous one because due to multifarious adaptations the systemic pressure shifts constantly.

2.2. What do boundedly instrumental rationalists learn in interactions?

The two keys to understanding how agents adapt in such a system are the notions of an *internal model* (Holland [1995: 57-60], Potts [2000: 173-179]) and of *credit attribution* (Holland [1995: 53-56], Axelrod and Cohen [2000: 136-143]). They are closely linked in the sense that the agent uses his internal model to undertake actions whose high or low credit value subsequently may lead him to modify the internal model.

An internal model is an evolving module in any agent's cognitive setup which represents that agent's more or less formalized, more or less exhaustive and more or less rigorously scientific,

view of the array of situations in which he believes he can find himself—either on the basis of his own past experience in the interaction or on the basis of the experience communicated by other agents through the interaction. A boundedly rational agent will obviously (*a*) not deduce his decision rules from an "overall model" of the whole economy, be it only because of the cognitive limitations he experiences or of the enormous cost of acquisition of expertise on this "total model," if it existed; and (*b*) not usually submit his "partial" internal model to the most extensive inter-subjective testing available at any moment, since experience is itself a way of "spreading" the testing costs over time, and economizing on testing as long as things don't go wrong.

"Going wrong" here means simply that the action undertaken does not fulfill the agent's aim as measured by his own interests. Such a failure implies that the action is given a low credit value—and if we conceive of the agent's interest as the accumulation of large or at least sufficient credit (as in Simon's approach of "satisficing") in order to survive according to the system's norms of evaluation, a low credit value signals to the agent that he must learn quickly if he wants to stay on board. There are essentially three ways in which learning will affect the internal model: (1) If the model is used by the agent as a *credit prediction device*, i.e., as a cognitive tool yielding an expectation of the credit value of his action in a perfectly defined momentary context, the failure implies that the internal model has to be modified in congruence with what the agent believes will be his future (also perfectly defined) context—in other words, he needs to change his internal model contingent on his expectation of the new situation with which it will have to help him cope. (The extent to which this will have to be done obviously depends in part on how "flexible" or "open-ended" the internal model is.) (2) If the internal model is used by the agent as an *situation interpretation device*, i.e., as a hermeneutic tool yielding a description of the momentary context in which he will have to maximize a perfectly defined credit function, the failure implies that the internal model has to be modified so as to provide a less incongruous description of the future context—independently of how the credit function to be maximized will evolve to the next period. (3) If, as is most likely, the internal model is used to do both (1) and (2) so that neither the situational context nor the expectation function are well-defined, subtle practical judgment of the Aristotelian *phronêsis* type is required to understand the reasons of the failure and, hence, the direction in which to look for improvement. The way in which either of the three improvements is done is through direct or

indirect interaction with other, similarly struggling agents—experiences are shared, "best practices" pooled, etc.

It is crucial to realize that the agent's perception of his action's credit value and of his consequent need for a modified internal model is clearly a *situated* perception: either he truly doesn't realize that the locally perceived signal comes to him from a subjectless, emergent aggregate phenomenon, or he realizes it but decides he has no choice anyway—in either case, the adaptive consequences are exactly the same, so that *the agent's awareness of the "origin" of the perceived failure or success in no way affects the subsequent adaptive sequence*. This is because not only the triggering signal but also the ensuing criteria for a better adaptation are considered to flow from the emergent phenomenon. This goes a long in explaining why in the introduction, you as my fictitious friend were prone to tell me that you didn't *want* to reflect on the "broader why" of your action—you simply felt it was useless, given that the subjectless emergent phenomenon (of which you were perhaps even acutely aware) was what it was. In particular, even if the manner in which all agents learn from others—sharing "best practices"—is deeply *enactive* in the sense of Varela, Thompson and Rosch (1991), i.e., even if each agent's "perceived social world" cognitively co-emerges with his "active engagement" with that social world (in particular, with the agents with whom he interacts) instead of being a pre-given, purely objective world, the agent's way of adapting to emergent signals from his enacted environment is nevertheless *passive* in a deep sense: he is purposefully, actively, "busily" adapting to emergent signals whose genesis he uncritically accepts because these signals are "given."

Axelrod and Cohen have defined "harnessing complexity" as "seeking to improve but without being able to fully control" and "a device for channeling the complexity of a social system into desirable change, just as a harness focuses the energy of a horse into the useful motion of a wagon or a plow" (Axelrod and Cohen [2000: xvi and 2]). The survival-oriented adaptations reviewed here, in which the emergent signal is taken as an imperative "message" from the environment that *you'd better learn something quick*, correspond to what I would call *opportunistic harnessing*: given the directed graph(s) defining the interaction pattern(s) in the system and the motivations that generate flows between nodes of the graph(s), take the courses of action which pursues your *unreflected* idea of freedom, i.e., the idea which the overall logic inherent in interaction patterns and motivations, as well as your location on the graph, spontaneously attributes to you. This category consists in each individual asking herself, "How

can I be freer in the social system as it operates?" But "free" here means, essentially, free to continue surviving or to reactively fit in with the emergent phenomena. These are not, as we will see shortly, the only ways of characterizing an individual's efforts to harness the complexity of the social system.

2.3. Aggregate behavior: Emergent properties and the puzzle of "critical mass"

An enduringly fascinating aspect of complex adaptive systems is that even such passive, if not mechanical, rule-following adaptation on the part of the individuals is apt to generate aggregate emergents whose behavior is all but simplistic, and certainly does not merely reproduce at the aggregate scale the mechanical rules used at the sub-aggregate scale. As is well known, this is in fact one of the hallmarks of social-complexity models:

> Agents adapt—they are not devoid of rationality—but they are not hyper-rational. They look around them, they gather information, and they act fairly sensibly on the basis of their information most of the time. In short, *they are recognizably human*. Even in such "low-rationality" environments, one can say a good deal about the institutions (equilibria) that emerge over time. In fact, these institutions are of precisely those that are predicted by high-rationality theories (…). In brief, evolutionary forces often *substitute* for high (and implausible) degrees of individual rationality when the adaptive process has enough time to unfold. (Young [1998: 5])

This is fine if one is hell-bent on explicative parsimony—especially if the behavior of the aggregate matters more than the understanding of individual particles' trajectories, which in some cases makes good sense (see Auyang [1998: 50])—and/or if one is not an Enlightenment humanist and one believes it is more important to understand the factors that can hold together a Hobbesian society (Ball [2004: 7-37]) than it is to understand how man can become enlightened so as not to reproduce antiquated patterns of violence (Varela, Thompson and Rosch [1991: 245-254], Dupuy [2002: 15-22]).

This Hobbesian minimalism is one of the cornerstones of contemporary economics, and it has clearly slipped over to complexity science. Put rather abruptly, its core question is, "What are the most idiotic and unsophisticated agents we can assume so that the aggregate phenomena we find important (sometimes called stylized facts, or recurrent patterns) can be accounted for?" Inherited from Laplace's eminently understandable anti-metaphysical stance—to clerics tempted to merge science with theology, he declared God an "unnecessary assumption"—this epistemological

move seems now to have been inverted into an almost ontological, or at least anthropological dogma. Simon's notion of bounded rationality, while certainly a credible alternative to neoclassical hyper-rationalism, hardly justifies an enduring fascination with emergence out of interacting automata. To the extent that complexity approaches have eschewed Hobbesian minimalism, they can provide a great potential for a *more reflexively oriented but still individualistic* economics, as I will claim in sections 3 and 4; to the extent, however, that agents are treated like sophisticated grains of sand or self-propelled billiard balls, the notion of "critical mass" that it conveys is fascinating but ultimately technocratic, since the mastering of scaling laws and pattern-generating mechanisms is mainly useful for city planners, park and museum designers, or other professional planning experts who have learnt that well set-up incentives are much less self-defeating than attempts at outright control (see Ball [2004 : 156-182]).

True enough, a whole bunch of planning tasks really *do* require little or no knowledge about "authentic human aspirations" and can be carried out much more easily—some would even claim, in a much more humanistic open-mindedness—if one can dispense with anything but coarse assumptions about individual motivations. If we accept that people with a broad range of motivations invest in the stock market, or that people with very diverse aims in mind walk through Central Park everyday, it seems indeed useful to be able to solve the puzzle of "critical mass" with the same tools as the ones used in the mechanics of gases or in condensed-matter physics. Krugman (1997) has shown that a purely mechanistic aggregate-level model of a two-sector region can generate critical-mass behavior due to nonlinearity—namely, due to the fact that the range of products produced locally is a function of the size of the locally market, the income-export base relationship displays two stable branches overlapping over a metastable interval: at some critical level of the export base, slowly falling income drops brutally; at another critical level, slowly rising income jumps up discontinuously. For a policy-maker concerned only with increasing aggregate income, this relationship may be sufficient knowledge, independently of how each individual inhabitant of the region feels about rising income. Brock and Durlauf (2001) have shown that the conformity component of an individual's interest may very well override her private component so that, on the sole basis of a very general knowledge about the statistical distribution of individual decision errors and about the size of imitation effects across individuals (and no precise knowledge of their motivations), it can be shown that there is a certain threshold beyond which one finds "the potential for collectively undesirable behavior, such as

out-of-wedlock birth rates, which is individually optimal" (Durlauf [1997: 88]). Note that while some of the results in that literature can be considered trivial in the sense that they remind us of classical results in equilibrium theory, here we have non-equilibrium theories in which critical mass leads to phase transitions or in which standard equilibrium results are generalized through what Durlauf calls "stochastic generalization."

The striking thing about today's complexity economics is that it insists *both* on the importance of modeling interaction (so that it rejects traditional equilibrium theories' strategy of independent-agent approximation) *and* on the possibility of dispensing with the interacting components' reasoning and reflexivity. This gives the whole research area an aspect indeed well captured by Ball's (2004) expression of "social physics," and society looks almost like a sophisticated pile of sand. The following two passages illustrate very well the potential and limitations of the approach:

> In a system of interacting agents, emergent properties are those that cannot be reduced to statements about the individual elements when studied in isolation. (…) One important aspect of emergence is that it breaks any logical relationship between methodological individualism and reductionism. What I mean is that emergent properties cannot be understood through the individual elements of a system, as they are intrinsically collective. This is so even though the behaviors of these elements determine whether or not emergent properties are present. (Durlauf [2002: 71-72])
>
> Think of a pile of sand on a table that has a continuous flow of sand falling on the top of the pile. For a while, the sand builds up into a large conical sandpile, but at periodic times, when the sandpile builds up to what Bak calls self-organized criticality, there is an avalanche or series of avalanches until the pile "relaxes" back to a state where avalanches cease. (…) The distribution of sizes and "relaxation times" of these avalanches follows scaling law patterns [e.g. Zipf's law, Pareto's law, power law, etc.]. The study of complexity tries to understand the forces that underlie the patterns or scaling laws that develop. (Brock [2000: 30])

Critical mass, here, is an aggregate phenomenon which the theory seeks to explain without endowing the grains of sand with more than the physical properties (attraction, repulsion, etc.) required to explain the observed aggregate. This means that complexity economics will be interested essentially in empirically observable and "persistent" patterns, as characterized by John Holland:

> Emergence occurs in systems that are generated. The systems are composed of copies of a relatively small number of components that obey simple laws. Typically these copies are interconnected to form an array (…) that may change over time under control of the transition function. The whole is more than the sum of the parts in these generated

systems? The interactions between the parts are nonlinear, so the overall behavior *cannot* be obtained by summing the behaviors of the isolated components. Said another way, there are regularities in system behavior that are not revealed by direct inspection of the laws satisfied by the components. (…) Emergent phenomena in generated systems are, typically, persistent patterns with changing components. Emergent phenomena recall the standing wave that forms in front of a rock in a fast-moving stream, where the water particles are constantly changing though the pattern persists (…). Persistent patterns often satisfy macrolaws. When a macrolaw can be formulated, the behavior of the whole pattern can be described without recourse to the microlaws (generators and constraints) that determine the behavior of its components. Macrolaws are typically simple relative to the behavioral details of the component elements. (Holland [1998: 225-227 *passim*])

In the following two sections, I want to scrutinize this approach for under-used potentials caused by this perhaps excessive macrolaw fetishism. As we shall see, the discussion will revolve principally around the re-introduction of a form of critical humanism which makes the grains of sand capable of reflecting on how they feel about arriving in a constant flow onto a pile and on what kinds of attraction, repulsion or other interactive properties they think would make their pile less avalanche-prone, or prone to "nicer" avalanches. In short, the idea is to exploit the complexity approach differently in order to go from a *critical mass* to a *mass of critics*.

3. The spread of critical ideas in complex economic systems : "Critical Mass" II

3.1. The economy as a self-criticizing interactive system: "Esprit critique" out of critical instrumental rationality

A deep-seated assumption in current complexity economics, and one which was already quite present in Hayek's discussion of cultural evolution, is that *imitation and conformity* are individually instrumental in the social process both to reduce cognitive overflow and to survive by trying to mimic or copy the strategies or internal rules which, given how others act, seem to yield a high credit attribution. This copying may be imperfect and error-prone (which introduces stochastic variation) but it is a key element in the large majority of complex adaptive systems. I want to suggest that all important social-learning mechanisms need not be imitative, and that some crucial ones may actually be a mix of *rational non-conformity*, *critical demystification* and *Socratic cohesiveness*. Much of section 3 is devoted to explaining these terms and showing how the complexity approach can make fruitful use of them.

As highlighted in section 1.3 above, what I call Hayekian Critical Theory centers around two key insights—namely, first, that individuals are endowed with *critical instrumental rationality*, so that they reflect on what is wrong with their current society and on how to act in order to improve it; and second, that a free society is one in which the illusion of top-down control is relinquished in key areas and replaced by the acceptance of emergent phenomena, called *characteristic emergents* because they are "tags" describing a particular society's way of being "free."

Let me begin by outlining a theory of critically instrumental rational action, which will mostly emphasize the "rational non-conformity" aspect. (Critical demystifcation and Socratic cohesiveness will be studied more in section 3.2.) Suppose a society can be represented as a directed graph G. Take an individual i, identifiable as a node of G and endowed with preferences over socially produced outcomes. In neoclassical theory, her preferences will be represented by a utility function U_i and, being concerned with socially produced outcomes, it will be affected by the individual's own actions denoted by a vector a_i and by the vector of vectors of actions of other individuals in some neighborhood n_i of agent i, denoted by a vector $\mathbf{a}_n$. Thus we would have the usual notation for the individual's utility, namely $U_i(a_i, \mathbf{a}_n)$. This could be refined in the way suggested by Durlauf (2002: 54) by assuming that i makes a statistical error e_i when determining her action and that what differentiates individual utility functions is a set of personal characteristics Z_i, so that

$$U_i(a_i, \mathbf{a}_n) = V(a_i, \mathbf{a}_n, Z_i, e_i). \qquad (1)$$

To take a simple example, suppose i is a consumer and one coordinate in a_i is her purchase of cream cupcakes; being perhaps a Veblenian, "keeping-up-with-the-Joneses" kind of gal, she tends to eye at her dieting neighbors' consumption and to set her consumption level so as to optimally balance her craving for sweets and her urge not to appear too bingy. Although the Brock model postulates a to be a binary variable, and although Durlauf (1997: 100) has emphasized that continuous-choice versions of this same problem are still in their difficult infancy, this need not concern me here because the point I want to make is at a different level.

The usual issue in these models is how the agents' statistical errors and their interactions on the basis of beliefs about each others' actions will generate a joint probability distribution for their actions $(a_1, a_2, \ldots, a_N)$, where N is the total size of the neighborhood. My question, rather, is how i's critical reflection on her social life can modify interactions in the neighborhood. Indeed,

suppose that this agent has, over time, developed a successful strategy of eating cupcakes while creating a neighborhood image of a fairly austere person. This may involve ostensibly exercising to burn away the calories, carrying certain expensive sporting clothes as "tags," speaking about health and dieting with key neighbors, and so on. Suppose, next, that the neighbors suddenly observe *i* no longer coming to talk to them much, wearing simple self-made clothes and walking around the neighborhood with Marcuse's *One-Dimensional Man* ostensibly tucked under her arm. Clearly, given the general state of mind in the neighborhood, this is *not* an adaptively optimal strategy—initially, it may well lead agent *i* to be questioned aggressively, or ignored and ostracized, or openly ridiculed. It is therefore irrational if U_i is *i*'s utility function; so why does she do it? The answer is simple, and totally compatible with instrumentally rational methodological individualism: she has *for some reason* added a term to her utility function, a term that makes her want to participate in her current society's values and "adapted" actions very differently—namely, through what Horkheimer (1937) has called "critical acceptance." She does not exit the industrial market society in which she is embedded, but she uses the market and the prevailing industrial mode of production to buy critical books and fair-trade cotton fabric instead of cream cupcakes and expensive sporting gear. Thus, we might no write her expanded utility function as

$$U_i(a_i, b_i, \mathbf{a}_n, \mathbf{b}_n) = V(a_i, \mathbf{a}_n, Z_i, e_i) + k_i W_i(d_i(S, S^*), b_i, \mathbf{b}_n),$$
$$k_i>0,\ W_i(0,.,.)=0,\ d_i(S, S)=0 \text{ for all } S,\ \partial W_i/\partial d_i<0,\ \partial W_i/\partial b_i>0. \qquad (2)$$

The function $d_i(.,.)$ is the agent's way of estimating the "distance" she experiences between the actual current society *S* and what she feels would be a better society, *S**. This latter variable can change with time because *i* might modify her critical views as time passes; the current state of society, on the other hand, is obviously the emergent property of the conjoined actions of all agents in the society: if *M* is the total population, we will have $S(a_1, a_2, \ldots, a_M)$. If $d_i>0$ and increases, agent *i* feels higher frustration and her "reflective-utility" index W_i decreases; if $d_i=0$, i.e., if $S=S^*$, there is no frustration at all and the second term of the utility function vanishes. Since our agent is a critical *instrumental* rationalist, she is not content with *feeling* frustration; she also seeks to *take concrete action* in order to improve society in the direction her ideal. This is represented by action vector b_i, which represents the means which *i* upon reflection and after interaction with other agents views as appropriate to approach *S**; how good she will feel implementing these means obviously also depends on whether her neighbors are also acting so as

to further the advent of S^*, as denoted by the vector $\mathbf{b}_n$. Thus we shoud actually write these variables as $b_i(S^*)$ and $\mathbf{b}_n(S^*)$. Clearly, if none of i's neighbors are subscribing to the ideal S^*, or if none of them choose to act towards that ideal, then $\mathbf{b}_n(S^*)=\mathbf{0}$.

We now have a straightforward—though by no means unique—way to explain the apparent irrationality of our agent's "non-adapted" behavior. One might have a simple step-function by which

$$k_i = 0 \text{ for } d_i \leq D$$
$$k_i = k > 0 \text{ for } d_i > D, \qquad (3)$$

where D can be thought of as a frustration threshold above which the reflective-utility component W_i starts "kicking in." Suppose now that while possessing a fixed social ideal S^* (a simplifying but not a necessary assumption) i experiences that, over time, the conjunct actions ($a_{1,t}$, $a_{2,t}$, …, $a_{M,t}$) leads to a sequence of social states $\{S_t\}$ such that $d_i(S_t, S^*)$ increases over time, and at some point jumps the threshold D. At that point it actually becomes *rational, i.e., U_i-maximizing but no longer V-maximizing*, for i to start taking actions bi designed to hasten the arrival of S^*, both through changes in her own lifestyle and through a modification of the "tags" which she displays to other agents, in the hope that they too will be moved to react and will start promoting S^*. Thus, actions bi are based in part on i's beliefs about how she can actively affect the functional form of the other agents' d-functions, and hence also k-functions, so as to make them feel the frustration earlier than if she had silently stopped buying cupcakes but had kept outwardly behaving in the same way. So, obviously, by W_i kicking in the whole profile of actions (b_i, a_i) is modified, and this is not irrational: indeed, the key to non-conformity in this approach is that what has proved adaptively optimal as a "best practice" over a long period may actually turn out to be what i believes must be *abandoned*—rather than reinforced—if $d_i(S_t, S^*)$ increases over time. The key mechanism by which a steady rise in d_i at some point puts to a strictly positive value may be called i's *social-criticism mechanism*; on its basis alone, the whole theory of critical instrumental rationality can get off the ground. Note carefully that since i's actions are instrumental, there is no principled anti-conformism involved in the agent's behavior. She doesn't disagree for the pleasure of disagreeing, but because her frustration at the experiential distance between her actual life and her desired life is too strong. If by chance S were to finally converge on S^*, the agent's non-conformity would vanish.

In this paper I cannot hope to analyze the immensely complex dynamics which this model can generate when applied to a graph of M interacting nodes, each of them embedded in an N-neighborhood. What I am interested in, rather, is the discussion of such a social system's characteristic emergents. Indeed, apart from the behavioral rules (2) and (3), each agent i, also uses a whole array of *interaction rules* aimed at various things: acquiring information about everyday-life aspects linked to "traditional" actions a_i; understanding as clearly as possible how the actual society S_t works and drawing up the social ideal $S_{i,t}$* as rigorously as possible, by communicating with others through reading, following lectures, seminars and conferences, and engaging in Socratic dialogue with like-minded or opposed but talkative neighbors; reflecting with like-minded others on the best means to approach their ideal in the current period; etc. etc. Of course, these interaction rules imply the existence and modification along the path $\{S_t\}$ of very numerous institutions, among which may figure markets, government, contracts, gift exchange, gratuity, etc. Unless all these institutions, including those which assist the agents in forming their behavioral parameters, are authoritarian planning institutions and unless all necessary interactions are strictly trivial (in the analytical sense of Koppel, Atlan and Dupuy [1983]), each momentary society S_t will be characterized by an array E_t of emergent phenomena (e^a_1, e^b_2, …, $e^{z(t)}_{R(t)}$), where room is made for the possibility that as society changes, the number—lower index (1, 2, …, R)—and types—upper index (a, b, …, z)—of these emergent phenomena might themselves change. E_t is what I call the array of charateristic emergents of society at time t.

It might of course be that the frustration threshold D is infinite for all i, which for all practical purposes mean that all agents are in all periods in full agreement with the current society so that S*=S and d_i=0. But unless this is so, a crucial characteristic of a free society is that *interactions between critically instrumental rationalists, each seeking to understand what is right or wrong about their society and how to either defend its successes or improve its defects, will be free, uncoerced interactions subject only to self-generated norms and constraints*—in other words, reflectively motivated interactions between critically instrumental rationalists must generate a specific emergent which, like all emergents, is irreducible to any individual's traits (here, to any individual's critical views about society) and at the same time reacts back onto the traits of each individual. I call this emergent the *esprit critique* of the society at time t and denote it as $\mathbf{C}_t$. At first it might seem a strange and somewhat ethereal emergent, almost like a spiritual entity; it is certainly not numerical or simply quantifiable (although it may be translatable into an array of

Boolean variables as, for instance, in Axelrod's [1997] of diffusion of cultural traits), but it is in fact a solid concept. To understand it, imagine an industrial market society in which at time *t* a certain proportion of agents are in a situation of critical acceptance: they play along in the market economy, but they feel it's fundamentally flawed and seek to act in nonstandard ways and to disseminate critical views about the market logic through writing, speaking out, and otherwise acting in visible ways "tagged" as "anti-market." In such a society, prices continue to serve as guides for the agents' subsequent economic decisions, but in parallel there is an *esprit critique* which represents, to speak in Hayekian terms, a "summary" of the overall critical atmosphere reigning in the society, and which conveys to all involved that they are—whether critical or uncritical accepters—living in a society where there is a tension between pro-market and anti-market views. This *esprit critique* clearly does not prescribe a definite view, but it serves as a nondirective guide to the agents' subsequent decisions on how to revise their critical views—for instance, in the sense that they will start reflecting on their critical position of next period on the basis of the prevailing tension rather than on the basis of other ideas left in the background or present in other societies.

With these two elements—critical instrumental rationality as a tool of "rational non-conformity" and *esprit critique* as a key "characteristic emergent"—we can go a long way in characterizing the method of Hayekian Critical Theory. Let me now discuss in more detail the dimensions of "critical demystifcation" and "Socratic cohesiveness."

3.2. What do critically rational instrumentalists learn in interactions?

Basically, these two other dimensions have to do with the agents' learning process. Both Bayeasian and non-Bayesian economic theories of social learning make up a growing body of literature (see e.g. Chamley [2004]) and my aim here cannot be to formalize the agents' interactions. However, some qualitative aspects can be discussed in the light of the above heuristic model.

As the Frankfurt School of critical theory has emphasized (Horkheimer [1937, 1946], Adorno [1962]), the way agent *i* goes about drawing up her social ideal S^* is through an attempt to gain critical distance from the current social process *S*, and this requires that *i* gather a type of knowledge about society which I would call *critical knowledge*: the idea is for *i* to come to understand current society as it is, but in such a way that this understanding carries within it the

possibility for *i* to understand *also* whether she wants to defend or change society as it is. We need not be Marxists to grasp this crucial idea through Max Horkheimer's somewhat antiquated, but still remarkable, way of putting things:

> Now, inasmuch as every individual in modern times has been required to make his own the purposes of society as a whole and to recognize these in society, there is the possibility that men would become aware of and concentrate their attention upon the path which the social work process has taken without any definite theory behind it, as a result of disparate forces interacting, and with the despair of the masses acting as a decisive factor at major turning points. Thought does not spin such a possibility out of itself but rather becomes aware of its own proper function. In the course of history men have come to know their own activity and thus to recognize the contradiction that marks their existence. The bourgeois economy was concerned that the individual should maintain the life of society by taking care of his own personal happiness. Such an economy has within it, however, a dynamism which results in a fantastic degree of power for some, such as reminds us of the old Asiatic dynasties, and in material and intellectual weakness for many others. The original fruitfulness of the bourgeois organization of the life process is thus transformed into a paralyzing barrenness, and men by their own toil keep in existence a reality which enslaves them in ever degree. Yet, as far, as the role of experience is concerned, there is a difference between traditional and critical theory. The viewpoints which the latter derives from historical analysis as the goals of human activity, especially the idea of a reasonable organization of society that will meet the needs of the whole community, are immanent in human work but are not correctly grasped by individuals or by the common mind. A certain concern is also required if these tendencies are to be perceived and expressed. (Horkheimer [1937: 212-213])

Clearly, a Hayekian approach to this kind of knowledge problem rejects the somewhat naive—and anti-complexity—idea that any social process could and should have "a definite theory behind it" without any "disparate forces interacting." On the contrary, the fact that "every individual in modern times has been required to make his own the purposes of society as a whole and to recognize these in society" can be an expression of a free society—if by "recognizing the purposes of society in one's own purposes" we understand what complexity theory means, i.e., that the society's characteristic emergents are necessarily fed back as systemic constraints into the further decisions of individuals. The problem, then, is not how to assume away or institutionalize away complexity by intronizing "a definite theory to be grasped by the common mind"—rather, and here Horkheimer puts his finger on something crucial, the problem is how to organize social institutions and the underlying interactions in such a way that "the role of experience" can be exploited in a maximally fruitful way. And this can be done only if all possibilities for mutually beneficial experience exchange are exploited *under the auspices of critical rationality*, i.e., in such a way that people can "come to know their own activity and thus to recognize the

contradiction that marks their existence" and question themselves to see whether they are indeed "keeping in existence a reality which enslaves them in ever degree." Since "thought does not spin such a possibility out of itself but rather becomes aware of its own proper function," a level playing field must at all costs (almost!—more on this in section 4.2) be created on which ideas can flow and be exchanged, including of course ideas hostile to free economic markets and unrestricted commercial flows…

But let us be a bit more precise about what these "ideas" are to be. The society's *esprit critique* is the emergent phenomenon of what critically instrumental rationalists teach each other and learn from each other—in their reflectively motivated interactions such as purposeful critical discussions, as well as in their everyday interactions marked by specific "tags" which make the *reflective origin* of those everyday actions tangible (the "way" one buys bread in a small bakery, the "respect" one shows to a low-degree employee, etc.). Underlying this multifarious network of reflective resources is *an array of so-called "critical theories of society,"* denoted again with double indices so as to allow for maximal dynamic flexibility in how critical views evolve over time ($c^a{}_1$, $c^b{}_2$, …,$c^{z(t)}{}_{N(t)}$). Each theory c is one way (a) of identifying the pro's and con's of the current society S_t and (b) of pinpointing certain typical actions $\mathbf{b}(c)$ that could be undertaken so as to improve society, given a social ideal $S^*(c)$ conveyed explicitly or implicitly by the theory. These ideal actions $\mathbf{b}(c)$ are those which induce the vectors b_i and $\mathbf{b}_n$ in eq. (2); the inducement, however, is not a perfect logical deduction between $\mathbf{b}(c)$ and eq. (2) because in between there is the crucial filter of i's subjective distance measure d_i and of her own untheorizable practical judgment about what actions on her part and on the part of others would be appropriate.

One of the crucial ways in which this practical judgment is not theorizable is that i bases her actions on her own understanding of how the current society S_t works, and this understanding may not be fully articulated in the critical theory c which i adopts. In other words, in addition to a critical theory c the agent may also, in some cases, make use of one or even several explicative theories available in the society, and listed in an array ($x^a{}_1$, $x^b{}_2$, …,$x^{z(t)}{}_{N(t)}$). Each theory x is one way of understanding the "objective" reasons why society S_t has certain pro's and con's and the way in which S_t generates "unavoidable" constraints linked to its own immanent operation. It generates a particular explanatory model $S_t(x)$.

So here is the stylized picture of what critically instrumental rationalists learn in interactions. Each i co-determines three things: her subjectively experienced distance $d_i(S,S^*)$; the critical

theory c which allows her to give a direction to her frustration by seeing how from S an ideal S^* could be approached; and the explicative theory x—if any—which allows her to understand the objective reasons why S has the properties she is experiencing. The way she co-determines these variables is through an interactive process in which (*a*) she faces defenders of S and seeks to gather wisdom (explicative, for S and/or against S) from them by purposively demystifying their deepest convictions and trying to understand their reactions; and (*b*) she encounters both defenders of S and denigrators of S in a global atmosphere of cohesion based on a mutual desire to end up with sound, argued reasons either for agreeing to agree or for agreeing to disagree. Thus, the process involves both critical demystification and Socratic cohesiveness: no critically instrumental rationalist takes anything for granted from the others and even cynically opposes their views in order to make up his own mind about what S^* to adopt, but none of them seeks to destroy the other or to brutally transform S into S^* by some global act of authority, because all of them view each other as engaged in the same reflective search for their ideal social world and this gives them a paradoxical kind of cohesion. In the end, at each period each agent i takes home a vector $(x, c, d_i[S(x), S^*(c)])$ which serves as an elaborate internal model to put into motion the maximization of (2) subject to (3). The actions flowing from this maximization then join all other agents' actions to generate the society's *esprit critique* as well as to transmit to each agent a credit attribution. The notion of "credit" underlying critically instrumental interactions is much more difficult to define than in a simpler complex adaptive system because, as was highlighted earlier, an agent will not merely be attempting to mimic others' high-credit rules and to avoid others' low-credit rules; the actions (a_i, b_i) and the internal model (x, c) on which they were based are evaluated by the agents in a much more complicated way—for instance, i's success may not coincide with a lot of agents $j \neq i$ having been converted or even made aware of i's options, but with certain key agents in i's neighborhood or in the society at large having adopted i's point of view. A significant amount of work is called for to get a better understanding of credit attribution in complex adaptive systems with critically instrumental rationalists.

Some readers might be worried that, in contravention to the options underlying the complexity approach, we might be reverting to the kind of hyper-rationalism extolled by Peyton Young earlier on. No doubt, the demands of socially reflective rationality are more stringent than the merely mechanical use of, say, "if/then" rules by a frog or by a human agent carrying out routinized activities in a constantly changing environment. However, the model suggested here

may be compatible with simplified versions of the c- and x-theories being used in "tagged" actions and reactions of the agents—as when an individual is identified as "a Marxist" or as "a neoliberal" through some typical actions, reactions or statements that do not require the statement of a whole underlying theory but could be traced back to one such theory. One could, for instance, define for each c and each x a set of characteristic or typical actions $A(c)$ and $A(x)$; what really matters is the way in which each such set translates into the value of d_i, and hence into the way (2) is maximized.

3.3. Aggregate behavior: Emergent properties and the puzzle of a "mass of critics"

The generalized learning process just sketched, along with the subsequent maximization of (2)-(3) by all agents, yields a complex adaptive system, one of whose characteristic emergents is the society's momentary *esprit critique*, $\mathbf{C}_t$. In other words, one way in which this complex adaptive system is a *free* society is that it allows for the above mechanism of interactive learning and reflection on individual action. Formally, this *esprit critique* could be reformulated as a function of all circulating critical theories:

$$\mathbf{C}_t = L_t(c^a{}_1, c^b{}_2, \ldots, c^{z(t)}{}_{N(t)}), \tag{4}$$

where the form of the time-dependent "emergence function" L_t might itself be determined by all the other parameters of the model, mainly the vector of explicative theories $(x^a{}_1, x^b{}_2, \ldots, x^{z(t)}{}_{N(t)})$ and the vector of individual actions $[(b_{1,t}, a_{1,t}), \ldots, (b_{i,t}, a_{i,t}), \ldots, (b_{M,t}, a_{M,t})]$, hence the state of society S_t itself.

Clearly, being a characteristic emergent does not imply—contrary what is imposed in the majority of "social physics" theories where emergent phenomena are studied—that $\mathbf{C}_t$ is a quantifiable concept; it is not even clear, at this point, to what space it ought to belong. (As Potts [2000: 11-54] brilliantly explains, in neoclassical theory market prices are located in R^k where k is the dimension of commodity space, and the qualities of other variables are similarly made to fit into real-numbered hypercubes, but this restriction is often too stringent for other social phenomena.) It is no one's critical theory in particular, but it "tags" the overall type of society in which *all* agents live. As such, it is not always an easily observable aggregate because, in the end, as an aggregate it only shows up in certain localized "tags" carried by agents, or in certain hidden

norms or heuristic guidelines for social criticism that circulate in the society. Much like Imre Lakatos's (1970) criteria for the positive and negative heuristics of a research program, the *esprit critique* is not written down or printed onto tags like prices or other "types"—it always need to be testified to by individuals in interaction, so that a social scientist himself wanting to "observe" a society's *esprit critique* would have to do so by interacting with the society's members. This crucial argument was raised by Geuss (1981: 92-94) in his discussion of the epistemic status of single critical theories, and it carries with it the risk that the very notion of *esprit critique* might be easily disparaged as "mystical" or "fuzzily inexact."

But in fact, as most of contemporary political philosophy has demonstrated, it is a crucial aspect of a free society because the whole learning process and the mutual adjustments occurring amongst sub-networks and networks of sub-networks of Socratic demystifiers are deeply *path-dependent* on $\mathbf{C}_t$: you can only critically interact by using the prevailing *esprit critique* as a globally endogenous but individually exogenous reference point. For example, criteria for what counts as criticism, as well as substantive contents of critical alternatives to the prevailing society, cannot be arbitrarily grasped out of the air; they are part of a society's evolving cultural heritage (if only because critical theories are always enacted by the agents with reference to the current state of society), just like other components covered by Hayek's notion of cultural evolution. One of the main attempts to offer a model using an implicit notion of *esprit critique* is Jürgen Habermas's (1981) model of communicative action and discursive rationality. This is not, however, a convincing model in the present framework, because its implicit assumption is too restrictive—namely, the assumption that in any interaction critically-minded agents must of necessity make implicit reference to the regulative ideal of a "free speech situation of uncoerced, undistorted communication." This regulative ideal is largely transcendental in Habermas's approach, and therefore it does not fit in with the present approach which sees the regulative ideals for critical interaction (this is one alternative way of characterizing the *esprit critique*) as themselves endogenously evolving within a complex adaptive system. In the present model, agents usually elaborate their (x, c, d_i) vectors in situations of *non-ideal* speech, so that Habermas's particular version of "$\mathbf{C}t=\mathbf{C}$ at all t" can only be a steady state—and perhaps only a metastable one, as much of political experience testifies—of a path-dependent trajectory towards collective emancipation.

In general, then, the time path $\{\mathbf{C}_t\}_{t=1, 2, \ldots}$ can be seen as a summary of a society's historical trajectory as driven by the reflective emancipatory interests of its individual interacting members. Being an emergent time path, it will obviously not visibly reflect any individual's normative options or any group's actual successes in freeing themselves from certain constraints; rather, it will be the compositional "trace" left in time by the often intractable inter-individual interactions of *a mass of critics*. In that sense, this trajectory is the key characteristic emergent that a Hayekian Critical Theory is concerned about—independently, that is, of the other characteristic emergents whose nature will vary depending on historical events such as the unintended emergence of certain economic regimes (self-managed socialism, Rhineland capitalism, neoliberal, etc.), political regimes (social democracy, liberal anarchism, Thatcherian ultraliberalism, etc.), and so on.

4. Can normative economics become a discipline that studies social emancipation?

4.1. "Knowledge," yes—but what knowledge? The critical harnessing of social complexity

We are now in a position to come back to some of the issues raised in the first section about the need to deepen the epistemic foundations of normative economics. What the above discussion has certainly emphasized is that in order to study complex social emancipation—i.e., the process by which interacting individuals make their society more free in often individually unintended ways—normative economics has to join positive economics in a radical recasting of its assumptions about the agents' use of knowledge. We have inherited both from Austrian economics and from the new institutionalism an epistemic model of the economic agent which, while being sophisticated on many counts, is too crude when it comes to the agent's critical capacity and her ability to formulate "emancipatory interests" (Habermas [1965]) and to act on them. In short, both normative and positive economics have yet to take full account of the agents' critical instrumental rationality—and, as we have seen, when they do so they are likely to find out that the positive/ normative split itself is in jeopardy.

To illustrate what I mean, let me contrast the notion of "opportunistic harnessing of social complexity" which was introduced at the end of section 2.2 with two other notions, that of

course-of-action-critical harnessing and that of *overall-structure-critical harnessing*. In one of his most-quoted essays of the mid-1940s, Hayek wrote that

> If we can agree that the economic problem of society is mainly one of rapid adaptation to changes in the particular circumstances of time and place, it would seem to follow that the ultimate decisions must be left to the people who are familiar with these circumstances, who know directly of the relevant changes and of the resources immediately available to meet them. We cannot expect that this problem will be solved by first communicating all this knowledge to a central board which, after integrating all knowledge, issues its orders. We must solve it by some form of decentralization. But this answers only part of our problem [because] the "man on the spot" cannot decide solely on the basis of his limited but intimate knowledge of the facts of his immediate surroundings. There still remains the problem of communicating to him such further information as he needs to fit his decisions into the whole pattern of changes of the larger economic system. How much knowledge does he need to do this successfully? (…) It is in this connection that (…) the "economic calculus" (or the Pure Logic of Choice) helps us, at least by analogy, to see how this problem can be solved, and in fact is being solved, by the price system. (Hayek [1945: 83-86 *passim*])

In this passage, the scope of the "knowledge" supposedly used by economic agents is kept rather narrow since it comprises two categories: first, the agent's localized, immediately accessible knowledge of his everyday situation, and second, a summary of all *market* interactions through the "price system." Clearly, within the present framework, Hayek turns out to be laboring within one very specific steady state of the cultural-evolution process—namely, a state in which no agent uses any critical knowledge anymore because, supposedly through previous rounds of critically-minded interaction, all agents have unanimously arrived at the conclusion that $S=S^*$, where S is the free-market democracy advocated by Hayek (see e.g. Hayek [1960]), so that $d_i=0$ for all i and the need for further social criticism is fully extinguished: this is a well-known lock-in phenomenon in which a reflectively adopted theory becomes self-fulfilling at the one-for-one scale—but even Hayek would have known that such steady states are, at best, metastable and, in any case, quite prone to sudden phase transitions as soon as some agents start reflecting again and wondering whether this is all the social world has to offer, even in the midst of plenty. A similarly narrow view of knowledge used by agents, within a broad view of economic history, has been offered more recently by Douglass North in a sort of summary of his seminal work:

> Competition forces organizations continually to invest in new skills and knowledge to survive. (…) Whether through learning by doing on the job or the acquisition of formal knowledge, improving the efficiency of the organization relative to that of rivals is the key to survival. (…) The immediate investment of economic organizations in vocational and on-the-job training obviously will depend on the perceived benefits, but an even more fundamental

influence on the future of the economy is the extent to which societies will invest in formal education, schooling, the dissemination of knowledge, and both applied and pure research, which will, in turn, mirror the perceptions of the entrepreneurs of political and economic organizations. The key to the choices that individuals make is their perceptions, which are a function of the way the mind interprets the information it receives. The mental constructs individuals form to explain and interpret the world around them are partly a result of their cultural heritage, partly a result of the local everyday problems they confront and must solve, and partly a result of nonlocal learning. (North [1997: 226-227])

What we have here is a portrait of economic theory as the positive study of interacting agents using uncritical, "everyday" knowledge perceived as information utilizable for survival purposes. Crucial as such survival considerations may be, by the end of the present paper they should clearly appear as insufficient. True, North pays some lip service to agents' "cultural heritage"—but observe that since beside that heritage they use nothing but "local everyday" knowledge and "nonlocal learning," the cultural dynamics are in fact conceived as purely exogenous to the agents' intentional interactions. Consistently with the neoclassical framework, therefore, North operates a reductionist move which by construction has to view culture not only as an *ex post* unintended result, but also as an *ex ante* non-intentional result of the agents' problem-solving interactions: culture is whatever supports the agents' problem solving, i.e., whatever rules of thumb or adaptational schemas they create by tinkering with everyday knowledge. In other words, the agents' internal models are never critical in the above sense—which is reflected in the fact that most of what North portrays his agents as doing is a by-product of an evolving industrial market structure: on-the-job training, performance-driven education and schooling, etc. These agents are assumed to engage in purely opportunistic harnessing of the social complexity which they unintentionally feed and have no intention of reflecting on.

Such positive economic theory can easily become part of the uncritical everyday knowledge of policymakers about the society they supposed to steer. If, as much of the new political economy conceives of them (see e.g. Basu [1999]), these policymakers themselves have survival-determined interests—be it only the interests in political survival through re-election—they will seek to use the positive economic theory provided by North in order to opportunistically harness the social complexity which they can *within limits* try to influence but cannot fully control. The limits, of course, are always given by the characteristic emergents of the system which they analyze through North's particular explicative theory. Thus, we may very well have a fully

opportunistic system in which both non policymakers and policymakers seek to exploit the existing complexity as they see it *ex ante*, using some of the tools suggested by Axelrod and Cohen (2000), in order to get the most out of it for their own respective survival within the system. No one is attempting to take critical distance from the system, because everyone is busy exploiting its (limited) potentialities from the inside.

Now one could very well imagine some non-policymaker or even policymaker suddenly wondering whether something else than mere survival might be a more important objective given the overall structure of the economy, or whether in fact that overall structure itself might have to be modified to increase *everyone's* survival opportunities and/or to increase the latitude for implementing nonsurvival objectives. In either case, that agent while remaining within the system suddenly turns to what I would call *critical harnessing of the social complexity*. This can be done in two ways, as my example suggests, and these two ways can be called "course-of-action (CA)-critical" and "overall-structure (OS)-critical":

i. *CA-critical harnessing:* Given the directed graph defining the society's interaction patterns and given the motivations that generate flows between nodes of the graph, take courses of action which will make use of these interaction patterns and those motivations to pursue a *reflected* and *intra-systemically coherent* idea of emancipation, i.e., an idea which the overall logic inherent society's structure, as well as your location on the graph, does not spontaneously attribute to you but which remains consistent with upholding the interaction patterns and the motivations.

ii. *OS-critical harnessing:* Take courses of action which use the directed graph defining the society's interaction patterns and given the motivations that generate flows between nodes of the graph to pursue a *reflected* and *extra-systemically coherent* idea of emancipation, i.e., an idea of emancipation whose coherent pursuit requires modifying the overall logic inherent in the interaction patterns and the motivations, i.e., requires modification of the overall structure of the society.

Now, in order to act in ways which advance one's CA- or OS-critical views, one needs the whole arsenal of critically-minded interaction which the above model has outlined. Essentially, this model amounts to assuming that agents seek to obtain knowledge beyond their everyday

situational knowledge—which both Hayek and North acknowledge they must—but also *beyond the uncritically driven characteristic emergents of the system*—which neither Hayek nor North allow for in their exceedingly narrow focusing of the agents' knowledge collection.

But if this is so, as Hayekian Critical Theory claims, then normative economics can only deepen its epistemic foundations by explicitly modeling the various processes of knowledge production—both for explicative and critical knowledge, since they mutually determine each other in a dynamical perspective—and of critical-knowledge acquisition by the agents in the economy. That knowledge-production processes may themselves have to be modeled as complex adaptive sub-systems of the larger system has been shown above in section 3.2 and has acknowledged repeatedly by science economists and by evolutionary microeconomists—with a constant focus, however, either on explicative knowledge aimed at improving the predictive capacity of internal models (Potts [2000: 155-180]) or on explicative knowledge confined to scientific communities (Brock and Durlauf [1999]). The specific topic of the production and acquisition of *critical knowledge designed to reflect on one's everyday actions to improve society as a whole* has not, to the best of my knowledge, received even cursory attention in the theory of complex adaptive systems.

4.2. Collective choice of a complex adaptive system: Which characteristic emergents do "we" want?

One of the key lessons of this whole discussion (see also Arnsperger [2004c, 2004d]) is that while a society's interaction structure unavoidably has to evolve certain complex adaptive systems embedded in its if it is to count as a free society at all, this by no means implies that we have by the same token resolved the question, *Which* complex adaptive system(s) should and can evolve?

This is, of course, a daunting question because to even begin answering it requires that we postulate a relatively clear functional relationship by which the apparition of any given characteristic emergent e as a relevant one for a society S can be traced back to the morphogenetic features of S—namely, to how interactive structures in general generate a spectrum of emergents and to how the interactive structure of S has changed over time—so as to be able to say what the current spectrum of characteristic emergents for S is. If this were possible, then by conceiving every explicative theory x and every critical theory c as a theory that models

society S or S^* as a system of complex adaptive systems, each critically instrumental rationalist could extract from her vector (x, c, d_i) a set of desired characteristic emergents which would serve as "tags" for her critical view on social emancipation. Her actions (b_i, a_i) would then be geared, among other things, towards the realization of these desired emergents through the appropriate complexity-creating institutional and legal framework. (One straightforward case is a free-market oriented agent militating for full wage flexibility in a society where there are currently minimum-wage laws and other legal obstacles such a near 100% taxation of very high labor incomes.) At this level of generality, unless one has an extremely dogmatic view that throughout history interaction patterns and associated motivations have been constant (North and also Coleman or Becker would appear to be close to this position), the question of which complex adaptive systems one believes should be fostered is an issue of OS-critical harnessing of social complexity: each agent reflects *ex ante* (i.e., without any guarantee that he can implement it *ex post*) on what overall structure would be most adequate to the pursuit of emancipation as she conceives it in her critical theory. And in fact, even if one had a very essentialist view that interaction patterns and associated motivations have been virtually constant throughout history, all the subsystems under scrutiny would be so differentiated that one could for all practical purposes consider them as quasi-structural and, again, the question of which subsystems should be fostered as complex adaptive systems would be virtually of the OS-critical kind.

The skeptical reader of the bounded-rationality perspective will again intervene at this point to warn of the danger of hyper-rationalism. And actually, this time, she would be right: for an agent to build an *ex ante* OS-critical vision of what characteristic emergents he believes "we" should want, moreover without any guarantee that his vision will have any *ex post* impact on the historical trajectory $\{S_t\}$, may seem a uselessly daunting task best delegated to a minority of experts or theorists while the agent himself goes back to pursuing his opportunistic harnessing of the prevailing social complexity. This might in turn be explained by the most standard microeconomic choice models, whether of the full-rationality variety such as Becker (1965) or of the bounded-rationality variety such as Simon (1955), as well as of the transaction-cost variety such as Williamson (1973). Effectively, this introduces a socially functional differentiation between at least three categories of agents:

- those for whom k_i=0 independently of the value of d_i, and who in the extreme limit simply eject d_i from their cognitive realm;

- those for whom *D* increases sharply due to the realization that huge cognitive and economic costs are involved in really shaping a fully fledged critical theory to guide their actions in society;
- those few professionals—Horkheimer (1937) calls them "*oppositional intellectuals*"—who take it upon themselves to produce critical theories of society as *both* their own specific way of opportunistically harnessing the prevailing social complexity—i.e., of making a living off a society of potentially dissatisfied but normally unreflective agents—*and* their own specific way of acting in a CA-critical or OS-critical way.

As Horkheimer and other members of the Frankfurt School have emphasized, the only way for these oppositional intellectuals not to become mere producers of pseudo-critical knowledge about society (that is, knowledge that all three categories of agents balk at during weekends and holidays but never actually put into practice) is to make their work *socially engaged:* one of their main tasks, which must actually be part of any critical theory of society they produce, is to constantly call on the two other categories to reflect not on what these agents themselves believe society should be like and how it could be improved (they have, by definition, given up this effort) but on the critical theories constructed by the oppositional intellectual with these agents in mind. Thus, as argued by Walzer (1987, 1988), the oppositional intellectual is in a position where she has to criticize her fellow society members a first time for not being reflective enough about some of the *de facto* results of their uncritical social interactions, and then a second time for not paying sufficient attention to that first criticism.

Whatever the case may be, however, two things seem clear after the discussion of the above model. First of all, even if the tripartition of society should occur, it is unlikely that standard top-down and general axiomatics would be the best way to work as an oppositional intellectual. Most likely, different sets of axioms would have to be specified for different particular social situations, each situation being identified by a particular critical theory of society which is not itself reducible to these axioms. Thus, axiomatic work may not generally dispense the oppositional intellectual from reflecting more deeply on which critical theory of society she adheres to, and to which agents in the population she is addressing her axiomatic analysis as a tool for their emancipation. Second, the division of society into these three categories of agents is not a kind of "natural" or "inevitable" process, as Hayek would have it implicitly when he claims

that epistemic limits inherent in the human brain make it impossible for any individual agent to "comprehend" the whole social process. Of course that impossibility is a fact, but no critical theorist has ever been deterred from theorizing a better society by his *de facto* inability to see all agents' actions in one gaze. But if this is so for someone who is already a theorist, then why should it be "naturally" different for any of the other members of society? Epistemic, psychological and economic costs may be more a datum of the prevailing economic system than a natural or constitutive feature of the human brain—and hence, every agent might *as part of his critical endeavor* start by reflecting on the society-wide factors which make him too limited, or too tired, ot too busy to learn about social complexity in the first place. He might quickly discover that it is the adaptive demands made on him by the current system's complexity which prevent him from learning enough about complex adaptive systems to be able to criticize these current adaptive demands… Surely, this would be a good reason for him to start questioning the current economic system, so that one of the first and main tasks of the oppositional intellectual might well be to reflect on ways of making "complexity teaching" as widespread as possible—a task which even the most contemporary work on economics and the teaching of complexity has tended to leave lying underneath the huge pile of more positivistic, descriptive and explicative tasks attributed to economics by those who value knowledge of "the real world" (see e.g. Colander [2000]).

4.3. Normative economics with a deeper foundation: The critical process and the intra-systemic emergence of "mass criticism"

Many conceptual and technical issues are left open in this paper, and will call for extensive research. However, I hope to have shown that Hayekian Critical Theory, apart from having a sound philosophical foundation which I have not rehearsed here (see Arnsperger [2004c, 2004d]), offers an important way to extend the theory of complex adaptive systems in the direction of a normative economics with renewed epistemic foundations. To be precise, Hayekian Critical Theory can be viewed as a version of complex adaptive systems theory where the agents' internal models, which drive their actions and reactions in interactions, are critical theories of society, that is, theories which purport to describe society while at the same time rejecting the fetishism of "factual information" or "society as it is" and replacing it with the probing of ways in which current society could be improved as to what regards emancipation. Obviously, the task of each

critical theory of society is to flesh out the idea of emancipation and to use it to formulate a critical judgment about the existing social system as well as a practical judgment about what ought to be done.

This view of the role of knowledge about society in society seems to me to jeopardize two established dichotomies:

- The first dichotomy is that between positive and normative economics, which has been questioned ever so often already in the history of economics. Here it is questioned yet again, but this time because any agent endowed with a critical theory of society—or at least with the behavioral and attitudinal "tags" which can be traced back to a fully-fledged critical theory—carries with him *both* a description *and* a condemnation of society as it is, all in one. In fact, the *reason* why the agent wants to have an accurate description of current society is that she wants to act to change it either CA-critically or OS-critically—that, in a nutshell, is the rationale for critical theorizing (see Horkheimer [1937], [1946]).
- The second dichotomy is located within what is still called normative economics, and concerns the split between the axiom-building theorist and the unknowing agents who supposedly "act out" these axioms. If the theorist is indeed doing *normative* work, he is supposedly (see Fleurbaey [1995]) offering axioms which he believes the agents would agree to if they knew them—or else, we are confined to the narrow paternalistic-technocratic view of how the theorist, the "ruler" and the agent coexist in a free society. But if the question is whether the agents would agree with the axioms if presented with them, there is no legitimate reason not to assume *as part of the rational-action model* that they are endowed with a capacity to reflect on, and perhaps even outright construct, the axioms they feel would be best given their "emancipatory interests." And this leads us directly into the discussion we have had here, i.e., how critically-minded, action-oriented rational agents will interact with their *ex ante* intentional critical theories so as to generate an *ex post* unintentional social situation. This is still, if you want, "normative" economics, but it is based on *an entirely different view as to where in society normativity is being produced and with what epistemic tools the agents are equipped.*

To repeat, much about the complex dynamics of such a system of interacting agents has been left unanswered here. In particular, we have no precise and plausible models of credit attribution, and we also do not yet know a lot about how such *ex ante* intentional theories will get "resolved" into an *ex post* state in which, perhaps, none of them is fully realized and most of them fund the actual society still as unsatisfactory as before. Will this very fact—i.e., the brutal drop from grand emancipatory intentions to mediocre concrete realizations—tend to erode the agents' desire to even work to look for and formulate a critical theory, which would explain much of contemporary *anomie*? Is such "mass criticism," in which all individual agents to a greater or lesser extent modify their everyday actions so as to try to change the world (see eq. (2) above) a recipe for massive disappointment—or are there certain conditions under which this self-perversion of mass criticism can be averted? These are deep issues, but they can only be formulated in a theoretical framework that portrays the good society as one in which *a permanent critical process is going on within the uncritical interactions usually modeled by economic theory.*

References

Adorno, Th. (1962), "Zur Logik der Sozialwissenschaften", reprinted in *Soziologische Schriften I*, Frankfurt, Suhrkamp, 1972, pp. 547-565.

Arnsperger, Ch. (1990), "Information Failures and Wandering Systems in Keynesian Economics: Clower-Leijonhufvud Revisited", in M. Sebastiani (ed.), *The Notion of Equilibrium in the Keynesian Theory*, London, Macmillan, pp. 189-217.

Arnsperger, Ch. (1994), "Envy-Freeness and Distributive Justice", *Journal of Economic Surveys* **8**(2), 155-186.

Arnsperger, Ch. (1997), "Reformulating Equality of Resources", *Economics and Philosophy* **13**, 61-77.

Arnsperger, Ch. (2000), "Methodological Altruism as an Alternative Foundation for Individual Optimization", *Ethical Theory and Moral Practice* **3**(2), 115-136.

Arnsperger, Ch. (2004a), "Raison des Lumières et critique sociale. 1° partie: Sur la tâche émancipatrice de la philosophie politique et de la théorie économique", *Cahiers du GRISE,* forthcoming.

Arnsperger, Ch. (2004b), "Raison des Lumières et critique sociale. 2° partie: Vers un modèle de rationalité instrumentale critique", *Cahiers du GRISE,* forthcoming.

Arnsperger, Ch. (2004c), "The Use of Knowledge About Society. Part I — Reopening the Road from Frankfurt to Vienna: Why 'Hayekian Critical Theory' is Not an Oxymoron", Chaire Hoover Document DOCH 139 (www.etes.ucl.ac.be) and GREC Working Paper 2004-04 (www.unites.uqam.ca/philo/pdf/Arnsperger_C_2004-04.pdf).

Arnsperger, Ch. (2004d), "The Use of Knowledge About Society. Part II — Critical Instrumental Rationality Between Spontaneity and Reflexivity: Spelling Out Hayekian Critical Theory", Chaire Hoover Document DOCH 140 (www.etes.ucl.ac.be) and GREC Working Paper 2004-05 (www.unites.uqam.ca/philo/pdf/Arnsperger_C_2004-05.pdf).

Arnsperger, Ch., and D. de la Croix (1996), "Envy-Minimizing Unemployment Benefits", *Economic Design* **2**, 119-146.

Arnsperger, Ch., and Y. Varoufakis (2003), "Toward a Theory of Solidarity", *Erkenntnis* **59**, 155-188.

Arthur, W.B., S. Durlauf and D. Lane (1997), "Introduction" to W.B. Arthur, S. Durlauf and D. Lane (eds), *The Economy as an Evolving Complex System II*, Cambridge, Perseus, pp. 1-13.

Auyang, S. (1998), *Foundations of Complex-System Theories in Economics, Evolutionary Biology, and Statistical Physics*, Cambridge, Cambridge University Press.

Axelrod, R. (1997), "The Dissemination of Culture", *Journal of Conflict Resolution* **41**, 203-226.

Axelrod, R., and M.D. Cohen (2000), *Harnessing Complexity: Organizational Implications of a Scientific Frontier*, New York, Basic Books.

Ball, Ph. (1999), *The Self-Made Tapestry: Pattern Formation in Nature*, Oxford, Oxford University Press.

Ball, Ph. (2004), *Critical Mass: How One Things Leads Into Another*, London, Random House.

Basu, K. (1999), *Prelude to Political Economy: A Study of the Social and Political Foundations of Economics*, Oxford, Oxford University Press.

Becker, G. (1965), "A Theory of the Allocation of Time", *Economic Journal* **75**, 493-517.

Bowles, S. (2004), *Microeconomics: Behavior, Institutions and Evolution*, Princeton, Princeton University Press.

Brock, W.A. (2000), "Some Santa Fe Scenery", in Colander (2000), pp. 29-49.

Brock, W.A., and S. Durlauf (1999), "A Formal Model of Theory Choice in Science", *Economic Theory* **14**, 113-130.

Brock, W.A. and S. Durlauf (2001), "Discrete Choice With Social Interactions", *Review of Economic Studies* **68**(2), 235-260.

Chamley, Ch. (2004), *Rational Herds: Economic Models of Social Learning*, Cambridge, Cambridge University Press.

Colander, D., ed. (2000), *The Complexity Vision and the Teaching of Economics*, Cheltenham, Edward Elgar.

Dupuy, J.-P. (2002), *The Mechanization of the Mind: On the Origins of Cognitive Science*, Princeton, Princeton University Press.

Durlauf, S. (1997), "Statistical Mechanics Approaches to Socioeconomic Behavior", in W.B. Arthur, S. Durlauf and D. Lane (eds), *The Economy as an Evolving Complex System II*, Cambridge, Perseus, pp. 81-104.

Durlauf, S. (2002), "A Framework for the Study of Individual Behavior and Social Interactions", *Sociological Methodology* **31**, 47-87.

Fleurbaey, M. (1995), *Théories économiques de la justice*, Paris, Economica.

Geuss, R. (1981), *The Idea of a Critical Theory: Jürgen Habermas and the Frankfurt School*, Cambridge, Cambridge University Press.

Habermas, J. (1965), "Erkenntnis und Interesse", reprinted in *Technik und Wissenschaft als "Ideologie"*, Frankfurt, Suhrkamp, pp. 146-168.
Habermas, J. (1981), *Theorie des Kommunikativen Handelns*, 2 volumes, Frankfurt, Suhrkamp.
Hayek, F.A. (1945), "The Use of Knowledge in Society", reprinted in *Individualism and Economic Order*, Chicago, University of Chicago Press, 1948, pp. 77-91.
Hayek, F.A. (1960), *The Constitution of Liberty*, London, Routledge.
Holland, J. (1995), *Hidden Order: How Adaptation Builds Complexity*, Cambridge, Perseus.
Holland, J. (1998), *Emergence: From Chaos to Order*, Cambridge, Perseus.
Horkheimer, M. (1937), "Traditional and Critical Theory", reprinted in *Critical Theory: Selected Essays*, New York, Continuum, 1972, pp. 188-252.
Horkheimer, M. (1946), *Eclipse of Reason*, New York, Continuum, 1974.
Koppel, M., H. Atlan and J.-P. Dupuy (1983), "Complexité et aliénation: Formalisation de la conjecture de von Foerster", in F. Fogelman Soulié (ed.), *Les Théories de la complexité: Autour de l'œuvre d'Henri Atlan*, Paris, Seuil, pp. 410-421.
Krugman, P. (1997), "How the Economy Organizes Itself in Space: a Survey of the New Economic Geography", in W.B. Arthur, S. Durlauf and D. Lane (eds), *The Economy as an Evolving Complex System II*, Cambridge, Perseus, pp. 239-262.
Lakatos, I. (1970), "Falsification and the Methodology of Scientific Research Programmes", in I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press, pp. 91-196.
Muller, J. (1993), *Adam Smith in His Time and Ours: Designing the Decent Society*, Princeton, Princeton University Press.
North, D. (1981), *Structure and Change in Economic History*, New York, Norton.
North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge, Cambridge University Press.
North, D. (1997), "Some Fundamental Puzzles in Economic History/ Development", in W.B. Arthur, S. Durlauf and D. Lane (eds), *The Economy as an Evolving Complex System II*, Cambridge, Perseus, pp. 223-237.
Platteau, J.-Ph. (2000), *Institutions, Social Norms, and Economic Development*, Amsterdam, OPA, Harwood.
Potts, J. (2000), *The New Evolutionary Microeconomics: Complexity, Competence and Adaptive Behaviour*, Cheltenham, Edward Elgar.
Simon, H. (1955), "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics* **69**, 99-118.
Varela, F., E. Thompson and E. Rosch (1991), *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MIT Press.
Waldrop, M.M. (1992), *Complexity: The Emerging Science at the Edge of Order and Chaos*, New York, Simon and Schuster.
Walzer, M. (1987), *Interpretation and Social Criticism*, Cambridge, Harvard University Press.
Walzer, M. (1988), *The Company of Critics : Social Criticism and Political Commitment in the Twentieth Century*, New York, Basic Books.
Williamson, O. (1973), "Markets and Hierarchies: Some Elementary Considerations", *American Economic Review* **63**, 316-325.
Young, H. P. (1998), Individual Strategy and Social Structure: An Evolutionary Theory of Institutions, Princeton, Princeton University Press.

NUMÉROS RÉCENTS

Yves Gingras: *What Did Mathematics Do to Physics?* (No 2001-01);
Daniel Vanderveken: *Formal Ontology and Predicative Theory of Truth. An Application of the Theory to the Logic of Temporal and Modal Propositions* (No 2001-02);
Peter J. Boettke, John Robert Subrick: *From the Philosophy of Mind to the Philosophy of the Market* (No 2001-03);
Robert Nadeau: *Sur l'antiphysicalisme de Hayek. Essai d'élucidation* (No 2001-04);
Steven Horwitz: *Money and the Interpretive Turn : Some Considerations* (No 2001-05);
Richard Hudson, Gisèle Chevalier: *Collective Intentionality in Finance* (No 2001-06);
Carlo Benetti: *Smith et les mains invisibles* (No 2001-07);
Michel B. Robillard: *Compte rendu critique de* Cognitive Adaptations for Social exchange *de Leda Cosmides et John Tooby* (No 2001-08);
Maurice Lagueux: *What does rationality mean for economists ?* (No 2001-09);
Gérard Duménil et Dominique Lévy: *Vieilles theories et nouveau capitalisme: Actualité d'une économie marxiste* (No 2001-10);
Don Ross: *Game Theory and The New Route to Eliminativism About the Propositional Attitudes* (No 2001-11);
Roberto Baranzini: *Le réalisme de Walras et son modèle monétaire* (No 2001-12);
Paul Dumouchel: *Règles négatives et évolution* (No 2002-01);
Jean Robillard: *La transsubjectivité et la rationalité cognitive dans la méthode de la sociologie cognitive de Raymond Boudon* (No 2002-02);
Daniel Vanderveken: *Attempt and Action Generation – Towards the Foundations of the Logic of Action* (No 2003-03);
Robert Nadeau: *Cultural Evolution True and False: A Debunking of Hayek's Critics* (No 2003-04);
D. Wade Hands: *Did Milton Friedman's Methodology License the Formalist Revolution?* (No 2003-05);
Michel Rosier: *Le questionnement moral : Smith contre Hume* (No 2003-06);
Michel Rosier: *De l'erreur de la rectification par Bortkiewicz d'une prétendue erreur de Marx* (No 2003-07) ;
Philippe Nemo : *La Forme de l'Occident* (No 2003-08);
Robert Nadeau: *Hayek's and Myrdal's Stance on Economic Planning* (No. 2003-09);
Guillaume Rochefort-Maranda: *Logique inductive et probabilités : une analyse de la controverse Popper-Carnap* (No. 2003-10);
Guillaume Rochefort-Maranda: *Probabilité et support inductif. Sur le théorème de Popper-Miller (1983)* (No. 2003-11);
F.P.O'Gorman: *Rationality, Conventions and Complexity Theory: Methodological Challenges for Post-Keynesian Economics* (No. 2003-12);
Frédérick Guillaume Dufour: *Débats sur la transition du féodalisme au capitalisme en Europe. Examen de contributions néo-wébériennes et néo-marxistes* (No. 2003-13);
Jean Robillard: *Théorie du sujet collectif et attribution des propriétés sémantiques individuelles* (No. 2003-14);
Philippe Mongin: *L'analytique et le synthétique en économie* (No. 2003-15) ;
Philippe Mongin: *Value Judgments and Value Neutrality in Economics. A Perspective from Today* (No. 2003-16) ;
Maurice Lagueux : *Explanation in Social Sciences. Hempel, Dray, Salmon and van Fraassen Revisited* (2003-17) ;
Learry Gagné : *Les fondements rationnels et émotifs des normes sociales* (2003-18) ;
Pierre Milot : *La transformation des universités dans le contexte d'application de l'économie du savoir* (2003-19);
Marc Chevrier: *Les conflits de savoirs en démocratie constitutionnelle: le cas du constructivisme judiciaire canadien* (2003-20);
Guillaume Rochefort-Maranda: *Confirmation et corroboration: accords et désaccords* (2003-21);
Nadeau, Robert: *Has Hayek Refuted Market Socialism?* (2004-01);
Mathieu Marion: *Investigating Rougier* (2004-02);
Frédérick Guillaume Dufour: *Historical Sociology and the Analysis of Social-Property Relations* (2004-03);
Christian Arnsperger: *Reopening the road from Frankfurt to Vienna: Why "Hayekian Critical Theory" is not an oxymoron* (2004-04);
Christian Arnsperger: *Critical instrumental rationality between spontaneity and reflexivity: Spelling Out Hayekian Critical Theory* (2004-05);
Christian Arnsperger: *The two meanings of "critical mass": Probing the new frontiers of economics in search of social emancipation* (2004-06).