Preprint: Phan D. (2009) "Other-dependent Preferences and Moral Sentiments in Cognitive Economics Models: First Steps Towards a More "Generalised" Rational Choice Model" - Accepted chapter for: M. Cherkaoui & P. Hamilton (Eds.) *Raymond Boudon: A Life in Sociology*, Oxford, The Bardwell Press

Other-dependent Preferences and Moral Sentiments in Cognitive Economics Models: First Steps Towards a More "Generalised" Rational Choice Model

Denis Phan[1]
GEMAS, Paris

Raymond Boudon's interest in individual rationality is closely related to the Weberian conception of understanding the individuals' motives — or reasons. From this point of view it is an important background in all his work. But since his paper in Rationality and Society (Boudon, 1989), he puts more directly the emphasis on the forms of individual rationality that can be used for an individual-based sociological theory of action. In his contribution of the 1993 special issue of International Studies in the Philosophy of Sciences devoted to this question, he criticizes the weakness of the two opposite explanatory conception of action in sociology: the collectivist — or structural — social causality and the Rational Choice Theory (RCT) advocated by Coleman (1990). He calls "synthetic" or "general" theory of rationality his proposed solution.

[1] I acknowledge sociologist Nathalie Bulle for relevant remarks and valuable suggestions on a preliminary version of this chapter; my partners in formalizing "other regarding" models, the physicists M.B. Gordon, J.P. Nadal and V. Semeshenko; the economist R. Waldeck for the emotional games, and M.Cherkaoui to provide stimulant intellectual background by allow me to work in the GEMAS. This work is part of the project ANR-08-SYSC-0008 (DyXi) supported by the *Programme Systèmes Complexes et Modélisation Mathématique* of the French *Agence Nationale pour la Recherche (ANR)*. I am a CNRS member. All errors and weakness remain my own.

A main feature of his program is to mix both cognitive and axiological dimension of rationality in order to encompass the RCT, but to go beyond its limitations. To go towards this direction of research, the first section of this chapter is devoted to the discussion of the six postulates identified by Boudon in the RCT from an economist's point of view, while the second and the third discuss the introduction of "other-regarding, other-dependent" and "emotional- moral" arguments into a very simple models of choice.

1. What kind of rationality do you need?

The new program of research on individual rationality in social context as been called by Boudon (1996) both: the "Generalised Rational (choice) Model" (GRM) and the "Cognitivist Model" (CM). It is synthesised in his book: *Raison, Bonnes Raisons*, published in 2003 (in French).

Chapter 1 is devoted to distinguish 6 postulates in the RCT model.

(P1) According to the *methodological individualism* paradigm, for Boudon all social phenomenon is the result of individual: "action, decision, attitude, behaviour, belief …" or ADABB.

(P2) The *understanding* of ADABB responsible of a social fact is a *momentum* of the sociological explanation (p. 20).

(P3) *Rationality*: if some actions (i.e. *habitus corporis*) can be explained without reasons, in the ADABB paradigm, actions are "principally the product of reasons" (p. 21). These reasons are for the individual "*good reasons*", and can be understood by the sociologist as such[2].

While these three postulates are the common basis of both CM and RCT approach, the later used three additional postulated that are as many restrictions with respect to a more general conception of rationality (i.e. CM). These postulates are

(P4) "*consequentialism*" or alternatively "instrumentalism" (p. 21);
(P5) "*selfishness"*
(P6) "*maximisation"*; an alternative to (P6) is satisficing (P6'), that allow the author to incorporate a Simonian world into the

[2] A question is how this "good reasons" are the real cause (ex ante) of a given action or (ex post) justification of this action? This question is compatible with non-material causality, but underlines the fact that the justification a posteriori sometimes differs from the effective a priori determinants of the action. This is always the case for Wittgensteiniens for which reasons rationalize actions, which is different from events causing other events. By opposite, for Davidson (1980), among various reasons to act, there is at least one that is the primary cause (or causal reason) of this action

picture, but both concern a particular modality of cost/advantage comparison.

Finally, Boudon claims that downward logical relations link these postulates. For instance P4 imply P1–P3 (but the converse is false).

This distinction of six postulates introduced by Boudon allows us to make clearer the epistemic debate on individual rationality. But the meaning of this postulates and the principle of a hierarchy of logical downward relation between its can be questioned. Neglecting the different forms of cost / advantage comparison (P6 or P6'), the main question concerns the hierarchical relationship between (P4) (P5) and (P6).

First of all, the full assimilation in P4 between consequentialism and instrumentalism is questionable from an economist's point of view. By definition of consequentialism, the consequences of an action outweigh other considerations: but what kind of consequences? for which one(s)? Then, the kind of consequentialism must be specified, this is for instance the case of the postulate (P5), that can be viewed as a form of ethical egoism and avoid the case of the ethical altruism form of consequentialism. Moreover, some confusion in the debate is introduced by a wide use by economists and RCT sociologists of utilitarism, according to which an action is right, if its maximizes utility, wrong, if it does not. In economics, the word "instrumentalism" has two meaning. The first is taking from epistemology (e.g. the instrumentalism of Friedman) and concerns the researcher's method. The second concern the way of modelling the agent's rationality, say the "instrumental rationality". It is defined by Walliser (1989) as referring to "the adequacy of pursued objectives to available means" — by opposition with cognitive rationality defined as "the adequacy of designed beliefs to available information". In that sense, instrumental rationality put the accent on the most efficient means to achieve a specific end (i.e. the cost/ advantage evaluation), but not necessarily on the value of that end by itself nor on the understanding of the reasons of the action [3]. Within this second perspective "instrumentalisation" is used to qualify the integration within instrumental rationality (e.g. cost/advantage evaluation) of new aspects of human decision-making process.

However, as used in conventional economics, instrumental

[3] The economist notion of "instrumental rationality" focus more on the efficiency relation between ends and means than the philosophico-social notion of "instrumental action" of Habermas, considering both the consequences of an action and the means to achieve it.

rationality generally operates as strong form of consequentialism, close to ethical egoism. In that way, utilitarism can be viewed as the opposite of deontological ethics, generally presented as putting the stress on the rightness or wrongness of intentions or motives (say: reasons) behind action such as respect for values and principles, rather than the consequences of action. For instance, for Weber, there is an "abysmal contrast" between the deontological "ethics of conviction" and the consequentialist "ethics of responsibility". The question is then about the difficulty to integrate ethical or moral behaviour and more consequentialist form of rationality within a single framework.

Some Scholar claims that consequentialist and deontological point of view are not necessarily mutually exclusive. In particular, some deontological principles can be justified with reference to their consequences; then, consequences are one of the important dimensions of moral evaluation. In a complex world it is often difficult to evaluate the multiples consequences of a particular action (including for oneself, then more generally for the others). It is then useful to make "reasonable conjectures"[4] to simplify this complexity. This seems to be coherent with Deontological principles can be viewed as useful guidelines for action, and deontological ethics as a weak form of consequentialism.

In a complementary way, Nozik (1974) invoke that he call "side constraints"; say, principles that interfere with pursuit of consequences. Side constraints could have priority over consequences. For an instrumentally rational approach, the question is: what kind of priority; absolute or conditioned; through some threshold? Accordingly some *axiologics* features could be intrumentalised under (P4). As example, in a contribution to the book on *social norms* edited by Boudon *et al.*, Wolfelsperger (2001) precisely underlines how the instrumental-rational models of economists can capture some critical features of moral sentiments and social mechanisms of rationality, both as constraints (*a la* Nozik) or as a goal in the preference function (i.e. a kind of mix of ethical egoism and ethical altruism). The works discussed hereafter follow this way. Then, a weak version of P(4) together with P(6) do not imply a "strong" version of P(5), and are compatible in some way with other regarding and/or pro-social determinants of action.

[4] For Boudon (1995): "Subjects try to solve complex questions by proposing reasonable conjectures. It is because these conjectures are reasonable that a great majority adopt them" (1995, translated from French).

Accordingly, since twenty years, an increasing number of economists are engaged in to integrate altruism, morality, pro-social behaviour and other forms of *other-regarding preferences* into an instrumentally rational theoretical framework. A challenge for theses approaches is to explain non selfish or other form of socially determined behaviour — like conformism — within an instrumental cost-advantage framework that can be viewed as only selfish in first instance, but including some instrumentalised aspects of ethical altruism — i.e. a weak version of (P5). However, in most if these models purely non-consequentialist moral values-driven behaviours are not taken into account or must be viewed as limit case. Consequently, these new approaches could be viewed by sociologists, just as a new quantitative and reductionist approach to axiological side of rationality. But results of these approaches open new ways and ask new questions that could be interesting for sociological concerns. If one can consider that the Coleman work is a way of get closer sociologists and economist's methodology and to improve dialogue between them, the works under considerations are — on different perspectives — getting closer the economists and the questions of interest of the sociological theory.

As Lévy-Garboua, Meidinger and Rapoport (2006) recall, the interest of economists for moral considerations and other-regarding other-dependent behaviour is not recent, since the Adam Smith's Theory of Moral Sentiments was published in 1759, almost twenty years before The Wealth of Nations (1776). But the formalizing of such other dimension regarding behaviour into the agent's preferences arises a little more than two hundred years after. The question of how social influence (i.e. externalities) affects individual choices and the resulting collective — possibly complex — issue, also known as the problem of the trade-off between "Individual Strategy and Social Structure'' (Young, 1998) is on the economists' agenda for at least two decades. Results of this research program are published in the Santa Fe Institute Studies in the Sciences of Complexity (Oxford University Press, see i.e. Arthur et al., 1997; Blume,. Durlauf, 2005) More on complex collective issues; the Economic Learning and Social Evolution collection (Brooking Institution & MIT Press; i.e. Durlauf and Young, 2001). The Roundtable in Behavioural Economics (Camerer 2003, Bowles, 2003) is more oriented by the series editor, Colin Camerer and Ernest Fehr towards experimental games and individual (behavioural) issues, but also towards evolutionary origins of behaviour and institutions. In France, similar research project is

know as "cognitive economics" (Bourgine and Nadal 2004, Walliser 2000, 2004).

This program includes in particular endogenous preferences[5] (Bowles, 1998) and, according to Wolfelsperger (2001), *other regarding dependent utilities* preferences can be viewed as the economist's way to model axiological rationality, moral sentiments or social mechanism of rationality. The Wolfelsperger's preferred approach of this question is the one of Rabin (1993), but many others follow the same way since this seminal work. This approach consist into intrumentalised morality, conformity (Bernheim 1994), and others pro-social behaviours (Gintis 2003).

Of course, from sociologist's point of view, following this way conducts to reduce drastically pro-social behaviours in many dimensions; but this allow economists to explore formally the *collective implications* of these other regarding and pro-social individual behaviours, when others disciplines do not. This is particularly useful in the cases summarized below, where very simple individual behaviours produce complex issues. In other words, in this matter, economists deal mainly with "conceptual exploration" (Hausman 1992). In the argumentative debate around a given empirical claim this allows economists to exhibit particular conditions of possibilities or counterfactual issues (i.e. counterexample) about *collective consequences* of the assumed form of individual behaviour and social structures. According to Sugden (2002), conceptual exploration cannot be viewed as a complete and definitive theory of the phenomenon, but as a partial and useful exercise, that provides a — if possible credible — candidate explanation to the related phenomenon, then, not "the" ultimate one. These models can be viewed as a tentative of (parsimonious) explanation of collective consequences of pro-social behaviour, but this does not means than they explain anything so far.

Anyway, such formalisms retain also the attention of few sociologists from Granovetter (1978) to Hedström et al. (2003). In the following, we summarize results of two closely related classes of models with heterogeneous agents (Kirman and Zimmermann 2001, Föllmer and Horst and Kirmann 2005). The first concerns social influence through positive externality with from the behaviour of the others, also know as "Neighbourhood Effects" and the second pro-

[5] On this question the conception of Becker himself has evolved between "De Gestibus Non Est Disputandu" (Stigler and Becker 1977) and recent works (such as: Becker 1996, Becker and Murphy 2000). See also Becker (1991)

social emotions.. In both cases, it is show that traditional dilemma of game theory (such as stag hunt or prisoner dilemma) can be viewed as particular case of richer interactive structure, when the agents are heterogeneous with respect to a particular idiosyncratic characteristic in their preferences, such as their willingness for to do (or not) a given action.

2. Other regarding dependent preferences, conformism and Social causality

The simplest form of "other regarding dependent" preferences is a linear model of individual's binary choice with social influence (externalities). The resulting collective behaviour is well known both by economists and sociologists since the seminal work of Schelling (1973, 1978), completed by the notion of individual and collective thresholds of adoption introduced later by the sociologist Granovetter (1978). In the example of a riot, used by Granovetter, agents have to choose between participating or not to a riot. Each agent has an individual threshold of adoption (a subjective cost for an economist), which corresponds to the number of people in the riot he considers to be sufficient to join the riot. In his simple example, Granovetter assumes a population of 100 people with a uniform (discrete) distribution of the thresholds. Each agent has an idiosyncratic threshold from 0 to 99; so that the thresholds are uniformly distributed and separate by the same increment, one.

The agent with threshold 0 is willing to participate at a riot even is any other does it, while the last agent having threshold 99 need that 99 other agents participate to do the same. The individual with threshold 0 is then the instigator of the dynamics, always adopting a deviant behaviour. As a consequence, the agent with threshold 1 joins him, because he need that there is at least one other to do it. In turn leads the agent with threshold 2 to modify his behaviour, and so on. Gradually, the riot grows by "domino effect" (or "avalanche" for the physicists) and reaches equilibrium where the whole population is affected. In above example, the distribution is uniform and the population is regularly distributed across discrete thresholds in the sense that there is one agent for each integer between 0 and 99. Then, the equilibrium avalanche size is equal to the total population. If the distribution of the thresholds is modified, the equilibrium may change. Let us suppose e.g. that the individual with the threshold 4 does not exist. Then, the dynamics is limited to the first four people with the thresholds 0 to 3. In this case of a finite size and discrete distribution, the revision of the situation of only one person in the

population may lead to a deep change in the global behaviour and exclude or at least reduce the size of avalanches.

As shown in the previous example, the collective behaviour of a population of interdependent agents is non trivial. The aggregate outcome may be characterized by multiple equilibria and complex dynamics with “tipping” and ”avalanches” (Phan, Pajot, 2006). As a result, in the deterministic context, the final equilibrium depends on the distribution of individual thresholds; the strength of social interactions, and in numerous cases with several equilibria, the selection of a particular equilibrium depends on the history of the collective dynamics.

In addition to Schelling and Granovetter, there are numerous other contributions to the modelling of social interactions that are related to this approach. Pioneering works in the *Journal of Mathematical Sociology* are Kindermann and Snell (1980), which identify social network as an application of random field; Galam et al. (1982) which identify the existence of a critical point in the neighbourhood of which the system’s behaviour is extremely sensitive to small changes in parameters as well as to the history of the system. Then, around this point, small microscopic changes in the initial conditions can lead to drastic changes at the macro level due to tipping effects. Others authors applied methods from statistical physics to sociodynamics. Weidlich and Haag, (1983) proposed a generic model of stochastic opinion formation with global social influence where social influence can lead durably the population within a situation where the majority of choices do not corresponds to the majority of idiosyncratic preferences of agents. This formalizes the idea that for pro-social reasons, agents in groups can have a different behaviour that the one they could have in isolation (this kind of approach will be reused further in economy, notably by Orleans (1998) for financial markets). To finish with sociological literature, further works appeared in the 1990s. For a survey see among others: (Marwell and Olivier 1993; Olivier and Marwell 2001), and more recently (Sampson, Morenoff and Gannon-Rowley 2002).

In economics, the real take-off for the models of individual choice with social influence was in the 1990s, in particular with Durlauf and coauthors (see among others Durlauf 1997, and 1999 for a synthesis dedicated to Social Sciences). The social dimension in the Beckerian tradition has been developed by Becker himself (Becker 1991; Becker and Murphy, 2000) and more formally by (Glaeser,

Sacerdote and Scheinkman 1996; Glaeser and Scheinkman, 2002). This approach has an empirical side that investigate econometrically he role of social — or "non market" — interactions in individual decision-making as well as on the resulting effects on the outcomes (see among others Manski 2000, Soetevent 2006.)

The model summarized here (hereafter GNP) has a partially a common structure with numerous of theses models, but it differs on some technical points, driving to specific results (discussed in particular in Phan, Semeshenko, 2009). Roughly, the main difference is in the *nature of the disorder* (e.g., the heterogeneity across agents and randomness). Many of these models belong to the class of Random Utility Models in the Luce (1959) tradition (agents have stochastic choices, but do not differ in their deterministic argument). By opposite, in the GNP model, agents have *heterogeneous idiosyncratic preferences*, without random argument. This means that agents are intrinsically - not accidentally - different, with respect to this point. The corresponding preference parameter of each agent is then called "Idiosyncratic Willingness to Adopt" (IWA). Across the population IWA is distributed according to a specific law, which could be identified empirically. Practically, for the simplicity of exposition, the result presented hereafter come from a bounded triangular law (Phan, Semeshenko, 2009); but it has been demonstrated by Gordon et al. (2007) that type of result is more generic, e.g. hold for a very large class of statistical distributions, including many bounded and unbounded law.

In the GNP model, agents have to make a generic binary choice under social influence; say "to adopt" ($\omega_i = 1$) or "not to adopt" ($\omega_i = 0$). Concretely, the choice may be to adopt or not a new technology, to vote for or against the new constitution, to buy or not a good, etc. But it can also to adopt (or not) a given posture, role or to conform to some behaviour.

According to Phan (2004) the agents' preferences are say to be both *intrinsically heterogeneous* (the agents have different idiosyncratic preferences) and *interactively heterogeneous* (others-regarding dependent part of preferences - e.g. surplus function- depend on the choices of their neighbours). The idiosyncratic preference parameter (IWA) of the "i" agent is denoted by H_i. All agents support the same cost of adoption denoted by C. This cost can be subjective (it may e.g. represent the subjective effort for to adopt a given behaviour) or objective, and represent for instance the price of one unit of a good to purchase. In this framework, an isolated agent adopts a particular

behaviour ($\omega_i = 1$) if his surplus of adoption, say ($H_i - C$) is positive.

The "social influence" or neighbourhood effect could be local (Phan 2007) or global depending on the neighbourhood structure under consideration. But this social influence is always assumed to have an additive impact on the surplus. This enforces a "conformity effect" (Bernheim 1994). Each agent is influenced by the (possibly expected) choices of his neighbours. The total social influence is a weighted sum of the neighbour's choices. In the simple case of homogeneous influence (of strength J for all agents and all neighbours) it is possible to write the social influence as a linear function of the rate of adoption in the neighbourhood of an agent i (η_i). Under the assumptions the surplus of this agent i if he adopted is: $(H_i - C) + J.\eta_i$. The total willingness to adopt: $H_i + J.\eta_i$ is an affine function of the rate of adoption within the neighbourhood (e.g. other regarding preferences). More people adopt in the neighbourhood, higher is the incentive to adopt for the agent.

Table 1. Payoff matrix for an agent i

	$\omega_n = 0$	$\omega_n = 1$
$\omega_i = 0$	0	0
$\omega_i = 1$	$H_i - C$	$H_i - C + J$

Player i in rows, field strategy in the neighbourhood (indexed n) in columns

In a *population game* framework, where each agent interacts at random with a given neighbour or against the field of neighbours, it is possible to represent the expected surplus of an agent as the payoff of game against an average player, representative of the neighbourhood of the agent (Table 1). Let define by ω_n the proportion of adoption in the neighbourhood. If nobody adopt in the neighbourhood, $\omega_n = 0$. The payoff of the agent i is then $(H_i - C)$ since $.\eta_i = 0$. If everybody adopt: $\omega_n = 1$; the payoff of the agent i is then $(Hi - C + J)$ since $\eta_i = 1$. If there is a proportion $1 > \eta_i > 0$ of adopters within the neighbourhood, the game works as a Bayesian Game (Harsanyi) and the (expected) payoff of agent is the weighted sum of these payoff (i.e. the mathematical expectation):
$(H_i - C)(1 - \eta_i)+(H_i - C+J).\eta_i$.

In the case of bounded distribution for IWA, the existence of the lower bound and the upper bound for the distribution may allow to identify the behaviour in the population in spite of heterogeneous IWAs. By dominance analysis (elimination of the dominated strategies) it is possible to identify the classes of agents with the same behaviour. For Instance if for all i: $(H_i - C) > 0$ everybody

always adopt, if $H_i < (C + J)$, nobody never adopt, and it is possible to identify in the same manner typical behaviours for intermediary cases ($1 > \eta_i > 0$), where adoption is conditional — e.g. depends on the rate of adopters in the neighbourhood

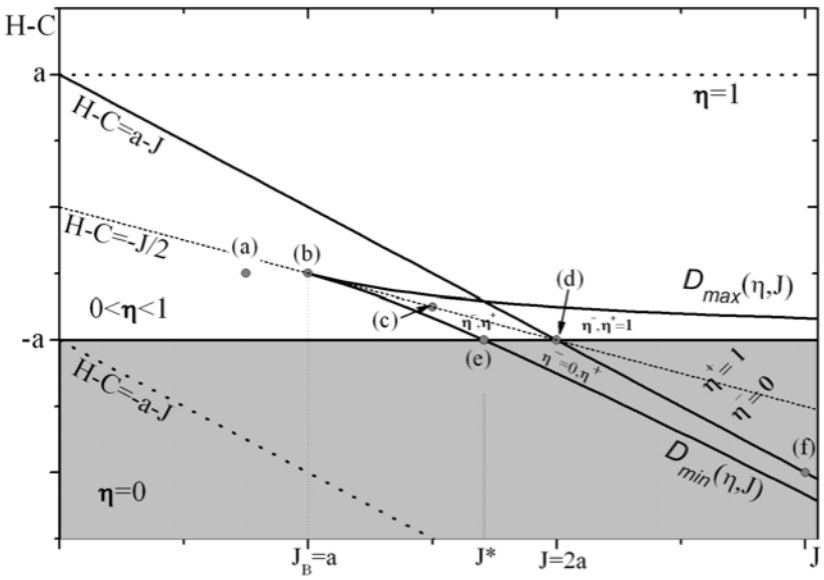


Figure 1. *Equilibrium regimes for symmetric triangular distribution on the interval [–a, a], with a = 2 in the plane (J, H − C)* .
Source: Phan, Shemeshenko, 2009

Accordingly, general results from (Gordon *et al.* 2007) can be graphically summarized on a *phase diagram* in the plane (J, H − C), where H is the average IWA. This phase diagram exhibits different regions of the parameters space corresponding to different regimes of Nash equilibria for the game. Then, abscissa represents the reward of the decision when there is no social effect e.g. the average surplus is: H – C. Ordinate represent the strength of social effect J, e.g. the other-regarding related reward of the decision. Greater is J, lower is the weight of idiosyncrasy in the decision, and greater is the social conformism. In Figure 1 the heterogeneity of the IWA is distributed according to a bounded, triangular-symmetric law on: [–a, a] (Phan, Shemeshenko, 2009), but others bounded distribution product similar result, with limited degenerate case when the distribution is uniform (see i.e. Gordon *et al.* 2005).

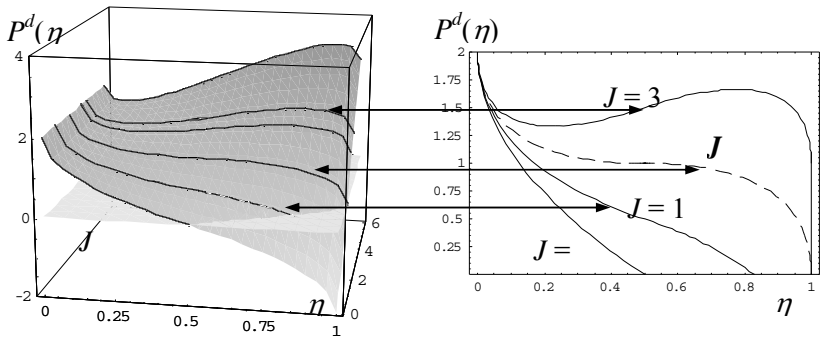
For high values of (H – C) > (a – J), in the northeast region (possibly low cost or/and sufficient enough strength of social influence), everybody adopts and η = 1. Conversely, in the south-

west zone, for weak social influence and low values of $H - C < -a$ (possibly high cost, or alternatively high rejection of the adoption), nobody adopts and $\eta = 0$. Analytical examination detailed in (Phan, Shemeshenko, 2009) allows us to identify a new region with two Nash equilibria between the curves $D_{min}(\eta, J)$ and $D_{max}(\eta, J)$. In the eastern region with:
$-a < (H - C) < (a - J)$ and $J<J_B$, for a sufficiently "moderate" social effect, there is a single *hybrid Nash equilibrium*, with both non-adopters and adopters coexist in proportion $0 < \eta < 1$. For $J > J_B$ there is a *multiple solutions region* delimited by the two frontiers $D_{min}(\eta, J)$ and $D_{max}(\eta, J)$ that merge at the singular point; $J_B = a$. In this region if: $D_{max}(\eta, J) > (H - C) > D_{min}(\eta,J)$, there are two stable Nash equilibria: one with a rate of adoption less than 50% (possibly 0%) and another with more than 50% (possibly 100%). Accordingly, the single *hybrid* Nash equilibrium zone has two extensions for $J_B < J < J^*$, one with $0 < \eta^- < 0.5$, if :
$D_{max}(\eta, J) > H - C > -a$; and another one:
$0.5 < \eta+ < 1$ if: $(a - J) > (H-C) > D_{min}(\eta,J)$, respectively.

Finally, a *coordination region* with two equilibrium: everybody adopts ($\eta = 1$) or nobody adopts ($\eta = 0$) is always included under the horizontal line: $(-a)$, and over the diagonal line: $H - C = a - J$.

Within this coordination zone, A "stag hunt" zone is below the dashed line, for $-J/2 \geq (H - C)$. For philosopher Brian Skirms (2004) this game configuration, fond in Rousseau's *Discourse* game configuration, fond in Rousseau's *Discourse on Inequality* is prototypal of many social dilemmas. Anyway, the introduction of heterogeneity shows that these paradigmatic stories in game theory do not cover all the field of possible regimes when individuals' decisions are others-regarding. Moreover, when the statistical distribution of IWA tends to become unbounded (that could means for instance that some individuals do not want adopt the considered behaviour - *a aucun prix*. In this case, the lower bound: $-a$, tends to negative infinity.

Accordingly, pure classical game theoretic configurations (such as Stag Hunt) tend to dis- appears from the empirically relevant region of the phase diagram and only mixed solution remains. Among many others consequences, this framework allows us to found on the diversity of interacting micro-behaviours (with social influence and heterogeneity of individual preferences) some specific results on demand for goods or services with consumption externality.



(a) $P^d(\eta, J)$ for $0 \leq J \leq 5$ (b) $P^d(\eta, J)$ for $J \in \{0,1\ 2,3\}$

Figure 2: *"Classical" downward-slopping demand curve as particular case for weak J*

Figure 2 (a) & (b) presents two views of an inverse demand function: $P^d(\eta, J)$ calculated from the model previously exposed, where: C = P, the market price. For: J > 2 figure 2 (b) is reminiscent to the article on "restaurant pricing and other examples of social influences on price", by Becker (1991, figure 3) However, Becker does not explicit how such a curve can be constructed, while the GNP model provides individually based conditions under which such construction is possible, then encompass the "traditional theory of demand as sub case. For $J < J_B$, the strength of the social influence is weak and we have a "classical" downward-slopping demand curve, according. Over $J = J_B$ (dashed line) the social influence is sufficiently strong to produce an upward-slopping segment. In this case, and for a given price, we have two stable candidates to be market equilibrium, then very specific properties for the corresponding market. (For a discussion in the case of a monopolist market, see i.e. Nadal et al., 2005; Gordon et al., 2005, 2009)

3. Pro-Social emotions and variety of strategic interactions

In everyday life and across all societies human individuals globally follow collective norms in order to enforce cooperation (see i.e. Henrich et al. 2003). These observations contrast with theoretical claims of Olson about *collective action* (1965): This author claims that rational selfishness drives to full defection in the public good dilemma (within "large" *unorganized* groups). By opposite, experimental economics shows that the tendency to conform to

norms also differs across individuals within experimental (e.g; unorganized) group (Fehr and Fischbacher 2004). Basic empirical observations on many non-profit making, political and other forms of benevolent-based organizations committed into collective action show a rough partition within these medium-sized and "open" groups between a little set of highly active benevolent individuals and a majority of less (often poorly) active individuals.

In economic terminology, collective action results in the production of some public good, and the more active benevolent act as strong cooperators, while others are more or less pure consumers of the public good, usually called *free riders* in game-theoretic literature. These heterogeneous behaviours are not in accordance with the standard selfish rational model of choices, but it can be observed for many forms of organization and for a variety of situations. In addition, both individual behaviours (heterogeneity) and proportion of strong cooperators seems to be contingent and differ with respect to institutional and cultural factor such as: law-abiding behaviour, customs and popularity of moral principles and others forms of "others-regarding" behaviour. Anyway, this kind of heterogeneous configuration seems to be (contingently) stable in many circumstances, even if individuals can change their behaviour. In particular, many individuals do not definitely act as co-operator or free riders, but change of their behaviour through time or for a given circumstance. For example, an individual can be an active political militant in a period of his life and just a simple supporter in another period. There is well known also that both categories growth in period of election: the *proportion* of active militant within a political organisation, and the number of members as well. Same observation stands for the activism in case of social emergency for charity organisations.

This suggests that the diversity of type of behaviour could be a metastable situation in a variety of circumstance (i.e. prevail mainly in the long run). This diversity can evolve in the middle run, but often remain stable in the short run. Neglecting in first approach these stability questions for to focus on the possible existence of equilibrium, the intention of (Gordon et al. 2005) was to provide a (game-theoretical framework compatible) model of individual binary choice in social context for to relate social regularity to this heterogeneity of individual attitudes. As in the previously reviewed model of GNP, idiosyncratic preferences (ie. individually distributed heterogeneity) do not vary with circumstances. Following Rabin (1993), an other-regarding, other-dependent argument in the

preference function is a tentative to include some aspects of fairness or altruism into an instrumental optimising style of framework (P4) and (P6). As result, two identical agents (with respect to their idiosyncratic type) may be heterogeneous in their pro-social behaviour, with respect to the behaviour of the others in their environment (e.g. their neighbourhood, as previously). Later works investigate respectively the stability question (Ma et al., 2006) and the behavioural game foundation of agent's decision (Phan, Waldeck, 2008). In the following, we focus on the later, as a way to formalize the "good reasons" that individuals could have to deviate from the standard selfishness linked with (P5) postulate used, according to Boudon, by the standard version of RCT.

Before to present the mains result of (Gordon et al. 2005; Phan and Waldeck, 2008), let us turn towards literature that advocates this approach. Since Olson (1965), a large body of literature tries to explain why cooperation can exist among pure selfish individuals (i.e. within (P4)–(P6) postulates). The usual selfishness approach explains cooperation by the expected future reward in repeated game (see i.e. Axelrod, Hamilton 1981, Novak Sigmund 1998 among others). However, this explanation does not works for anonymous one-shot interactions, and the large experimental program of (Henrich et al. 2003) shows that cooperation occurs also in these situations. The idea of sanction to apply to the free riders in the public good game is a complementary — but possibly alternative — plausible explanation. The case of enforcement by the way of external sanction by a third party is well known in sociological RCT, as largely discussed in Coleman (1990). Experiments by Fehr and co-authors show that if individuals have punishment opportunities some of them have pro-social behaviour and are willing to support the cost of punishment even without monetary rewards from doing so (Fehr and Gachter 2000, 2002, Fehr and Fischbacher 2004).

One other way (complementary, but sometimes alternative) is to consider pro-social emotions (Bowles and Gintis 2005) and then weaken the (P5) postulate. Pro-social emotions like shame, guilt, empathy, work like an internal self-punishment when people deviate from the social norm. In the case of pro-social emotions, the value of conformance to the norm triggers the strength of these emotions. In our case, the appropriate behaviour is defined as the socially desirable outcome for the group. Accordingly, in Phan, Waldeck, (2008), emotions will thus be generated from deviations from this desirable outcome and be proportional to the proportion of

cooperators. This assumption is in accordance with experimental results by Gächter and Fehr (1999), showing that emotionally prompted social approval or disapproval and peer pressure on defection leads to overcoming free-rider incentives in collective action. This provides empirical evidence and basis for formal models of social incentives within group such as in Akerlof (1980) among others.

In (Gordon et al. 2005), agents have a simple linear preference function with 4 arguments, and take two decisions. The first decision is to be member of a given group ($\omega_i = 1$) or not ($_{\omega i} = 0$). For members of the group, the second is to cooperate ($s_i = 1$) or to be a free rider ($s_i = 0$). Let us denote by H_i the idiosyncratic valuation of the group by agent i. This parameter is variable between individuals, going from misanthropic (strongly negative H_i) to highly socialized persons (strongly positive H_i). Between both extremes, $H_i < 0$ means that participate to the group represents a cost, but this later can be also interpreted as the value of an outside opportunity. For those people who choose to be within the group, the second decision is to cooperate or not. The second argument is then G, the social reward from cooperation (supposed to be the same for all members of the group). It is weighted by η_c, the proportion of cooperators within the group. The higher is G, the stronger should be the incentive to enter the group, to stay within and to cooperate. The third and fourths arguments are the costs of the individual behaviour with respect to the cooperation. The former is the cost C supported by cooperators for the time spent and the efforts involved in cooperation. The Later is viewed by (Phan, Waldeck 2008) as an (assumed idiosyncratic) emotional moral cost, linked to the concept of pro-social emotions (Bowles and Gintis 2005). This cost is borne by an individual (a free rider), when he unilaterally deviates from a socially desirable norm of behaviour. Then, free riders bear this cost, which is assumed to be an idiosyncratic argument X_i weighted by the fraction of cooperators within the group. One can suppose that the group members obtain a higher distribution of moral costs within the group when the behaviour is more observable by the group members (i.e. less privacy and anonymity) or if there is a lower social distance among the group members.

The resulting preference function, is: $(H_i + G.\eta_c) - (s_i.C + (1 - s_i).X_i\ \eta_c)$; to be compared to those of section 1. In the following discussion, we focus on the effect of the emotional moral cost alone and restrict to the simple case of identical $H_i = H$ for all agents. In this case, the surplus among member of the group can be rewritten as: $H + G.\eta_c - C$

for a cooperator and as $H + (G - X_i).\eta_c$ for a free rider. As in section (1) if nobody cooperate in the neighbourhood ($s_n = \eta_c = 0$), the payoff is H for a free rider, and H – C for a cooperator. If everybody cooperate in his neighbourhood ($s_n = \eta_c = 1$), a free rider benefits from the social reward G, but support an idiosyncratic moral cost for pro-social emotions say X_i.

Table 2: Payoff matrix for an agent i

	$s_n = 0$	$s_n = 1$
$s_i = 0$	H	$H+G-X_i$
$s_i = 1$	H − C	H +G- C

Player i in rows, field strategy in the neighbourhood - indexed n - in columns

Accordingly, the payoff for a free rider is $(H + G - X_i)$ By opposite, a cooperator supports the cost of cooperation, and the corresponding payoff is (H + G – C). If the proportion of cooperators in the neighbourhood is $1 > \eta_c > 0$, the expected payoff is the weighted sum of both, and this game has the following Bayesian payoff matrix (Table 2).

As previously, this discrete choice population game model can be fruitfully compared to standard game theory when the values of the idiosyncratic emotional argument X_i are spread over a bounded support $[X_{min}, X_{max}]$. As a consequence, bilateral games are asymmetric with respect to the payoffs: all the players have different preferences over the same strategic set. Because agents play against the field, it is nevertheless possible to identify in the phase-diagram some sub-domains where symmetric games concepts and results apply. Then, if all the agents play the same strategy, we can analyse the structure of best response despite the heterogeneity of the individual payoffs, and relate some of our results to well known concepts in a symmetric population games. In the game corresponding to this model, we identify the following parameter configurations (Figure 3), where all parameters are normalized by: $d \equiv (X_{max} - X_{min})$; then: $x_i = X_i/d$; $c = C/d$ (and so on…).

1. If $c > x_{max} = x_{min} + 1$ then all agents have $xi \leq c$ and are *intrinsic free-riders* for all values of η. This zone is located in the northwest of the phase-diagram above the line $c = x_{min} + 1$. The only one equilibrium is $s_i = 0$ for all agents, that is, $\eta_c = 1$. The corresponding game is a "classic" *prisoner's dilemma*

2. If $c < x_{min}$, the best reply depends on the choice of all the population. This choice is homogeneous, despite the difference in the

payoffs that are all positive. The game belongs to the class of *coordination games*. We obtain two Nash equilibria, $s_i = 0$ for all i or $s_i = 1$ for all i. Depending on the value of c we identify several regions, including a sub-region where the coordination game belongs the class of the so called: "*stag hunt*" game, mentioned previously (for $x_{max}/2 < c < x_{min}$).

3. In the intermediate zone: $x_{min} < c \leq x_{max}$, the equilibrium strategy of agents is heterogeneous with respect to the dominance structure of their payoff matrix. Some of them are *intrinsic free riders*, while others as not. The relative weight of intrinsic free riders is given by the relative values of the parameters x_{min} and c. Inside this region with heterogeneous dominance structure across agents, one of the two Nash equilibria admit heterogeneous (mixed) behaviour and two different strategies (cooperation and free riding) coexist (tri-angular grey zone, in the south west). An interesting point is that the observed behaviour correspond generally to this small zone. Complementary statistical investigations have to be developed to assess this point (Manzo, 2007).

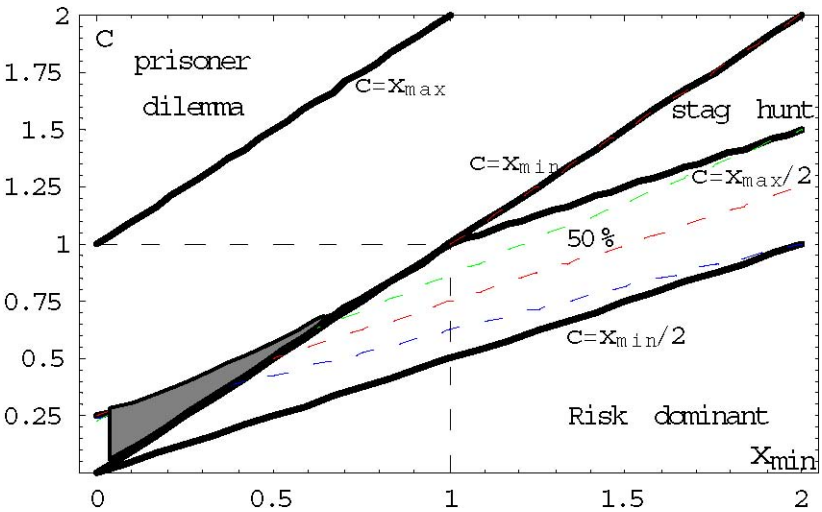


Figure 3. *Heterogeneity of pro-social sentiments and regimes of game theory - Source: Gordon et al. (2005)*

To summarize, the phase diagram in Figure 3 presents the equilibrium regimes of this "emotional" game. Depending on the parameters these may correspond to full cooperation, pure free riding, or coexistence of two different strategies (cooperation and free riding). The existence of the latter, which corresponds at numerous empirical evidences, is our most interesting result. It is a

direct consequence of the heterogeneity of the idiosyncratic characteristic X_i, that can be interpreted as the agent's emotional reaction to free riding (Phan Waldeck 2008) or by the expectation of some punishments inflicted by the cooperators to free-riders in a Colemanian' approach.

4. Conclusion

For a half secular, Boudon has make a significant effort into promote usage of mathematics in sociology (Boudon, 1967, 1986), to take into account the complex nature of the *composition effects* in population of interacting individuals, e.g. the *unintended consequences of actions* (see Cherkaoui, 2007) — or non intentional collective consequences of intentional individual action (Boudon 1977, 1979) and to account for a more general notion of rationality, as underlined in introduction. As editor of the collection "sociologies" he promoted the translation into French of seminal works in the field of interest of this chapter, such as Schelling (1978) and Olson (1965), and more recently many works in the field of a cognitive rationality, such as the Livet book on emotions (2002) among others. To formalize some of these questions into the mathematical framework of an extended approach of game theory make us able to identify and proposed candidate explanation for many social dilemma, such as prisoner dilemma or stag hunt, as particular sub-case of a more general social landscape taking into account both heterogeneity of preferences and strength of social influence. Then, some limitations of "classical" RCT can be encompassing by formalizing other regarding dependent preferences and moral attitudes, even if these models postulate and not understand / explain the individual behaviour as such.

5. References

Akerlof G. (1980) "A Theory of social custom, of which unemployment may be one consequence", *Quarterly Journal of Economics*, 94: 749–775.

Arthur B., Durlauf S., Lane D. (eds) (1997) *The Economy as an Evolving Complex System II*, Santa Fe Institute Studies in the Sciences of Complexity, Volume XXVII, Addison-Wesley.

Axelrod R. Hamilton W.D. (1981) "The evolution of cooperation". *Science*, 211: 1390–1396.

Becker G.S. (1991) "A Note on restaurant pricing and other examples of social influences on price", *Journal of Political Economy*, 99(5): 1109–1116.

Becker G.S. (1996). Accounting for Tastes. Harvard University Press.

Becker G.S. Murphy K.M. (2000) *Social Economics. Market Behavior in a Social Environment*, Cambridge MA, The Belknap Press & Harvard

University Press 2000.
Bernheim B.D. (1994) ”A theory of conformity”, *Journal of Political Economy*, vol.102/5: 841–877.
Boudon R. (1967) *L'analyse mathématique des faits sociaux*, Paris, Plon.
Boudon R. (1977) *Effets pervers et ordre social*, Paris, PUF (Reprinted, “Quadrige”, 1993).
Boudon R. (1979) *La Logique du social*, Paris, Hachette, (Reprinted, Pluriel, 1983).
Boudon R. (1986) “Mathematical and statistical thinking in the social sciences”. In: Deutsch K., Markovits A. and Platt J. (eds), *Advances in the Social Sciences, 1900– 1980: What, Who, Where, How?* New York, University Press of America.
Boudon R. (1989) “Subjective rationality and the explanation of social behavior”, *Rationality and Society*, 1(2): 173–196.
Boudon R. (1993) “Towards a synthetic theory of rationality”, *International Studies in the Philosophy of Sciences*, 7(1): 5–19.
Boudon R. (1993) “More on ‘Good Reasons’: reply to critics”, *International Studies in the Philosophy of Sciences*, 7(1): 87–102.
Boudon R. (1995) *Le juste et le vrai. Etudes sur l'objectivite des valeurs et de la connaissance*, Paris, Fayard.
Boudon R. (1996), “The ‘rational choice model’: a particular case of the ‘cognitivist model’”, *Rationality and Society*, 8(2): 123–150.
Boudon R. (2003) *Raison, Bonnes Raisons*. Paris, Presses Universitaires de France, 2003.
Bourgine P., Nadal J.P. Eds (2004) *Cognitive Economics, An interdisciplinary Approach*, Berlin, Springer.
Bowles S. (1998) “Endogenous preferences: the cultural consequences of markets and other economic institutions”, *Journal of Economic Literature*, 36: 75– 111.
Bowles S. (2003) *Microeconomics, Behavior, Institutions and Evolution*, Princeton, Princeton University Press.
Bowles, S., Gintis H. (2005) “Prosocial emotions”. In: Blume L. Durlauf S. (eds), *Economy as a Complex Evolving System, III; Current Perspectives and Future Directions*. Col. Santa Fe Institute Studies in the Sciences of Complexity, Oxford University Press: 339–366.
Camerer C.F. (2003) *Behavioral Game Theory , Experiments in Strategic Interaction*, Princeton, Princeton University Press.
Cherkaoui M. (2007) *Good Intentions: Max Weber and the Paradox of Unintended Consequences*, Oxford, The Bardwell Press.
Coleman J.S. (1990) *Foundations of Social Theory*, Cambridge: Harvard University Press.
Davidson D. (1980), *Essays on Actions and Events*, Oxford, Oxford University Press.
Durlauf S.N. (1997) “Statistical Mechanics Approaches to Socioeconomic Behavior”. In: Arthur et al. op. cit.: 81–104.
Durlauf S.N. (1999) “How can statistical mechanics contribute to social

science?", *Proceedings of the National Academy of Sciences*.
Fehr E., Fischbacher U. (2004a) "Third-party punishment and social norms", *Evolution and Human Behavior*, 25: 63–87.
Fehr E., Fischbacher U. (2004b) "Social norms and human cooperation". *Trends in Cognitive Sciences*, 8(4): 185–190.
Fehr E., Gächter S. (2000) "Cooperation and punishment in public good experiments", *American Economic Review*, 90: 980–994.
Fehr E., Gächter S. (2002) "Altruistic punishment in humans", *Nature*, 415: 137–140.
Gächter, S., Fehr, E. (1999) "Collective action as a social exchange", *Journal of Economic Behavior and Organization,* 39: 341–369.
Galam S., Gefen Y., Shapir Y. (1982) "Sociophysics: a mean behavior model for the process of strike", *Mathematical Journal of Sociology*, 9: 1–13.
Gintis H. (2003) "Solving the puzzle of prosociality.", *Rationality and Society*, 15(2): 155–187.
Gordon M.B., Nadal J.P., Phan D., Vannimenus J. (2005) "Seller's dilemma due to social interactions between customers", *Physica A* , 356(2–4): 628–640.
Gordon M.B., Nadal J.P., Phan D., Semeshenko V. (2007) "Discrete choices under social influence, generic properties", HAL-SHS 07/03/2007, accepted in *Annual Reviews on Complex Systems — Modelling and Mathematical Methods* in the MSSET Series. Forthcoming in 2009.
Gordon M.B., Phan D., Waldeck R., Nadal J.P. (2005) "Cooperation and free riding with moral cost". In: Kokinov Boicho (ed.), *Advances in Cognitive Economics*, Proceedings of International Conference on Cognitive Economics (ICCE), Sofia, NBU Press: 294–304.
Granovetter M. (1978) "Threshold models of collective behavior", *American Journal of Sociology*, 83(6): 1360–1380.
Harsanyi J. (1977) Morality and the theory of rational behavior". *Social Research*, 44: 623–656.
Hausman D.M. (1992) *The Inexact and Separate Science of Economics*, Cambridge MA, Cambridge University Press.
Hedström P, Kolm A-S, Åberg Y. (2003) "Social interactions and unemployment", Working Paper Series November *2003*:15, Uppsala University, Department of Economics.
Henrich J., Boyd R., Bowles S., Camerer C., Fehr E., Gintis H. (2003) *Foundations of Human Sociality: Ethnography and Experiments in Fifteen Small-scale Societies*, Oxford, Oxford University Press.
Kindermann R., Snell J.L. (1980) "On the relation between Markov random fields and social networks", *Journal of Mathematical Sociology*, 7: 1–13.
Kirman A.P., Zimmermann J.B. Eds. (2001) *Economics with Heterogenous Interacting Agents*. Heidelberg, Springer Verlag.
Föllmer H., Horst U. Kirmann A.P. (2005) "Equilibria in financial

markets with heterogeneous agents: a probabilistic perspective", *Journal of Mathematical Economics*, 41(1–2): 123–155.

Lévy-Garboua L., Meidinger C. Rapoport B. (2006) "The formation of social preferences lessons from psychology and biology". In: S.C. Kolm and J. Mercier-Ythier (Eds), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Volume I. Elsevier B.V.: 545–613.

Ledyard J.O. (1995) "Public Goods: a survey of experimental research". In: Hagel, Roth (Eds), *Handbook of Experimental Economics*, Princeton, Princeton University Press: 111–194.

Livet P. (2002) *Emotions et rationalité morale*, Paris, Presses Universitaires de France.

Luce R.D. (1959) *Individual Choice Behavior, A Theoretical Analysis*, New York, Wiley.

Ma Y., Gordon M.B., Nadal J.P. (2006) "Dynamics of social systems: cooperation and free-riding", *Physics of Socio-Economic Systems*. AKSOE, Dresden, 27–30 March 2006.

Manski C.F. (2000) "Economic analysis of social interactions", *Journal of Economic Perspectives*, 14(3): 115–136.

Manzo G. (2007) "Variables, mechanisms, and simulations: can the three methods be synthesized? A critical analysis of the literature", *Revue Française de Sociologie*, Annual English Selection, 48 (Supplement): 35–71.

Marwell G., Olivier P.E. (1993) *The Critical Mass in Collective Action, a Micro-social Theory*, Cambridge, Cambridge University Press.

Novak M., Sigmund K. (1998) "The dynamics of indirect reciprocity", *Journal of theoretical Biology*, 194: 561–574.

Nadal J.P., Phan D., Gordon M.B., Vannimenus J. (2005) "Multiple equilibria in a monopoly market with heterogeneous agents and externalities", *Quantitative Finance*, 5(6): 1–12.

Nozick R. (1974) *Anarchy, State and Utopia*, Oxford, Blackwell.

Olivier P.E., Marwell, G. (2001) "Whatever happened to critical mass theory? a retrospective and assessment", *Sociological Theory*, 19(3): 293–311.

Olson M. (1965) *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge, MA, Harvard University Press.

Phan D. (2007) "Heterogeneous Agents with Local Social Influence Networks: Path Dependence and Plurality of Equilibria in the ACE Noiseless case". In: Consiglio, A. (ed.), *Artificial Markets Modeling Methods and Applications*, Lecture Notes in Economics and Mathematical Systems, Vol. 599, Berlin, Spinger: 181–194.

Phan D., Pajot S. (2006) "Complex behaviours in binary choice models with global or local social influence". In: Bruun C. (ed.), *Advances in Artificial Economics The Economy as a Complex Dynamic System*, Series: Lecture Notes in Economics and Mathematical Systems , Vol. 584, Berlin, Spinger: 203–219.

Phan D., Semeshenko V. (2009) "Equilibria in models of binary choice

with heterogeneous agents and social influence". MASHS 2007 Computational Methods for Modelling and learning in Social and Human Sciences Brest, 10–11 May 2007. Revised version forthcoming in *European Journal of Economic and Social Systems*.

Phan D., Waldeck R. (2008) "Social norms, Emotions and cooperation in groups". To be presented at: CEF '08, 14th International Conference on Computing in Economics and Finance, 26–28 June 2008, Sorbonne, Paris & at ESHIA/WEHIA 2008, 9–21 June 2008, Warsaw University of Technology, Warsaw, Poland.

Rabin M. (1993) "Incorporating fairness into game theory and economics", *American Economic Review*, 83/5: 1281–1302.

Sampson R., Morenoff J., Gannon-Rowley T. (2002) "Assessing 'neighborhood effects': social process and new directions in research", *Annual Review of Sociology*, 28: 443–478.

Schelling T.C. (1973) "Hockey helmets, concealed weapons and daylight saving a study of binary choices with externalities", *Journal of Conflicts Resolution*, 17(3): 382–428.

Schelling T.C. *(1978) Micromotives and Macrobehavior*. New York, Norton and Company.

Skirms B. (2004) *The Stag Hunt and the Evolution of Social Structure*, Cambridge, Cambridge University Press.

Stigler, G. J. Becker, G. S. (1977) "De gestibus non est disputandum", *American Economic Review*, 67(2): 76–90.

Soetevent A.R. (2006) "Empirics of the identification of social interactions; an evaluation of the approaches and their results". *Journal of Economic Surveys*, 20(2): 193–228.

Sugden R. (2002) "Credible worlds: the status of theoretical models in economics". In Mäki, U. (ed.), *Fact and Fiction in Economics. Realism, Models, and Social Construction*, Cambridge, Cambridge University Press: 107–136.

Young P. (1998) *Individual Strategy and Social Structure*, Princeton, Princeton University Press.

Walliser B. (1989) "Instrumental rationality and cognitive rationality", *Theory and Decision*, 27: 7–36.

Walliser B. (2000) L'économie cognitive Paris, Odile Jacob.

Walliser B. (2004) "Topics on Cognitive Economics". In: Bourgine, Nadal. (eds), op cit.: 183–198.

Weidlich W., Haag G., (1983) *Concepts and Models of a Quantitative Sociology, the Dynamics of Interacting Populations*, Berlin, Springer Verlag.

Wolfelsperger A. (2001) "La modélisation économique de la rationalité axiologique. Des sentiments moraux aux mécanismes sociaux de la moralité". In. Boudon, Demeulenaere, Viale (Eds), *L'explication des normes sociales*, Coll. Sociologies, Paris, Presses Universitaires de France.