

HOW SOCIAL MECHANISMS MATTER FOR SOCIAL JUSTICE

José A. Noguera
Department of Sociology
Group of Analytical Sociology and Institutional Design
Universitat Autònoma de Barcelona
08193 Cerdanyola (Barcelona), SPAIN
jose.noguera@uab.es
<http://gsadi.uab.cat>

Very provisional draft: comments are most welcome

Abstract

One of the main issues in non-ideal theories of social justice has to do with how to apply abstract normative principles in real world circumstances, and what kind of institutional designs are more adequate for that end. This paper aims to show how the ‘social mechanisms approach’ (developed by some social scientists as Elster, Hedström, or Boudon) may be of great help in order to examine carefully the causal assumptions involved in some ideal normative theories, as well as the causal microfoundations that make possible (or not) its application in non-ideal circumstances.

The first section will list and define some useful criteria in order to evaluate the degree of applicability of a normative theory in terms of concrete institutional designs. Secondly, it will be argued that the ‘social mechanisms’ approach provides the appropriate ‘causal grammar’ to account scientifically for the nature of the links between a given institutional design or policy and its results in a particular social context. The third section will show how that approach may give answers to many of the questions posed by the criteria of applicability defined in section one. Section four will try to illustrate this with some examples in the field of tax and transfer policies. The final section will briefly list some problems that the social mechanisms approach may raise from a normative point of view.

0. Introduction: sociology, institutional design, and non-ideal normative theory

There are few who would ignore or deny the relevance of recent developments in the theory of institutional design for political philosophy and non-ideal theories of justice. However, and surprisingly, none of these two fields of academic scholarship has today a close relationship with sociology and sociological theory (no doubt, these latter disciplines are also to blame for that situation). In this paper I will argue for such a relationship, and I will focus particularly in the usefulness for non-ideal theories of justice of a theoretical and methodological orientation in sociology, *analytical sociology*, which has raised great interest in recent years.

The distinction between ideal and non-ideal theory in political philosophy may initially seem a sharp one, but to my view there are usually two analytically different issues merged in it: first, the distinction between ideal (in the sense of *theoretical*) reasoning about justice and the problems of the *application* of normative ideals; and, second, the distinction between the application of those ideals in *idealized circumstances*, and its application in non-idealized ones, like those of the real world. These two distinctions point to two conceptually different problems: on the one side, the *application* or implementation problem, which poses issues of *institutional design*, and on the other side, the *contextual* problem, which emerges from the contextual and specific nature of any concrete situation to which ideals of justice may be applied (be that situation real or not). The first issue has to do with which kind of institutional design captures best a given normative principle, while the second refers to how that principle may be modified or relaxed when confronted with different real circumstances. Although most of the problems we usually face in non-ideal theory have to do with both issues, that is, with *application in 'non-ideal-because-of-real' circumstances*, we may also face problems of application in ideal circumstances (for example in some thought or real experiments), and problems in adapting ideal theory to cope with some contextual circumstances, even if we lack a concrete institutional design to apply it.

My focus here will be with application problems more than with contextual circumstances (although most of the issues that I will discuss may be also of interest for

the contextual problem). The reason is that the application problem is (or so I believe) more directly related with the well-known rule in non-ideal theory which says that ‘ought implies can’. I agree with Brighouse (2004:26-28), and disagree with Cohen (2003), that ‘ought implies can’ is a good methodological principle for a theory of justice (and not just for ‘regulative principles’ as Cohen argues). But that means that a good knowledge of sociology, and specifically of analytical sociology, which tries to uncover causal mechanisms that explain social phenomena, is essential for the normative theorist. What I claim is that ‘ought implies can’ is in fact a mandate for the non-ideal theorist of justice to become a putative social scientist, and that the best approach available today in the social sciences is the ‘social mechanisms approach’ advocated by analytical sociologists. Normative theories involve causal assumptions and hypothesis, and social mechanisms are the ‘grammar’ of causality in the social world which allow us to test the plausibility of those assumptions and hypothesis.

In a very general way, I see at least five reasons to place the study of social mechanisms (henceforth, SM) at the core of non-ideal theories of justice:¹

1. SM may explain the factual assumptions implied by normative principles.
2. SM are necessary to test and explain the degree of applicability of normative principles.
3. SM are necessary for the comparative evaluation of different institutional designs which try to apply normative principles.
4. SM underlie the formation of normative intuitions and feelings of justice (Boudon 2003; Konow, 2003; Frölich & Oppenheimer, 1992).
5. SM may suggest how to bring about the social conditions to apply ideal theory in non-ideal circumstances; they may help us to ‘push’ those circumstances closer to ideal ones, or to make more irrelevant their non-ideal character.

¹ I will not explore here the other side of the relationship, that is, the possible relevance of philosophical theories of justice for sociological theory and research; see, for that issue, the illuminating work of Swift (1999).

Here I will focus on points 2 and 3 (and somewhat on point 5). Although points 1 and 4 are also of great interest, I think that points 2, 3 and 5 are the key ones from the point of view of institutional design and ambitious social reform. If we do not resign ourselves to accept a ‘second-best’ policy, we should better think how we can use the best available means (even if not obvious, direct or immediate) to achieve a ‘first-best’ policy in terms of the selected normative principles, instead of thinking from the beginning how circumstances may almost always force us to accept ‘second-best’ policies. To my view, non-ideal theory of justice should not be concerned only with the question of how facts constrain the application of normative principles, but also, and perhaps primarily, with the question of which social processes may ‘push the facts’ to a situation in which the applicability of those principles would be maximized.

Of course, non-ideal theories of justice face the fact that *there is* a problem of institutional design when thinking on how to put into practice principles of justice: it is not enough to ‘be right’ in normative terms if we lack feasible and concrete proposals to implement our normative ideals (Goodin, 1996; Le Grand, 2003). We have to implement them in a playing field where the players and the context have already some specific features, some ‘rules’ of the game are already in place, and specific strategic and non-strategic interaction patterns are generating particular macrosocial effects. Some of these features, rules and patterns may be subject to change and to intentional design or political and administrative influence, and some may not; some times, trying to influence them may cause more harm than good, and some other times we may have good chances to alter the situation successfully (Elster, 1983; Le Grand, 1997, 2003). The two main hypothesis of theories of institutional design (Goodin, 1996; Pettit, 1996) are precisely that individuals’ behaviour is sensitive to the available opportunities and incentives of a given social context; and, second, that these opportunities and incentives may be institutionally designed, having in mind that certain macrosocial outcomes are more desirable than others (to a great extent because they are more just).

The preceding considerations show the need of knowing plausible and relevant mechanisms which give empirical social substance to non-ideal normative theories, and bring about the institutional designs and the macrosocial outcomes we are interested in. Here I intend to offer a first tentative, exploratory, and, of course, not exhaustive

account that aims to serve as a general framework for the research project mentioned above (see the explanatory note).

1. Some *desiderata* for institutional applicability

In this section I will list and define some useful criteria in order to evaluate the degree of applicability of a normative theory in terms of concrete institutional design. Later on, I will claim that in order to study whether they are satisfied or not, most of them require the kind of information that only the SM approach can give to us.

1. Informativity

In order for a normative theory to be applicable, it has to be an informative theory. This *desideratum* may be understood in two different ways. First, informativity may refer to how many possible worlds are excluded as unjust by the normative theory (in the same way that the informativity of a scientific theory or hypothesis is measured by the number of possible empirical worlds it excludes). Second, informativity points to a further question that may be also called ‘determination’ or ‘manageability’ of a normative theory: it should be possible to establish from empirical data when the principles of the theory are fulfilled and when they are not. For example, Rawlsian ‘difference principle’ involves counterfactual situations which makes it difficult (though perhaps not entirely impossible) to tell when it is properly fulfilled (cfr. Sen, 2006); the difference principle, it may be argued, excludes all possible worlds except one, but it is difficult to find the relevant empirical evidence to decide whether we are in that world or not. The same could be said about some well-known ‘information problems’ in normative theory, such as the problem of knowing individuals’ efforts (Brighouse, 2004:23ss) or willingness to work (De Wispelaere, 2000). Theories of justice, then, should provide criteria and identify suitable indicators that allow us to know when their principles have been satisfied, and/or to what extent.

2. *Efficacy*

The criterion of efficacy points to the problem of policy recommendation or policy choice: given a theory of justice or a normative principle, which policy or set of policies satisfy it in practice? This is one of the main problems of institutional design: arguably, to ask how effective an institutional design is for a given normative theory is to ask how to design institutions which deliver the desired outcomes when they are properly functioning, and which do so *because of its design* (because of the rules and controls they implement), and not because of eventual luck or good will of the agents who make them function.

3. *Efficiency*

Generally speaking, a given institutional design is more efficient than other when it satisfies the same principles at a lower cost (where the term ‘cost’ may have very different meanings: economical, political, and so on), or when it satisfies those principles to a greater degree at a similar cost. The problem of how efficient an institutional design is in realizing some normative principle is, then, different from the problem of how effective it is. To see that, it is enough to realize that different institutional designs may satisfy the same principle or theory; that is to say that theories of justice are *institutionally underdetermined* (De Wispelaere & Stirton, 2007), so criteria such as comparative efficiency of each effective institutional design may be decisive in order to make the final choice. It is worth to note that pure principled policy choice is unlikely to occur ever, if only because justice is not the only valuable thing when making that choice. Theories of justice define the *objectives*, but not the *mechanisms* through which those objectives will be achieved: that is the job for institutional design (and in doing so it will need the SM approach).

Efficiency may also be an issue when we are not pursuing just one normative ideal, but different ones at the same time. Then we should try to select the institutional design that scores better across all the objectives, even if it scores worse than others in some objectives taken in an isolated way.

4. Feasibility

There are many senses in which we may speak of the feasibility of a given institutional design or non-ideal normative theory. Trivially, *physical* and *biological* feasibility are obvious requirements of the rule ‘ought implies can’. *Economic* feasibility is also an important point, and we have good methods available in order to evaluate it. The less obvious and more difficult fields of study are those of psychological and political feasibility.

Political feasibility is probably the main area of problems for reform-oriented social science and institutional design in our societies, but probably the less studied from a scientific point of view. We may distinguish two issues here. First, *social approval* of a principle, ideal, or institutional proposal, and which kind of strategies are best in order to achieve that approval. Second, the feasibility of its *implementation* from an administrative and institutional point of view in a given political and legal context. On the second one I will say more later, when I address criteria such as robustness or resilience. Now, the first issue may be sub-divided in two: social approval by powerful individuals and groups (such as government officials and agencies, as well as other economic and social agents who have direct opportunities and influence to introduce institutional changes) and approval by ‘public opinion’, citizens, taxpayers, or voters in general, who, in a democratic society, have some power to demand and legitimate political measures. This is no doubt a tricky field for social scientists with normative sensitivity, since, as Swift (1999:347) points out, social approval of a given state of affairs or reform may not depend only on whether people think it is just or unjust, since they may think it is unjust but not want it to change for other reasons (or the opposite). Moreover, one should distinguish between feasibility here and now, and feasibility ever (Brighouse, 2004:27).

As for *psychological* feasibility, it seems reasonable to demand from a normative theory a high degree of consistency with what we know today from evolutionary psychology and cognitive sciences about human motivations and psychological dispositions. A non-ideal normative theory should not count on human beings capable of getting rid from one day to another of deeply rooted psychological dispositions which are the product of millions of years of evolution (that is specially important, for example, when facing gender or ethnic inequalities). Particularly, a given

institutional design should not assume as a given fact some psychological dispositions which are the condition of possibility of such design (allegedly, that is what some versions of socialism did, and one of the main reasons of their failure).

5. Motivational or categorical force

The degree of psychological feasibility of an institutional design is very closely related with a weaker requirement for the application of normative theories: its motivational or categorical force, understood as the capacity to mobilize the relevant emotions and reasons, or to generate the appropriate desires in the individuals, in order to produce the kind of behaviour that fulfils the principles of the theory. (Maybe we should call this criterion ‘psychological robustness’ or, better, ‘psychological efficacy’).

6. Robustness

I understand ‘robustness’ as the capacity of a given institutional design to function properly within reasonable levels of change of the initial conditions: it is a sort of internal stability of the design. The more robust an institutional design is, the higher its ability to resist unchanged even in the worst possible scenario. Some mechanisms in order to enhance robustness have been discussed in the literature: pre-commitment (Elster, 2000) or self-regulating strategies such as indexation of benefits, or surveillance by evaluation agencies, are only a few examples. As it was suggested, we might speak also of ‘psychological robustness’, when an institution is designed to avoid giving incentives to agents to behave against the basic principles of the normative theory (for example, game theory shows how in a Nash equilibrium no player has an incentive to change his strategy even if that equilibrium is suboptimal for all; if the equilibrium situation is a normatively rejectable one, then to design interaction arenas which systematically lead rational agents to it is not a psychologically robust strategy). One problem is that a robust strategy may not be the most effective or efficient in the short term, and that may affect its political feasibility.

7. Resilience

We may understand ‘resilience’, as De Wispelaere (2007) does, as a sort of ‘political robustness’, that is, the capacity to resist political pressures for change. A

resilient strategy of institutional design ensures some level of political support ‘no matter what’ happens in the political agenda and which political forces come to office after an election (an example of a very resilient policy, at least in Spain, would be the public contributory PAYG pension system).

8. *Consistency*

Finally, we may also speak of the ‘consistency’ of an institutional design or policy option, meaning that the institutional tools which are used or implemented do not violate the normative principles they try to put into effect. This would be a version of the traditional moral view that the ends do not justify the means, and it is related with the problem of the moral neutrality of SM, which will be mentioned later.²

2. The social mechanisms approach and analytical sociology

The set of desiderata listed above call for extensive social-scientific research in order to determine to what extent they are satisfied in a given context by a particular institutional design. Here I want to claim that Analytical Sociology (hereinafter, AS) provides the best means to do so. AS is one of the trends that has attracted most attention on the part of social scientists over the past years.³ AS constitutes an attempt to clarify concepts and practices, and to optimize and systematize good social-scientific work; it aims explicitly at paradigm unification in the social sciences, and at their integration with the rest of the contemporary scientific disciplines. The adjective “analytical” refers to the separation of the elements of a “whole” to study how they make it up. As Hedström says, “‘analytical sociology’ seeks to explain complex social processes by carefully dissecting them and then bringing into focus their most important constituent components” (Hedström, 2005:1).⁴

² I have tried to discuss some problems of consistency which are often faced by republicanism (Noguera, 2006b) and reciprocity theories (Noguera, 2007).

³ See Barbera (2004, 2006) and Noguera (2006a), for overviews; see Coleman (1990), Goldthorpe (2000), or Hedström (2005) for canonical texts.

⁴ Obviously, this denomination also implies an intellectual link with the tradition of analytical philosophy, which took the logical analysis of language as its core task. The uses of this tradition in the social sciences can be found mainly in economics, somewhat less in psychology and in political science, and very rarely in sociology and anthropology. This is, perhaps, the origin of the confusion and prejudice hold

One of the fundamental principles of AS is that the aim of sociology is the causal explanation of social phenomena. This may seem obvious, but as Boudon (2002) shows, most of what is given the name of “sociology” does not have primarily this *cognitive* function. Instead, it should be better seen as *expressive* sociology (which seeks to raise certain emotions or impressions by means of narrating the quasi-literary subjective experiences of the social agents or the sociologists themselves), *informative* sociology (which merely describes certain social phenomena or accumulates data), or *critical* sociology (which denounces and criticizes from a normative or ethical-political point of view certain states of the social world). In contrast, cognitive sociology as defended by AS seeks to explain in detail enigmatic or puzzling social phenomena by means of the construction of formal theories that specify intelligible causal mechanisms, thus providing microfoundations for the explanation. For AS, explaining is not describing, nor establishing taxonomies or typologies of phenomena (Hedström, 2005:12ss), nor engaging in sophisticated journalism, nor “empathising” with social actors or referring in a literary manner to social reality.

A remark should be made in this respect: the so-called “interpretivist” or “hermeneutic” sociologists have often tried to differentiate between the “causal explanation” of social phenomena and their “understanding”. For AS, however, both things have no sense as different methods; as Max Weber pointed out, “understanding” the meaning the actions have for the agents is part of the causal explanation of these actions. To “understand” is nothing less than attributing mental states with a certain propositional content on the part of the individuals. Therefore, AS shares the classical point of view set out by Davidson, according to which the *reasons* that individuals have for acting as they do can be legitimately understood as *causes* of their action, giving rise to intentional explanations (Davidson, 1963; Elster, 1983; Boudon, 2003). Hedström, for his part, speaks of *DBO theory*, that is, the theory that explains actions as being caused by individuals’ desires (D), beliefs (B) and opportunities (O). However, it is still frequent to hear some sociologists say things so inconsistent as “I don’t want to discover the causes of this phenomenon or action, I just want to understand it”. But any attempt to “understand” the actions of an agent necessarily involves some causal model

by many sociologists who accuse analytical sociology of being “economistic” (or, some times, “psychologistic”).

or assumption; the real choice is between making it explicit as such, or leaving it in the limbo of the unthought and vague.

To explain is to provide patterns of intelligibility (Van Parijs, 1981). Hence, a requirement for satisfactorily explaining a phenomenon is to specify its microfoundations (Elster, 1989, 1999). In the case of social phenomena, AS seeks to specify the *social mechanisms* that underlie them. According to Hedström (2005:2), “the core idea behind the mechanism approach is that we explain a social phenomenon by referring to a constellation of entities and activities, typically actors and their actions, that are linked to one another in such a way that they regularly bring about the kind of phenomenon we want to explain”. Explanation based on mechanisms avoids both the nomological model of the covering law, and also mere statistical association. Mechanism explanations are a sort of ‘sometimes-true theories’ (Hernes, 1998:76), in the sense that they uncover typical causal processes which work only in certain contextual conditions. However, these causal processes may be identified in several different contexts which give them some generality, and knowledge of them provides intelligible and ‘final’ explanations of social phenomena (Barbera, 2006; Boudon, 1998; Hedström & Swedberg, 1998).

The social mechanisms approach has several advantages. First, specifying the social mechanisms that cause phenomena avoids pseudo-explanations that rely on “black boxes” (many of which refer to vague entities and processes such as “socialization”, “culture”, “values”, etc.; cf. Boudon, 1998). Second, it excludes the postulation of any sort of *sui generis* causality or logic (different from the conventional one) such as “dialectics”, “structural homology”, non-intentional “meanings”, “reasons” without agent, or “objective” teleologies (Elster, 1983, 1989). Third, it rules out the confusion between “explaining” and “naming” or “labelling” phenomena (something that social theorists like Bourdieu, Giddens, Beck, Habermas and others are fond of; see Van den Berg, 1998). For AS, sociology should learn to accept that social phenomena may be explained causally (in the conventional sense of specifying a mechanism that connects two events), but not “understood” according to an alleged “deep meaning” that goes beyond this ordinary causality, or beyond the meanings that are intentionally projected by human beings. Finally, as remarked by Barbera (2004), another advantage of explanations by means of mechanisms is that they are open to empirical testing. The

best way of distinguishing a scientific theory from a pseudo-theory, a dogma, or an ideology is that the latter are able to reinterpret any imaginable event according to their principles, while the former cannot. A theory has informative power, and is therefore open to empirical testing insofar as it excludes possible worlds.

Social mechanisms may be classified mainly in three types (Hedström & Swedberg, 1998): *situational* mechanisms (when they account for belief and desire formation from a given context of opportunities and constraints); *action-formation* mechanisms (when they explain how agents act departing from a set of opportunities, beliefs and desires), and *transformational* mechanisms (when they explain how individual actions generate aggregation and composition effects at the macro-social level). There are also four basic principles that underlie the SM approach (Barbera, 2006; Hedström & Swedberg, 1998): (1) *Action*: it is actors and their intentional actions who ‘make social mechanisms work’. (2) *Precision*: SM explanations focus on concrete areas and phenomena, and do not claim to discover ‘universal laws’ of macrosocial phenomena; they are better understood as Merton’s ‘middle-range theories’. (3) *Abstraction*: SM explanations select some aspects of the phenomena to be explained and discard the rest as irrelevant, and so they avoid the mistake of confusing the map and the territory. (4) *Reduction*: SM explain phenomena in some level in terms of phenomena of a less complex level, so they can open ‘black boxes’ in order to see the causal microfoundations of complex macro-phenomena.

Since a great part of the tradition of sociological theory has been inspired by holist, structuralist, or collectivist assumptions, it may be useful to stress the first principle. The explanation by means of mechanisms is typically based on individuals’ *actions*, and preferably on their *intentional* actions (Van Parijs, 1981). The abovementioned *DBO theory* may be put this way: “Individuals act; they are not merely pushed around by anonymous social forces; and in order for a theory to be explanatory it must consider the reasons why individuals act as they do” (Hedström, 2005:36-37). Sociological explanation “should avoid the ‘atomized’ and ‘heroic’ assumptions of traditional economics as well as the ‘causalism’ of traditional sociology” (*ibid.*). As the quote shows, along with the central role of actions and intentions, AS also gives an explicitly fundamental role to *rationality* (Boudon, 2003). The reasons agents have to act, and not only blind force or irrational intentions (instincts, impulses, inertia or

similar), are often causes of their actions. When individuals believe that they have “good reasons” to act in a particular way, these reasons are the most plausible explanation for their action (and the beliefs and desires that form these reasons may, in turn, have been formed on the basis of reasons: this is what is stated by Boudon’s theory of *cognitive rationality*). The advantage of this point of view, apart from its plausibility in phenomenological and common sense terms, is that rational action is its own explanation (be it cognitively or instrumentally). As Weber well knew, irrationality or resort to “hidden” forces used to lead to pseudo-explanations, or just be the implicit recognition of the inability to explain an action. Irrationality on the part of agents should only be assumed when we have run out of all other possible rational explanations for their action (this is what is claimed by the “charity principle” for which Davidson has argued compellingly). For this reason, AS mistrusts “culturalist” or “sociological” explanations that do not specify understandable psychological mechanisms to account for actions: this is the case of pseudo-explanations of the “black box” type that affirm that “X carried out action Y because he is a member of society S or culture C”, or “because he has been socialized in the tradition T”, or “because he has the cultural identity I”.

However, it may be important to do some remarks: (a) AS does not usually accept the “standard” version of rational choice theory (as it is used, for example, in neo-classical economics), but rather softens some of its assumptions and enriches such version with empirical results from experimental psychology and economics and with wide concepts of rationality which go beyond the purely instrumental one (Elster, 1983; Boudon, 2003); (b) AS assumes, therefore, a motivational pluralism on the part of individuals (who may be selfish, altruistic, or both things in different degrees depending on the action context) without giving up its preference for rationality as a general driving force of action; (c) AS’s intentionalism and rationalism do not imply that social phenomena respond to an intentional design on the part of some agents; on the contrary, AS shows how these phenomena are often the unintentional and unexpected result of many intentional actions (Boudon, 1977, 1984; Axelrod, 1997); (d) obviously, AS also admits (and in fact uses profusely) other explanatory instances apart from rationality, mainly social norms and evolutionary mechanisms of selection and reinforcement (Hedström, 2006; Van Parijs, 1981; Elster, 1999; Axelrod, 1997).

3. Social mechanisms matter: their relevance for non-ideal theories of justice

Having summarized the core ideas behind AS and the SM approach, in this section I will give some reasons in favour of the necessary commitment of non-ideal normative theorizing with *cognitive* sociology of the type that AS defends (and also with *descriptive* one⁵), a commitment that does not appear so strongly in the case of other types of sociological discourses. I will depart, therefore, from Boudon's abovementioned distinction between four types of sociological endeavour: cognitive, descriptive or informative, critical, and expressive sociology. I claim that the relevance of each one for normative concerns is of a very different degree and nature.

Critical and expressive sociology and non-ideal normative theory

Unlike cognitive and descriptive types, the other two types of sociology are not specially relevant for the theories of justice. *Critical* sociology may be understood in most cases as just an attempt to defend a previous normative theory in the light of some facts described by the 'critical' sociologist, who often claims that such a description is somehow 'guided' by his normative commitments (something that, to say it gently, seems not precisely an example of scientifically controlled data gathering and selection). Since the only distinguishing feature of critical sociology is to criticize or denounce as unjust some states of affairs, then it can be reduced, at best, to descriptive sociology, *plus* some personal moral-political evaluations added by the sociologist to his descriptions. Even assuming that the description and the data selection are not biased by ideological pre-commitments, and that the normative principles underlying his criticisms are presented in a sharp and explicit way, the 'critical sociologist' still owes a rational justification of why he selects some particular normative principles and not others. Unfortunately, what is commonly regarded as 'critical sociology' in the tradition of sociological theory does not usually succeed in any of these three tasks, and, in fact, seem to be quite unconcerned by them; it rather tends (to use Popper's crude but

⁵ I take for granted the obvious fact that cognitive sociology always requires some kind of previous descriptive sociology in order to identify the social patterns or regularities which are to be causally explained (Goldthorpe, 2004).

direct expression) to “generate dark Cassandra’s cries about the bad times we are living and the perversion of our culture”⁶, getting very close to expressive sociology.

As for this latter, it does not seem to be specially relevant for normative theory either. Expressive sociology is just one more way (perhaps sometimes a sophisticated way) to raise emotions which are behind some of our normative intuitions, or maybe one more source of inspiration for those intuitions and the principles that try to capture them. Each of us may have some personal opinion as for whether good literature or good cinema are usually much better and more effective ways than expressive sociology in doing so (and even more honest ones, since they do not disguise themselves with academic or intellectual jaergon in order to pretend some ‘scientific’ or authoritative stance).

Norwithstanding the preceding considerations, we may concede that critical and expressive sociology may have some relevance for non-ideal normative theorizing, at least in two ways: first, when they include good descriptions of some social phenomena, they may be useful for evaluating the *desideratum* of *informativity* (as a possible sub-case, they may provide normative theorizing with some case studies or real-world examples which are relevant for justice concerns). Second, to the extent that they raise and mobilize some emotions which are at the origin of some shared normative intuitions, they may have some indirect relevance for the evaluation of the *motivational force* of some normative ideals.

Descriptive sociology and informativity

It is intuitively very clear that good descriptive sociology (Goldthorpe, 2004) is relevant for normative concerns. In an obvious way, the mere possibility of applying theories of justice and normative judgements to the real world requires some descriptive and reliable information about that world, and more information the more we move from a ‘trascendental’ approach to a ‘comparative’ one, to use Sen’s recent distinction (Sen, 2006): the first approach would distinguish just from unjust situations in an ideal-theoretical way, while the second would tell us if a given society or situation is more just than another one. Of course, in this latter case we need a sharp and reliable description of the two situations to be compared, at least if they are real and not just

⁶ Karl Popper, quoted in *Conversaciones con Marcuse* (Barcelona, Gedisa, 1978), p. 159.

theoretically built. Moreover, as noted by Brighouse (2004:16s), theories of justice often use real cases and case studies to progress in normative reasoning; good descriptive sociology is then needed to provide empirical material for that task.

Perhaps more interestingly, descriptive sociology is needed in order to solve *information problems* which are frequent in normative theorising. A lot of normative judgements which use categories such as *chance, choice, responsibility, willingness, preferences, interests, merits, or inequality*, depend on the information we may obtain through social science techniques. Some examples may be:

- Telling adaptive preferences from non-adaptive ones.
- Distinguishing between constraint and choice.
- Identifying willingness to cooperate or to work.
- Telling instrumental self-interested actions from deontological-normative orientations towards social norms.
- Identifying merits or efforts done by individuals.
- Uncovering discriminations and past injustice.
- Establishing whether a given kind and level of inequality is the case.

So it seems that what Boudon calls descriptive or ‘cameral’ sociology is not only an important part of social research (Goldthorpe, 2004), but also a highly relevant one for normative concerns. However, in the rest of this section I want to focus on how *cognitive* sociology may be as much (or even more) useful for normative theory than descriptive sociology.

Cognitive sociology and feasibility

The strongest argument I can find to defend the relevance of cognitive sociology for normative theory is the following one: if the principle ‘ought implies can’ is valid in normative theory (at least in its non-ideal form), then we should strive for the maximum amount and quality of information about *what* ‘can’ and ‘cannot’ be the case in the social realm (what is the ‘feasible set’ of social states), and about *why* and *how* it is so. In order to obtain that kind of information, descriptive sociology is not enough; we must go beyond describing concrete phenomena, social patterns or empirical regularities. Cognitive sociology and the SM approach, by identifying the causal mechanisms which

underlie social phenomena, are a key instrument to study the feasibility of normative principles in terms of institutional design.

Some rough examples may be easily presented. It is obvious that knowledge of social mechanisms (of the *situational* kind) on how public opinion perceives and reacts to some political proposals is important for the study of their political feasibility. Additionally, the knowledge of social mechanisms (of the *action-formation* kind) which govern the balance of power in a given society and its possible change may serve as a tool in order to design strategies that influence that balance in one way or another (and to identify or generate some other mechanisms of the *transformational* kind). Some developments in social science which are useful for those ends are: the theory of collective action, the study of political and social coalitions, the power resources theory (Korpi), the study of the ‘civilizing force of hypocrisy’ (Elster), the social consequences of the falsification of preferences (Kuran), the effects of the temporalization of reforms, and mechanisms such as ‘pluralistic ignorance’ or ‘rational imitation’ (Hedström, 1998).

It is important to notice that the SM approach may be of great help in order for non-ideal normative theory to avoid two common mistakes:

a) Discarding too fast as ‘unfeasible’ some social states of affairs which may be feasible by virtue of some unthought, unknown, or counter-intuitive social mechanism; (recall Brighouse’s distinction between ‘feasible here and now’, and ‘feasible in the best possible circumstances’). The counter-intuitive character of many social mechanisms is an important point to bear in mind here: social simulation techniques and formal interaction models have shown how unpredictable and paradoxical macro-social effects and compositions may be when a certain amount of agents are interacting simultaneously (Boudon, 1977, 1984; Hedström, 2006). This suggests that even if a normative principle may not seem feasible in a given context, a closer and detailed look to the social mechanisms involved may provide with strategies of institutional design that ‘push’ forward its degree of feasibility. We should be very careful before accepting some social circumstances or constraints as ‘given facts’ which have to remain unaffected by institutional design; to do that may lead us to accept ‘second-best’ policies where ‘first-best’ ones are possible, and, henceforth, to make suboptimal institutional choices. To this end, the potential of social simulation and experimental

techniques to reveal concealed factual possibilities or causal social processes we were unaware of, is still to be fully explored.

b) The opposite mistake is to regard too fast as ‘feasible’ some state of affairs whose production is more difficult than it seems intuitively, because they involve the concatenation of some social mechanisms that may not be adequately produced in a given context, or that cannot be the result of an intentional strategy (such as the states which are essentially by-products studied by Elster, 1983).

For all these reasons I contend that some knowledge of sociology, and specifically of the analytical type, is essential for the normative theorist. ‘Ought implies can’ is in fact a mandate for the theorist of justice to become a putative social scientist. Normative theories involve causal assumptions and hypothesis, and social mechanisms are the ‘grammar’ of causality in the social world.

Assumptions and social mechanisms

The SM approach may be useful not only for evaluating the feasibility of desired states of affairs, but also to evaluate the *plausibility of the assumptions* a normative theory makes in order to justify its principles. The most known example at hand is of course the debate on Rawls’ ‘difference principle’: it has been widely discussed to what extent the basic assumption in which the principle relies is plausible, that is, if it is frequently true that in order to maximize the endowment of the worst-off we have to allow for some degree of inequality which works as an incentive for the talented and laborious, who will then produce a larger cake (Cohen, 1991). This is clearly an assumption about SM (and one involving the three types of them).

SM are also involved when normative theories use *counterfactual* situations in order to reason for or against some principle. Let us consider again the ‘difference principle’, in the demanding (but compelling) interpretation given by Van Parijs (2002). In order to know if a given inequality is required to maximize the position of the worst-off, we have to provide some counterfactual reasoning to show that there is no other possible situation in which that position is at least as good in absolute terms, *and* inequality is lower. In order to do that in a grounded and rigorous way, we need the SM approach. This suggests that SM are also important for *informational* reasons in normative theory, and not only for feasibility ones.

Efficacy, efficiency, and social mechanisms

Again, mechanisms matter in order to grasp the degree of efficacy and efficiency of institutional designs in satisfying normative principles. For example, SM may be essential to determine whether incentives, punishments or value commitments are the appropriate means to achieve some institutional outcomes in an effective and efficient way. How the contextual logic of each social situation triggers some specific mechanisms when some type of reform is introduced, or what kind of perverse effects may be generated by it, are key questions for efficacy and efficiency of institutional designs.

Regarding both efficacy and efficiency, we face a difficult question for political philosophy and for political action as well, a question that has often confronted radical and moderate or reformist positions across the political spectrum: when we face a systematic or 'structural' injustice, should we identify and abolish the very social structure, institution or mechanism that brings it about (for example, 'capitalism', 'markets', 'patriarchal values' and so on), or might we be satisfied just by implementing policies that correct, compensate, or reverse that outcome once produced (or while it is being produced)? The first strategy may sometimes be more effective but less efficient, while the second may have the opposite problem (of course, feasibility problems often arise as a strong reason to decline the first kind of strategies and look for one of the second kind). Anyway, the scientific study of the social mechanisms in place may be of great help (in fact, may some times be the only reliable way) to decide these kind of questions.

Robustness, resilience, and social mechanisms

The SM approach may also be used to study properties of institutional designs such as robustness or resilience. For example, mechanisms of *path dependence* may be important in this regard: they account for the impact of early events or decisions on the deterministic nature of some institutional patterns, and explain why some institutional design decisions may direct or guide social evolution in a quite irreversible way (think of some programs of a welfare system like PAYG pension schemes, or of how tax reductions may be very difficult to undo). Other evolutionary mechanisms have to do,

for instance, with collective action in iterated social dilemmas: think of *threshold effects* which may determine how many people have to adopt a norm of behaviour in order for it to become stable (contrary to some common sense assumptions about political feasibility, the threshold may be far from being ‘the majority’ of the population).

Robustness and resilience may be also directly affected by psycho-social mechanisms whose existence has been proven by laboratory experimental research. To name just a few: *status quo bias* (we want what we already have just because it is what we have), ‘*social proof*’ *heuristics* (if I see that everyone believes that X is good, then I may think that they know something I do not know for believing it, so I better believe that X is good too, just in case; see Kuran, 1995:163), or *signaling effects* (if a particular behaviour is associated with some kind of positive property or information about the person who performs it, then more and more people will be attracted to that behaviour, in order to avoid ‘making a bad signal’).

Motivational force and social mechanisms

Finally, we should be aware of evidence from behavioural and experimental social sciences that has shown the existence of some mechanisms which regulate human motivations towards altruistic or egoistic conduct. Non-ideal theories of justice and institutional design have to be sensitive to motivational complexity and to cognitive and moral weaknesses of human beings, trying not to create or reinforce coordination problems or collective action dilemmas which are linked with them. Particularly, designs should avoid providing agents with incentives that undermine the desired behaviour (that is what happens, for example, when even if some punishments are at hand, it is openly declared that they will hardly be enforced, or when wide sets of exceptions to the rules are allowed: think of ban or restrictions on smoking in selected places, or of regulations to harden penalties for traffic infringements).

As a summary of what has been said in this section, Table 1 shows the different kinds and degree of relevance that the four types of sociology distinguished by Boudon have for non-ideal theories of justice.

Table 1
Relevance of the four types of sociology
for the desiderata on non-ideal theories of justice

Desiderata	Types of sociology			
	<i>Descriptive</i>	<i>Cognitive</i>	<i>Critical</i>	<i>Expressive</i>
<i>Consistency</i>				
<i>Informativity</i>	X	X (counterfactual assumptions)		
<i>Feasibility</i>	X (if A exists, it is feasible)	X (why A is feasible)		
<i>Motivational force</i>		X	X (raising moral intuitions)	X (raising moral intuitions)
<i>Efficacy</i>		X		
<i>Efficiency</i>		X		
<i>Robustness</i>		X		
<i>Resilience</i>		X		
<i>Case examples available</i>	X	X	X	X

A note on research techniques

I have repeatedly mentioned the particular usefulness of social simulation models (specially multi-agent ones) and experimental behavioural techniques in order to undertake the kind of empirical social research which may be needed in order to account for the satisfaction of the desiderata of non-ideal theories of justice⁷. It may be appropriate at this point to stress briefly some advantages of these two techniques:

1) This kind of techniques allow to study and uncover the *microfoundations* of behaviour and of macrosocial outcomes, since they permit to design artificial environments (which may include ‘social norms’) that allow to *isolate specific causal effects*, something that would not be possible in the real world. For example, Camerer & Loewenstein (2004:7) report how laboratory experiments can rule out cognitive confusion or reputation seeking as possible causes of some behaviors which seem

⁷ For multi-agent simulation, see Hedström (2005, 2006), or Axelrod (1997); for behavioural experimental research in the field of social justice and institutional design, see for example De Wispelaere (2007), Noguera & De Wispelaere (2006), Konow (2003), Dawes & others (2007), Roth (1995) or Camerer & Loewenstein (2004).

driven by a concern with fairness (as in the “ultimatum” game); those causal factors can hardly be present in laboratory designed environments where the situations are enough simple to be understood by everyone (which rules out cognitive confusion) or the players remain anonymous (which rules out reputation seeking as the cause of the ‘fair’ behaviour).

2) It is possible to do *series* of experiments (either in laboratory or with simulation models). This opens a wide margin to replicate, confirm, or reject some results, and to modify some features of the experimental design in order to refine its internal validity or to compare different situations. This approach allows to plan systematic series of experiments at a low cost and in a short period of time, without requiring additional financial or political support.

3) Both techniques are specially well suited to test *counterfactual* situations. Counterfactuals, as we saw, are important for normative theory, and are implied by most of the thought experiments that are frequently discussed in it (Brighouse, 2004:14ss). Multi-agent simulation is a very good technique for designing thought experiments and testing them in a carefully designed artificial environment (Axelrod, 1997), and so are laboratory experiments. Both consist in putting in relation agents which have some characteristics with a given context or structure of rules and see what outcome results. Needless to say how much important these experiments may be in order to evaluate the feasibility of some states of affairs or the possible effects of the introduction of some institutional designs.

4. Some illustrations: the case of tax and transfer policies

It has been claimed that the SM approach may be of great help to optimize the design of public policies in order to bring about just social outcomes. An obvious field where this claim should be illustrated is tax and transfer policies, the main instrument used by welfare states in order to redistribute income and wealth in a more equitable way. This is a case of social norms that are also official legal norms. But, as Axelrod (1997) has well pointed out, legal norms alone are often not enough to ensure the desired cooperative behaviour, and informal norms are very important in order to

explain levels of compliance: for example, informal norms about the level of tax burden that the government can legitimately impose (moral economy norms), or about whether one has the moral obligation to pay tax or not and to what extent, or whether one has to punish or actively detect tax evaders, or about the justice of social benefits and the deservingness of the recipients, are key in order to explain the political feasibility of tax and transfer policies (Bowles & Gintis, 2002, 1999). Law may operate as a support or reinforcement for informal norms, but it may be much more difficult for it to substitute or eradicate the latter (think of legal ban on alcohol or drugs). Axelrod's hypothesis is that laws may work better to prevent huge but infrequent defections, while informal norms work better against small but frequent ones. This is very important for tax and transfer policies, since usually anti-fraud laws and policies are not enough to fight small fraud, and since the feasibility of certain welfare schemes starts to crumble when a lot of middle class taxpayers develop a weak but persistent feeling that 'they give too much' to the poor, unemployed, or handicapped.

In order to see in which ways the SM approach may be applied to this field, let us make some remarks and look at some examples in a very unsystematic and exploratory fashion (consider it as a sort of brain-storming):

1) To be sure, present distribution of income and wealth is clearly far from *any* egalitarian theory of justice. Thus, one of the main concerns for egalitarian policy designers is how to achieve a greater degree of redistribution, something that the traditional machinery of classical welfare states seems to have problems to do beyond some point (Goodin & Le Grand, 1987). In order to think of innovative solutions to this problem of low redistributive effects, we should take into account the results of experiments on reciprocity (Bowles & Gintis, *ibid.*) and recent behavioural evidence of 'inequality aversion' (Dawes & others, 2007, show that given certain distributions, individuals develop strong negative emotions which lead to 'Robin Hood'-like behaviour). The work done in the field of basic income research (which is one of the most innovative proposals for a radical income redistribution) should be also studied carefully (De Wispelaere & Noguera, 2006).

2) Fiscal policy is one of the most interesting arenas in order to study the functioning of some SM. Take, for example, some reforms of the Spanish tax system during the last decade. There is a widely shared feeling among politicians and scholars

that it is not possible to raise income tax without losing an election as a consequence. But Aznar's right-wing first government (1996-2000) did exactly that. Moreover, he managed not only to raise income tax but to convince the majority of citizens that he had lowered it. Aznar came to office in 1996 having promised to lower income tax; Aznar's party (the Popular Party) did not have a majority in the Spanish Parliament, so he felt very compelled to be loyal to his electoral commitments. His government effectively lowered tax rates in the income tax. But (leaving aside they raised indirect taxes) this reduction was fictitious, since indexation and annual adjustment of the tax band limits with inflation was cancelled, and the same happened with the level of tax reliefs and the 'life minimum' (the amount of income which is not taxed). The global effect was a concealed income tax raise. So the tax burden as such was raised, and the distribution of the income tax burden was made more favourable to the better-off than ever before since the democratic transition. But surprisingly, the majority of people thought that taxes had been lowered! The lesson seems clear: contrary to the commonly shared view, it is perfectly feasible to raise income tax using some mechanisms which are cognitively opaque for citizens, make them believe that income tax has been reduced, and win the next election (and it is worth knowing how this is possible even if you think that it is not a morally acceptable strategy).

3) Another example from fiscal policy in Spain: during the second Aznar's government period (2000-2004), the Spanish Socialist Party, in the opposition, launched a public debate on one of his fiscal proposals, the introduction of a flat tax (together with some reforms of social benefits in the direction of a basic income). The proposal was immediately accused from everywhere in the political spectrum of ending progressivity in the Spanish tax system; opinion surveys showed voter's distrust with it, and the Socialist Party did not finally include it in his electoral program. But, of course, the accusation was confusing nominal with real progressivity. A flat tax plus a dramatic raise in the threshold of non-taxed income (which was the original proposal) would have brought more progressivity to Spanish income tax than a graduated tax rate with six bands as we have now. A graduated tax rate, when combined with other measures like tax reliefs and exemptions or special treatment for capital rents may be much less progressive than a flat tax for all incomes without exceptions. But the mechanism is very different from a cognitive point of view, and a flat tax will be easily perceived (and

politically presented by opponents) as unjust: “everyone pays the same rate irrespective of his income!”.

4) A third example is at hand: imagine that the present Spanish government announced that every owner of a dwelling will receive a generous social benefit, starting tomorrow. The announcement would no doubt result in a political (and social) scandal and fierce discourses about the injustice of such measure would arise. However, in a certain way, that is exactly what the Spanish income tax is really doing from many years ago, through a generous tax relief to everyone who is paying for a dwelling (no relief exists for those who rent a dwelling).

5) All the previous examples may be regarded as illustrations of what has been called *fiscal illusion*. Generally speaking, people tend to underestimate the amount of the taxes payed when (a) the tax system is complex, there are many taxes and their administration is fragmented in different levels of government; (b) the cost of the information about taxes is high; (c) taxes are less visible (for example, indirect or consumption taxes are less visible than direct or income taxes); (d) taxes are retained at source (and refunded if necessary), not payed through an income declaration (that is why self-employed people often have the impression to pay more taxes than the employees); (e) taxes are used for objectives approved by the taxpayers (this suggests that ear-marked taxes are easier to enforce and to accept than general ones; we are more willing to pay when we know the finality or the effective use of tax revenues; for example, we are more willing to pay social security contributions than general income tax). Of course, the opposite is also true, so people tend to overestimate taxes when the opposite circumstances are in place. (It would be interesting to test the hypothesis of the existence of a parallel ‘welfare illusion’, if we could show that social expenditure is overestimated or underestimated under some circumstances, and how this affects to political support and disposition to pay for benefits).

6) Some researchers have shown that *publicity* may be a powerful mechanism for the promotion of pro-social strategies. It is a fact that in small groups or societies where everyone can see if everyone cooperates or not, collective action dilemmas hardly result in high levels of defection. Of course, this is much more problematic in complex societies, but, would it be unreasonable to think of mechanisms of publicity for tax evaders, such as publishing their names, and let anyone know who they are and

how much they evaded? (incidentally, some regions in Spain publish the names of women abusers). Social mechanisms which govern shame, trust, guilt and reputation would then be put into work in favour of the common good.

7) The question of ‘strict’ vs. ‘tolerant’ norm enforcement is also worthy of study: to what extent some tolerance with small or partial defection may be useful for the extension of a more ‘general’ or ‘lax’ compliance or cooperation? To turn a blind eye on ‘small’ defection may be good to extend ‘lax’ compliance. This may be the case for traffic norms, labour market regulations, environmental laws, or academic performance. But, would that be also the case in the field of tax compliance? In which conditions the policy of ‘zero tolerance’ with defection leads to greater tax compliance? Some experimental results on this issue but on a quite different field, that of work compliance, are reported by Gneezy & Rustichini (2004); they find that work compliance may not satisfy that hypothesis: when incentive payments or fines are low, work performance declines. So maybe this is an argument for strong workfare (strong “fines”) or basic income (strong “incentive” to work, since there is no withdrawal of the benefit when one finds a job), vs. weak workfare or weak-conditional benefits.

8) We have seen how some unpopular or risky reforms (such as raising income tax) may be implemented by using some cognitive or framing strategies. This is important when we want to implement just reforms that face important social opposition. Another way may be to use ‘back door’ strategies (Vanderborght, 2002). One may be certainly surprised when some social policy scholars who are openly critic towards basic income declare to agree with most of the steps towards it, if they are presented one by one: for instance, to accept proposals such as a universal basic pension, a universal child benefit, a generous conditional minimum income or negative income tax, and even some in-work benefits like tax credits or sabbatical leaves, is to back one by one all the pieces that conform the puzzle of a universal and unconditional BI. From a logical point of view, it would be certainly irrational for one person to be against the reform R (which is equal to $a + b + c$), and to be, at the same time, in favour of a , b and c if taken separately. But from a psychological or political point of view this is often the case. As Marx once said, we should not confuse the logic of the matter with the matter of logic.

5. Some problems with the normative use of the social mechanisms approach

I would not like to finish this paper without at least mentioning some problems that we may find when working with the SM approach for assessing non-ideal theories of justice. For space reasons I cannot deal at length with them here, but it is worth to list a few questions which remain to be answered:

1. Does the use of the SM approach in non-ideal theories and in institutional design lead to a sort of moral cynicism?
2. Are SM morally 'neutral' themselves?, or does the pursuit of the application of a just principle justify the use of *any* SM at hand?
3. Are mechanisms themselves 'given' and unchangeable?
4. Might the fact that a SM (or the intention to rely on it) is made public affect its proper functioning? (this would point to what Hacking, 1999, calls 'loop effects').
5. Is it possible for the SM approach to generate axiologically neutral social science?, and, if not, aren't we falling into a logical circle?

These are important questions, and I am not resisting to say a few words about the first one. Am I arguing that we should deceive people by using framing, back door, and other similar strategies? Definitely no. To begin with, *deceiving* is quite a different thing from *framing* a social reform in some specific way: framing is about describing reality in a certain way, not about hiding or falsifying it (and the same may be said of a 'back door' strategy which simply adopts an incremental approach to implement reforms in time). With that distinction in mind, what I have tried to argue is that *if* an institutional reform is just *and* enough citizens oppose to it on the grounds of egoistic interests or misconceived judgements or perceptions, *then* knowing and using social mechanisms like the abovementioned may be perfectly justified and wise. Of course, when we face people's cognitive mistakes or false perceptions, information, rational deliberation and persuasion would be the ideal strategies, but, leaving aside their efficacy is not granted, they often take time and hard effort, which we may not have

when urgent social needs and demands are in place. It may be then perfectly wise and not particularly rejectable from a moral point of view (or so I think) to apply the reform first and persuade later. We should avoid confusing the logic of the normative justification of proposals with the logic of its practical realization. Finally, a caveat is perhaps due: once the normatively relevant objects have been selected for study, research on the SM involved should progress following an inner logic, and avoiding any external manipulation based on eventual political demands.

References

- Axelrod, Robert (1997). *The Complexity of Cooperation*. Princeton, Princeton University Press.
- Barbera, Filippo (2004). *Meccanismi sociali. Elementi di sociologia analitica*. Bologna, Il Mulino.
- Barbera, Filippo (2006). “A Star Is Born? The Authors, Principles and Objectives of Analytical Sociology”, *Papers*, 80, pp. 31-50.
- Boudon, Raymond (1977). *Effets pervers et ordre social*. Paris, PUF.
- Boudon, Raymond (1984). *La place du désordre*. Paris, PUF.
- Boudon, Raymond (1998). “Social mechanisms without black boxes”, in Hedström, P. & Swedberd, R. (eds.), *Social mechanisms. An Analytical Approach to Social Theory*. Nueva York, Cambridge University Press.
- Boudon, Raymond (2002). “Sociology that Really Matters”, *European Sociological Review*, vol. 18, n° 3, pp. 371-378.
- Boudon, Raymond (2003). *Raison, bonnes raisons*. Paris, PUF.
- Bowles, Samuel & Gintis, Herbert (2002). “*Homo Reciprocans*”, *Nature*, n° 405
- Bowles, Samuel and Gintis, Herbert (1999) “Is Equality *Passé*? *Homo Reciprocans* and the Future of Egalitarian Politics”, *Boston Review* 23 (3).
- Brighouse, Harry (2004). *Justice*. Cambridge, Polity Press.
- Camerer, C. F. and Loewenstein, G. (2004). “Behavioral Economics: Past, Present, Future”, in Camerer, C. F.; Loewenstein, G. & Rabin, M. (eds.), *Advances in behavioral economics*. New York, Russell Sage Foundation.
- Cohen, Gerald A. (1991). “Incentives, Inequality, and Community”, in Grethe B. Peterson (ed.), *The Tanner Lectures on Human Values* (vol. 13). Salt Lake City, University of Utah Press, 1992.
- Cohen, Gerald A. (2003). “Facts and Principles”, *Philosophy & Public Affairs*, vol. 31, n° 3.
- Coleman, James (1990). *Foundations of Social Theory*. Cambridge, The Belknap Press.
- Davidson, Donald (1963). “Actions, Reasons and Causes”, in *Essays on Actions and Events*. Oxford, Oxford University Press, 1980.
- Dawes, Christopher T.; Fowler, James H.; Johnson, Tim; McElreath, Richard & Smirnov, Oleg (2007). “Egalitarian Motives in Humans”, *Nature*, vol. 446, 12 April, pp. 794-796.
- De Wispelaere, Jurgen (2000) “Sharing Job Resources: Ethical Reflections on the Justification of Basic Income”, *Analyse und Kritik* 22 (2), pp. 237-256.
- De Wispelaere, Jurgen (2007) “The Political Resilience of Universal Basic Income”, unpublished paper.
- De Wispelaere, Jurgen & Stirton, Lindsay (2007) “The Public Administration Case Against Participation Income”, forthcoming in *Social Service Review*.

- Elster, Jon (1983). *Sour Grapes. Studies in the Subversion of Rationality*. Cambridge, Cambridge University Press.
- Elster, Jon (1989). *Nuts and Bolts for the Social Sciences*. Cambridge, Cambridge University Press.
- Elster, Jon (1999). *Alchemies of the Mind. Rationality and the Emotions*. Cambridge University Press.
- Elster, Jon (2000). *Ulises Unbound. Studies in Rationality, Precommitment, and Constraints*. Cambridge, Cambridge University Press.
- Frolich, N. and Oppenheimer, J. A. (1992). *Choosing Justice. An Experimental Approach to Ethical Theory*. Oxford, University of California Press.
- Gneezy, Uri and Rustichini, Aldo (2004). "Incentives, Punishment, and Behavior", in Camerer, C. F.; Loewenstein, G. & Rabin, M. (eds.), *Advances in behavioral economics*. New York, Russell Sage Foundation.
- Goldthorpe, John H. (2000). *On Sociology. Numbers, Narratives and the Integration of Research and Theory*. Oxford, Oxford University Press.
- Goldthorpe, John H. (2004). "Sociology as Social Science and Cameral Sociology", *European Sociological Review*, 20, pp. 97-105.
- Goodin, Robert E. (ed.) (1996). *The Theory of Institutional Design*. Cambridge, Cambridge University Press.
- Goodin, Robert E. y Le Grand, Julian (1987). *Not Only the Poor. The Middle Classes and the Welfare State*. London, Allen & Unwin.
- Hacking, Ian (1999). *The Social Construction of What?*. Cambridge (Mass.), Harvard University Press.
- Hedström, Peter (1998). "Rational Imitation", in in Hedström, Peter & Swedberg, Richard (eds.): *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge, Cambridge University Press.
- Hedström, Peter (2005). *Dissecting the Social. On the Principles of Analytical Sociology*. Cambridge, Cambridge University Press.
- Hedström, Peter (2006). "Explaining Social Change: An Analytical Approach", *Papers. Revista de Sociologia*, nº 80.
- Hedström, Peter & Swedberg, Richard (1998). "Social Mechanisms: An Introductory Essay", in Hedström, Peter & Swedberg, Richard (eds.): *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge, Cambridge University Press.
- Hedström, Peter & Swedberg, Richard (eds.) (1998). *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge, Cambridge University Press.
- Hernes, Gudmund (1998). "Real Virtuality", in Hedström, Peter & Swedberg, Richard (eds.): *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge, Cambridge University Press.
- Konow, James (2003). "Which is the Fairest One of All? A Positive Analysis of Justice Theories", *Journal of Economic Literature*, 41 (4), pp. 1188-1239.
- Kuran, Timur (1995). *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, MA: Harvard University Press.
- Le Grand, Julian (1997). "Knights, Knaves and Pawns", *Journal of Social Policy*, vol. 26, nº 2.
- Le Grand, Julian (2003). *Motivation, agency, and public policy: of knights and knaves, pawns and queens*. Oxford, Oxford University Press.
- Noguera, José A. (2005) "Citizens or Workers? Basic Income vs. Welfare-to-Work Policies", *Journal of Law and Urban Studies*, 2 (1), pp. 103-124.
- Noguera, José A. (2006a). "Why We Need an Analytical Sociological Theory", *Papers*, 80, pp. 7-28.
- Noguera, José A. (2006b). "Freedom, Justice, and Work: Why Republicans Cannot Be Against Workfare", paper presented in the 2006 ALSP Conference on Social Justice, Dublin.
- Noguera, José A. (2007). "Why Reciprocity Theories Are Inconsistent", *Basic Income Studies*, 2 (1), forthcoming.

- Noguera, José A. & De Wispelaere, Jurgen (2006). “A Plea for the Use of Laboratory Experiments in Basic Income Research”, *Basic Income Studies*, 1 (2).
- Pettit, Philip (1996). “Institutional Design and Rational Choice”, in Robert E. Goodin (ed.), *The Theory of Institutional Design*. Cambridge, Cambridge University Press.
- Roth, A. E. (1995). “Introduction to Experimental Economics”, in Kagel, J.H. y Roth, A.E. (eds.), *The Handbook of Experimental Economics*. Princeton, Princeton University Press.
- Sen, Amartya (2006). “What Do We Want from a Theory of Justice?”, *Journal of Philosophy*, 103 (5), pp. 215-238.
- Swift, Adam (1999). “Public Opinion and Political Philosophy: The Relation Between Social-Scientific and Philosophical Analyses of Distributive Justice”, *Ethical Theory and Moral Practice*, 2, pp. 337-363.
- Van den Berg, Axel (1998). “Is Sociological Theory Too Grand for Social Mechanisms?”, in Hedström, Peter & Swedberg, Richard (eds.): *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge, Cambridge University Press.
- Van Parijs, Philippe (1981). *Evolutionary Explanation in the Social Sciences*. Totowa (New Jersey): Rowman & Littlefield.
- Van Parijs, Philippe (2002). “Difference Principles”, in Samuel Freeman (ed.), *The Cambridge Companion to John Rawls*. Cambridge, Cambridge University Press.
- Vanderborght, Yannick (2002). “Basic Income in Belgium and the Netherlands: Implementation Through the Back Door?”, paper presented at the IX Congress of the Basic Income European Network (BIEN), Geneva.