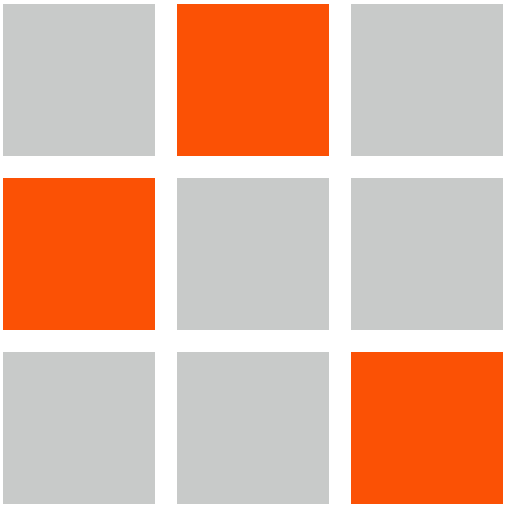
Universitat Autònoma de Barcelona. Departament de Sociologia

Octubre-desembre 2014, vol. 99, núm. 4
http://papers.uab.cat
ISSN 2013-9004 (digital)
ISSN 0210-2862 (paper)

PAPERS

Revista de Sociologia

99/4



Raymond Boudon

UAB
Universitat Autònoma de Barcelona
Servei de Publicacions

Redacció
Universitat Autònoma de Barcelona
Departament de Sociologia
08193 Bellaterra (Barcelona). Spain
Tel. 93 581 12 20. Fax 93 581 24 37
r.papers.sociologia@uab.cat

Intercanvi
Universitat Autònoma de Barcelona
Servei de Biblioteques
Secció d'Intercanvi de Publicacions
08193 Bellaterra (Barcelona). Spain
Tel. 93 581 11 93
sb.intercanvi@uab.cat

Administració i edició
Universitat Autònoma de Barcelona
Servei de Publicacions
08193 Bellaterra (Barcelona). Spain
Tel. 93 581 10 22. Fax 93 581 32 39
sp@uab.cat
http://publicacions.uab.cat

ISSN 2013-9004 (digital)
ISSN 0210-2862 (paper)
Dipòsit legal: B. 25.307-1983

Composició
Binorama, SCP

Bases de dades en què PAPERS està referenciada

- ARCE-FECYT
- CARHUS+
- CIRC (Clasificación Integrada de Revistas Científicas)
- Compludoc
- Dialnet (Unirioja)
- DICE (Difusión y Calidad Editorial de las Revistas Españolas de Ciencias Sociales y Jurídicas)
- DOAJ (Directory of Open Access Journals)
- Educ@ment
- Índice Español de Ciencias Sociales y Humanidades (ISOC-CSIC)
- IN-RECS (Índice de Impacto de Revistas Españolas de Ciencias Sociales)
- International Bibliography of the Social Sciences (IBSS)
- Latindex
- MIAR (Matriu d'Informació per a l'Avaluació de Revistes)
- RESH (Revistas Españolas de Ciencias Sociales y Humanas)
- RACO (Revistes Catalanes amb Accés Obert)
- SCOPUS de SciVerse
- Social Services Abstracts
- Socindex, Socindex Full Text
- Sociological Abstracts
- TOC Premier
- Ulrich's

PAPERS és una publicació del Departament de Sociologia de la Universitat Autònoma de Barcelona fundada l'any 1972. El seu objectiu és servir de mitjà de difusió d'idees i d'investigacions originals, en el camp de la sociologia i altres ciències socials afins (psicologia, ciència política, economia, antropologia).

L'acceptació d'articles es regeix pel sistema de censors. Es poden consultar les normes del procés de selecció i les instruccions per als autors a http://papers.uab.cat/about/submissions#authorGuidelines.

Aquest volum ha rebut un ajut econòmic de: Fundació BancSabadell

Sumari

Papers. Revista de Sociologia
Octubre-desembre 2014, vol. 99, núm. 4, p. 421-629
ISSN 2013-9004 (digital), ISNN 0210-2862 (paper)
Les paraules clau són en llenguatge lliure
http://papers.uab.cat

Raymond Boudon

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2111

Raymond Boudon: An analytical social theorist

Jordi Tena-Sánchez
Universitat Autònoma de Barcelona. Department of Sociology
Analytical Sociology and Institutional Design Group (GSADI)
jordi.tena@uab.cat

Received: 11-06-2014
Accepted: 04-07-2014

Summary

My own ideal was […]: to say complicated things as simply and clearly as possible.

Raymond Boudon.

1. Raymond Boudon, a giant of the social sciences

On 10 April 2014, Raymond Boudon – one of the most prominent social theorists in the second half of the 20th century and the first part of the 21st – died in Paris. His death marked an enormous loss for the social sciences in general and for sociology in particular.

With many of his works already authentic classics of the social sciences, Boudon's *oeuvre* is immense, covering fields as diverse as the sociology of education and social mobility, social theory, methodology, the analysis of classics of these disciplines, ideologies, beliefs and moral values, political theory, rationality, and a long etcetera. Obviously, this introduction is not the place to discuss his work as a whole, or even superficially, so I will focus on a few contributions which, from my point of view, are among his most important and share a thread which I will refer to later.

2. Methodological individualism and social mechanisms

First of all, one of Raymond Boudon's most important contributions to social sciences was his defence of *methodological individualism* and of causal explanations based on *social mechanisms* (also called *causal* or *generative mechanisms*). That is to say, the principle according to which any macro-social phenomena must be explained as a result of individuals' behaviour, which in turn is the result of these individuals' reasons and motivations. Also in turn, these reasons and motivations can only be understood in reference to the initial social situation of these individuals.

Despite that, as the author himself stated (see his paper in this issue), this is an old idea, Raymond Boudon (together with other very prominent social theorists such as Thomas Fararo, James Coleman or Jon Elster) made a decisive contribution to making this principle the basis of what constitutes an appropriate explanatory strategy in social science. Boudon developed this strategy in opposition to *pseudo-explanations*, both *functionalist* or *structuralist* ones as well as exclusively statistical ones, whose paradigmatic example was *positivism* (see Cherkaoui in this issue).

It is necessary to clarify this point somewhat in order to avoid misunderstandings. Boudon was always a promoter of empirical sociology and a staunch defender of using statistical tools and formal models. However, he was critical to the same extent of what has been called *variable sociology* (Esser, 1996), that is to say, the approach by which a phenomenon is explained when we identify a set of independent variables that predict the variance of dependent variables.

3. The scientific nature of social sciences and the study of educational inequalities

For Raymond Boudon, the purpose of sociology (and of social science in general) should not be to move the reader or to make him/her enjoy (as literature can do), nor to transform society through political activism, or even to produce data and analyses aimed at making decisions. As he masterly maintained in his "Sociology that really matters" (2001), the main purpose of sociology must be causal explanation of enigmatic social phenomena. From this point of view, sociology has a scientific nature and must be ruled by the same principles of formal and methodological accuracy like in any other discipline.

L'inégalité des chances (1973) constituted the first of Raymond Boudon's major work and what finally placed him in the foreground of the international scene. In this work, which today is an unavoidable starting point for the best contemporary research on social and educational inequalities, the Frenchman put into practice the principles of his social science to show how agents' decisions at the micro level, given their different starting points, result in differentiated scholarly careers and reproduce existing inequalities. In this way, Boudon was successful in showing how actions and interactions at the micro-level can produce aggregated outcomes at the macro-level that nobody expects or wishes (*perverse effects*) without resorting to obscure *teleological arguments* or employing mere descriptive labels such as "socialization" with explanatory aims (see León in this issue).

The explanation offered by Boudon contrasted with the *pseudo-explanation* in fashion at the time (and still today) of Bourdieu and Passeron (1970), for whom school is, in reality, a tool for the reproduction of social inequalities. Boudon himself described Bourdieu and Passeron's work as rhetorical, pedantic and nebulous (see his paper in this issue) because the fact is that Boudon's work is at the other extreme of the *bullshit*, so sadly habitual in some contemporary intellectual circles.[1] As Jean Cazeneuve stated in the speech he gave upon Boudon's election to the *Académie des Sciences Morales et Politiques* (see Boudon in this issue), Boudon's style was the opposite: to say complicated things as simply and clearly as possible.

4. Ordinary rationality vs. rational choice theory and explanations with black boxes

Another of Raymond Boudon's major contributions was his *ordinary rationality theory* (also known as *cognitive* or *subjective*). The Frenchman showed himself to be lucidly critical of *rational choice theory* and, among other things, pointed out that human beings do not always act in an instrumental way, so this theory can only explain a more or less restricted part of human behaviour (especially if the *universal self-interest* principle is assumed). In spite of its limited explanatory power, however, Boudon still recognized the methodological goodness of *rational choice theory*. Explanations based on this theory are *final*, without *black boxes.* When we can show that somebody has done something because it was in his/her interest, this person's behaviour becomes understandable for us, we do not have additional questions.

In this sense, explanations based on *rational choice theory* are better than employing (so habitual in the social sciences) mere descriptive labels such as "socialization", "enculturation", "habitus", etc. with explanatory aims. According to the Frenchman, when we say that somebody has done X because he/she has been socialized to do X, in reality we are not explaining his/her behaviour, we are just using a technical label to name a phenomenon whose workings we ignore.

1. On the concept of *bullshit,* see Cohen (2002).

So, according to Boudon, we need a theory with the methodological power of rational choice theory (which avoids black boxes in the explanations) but which increases its explanatory capacity. This theory is, according to Boudon, ordinary rationality theory. Boudon argued that, in principle, we must assume that, in a specific cognitive context, individuals always have good reasons to do what they do or to believe what they believe. These reasons can be instrumental as well as cognitive or axiological. In this way, Boudon solved, for example, the well-known problem (for rational choice theory) of why most people vote in elections when going to vote has costs for individuals and they know that the effect of their vote on the outcome will be almost null. According to the Frenchman, people vote when they believe that democracy is something valuable (cognitive reason), that they must contribute to preserve it (axiological reason), and that one party is better than the others to govern (cognitive reason). Again, Boudon solved the problem without having to turn to black-box pseudo-explanations such as people vote because they have been socialized to vote.

5. Anti-relativism

Finally, the last of Raymond Boudon's contributions to the social sciences I would like to refer to in this introduction is his critique against *constructivism* or *relativism*, not only against *epistemic relativism*, but also against *moral* or *political relativism*, though here I will exclusively focus on the first one. In spite of the fact that this topic is clearly implicit in what I have dealt with above, Boudon dedicated some major works (i.e. *The poverty of relativism*) to discuss it in detail. The Frenchman denounced the thesis of the *avant-garde* of the sociology of science according to which social-scientific knowledge, like any other form of knowledge or discourse, is relative to its social, cultural or historical context of production (Berger), is the product of some kind of interests of power (Foucault), or constitutes an exercise of "symbolic violence" (Bourdieu), so that the objective knowledge pursued by social sciences with a scientific vocation is impossible. Among many other problems, this kind of arguments are inconsistent and self-nullifying. Note that when somebody argues that any discourse constitutes, for example, a discourse of power and that for this reason its validity claims cannot be accepted, we can immediately ask why we should accept the validity claims of his/her argument if it is just a discourse of power whose validity claims cannot be accepted. That is to say, our interlocutor asks us to accept his/her reasons of why there are no reasons, just power exercises (Noguera, 2006a).

6. Raymond Boudon, an inspirer of analytical sociological theory

The contributions and characteristics of Raymond Boudon's social science to which I have made reference (as well as many others discussed in the papers of this issue) constitute some of the central components of what has been called *analytical sociological* (or *social*) *theory* (AST).

AST should not be understood as the thousandth school or "paradigm" in social sciences. It constitutes, rather, an attempt, where several theories, methodologies, positions and social research traditions converge, to establish reasonable "rules of the game" in social sciences: to reorder and clarify the nature of the contributions of sociology with the aim of making it a rigorous scientific discipline that provides empirically grounded explanations of relevant social phenomena[2] (Noguera and Tena-Sánchez, 2013). Clearly, Raymond Boudon, together with many other social scientists and philosophers, has been one of the main inspirers of this movement.

7. Short comment of the papers in this issue

"Why I became a sociologist" is an autobiographic text published in 2009 where Raymond Boudon reviews his career, from his beginnings as a student at the *École Normale Supérieure* to the last phase of his trajectory. Boudon describes what were his main intellectual influences and interests throughout his life and how these were embodied in his works. The text has the virtue of offering a summary of Boudon's work written by the author himself.

Mohamed Cherkaoui's paper sets Boudon's proposal against empiricist approaches, whose most prominent expression was *positivism*. Cherkaoui makes clear the weak spots of this approach, as well as how the Boudonian strategy of *generative mechanisms* contributes to overcoming them. In this way, the paper contributes to answering the usual criticism which brands authors such as Raymond Boudon (and AST in general) as *positivists* (as if this accusation – were it true – would automatically disqualify an entire approach).

Ángeles Lizón, on the other hand, offers a retrospective of Boudon's academic contribution through the three major stages of his academic career (path regression models, game theoretical mechanisms and ordinary rationality). In this sense, like Cherkaoui, Lizón's paper also helps to understand how the *generative mechanisms* approach constituted, among other things, an overcoming of the previous atheoretical empiricist approaches and to refute the hasty criticism of positivism against Raymond Boudon's approach (and AST in general).

In the last part of the paper, Lizón criticizes Boudon's conception of *methodological individualism* because, according to her, he is not clear enough about the ontological assumptions underlying this explanatory principle.

Karl-Dieter Opp criticizes what he considers a major problem of Raymond Boudon's rationality theory: its relatively low explanatory power (given that Boudon does not establish a criteria to select the causally relevant reasons for an *explanandum*). Opp inquires into the validity of the theory, that is, if it is plausible that a single theory can explain the wide range of phenomena Boudon focuses on.

On the other hand, Opp criticizes the fact that Boudon rejects the *utility maximization* principle and makes a powerful defence of it as well as of *rational*

2. For a deeper explanation of AST, see Demeulenaere (2011), Hedström (2005), Hedström & Bearman (2009), Manzo (2010, 2014) or Noguera (2006b, 2010).

choice theory, arguing that, in fact, Boudon's theory is compatible with a *wide* version of rational choice theory and can be understood in this way.

Pierre Demeulenaere's paper focuses on Boudon's *ordinary rationality theory*. First of all, he presents the general theories of rationality that currently exist in social sciences. According to the author, rationality has been understood historically as 1) *intentionality* (in the weak sense that people have reasons to do what they do), 2) *preference consistency*, 3) adequacy of the choice of the means to reach an end, and 4) *self-interest.* Demeulenaere argues that, in spite of the fact that, in practice, these four criteria (or some of them) are usually found together in the concepts of rationality that social scientists use in their work, they are in fact four analytically distinguishable concepts of rationality. According to the author, Boudon's ordinary rationality theory has some major advantages with respect to previous theories, but Demeleneure states that it can be enhanced and completed in several dimensions in order to achieve a sound theory of the interpretation of human actions in the social world. In this regard, he points to the need to find a general basis for defining something as rational.

Gianluca Manzo's paper deals with several topics related to Boudon's work. First of all, Manzo argues that because Bourdon's *ordinary rationality theory* broadens the concept of rationality of rational choice theory, it loses its predictive capacity.

In this sense, he points out that, to recuperate it, it is necessary to identify reason trigger factors, that is to say, mechanisms that systematically tend to trigger certain sets of reasons. In this way, we would have access to a set of regularities that enable us to formulate *ex-ante facto* clear expectations on what micro- and macro-level outcomes are more likely to be observed. Next, Manzo discusses three kinds of factors which, according to him, are good candidates to progress in that direction: social-identity, emotions and heuristics.[3]

In the second part of the paper, Manzo discusses Boudon's point of view on AST.[4] Manzo shows that AST places the methodological principles defended by Boudon explicitly, consciously and programmatically at the core of the approach. In fact, it could not be any other way given that, as said previously, Boudon is one of the main inspiring intellectual sources of the analytical sociology movement. Finally, Gianluca Manzo discusses the potential of an innovative type of formal models for AST: agent-based models (ABM).

Precisely, Francisco Linares' paper shows one of the main utilities of ABM, its potential as a formal tool to test the plausibility of the assumptions of a theory. Linares builds a model where he explores the role of homophily and social contagion in the search for a partner. Linares argues that his exercise illustrates, in practice, Boudon's proposals, as he formulated them in works such as *La logique du social.* Besides its intrinsic quality, the exercise is interest-

3. Heuristics are cognitive shortcuts adopted by actors when they have to decide and solve problems. Specifically, Manzo make reference to Gigerenzer's *fast and frugal heuristics theory* (2008).
4. See Boudon (2012b).

ing because Boudon (2012b) referred indirectly to ABM as "secondary technical details". In this sense, Manzo and Linares' papers complement one another in showing that ABM is not a secondary detail but a fundamental tool for the advance of AST as well as a good instrument to put in practice the methodological principles defended by Raymond Boudon himself.

Fernando Aguiar starts by asking about the possible reasons Boudon did not develop a theory on social identity or pay special attention to the concept. Boudon himself answered this question in part when, as Aguiar states, he explicitly rejected the concept of identity which he saw as a product of a hypersocialized portrayal of the actor. However, as Aguiar argues, it does not necessarily derive from Boudon's criticism that social identity is not an important factor for explaining social behaviour, but that, again, it is usually employed as a mere label without explanatory value. In his paper, Fernando Aguiar offers an interesting avenue to overcome this problem, that is, by employing Boudon's *ordinary rationality theory* to build a concept of social identity consisting of a set of positive and normative beliefs about ourselves that give us reasons to act.

Finally, Francisco León discusses Boudon's well-known argument that psychological or biological concepts should not play a role in the explanation of social phenomena given that – exactly as what happened with concepts such as "socialization" – these concepts reintroduce black boxes into explanations. Again, according to Boudon, in order to offer an explanation without black boxes, it is necessary is identify the reasons that drive individuals' behaviour.

As León states, it now seems impossible to maintain that argument and the vast amount of evidence coming from the different behavioural sciences (neuroscience, primatology, experimental economics, behavioural genetics, evolutionary psychology, etc.) in recent years cannot be ignored without further ado. Of course, accepting the role of "biological causes" does not entail that there are two kinds of phenomena, those that should be explained with reasons and those that should be explained with "biological causes", nor (at least for the moment) that reasons should not play a key role in explanations. The question is rather, as Manzo also points out in his paper (referring specifically to *fast and frugal heuristics*), that both kinds of phenomena work in some kind of complex interrelation that is still to be unravelled. In conclusion, as León states, social science nowadays needs a new analytical turn: in the future, the explanatory strategy based on social mechanisms will have to be founded on contributions from the behavioural sciences.

8. Bibliographic references (Boudon's references are in the next section)

Bourdieu, Pierre and Passeron, Jean Claude. (1970). *La reproduction.* Paris: Éditions de Minuit.

Demeulenaere, Pierre (ed.) (2011). *Analytical Sociology and Social Mechanisms*. Cambridge: Cambridge University Press.

Cohen, Gerald A. (2002). "Deeper into bullshit". Buss, Sarah and Overton, Lee (eds.). In: *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge (MA): MIT Press.

Esser, Hartmut (1996). "What is wrong with 'variable sociology'?". *European Sociological Review*, 12 (2), 159-166.
<http://dx.doi.org/10.1093/oxfordjournals.esr.a018183>
Gigerenzer, Gerd (2008). *Rationality for Mortals. How People Cope with Uncertainty*. Oxford: Oxford University Press.
Hedström, Peter (2005). *Dissecting the social: On the principles of analytical sociology.* Cambridge: Cambridge University Press.
Hedström, Peter and Bearman, Peter (eds.) (2009). *The Oxford Handbook of Analytical Sociology.* Oxford; New York: Oxford University Press.
Manzo, Gianluca (2010). "Analytical Sociology and Its Critics". *European Journal of Sociology*, 51 (1), 129-170.
<http://dx.doi.org/10.1017/S0003975610000056>
— (2014) (ed.). *Analytical Sociology: Norms, Actions, and Networks.* London: Wiley.
Noguera, José A. (2006a). "Why We Need an Analytical Sociological Theory". *Papers. Revista de Sociologia*, 80, 7-28.
— (2006b) (ed.). *Analytical Sociological Theory*. *Papers. Revista de Sociologia*, 80.
— (2010) (ed.). *Teoría sociológica analítica*. Madrid: Centro de Investigaciones Sociológicas (CIS).
Noguera, José A. and Tena-Sánchez, Jordi (2013). *Jon Elster. Un teòric social analític.* Barcelona: Universitat Oberta de Catalunya.

9. Selected bibliography of Raymond Boudon[5]

Boudon, Raymond (1973). *L'inégalité des chances: la mobilité sociale dans les sociétés industrielles.* Paris: Armand Colin. [1974. *Education, Opportunity and Social Inequality.* New York: Wiley].
— (1977). *Effets pervers et ordre social*. Paris: PUF. [1982. *The unintended consequences of social action.* London/Basingstoke: Macmillan].
— (1979). *La logique du social*. Paris: Hachette. [1981. *The logic of social action.* London/Boston: Routledge & Kegan Paul].
— (1979). "Generating Models as a Research Strategy". Merton, Robert K.; Coleman, James S. and Rossi, Peter H. (eds.). In: *Qualitative and Quantitative Social Research*. New York: The Free Press.
— (1995). *Le juste et le vrai: études sur l'objectivité des valeurs et de la connaissance*. Paris: Fayard. [2001. *The origin of values.* New Brunswick/London: Transaction.]
— (1998). "Social mechanisms without black boxes". Hedström, Peter (ed). In: *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
— (1998). "Limitations of rational choice theory". *American Journal of Sociology*, 104 (3), 817-828.
<http://dx.doi.org/10.1086/210087>
— (2001). "Sociology that really matters". *European Sociological Review*, 18 (3), 371-378.
<http://dx.doi.org/10.1093/esr/18.3.371>
— (2005). *The poverty of relativism.* Oxford: The Bardwell Press.
<http://www.bardwell-press.co.uk/publications/more.htm>
— (2006). "Homo sociologicus: Neither a Rational nor an Irrational Idiot". *Papers. Revista de Sociologia*, 80.

5. For an exhaustive list of Boudon's works (until 2009), see Cherkaoui and Hamilton (2009).

— (2006). *Tocqueville for Today*. Oxford: The Bardwell Press. <http://www.bardwell-press.co.uk/publications/tocqueville.htm>
— (2012). C*roire et savoir: penser la politique, le moral et le religieux.* Paris: PUF.
— (2012b). "'Analytical sociology' and the explanation of beliefs". *Revue européenne des sciences sociales*, 50 (2), 7-34. <http://dx.doi.org/10.4000/ress.2165>
— (2013). *Sociology as Science: An Intellectual Autobiography*. The Bardwell Press. <http://www.bardwell-press.co.uk/publications/sociology-as-science.htm>

10. Selected books on Raymond Boudon

Cherkaoui, Mohamed and Hamilton, Peter (eds.) 2009. *Boudon. A life in sociology*: *Essays in honour of Raymond Boudon.* Oxford: The Bardwell Press. <http://www.bardwell-press.co.uk/publications/boudon.htm>

11. Selected obituaries

Demeulenare, Pierre (2013). "Raymond Boudon". Obituary published on the webpage of the *Groupe d'Étude des Méthodes de l'Analyse Sociologique de la Sorbonne* (GEMASS) on the occasion of Raymond Boudon's death. Available at: <http://www.gemass.fr/spip.php?article3770&lang=fr> [last search: 28/05/2014][6].
Lindenberg, Siegwart (2013). "Raymond Boudon. A personal view". *L'Année sociologique*, 63 (2), 297-300. <http://dx.doi.org/10.3917/anso.132.0295>
Tena-Sánchez, Jordi (2013). "In Memoriam. Raymond Boudon (1934-2013): un teórico social analítico". *Revista Española de Sociología*, 20, 13-17.

Acknowledgements

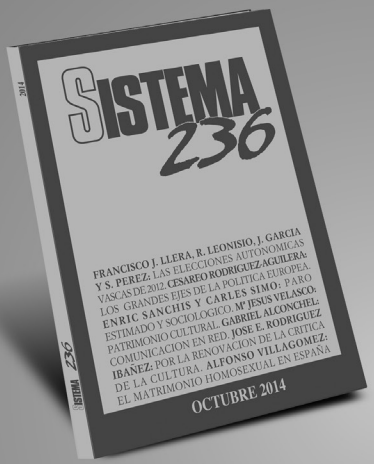
First of all, I would like to thank José A. Noguera, co-editer of *Papers. Revista de Sociologia* for having proposed that I coordinate this special issue, as well as for his help during its preparation. Second, I would like to thank the authors who have contributed to the issue. Third, I would like to thank Peter Hamilton for so kindly allowing me to republish Raymond Boudon's chapter in this issue. And last but not least, I would also like to thank Pep Sansó and Martha Gaustad for their efficient work in editing this special issue.

Funding

This work has been supported by the National Plan for R&D and the CONSOLIDER-INGENIO Programme of the Spanish Ministry of Economy (Grant Nos. CSO2012-31401, CSD2010-00034).

6. A long list of Raymond Boudon's obituaries can be found on this same link.

SISTEMA



Editor: **José Félix Tezanos**
Director: **Elías Díaz**
Secretaria de Redacción: **María Rosario Sánchez Morales**

ISSN: 0210-0223
P.V.P.: 15 €

Francisco José Llera, Rafael Leonisio, Jonatan García Rabadán y Sergio Pérez Castaños: Las elecciones autonómicas vascas de 2012. Alternancia en un nuevo ciclo político.
Cesáreo Rodríguez-Aguilera de Prat: Los dos grandes ejes de la política europea en los manifiestos de los europartidos para las elecciones de 2009.
Enric Sanchis y Carles Simo: Paro estimado y paro sociológico.
María Jesús Velasco Gutiérrez: El patrimonio cultural inmaterial desde una visión totalizadora de la cultura.
Gabriel Alconchel: Comunicación en red y tendencias de cambio en las culturas políticas.
José Enrique Rodríguez Ibáñez: Por la renovación de la critica de la cultura: una propuesta programática.
Alfonso Villagómez: El matrimonio homosexual en España: una opción legislativa de ampliación de un derecho fundamental.

Edita **EDITORIAL SISTEMA** **www.sistemadigital.es** **www.fundacionsistema.com** Teléfono: 914487319

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2126

Why I became a sociologist*

Raymond Boudon

Abstract

In this autobiographic work, Raymond Boudon reviews his trajectory from his beginning as a student at the *École Normale Supérieure* to the last stages of his career. Boudon describes his main intellectual influences and concerns throughout his life and how they were displayed in his works.

Keywords: Raymond Boudon; sociology; methodological individualism; education; social mobility; objectivity; relativism; rationality.

Resumen. *¿Por qué me convertí en sociólogo?*

En este trabajo autobiográfico, Raymond Boudon pasa revista a su trayectoria, desde sus inicios como estudiante de la *École Normale Supérieure* hasta la última etapa de su carrera. Boudon describe cuáles fueron sus principales influencias e inquietudes intelectuales a lo largo de su vida y cómo éstas se acabaron plasmando en sus obras.

Palabras clave: Raymond Boudon; sociología; individualismo metodológico; educación; movilidad social; objetividad; relativismo; racionalidad.

Summary

* This article was first published in: Cherkaoui, Mohamed & Hamilton, Peter (eds) (2009). *Raymond Boudon: A Life in Sociology. Essays in Honour of Raymond Boudon*, Vol. I. Oxford: Bardwell Press, pp. 53–75. Reprinted with permission from The Bardwell Press.

Lehrjahre

As the saying goes, psychologists become psychologists because they have problems with themselves, anthropologists because they have problems with the world, sociologists because they have problems with their society. This is not true in my case. I have the feeling I did what I wanted to do. I had the privilege of not being directly involved in any war, either in my own case or through my family. I had a direct and rapid career. I was appointed professor at the Sorbonne at the age of 33. France has not been in the best of shapes for years or even decades because its governments have done too little to correct the negative effects of centralization and the cult of the State. But I experienced the *Trente glorieuses* and I have always enjoyed the French *art de vivre*. So does my wife. I share with her a deep intellectual, moral and political complicity. She was born in Thüringen in Eastern Germany, and had to flee with her family to Bavaria in order to avoid the Soviet army. She studied law in Munich and taught German in a French college after we married. Thanks to her, I have the feeling that I enjoy a binocular view of French society and as a result practice comparative sociology all day long.

My wife's father was a doctor. After the War, he managed to obtain the indispensable certification that he was never involved in Nazism from the East German authorities and was allowed to practice his job for some months, though exclusively in the Soviet zone, until he too fled to Western Germany to join his family. As to my parents, both came from families of modest craftsmen. My father's passion was music. He played oboe in an orchestra in his youth. I admired his ability to read an orchestra score fluently. He reached a moderately senior position in a big Parisian commercial firm and gave his family a comfortable standard of living. He disliked the communists as deeply as the Nazis. Hitler's *Mon combat* (*Mein Kampf*) and Kravtchenko's *J'ai choisi la liberté* (*Choose Freedom*) were the books that had made the greatest impression on him. My mother was an excellent cook. So is my wife: her reputation as a cook is well known among our friends.

After my secondary school studies, I gained entrance to the *École Normale Supérieure* in 1954. I was proud of this success and benefited from the advantages provided by the institution. I obtained a grant that led me to Freiburg in Breisgau for one year, where I have the great opportunity to listen to Martin Heidegger, although I was not greatly impressed by his course on *Der Satz vom Grund*. I had the distinct impression that he was playing with words and observed that he was held in greater reverence by the many students who came from Latin America or Iran than the German students. So I remained faithful to my earlier philosophical masters, Kant and Hegel. But, in the fifties, most French professional philosophers spent their time more or less exclusively on discussions of classical philosophers. I wanted to orient myself more towards a discipline dealing with the concrete human world. Economics attracted me, more because it seemed to be the most rigorous of the human sciences, than by the topics it dealt with. Psychology appeared to me as artificial in its experimental version and verbose in its clinical version. I was always inter-

ested in history, but never liked the discipline, exactly for the same reasons as, I learned later, Bronislav Malinowski: because he found human history too dismal, especially that of his native Poland, he created a discipline which dealt with societies but ignored the sound and fury of history. His "functionalism" erased the image of history being a tale told by an idiot and replaced it by the study of the rationality and mutual complementarity of institutions. The history of both World Wars had given me the feeling that history is dark and that I would run the risk of becoming depressed if I tried to become a historian: so many wrong decisions and ideas had led to catastrophes which could have been easily avoided, when considered *a posteriori* at least. The French Revolution seemed to me to have been quite barbarian in its final phase. I regretted that France had been made to suffer so many political convulsions since the Revolution. Thus sociology and economics remained the only possible choices as the outcome of this exclusion process. Sociology attracted me more, in principle at least, because of the broader field it aimed at covering and also because I saw it as a modern and hence more attractive version of philosophy. It also dealt with values, beliefs, ideas, institutions, though in a more concrete way. But I was unconvinced by the books published under the label of sociology in France in the late 1950s, for I found them too rhetorical. I saw economics as narrow, but solid because of its use of mathematics. My mentor Raymond Aron, whom I consulted on my difficulties in choosing between economics and sociology, told me: "you should choose sociology: for a young man, there is more potential in sociology than economics".

I was easily convinced, since I saw that my lasting interest in philosophical questions was more easily satisfied by sociology than by economics. While browsing among sociological books, I had found Paul Lazarsfeld's and Morris Rosenberg's *Language of Social Research*. This book gave me the impression that the type of sociology it advocated was much more scientific than the laborious and boring classifications produced by the great sociological star of the fifties and early sixties in France: Georges Gurvitch. At my request, Raymond Aron recommended me for a grant to study in an American University and I opted for Columbia University in New York, attracted by the prestige of Robert Merton and Paul Lazarsfeld. My wife and I spent an unforgettable year there. At that time, few French intellectuals went to the United States. Many of them were close to the Communist party or at least sympathetic to its ideas. They saw the United States as the Empire of Evil. But thanks to our stay in New York, we discovered the great gulf between American and European Universities, in terms of budget, organization, diversity, facilities for the students, dynamism, and also rejection of rhetoric. As to the sociology that had developed around Lazarsfeld and Merton, it seemed to me that it was inspired by the scientific *ethos*.

Early Works

I came back from the States with a project for my doctoral dissertation: as economics had seemingly become more scientific as it became more math-

ematical, I decided to combine my interest in sociology and economics by trying to clarify the question of the uses of mathematics in sociology. I see my dissertation now as much too broad, and original only to a very limited extent. It brought little new, but helped me in seeing clearly that mathematics could only have a limited impact on sociology. One chapter alone was original: the one where I used a very simple simulation model to explain statistical data in the field of the sociology of law. The proportion of cases that were abandoned rather than sent to a court had according to the statistical data been regularly growing since the beginning of the 19th century. Why? Gabriel Tarde had asked a similar question about another trend: why had the proportion of trials ending in a verdict of not guilty regularly decreased over the long term? As with Tarde, I tried to make the trend an outcome of the strategy developed by the actors of the judicial system in order to be seen as successful by their peers. This exercise convinced me of two things: firstly, that macrophenomena should be explained as the effects of individual behaviours; and secondly, that a central sociological problem is consequently to find out the reasons and motivations of individual actors. All my later works are elaborations of these basic insights. I did not know then that Max Weber and Joseph Schumpeter had christened this approach *methodological individualism*.

The academic regulations in France at that time insisted that candidates to the *doctorat d'Etat* had to present a second dissertation, on a subject different from the subject of the main one. After discussing the matter, in particular with Paul Lazarsfeld, I decided to work on structuralism. Under the influence of Claude Lévi-Strauss, structuralism had become popular at that time. Structuralism was born in the field of phonetics. The core idea of structural phonetics was that the phonemes of a language constitute a system of sounds aiming at using a minimal set of elementary distinct sounds to make the communication of any message as unambiguous and economical as possible. The idea of structuralism is clear and distinct concerning phonetics, less so concerning the more complex dimensions of linguistics, such as grammar, rather less so concerning anthropology, and even less so in the analysis of literary texts. French structuralists succeeded though only for a while in convincing a number of professionals in the human sciences that structuralism was a method able to make all human sciences, from anthropology and sociology to grammar or even literary criticism for the first time genuinely scientific and moreover to unify them. Previous decades had seen Marxists create the fallacy that so-called scientific materialism could unify and make all human sciences scientific, from economics to literary criticism. This fallacy was slowly dissipated and replaced by the structuralist fallacy. The new fallacy endured until the early years of the 21st century, long after it was discredited in academic circles, because structuralist ideas were diffused from one generation to the next by college and secondary school teachers. I was convinced that these ideas were wrong and started wondering why false ideas were so easily introduced to the market by brilliant writers.

My monograph on structuralism (*À quoi sert la notion de structure? Essai sur la signification de la notion de structure dans les sciences humaine,* 1968) was well received in Britain, as Duncan MacRae's preface to the English translation (*The Uses of Structuralism*, 1971) shows, and in the US, where George Homans informed me that "at last somebody is telling the truth in France about structuralism". The book was translated into German and several other languages. But McRae saw rightly that it was "un-French". Hence my unpopularity — which was going to last a while — among many rank-and-file French sociologists. I had shown much too early that structuralism was a dead-end, and one which moreover had the effect of discrediting the far more serious and fruitful orientations which had been developed in the social sciences in the past.

Education and Social Mobility

After this critical work on structuralism, I wanted to deal with a challenging sociological question. I always believed that a good educational system and a high collective level of education is the key to progress and to collective success as well as a condition for the development of democracy and human freedom. The early sixties were characterized by a massive expansion within all Western educational systems. This was a form of progress. Democratization raised the general educational level of the population, but had little or no effect on the "equality of opportunity": the correlation between social origins and educational level, as well as the correlation between orientation status — the status of the orientation family — and destination status — the social status of the subject — was hardly reduced by the democratization of the educational system. Moreover, the inertia created by the inequality of opportunity affected all Western countries. So the topic was attractive for several reasons: this inertia was a stain on the image of democracies, since, while inequalities can be justified particularly when they are functional, inequality of opportunity cannot. Moreover, given the political and social interest in the subject, a huge body of statistical data was available. Thirdly, the problem was intellectually challenging: why was there such inertia? Fourthly, it gave me the opportunity of testing my twin ideas about how macrophenomena should be analysed as the aggregated effects of individual actions and individual actions as the effects of understandable motivations and reasons. Fifthly, the explanations then available on the market seemed to me deeply unsatisfactory. I considered that Pierre Bourdieu's explanation was rhetorical: he explained in a pedantic and tortuous style which evoked in my mind Molière's *Précieuses ridicules* that the situation was as it was because it could not be otherwise. Bourdieu and Passeron had sent the manuscript of their *Reproduction* to my friend François Bourricaud. As he told me, his first impression was that their parody of Spinoza's deductive pseudo-mathematical style was a joke or "hoax" typical of those known as *canulars*, which were traditionally in favour among the students of the École Normale Supérieure. It was not a hoax. The authors had thought that presenting their nebulous ideas in a pedantic fashion was a good strategy. My own

ideal was rather the opposite: to say complicated things as simply and clearly as possible, as Jean Cazeneuve was to state humorously in the speech he gave in 1991 at the occasion of my election to the *Académie des sciences morales et politiques*. But my main objection to Bourdieu's so-called reproduction theory was that it was fatalistic and useless from a political viewpoint.

My own theory of the inequality of opportunity proposed by contrast a practical way of lowering the inequality of opportunity. I diagnosed that reinforcing the evaluation of pupils and students, above all diversifying the educational system, insisting on the main function of schools, i.e. the transmission of knowledge, should reduce the inequality of opportunity. Amongst several other studies, a German article by V. Müller-Benedict using data drawn from the PISA study, has recently provided another confirmation of my views (*Kölner Zeitschrift für Soziologie,* dec. 2007: 615-38). Needless to say, the policy direction which my work recommended was hard to follow for political reasons in the political climate where it was published and I had few illusions about this, given that the intellectual climate of the 1960s and the two following decades was well impregnated by Rousseau's ideas on education: the child should enjoy the school, discover mathematical theorems and grammatical rules by himself, choose his values freely. Teachers were no longer allowed to teach. They could only assist children modestly in their discovery of the world. They were not allowed to evaluate the performance of the pupils.

The theory I had developed in my *Education, Equality and Social Opportunity* (first published as *L'inégalité des chances* in 1973) started from the simple idea that the educational and social ambitions of children and teenagers had their parameters set by their social milieu. For instance a person coming from a family of successful lawyers would in normal circumstances perceive the prospect of becoming a low level clerk as a social demotion, while a person from a modest workers family would see the same prospect as a success. I was proud to see that, once this hypothesis and others in the same vein were modelised, they reproduced correctly — though in a rough way — a considerable number of aggregated macrosociological data. My theory explained in particular the inertia of the level of educational and social opportunity. So, straightforward psychological assumptions, once properly formalized, were able to explain the statistical data available on the relations between educational level and social origins, as well as many other forms of data.

Many scientists in Britain, Scandinavia and the US, and some in France, such as Raymond Aron and my other French mentor, Jean Stœtzel, recognized the relevance of these ideas. Stœtzel had introduced opinion polls in France before the 1939-45 war , established the *Institut Français d'Opinion Publique* and was very active in the development of empirical sociology in France after the War. Stein Rokkan, a leading sociologist from Norway, organized a brilliant symposium on my book which produced a number of important contributions which were published in *Social science information*. A paper by Tom Fararo mathematised the first part of my simulation model. It remains a classic. Others swore exclusively by standard statistical methods. They saw *variables*

and not people as the units of sociological analysis and considered — in line with a long lasting positivistic tradition — that one should not be concerned with what people have in mind and why they do what they do, at least when it comes to scientific analysis. So, they rejected my analyses as moving heretically away from the authorized methods. I think that I had shown that *methodological individualism* was a much more natural and powerful approach to the analysis of social facts than multivariate analysis. The former enjoys an *explanatory*, the latter a mere *descriptive* power. On the whole, my approach attacked a tacit dogma and was perceived as a threat by *statistical zealots*, as Paul Lazarsfeld called them. Lazarsfeld had introduced multivariate analysis to social scientific circles through a seminal article in his *The Language of social research* and was later to inspire its sophisticated versions, such as the now fashionable log-linear analysis. But he was at the same time deeply unhappy with the mechanical methods collected under the label *data analysis*. He had received a solid scientific education, and as a result he saw clearly that *data analysis* and *explanation* are two widely different ideas. So, he welcomed my work without hesitation and told me, with his famous Viennese Jewish humour, that I had shown the Promised Land to him. In general terms my work had met with strong interest among sociologists at the international level. But in France, it also met with strong opposition from the self appointed experts in education. There was an interesting if familiar effect of such success: as Jean-Michel Morin (2007), Michel Dubois (2000) and Michel Vautier (2002) have written, it defined my scientific image: for many social scientists, I became identified as the author of this book alone.

Ideas and Beliefs

For reasons easy to understand — given the general Rousseau-esque intellectual climate that I have described above — my ideas on education had to wait before they influenced educational politics in France. In fact they never had much of an influence in political circles. Rather they emerged independently within the political sphere — but not before the last decade of the 20th century — as one effect of the reaction against the patent failure of the Rousseau-esque theories which had prevailed in the previous decades. It became more and more evident that, in conjunction with other societal factors, these ideas had produced under-education, anomie, under-employment and school violence. But as I despaired of the evolution of educational systems, which had sacrificed the traditional functions of education for the sake of enhancing the equality of opportunity, although they actually had not succeeded in raising it at all, I decided to turn to another topic: ideologies. This with the basic idea in mind that it would be better to understand why people endorse false ideas than to work on the great ideologies, such as Nazism or communism. Nazism had disappeared. Many countries were ruled by Communist parties. But it was easy to see that communist ideology was disappearing. Moreover, I saw the *great* ideologies as subjects for historians rather than social scientists, for

their implantation cannot be explained without taking all kinds of contingencies into account. As to *small* ideologies they will never disappear and are a normal component of societies, such as those which influenced in particular the politics of education and many other aspects of politics as well, such as the cult of the State or the cult of centralization, both which I saw as powerful brakes on the modernization of France. I considered these *small* ideologies to be a major sociological topic.

With his *Opium des intellectuels* (1955) Raymond Aron produced a brilliant and welcome essay on ideologies, but his contribution to their explanation was limited and had little to add to the question of why people, and in particular intellectuals and politicians, embrace false or suspect ideas so readily. Vilfredo Pareto in particular had been much more creative on the subject in his theory of *derivations*. I left aside the irrational side of the question: we have always known that passions and interests are apt to generate biased views of the world. I wondered instead whether false ideas are more likely to be generated by the normal operation of our cognitive capacities. This led me to raise the basic question as to how and why we become convinced that a theory or a statement is true or false. I started from an assumption directly opposed to Pareto's. He contrasted the true ideas derived from sound "logico-experimental" procedures to the "non-logical" ideas caused essentially, he maintained, by unconscious forces operating in the minds of people. I always felt deeply uncomfortable with the notion of unconscious forces and instead began from the viewpoint that beliefs, false and fragile ideas are generated by the same cognitive processes as those which generate true ideas. This conjecture was implicitly contained in Pareto's sarcastic statement that the history of science is a graveyard of false ideas which have been accepted for a while under the authority of scientists. Now, nobody would maintain that the numerous false ideas proposed by scientists in the past and also in the present are exclusively the result of unconscious affective, cultural and social forces. They were not produced by passions and interests either. Why should the many false ideas produced in ordinary life be the product of such forces? I felt deeply uncomfortable with such assumptions, because the existence of the hidden forces in question could only be confirmed through the effects they were supposed to produce. I saw such circular explanations as rhetorical rather than scientific.

In order to explore these questions in the light of empirical data, I used several approaches. I turned to cognitive social psychology because this discipline had established through experiment that human intuition could be deeply unreliable. I re-analysed data from this discipline and was able to show through many examples that the false answers given by subjects to the cognitive traps they were exposed to by experimenters can actually be explained as the effect of a strategy of cognitive *muddling through*. I showed in other words that, in order to explain failures of intuition, it was not necessary to assume the existence of hypothetical hidden forces, e.g. that the human brain might be wired in the wrong way as the result of some deviant evolutionary process, as some researchers have proposed. I tried to generalize the strategic interpre-

tation proposed by Daniel Kahnemann of the cognitive *biases* revealed by his experiments.

I realized then that such questions about the origins of false beliefs could be clarified not only by the experiments of cognitive social psychology, but also by sociology and anthropology. Many anthropologists and sociologists see false beliefs as explainable by the action of the hidden forces of socialisation. Subjects will accept what we see as superstitions or doubtful ideas because they have been exposed to them in their childhood, and because everybody around them accepts them as true.

Against these facile explanations, I discovered to my great satisfaction that Alexis de Tocqueville, Max Weber and even Émile Durkheim had instead put forward a rational interpretation of beliefs the observer may automatically consider to be irrational. For Weber, magical beliefs — the canonical example of the beliefs most likely to be considered as the irrational effect of hidden social forces — are actually rational. People accept them because they are grounded on an interpretation of the world, which in many of its aspects appears to them as credible and compatible with the real world, and because this interpretation has no serious competitor in their eyes. In the same fashion, Durkheim sees magical beliefs as rational. He goes as far as to say that the *primitive* — as the members of traditional societies were called in the 19th century — uses the same cognitive strategies as modern scientists. They dislike contradictions between their beliefs and the facts they observe and they try, like modern scientists, to develop auxiliary hypotheses to explain these contradictions. In the same way as modern Westerners, they ground many of their beliefs on correlations. These correlations may eventually prove spurious. As they practice rain dances, for instance, in the periods when rain is more likely to fall, they are more likely to observe a correlation between the rituals being practised and rain falling. But modern Westerners do the same. Even scientists base their beliefs on spurious correlations quite frequently. It was long thought, on the basis of spurious correlations that stress is the cause of stomach ulcers, until it was shown that it is more likely of bacterial origin.

As with magical rituals, scientific truths are currently artificially protected against scepticism and criticism by various strategies. For instance, according to an authoritative monograph on the subject, it was long considered an uncontroversial truth that bees have their own language: through their dances they were able to inform their sisters about locations where pollen is available. A systematic analysis of the scientific meetings where these questions were debated revealed however that many entomologists thought that bees are guided, like most other insects, by chemical stimuli rather than by the dance of their sisters (Wenner & Wells 1990). But the assumption that bees have a language was of course much more attractive. This hypothesis was triumphant for a while because those who were against it were not invited to the meetings where this type of question was discussed. Similarly, Lysenkoism was made credible by strategies also used in normal science. The difference is that it was

protected against criticism by the Soviet State itself, a State with powerful resources of social control.

A general assumption then could be formed, in opposition to the conceptions widely held among cognitive psychologists, sociologists and anthropologists, that in fact ordinary and methodical thinking differ from one another only in degree rather than nature. The *primitive* are no more irrational than modern Westerners. The common man struggling with a question he is not familiar with uses the same cognitive strategies as scientists, only in a less methodical way. Differences in what people know or don't know explain the differences in what they believe rather than highly hypothetical differences in the rules of inference they use, differences that would be due themselves to highly hypothetical unconscious forces.

The difference between superstitions and scientific beliefs derives from the fact that they are produced in different contexts. In a context where the laws of the transformation of energy are unknown, no difference can be detected between fire-making and rain-making. Because they do not know the laws of the transformation of energy, the *primitive* do not see any difference between the two practices and treat them as effective because they are based in their mind in the will of some spiritual forces. By contrast, to the Western observer who knows these laws fire-making appears as rational, i.e. as grounded in established laws, while rain-making appears to him as objectively groundless and for this reason objectively ineffective.

In general, spontaneously irrational explanations of beliefs should in most cases be replaced by explanations showing that these beliefs are grounded in intelligible reasons. By irrational explanations I mean those which see impersonal social, cultural, psychological or biological causes as the causes of these beliefs, instead of seeing the *reasons* people have to believe what they believe as the genuine *causes* of their beliefs. At the same time, it should be recognized that different contexts can produce different reasons. In a context where the notion of the *laws of nature* is taken for granted, unexpected and unexplainable phenomena are perceived as *miracles* by some people or as illusory phenomena by other people. In a context where the notion of the laws of nature has no significance to anybody, events can be unexpected and unexplainable, and still be perceived as normal and arouse no real surprise. To the people of the historical Middle East, miracles were an unsurprising and to this extent *normal* event because they had understandable reasons to think of them as such.

From the 1960s and even now, the *avant-garde* in the sociology of science espoused the idea that scientific beliefs cannot be considered to be objectively grounded. The "new" sociologists of science maintained that science rests on undemonstrated and non-demonstrable assumptions; that it is made of conceptual elements produced by the human mind; that human minds are moulded by the social context. Some of these arguments are true, at least in part. But they do not imply the *relativistic* conclusion the new sociologists of science drew from them. In his provocative style, Paul Feyerabend (1975) stated that the scientific vision of the world is a *fairy tale*. Following

his lead, constructivism became the ultimate truth in relation to knowledge and beliefs. As constructivism described truth as *constructed*, the distinction between grounded representations of the world and objectively groundless beliefs disappeared. The very notion of objectivity became meaningless.

This relativistic message is very far from my own views on beliefs and knowledge. While a conviction can have its parameters set by context, as when the ignorance of the laws of transformation of energy makes it possible for a person to believe that fire-making and rain-making are both produced by the interventions of spiritual forces, it can also be rationally discussed by an outside observer belonging to another cognitive context. As already mentioned, once the laws of the transformation of energy have been discovered and verified, the technique of fire-making can be rightly considered as using real natural forces, while this is not the case for techniques of rain-making. Against relativism, the views of the Western observer on the efficiency of the two types of techniques are *objectively* better grounded than those of the *primitive*. So, the relativistic message contained in the "new" sociology of science is groundless.

Being critical — in the Kantian sense — toward the "new sociology of science" seems to me very important, not only from a philosophical or sociological viewpoint, but from a political one as well. If scientific truths were the mere product of convention and construction, moral and political truths should *a fortiori* be treated as objectively groundless conventions and constructions. It seemed to me at this point that a leading cause of the political and moral disarray which characterizes many modern Western societies is the theory of knowledge and beliefs which has been developed and legitimated over the last four or five decades by the social and human sciences more generally. If scientific truths are mere conventions, why would, for instance, the idea that democracy is a better political regime than others be objectively grounded? For a number of years I have been worried, not only by the development of undemocratic practices and public decisions in democratic societies, and the many laws adopted in France recently which violate the principle of the freedom of expression, but also by the fact that, as a consequence of a growing relativism especially among intellectual and political elites, a new wave of criticism against democracy is developing among conservative intellectuals and politicians on both the left and the right. We experienced the Marxist phase and its criticism of so-called *formal democracy*, the Fascist phase which derided parliament as a *Quasselbude* (Gossip-shop), the libertarian phase of the nineteen sixties with its motto that *anything goes* and that all institutions are repressive. Now, we have the idea that democracy is just one sort of regime among others, and that it generates all kind of evils. We also have the idea that the belief in human, political and social progress has been discredited by the horrors of the 20th century, that the notions of truth and objectivity are illusions, and that the notion of the public interest merely conceals private interests.

Maybe this is the point to say that my interest in education, my great disappointment with the educational policies practised for many years in France and

elsewhere, my interest in beliefs and values were probably rooted in strong convictions based on my admiration for the philosophers of the Enlightenment, and especially for Voltaire and Kant. Both maintained that the general interest is threatened above all by false ideas. Voltaire, Kant, Tocqueville and Weber all believed that ideas are at least as important as interests for explaining social and political phenomena — and perhaps even more so. For this reason I never felt very receptive to Marx, Nietzsche and Freud, for these giants seemed to me to have clay feet. What they have in common is a belief in ideas as *dependent variables*: as the effects of unconscious social, psychological, cultural or biological forces. Needless to say, I felt even less receptive to those colleagues who took their inspiration from some vulgarized version of Marxism, Nietzscheism or Freudism and I never tried to hide it.

Finally, this theoretical reflection on the explanation of beliefs, from the false beliefs generated by the experiments of cognitive socio-psychology to the beliefs recorded by anthropologists and sociologists convinced me that the principles of *methodological individualism* were a valid method for explaining not merely statistical data of the type I had met in my work on education and mobility, but also other types of data, and especially those dealing with collective beliefs. The topic seemed highly important. I was pleased to discover that in his *Elementary Forms of the Religious Life*, Durkheim had defined collective beliefs as the main topic for sociology to explore.

Moral Feelings and Values

My ideas on the origin of beliefs attracted some attention for, alongside *L'inégalité des chances*, my book on *Le Juste et le Vrai* is mentioned in the *Petit Larousse*, an age-old venerable dictionary much used in France in schools and at home, notably to help crossword addicts to solve their puzzles. This was the starting point for developing my ideas on knowledge, beliefs, moral feelings and values.

I raised two questions in the book: one was about the origins of our *representational beliefs*, the other about our *normative beliefs*. I have just noted that the relativistic message sent by the social and human sciences in many of its publications has probably had highly negative and lasting political and social effects. Relativism ensured that many teachers no longer knew what to teach, and how to teach it, that youngsters and adults no longer knew what to think about many subjects. It ensured that social and political life was pictured as being just a confrontation of interests, and that the notion of the *general interest* was seen as a fallacy whose function was to cover up the interests of classes and corporations. It ensured that fundamental principles such the freedom of expression were violated in Europe, while others were violated in the USA after 9/11. Fortunately, the relativistic message of the social sciences is fundamentally wrong. It rests on dubious theories of knowledge and of norms.

In looking at normative and axiological beliefs, I started, (as I had with representational beliefs) by examining a basic question. Why does an individual

belonging to some context in the broadest sense of the word accept or reject any given normative or axiological belief? I started from this basic but difficult question because I had the strong feeling that, as in the case of representational beliefs, there was a lot of confusion in the field of normative and axiological beliefs. Philosophers remained mostly Kantian or neo-Kantian, while sociologists seemed more inspired by the Marxian and Nietzschean traditions. The Kantian tradition was able to explain why we accept *general* normative statements. Even before Kant, Voltaire's answer to Pascal, who doubted whether stable and objectively grounded rules can inspire normative behaviour, was that there is a single powerful rule, "followed by all nations". Do not do to others what you would not like others doing to you.

But the universe of normative and axiological feelings and beliefs is far from being exhausted by such general rules. We spend a good part of our life evaluating things, behaviours, institutions and more generally many kinds of situations. Social action is continually motivated by these evaluations. I started once again from the idea that by scrutinizing how the simplest among the myriads of prosaic evaluations we accept are grounded, we could shed some light on my basic question about normative and axiological beliefs. In a non-systematic fashion, as I had done in the case of representational beliefs, I explored a number of experimental data and theoretical explanations of evaluative data in order to answer my two questions: why do we accept or reject a given normative statement? Why and how can consensus emerge on normative issues?

During my research on this, I found that one of the most illuminating parts of Adam Smith's *Wealth of Nations* is where he wonders why his contemporaries seem to take it as self-evident that some occupations should be higher or lower paid than others. For instance he wondered why people in 18th century Britain took it as self-evident that miners should be paid more than soldiers. His answer was that this evaluative feeling is the conclusion of an implicit system of reasons containing widely accepted principles and factual uncontroversial statements. I saw the implicit theory contained in the particular analyses presented by Smith as proposing *in ovo* a general theory of normative and evaluative feelings and beliefs. I tried to develop this theory, to make it analytical and to apply it to various data, such as the empirical data I had collected. I reached the conclusion that normative and evaluative feelings and beliefs should be analysed as deriving from the systems of reasons that social actors accept more or less implicitly because they are unable to perceive a serious competing system of reasons which appear to them to be equally valid. Of course social actors are in many cases unable to arrive at such a convincing system of reasons. This is true of normative and evaluative beliefs as well as of representational beliefs.

I used these theoretical ideas to explain all kinds of phenomena and in particular to analyse a body of data I extracted from the Inglehart *et al* (1998) survey on *World Values*. I could see that on many normative questions, the English, French, Germans, Italians, Norwegians, Americans and Canadians who had been sampled gave converging answers and that the variations in their

answers were highly structured as a function in particular of age and educational level. I attempted then to explain the statistical structures characterizing the data by making them the outcome of systems of reasons in the mind of the sampled individuals. The strategy was on the whole the same as the one I had used in my work on education and mobility. My ambition had been again to transcend the *descriptive* level and to try to reach the *explanatory* level, in a context where the data used were much more raw than in my earlier study. I concluded from my analysis that many features of the data could be explained by a *rationalization* effect, in Max Weber's sense of the term. Thus, from one generation to the next, the sampled individuals displayed a more rational view of morals, religion, authority and of many other issues. These findings reveal one of the main functions of the social sciences: showing that long term trends are at work although they seem contradicted by the contingencies operating in the short term. *Rationalization processes* are threatened or thwarted by *historical forces*, i.e. by unfavourable conjunctures, stated Max Weber. The 9/11 events and the consequences they have generated have produced the impression that God was back in the Western World and that the teachings of the Enlightenment were forgotten. This impression was further reinforced by the success of the Evangelicals around the world, notably in those parts of the world where human misery, injustice or daily difficulties affect people most severely. In China itself, the government seems to have rediscovered that religion is a useful opium for the people and displays an increasing tolerance toward the many Christian or Taoist sects which are proliferating. These hard facts do not invalidate the *rationalization theory*. There is no chance that the so-called theory of intelligent design would really be accepted in the West, except by a minority of naïve believers and by the few politicians who take the idea seriously that fundamentalism can only be defeated by another form of fundamentalism.

In other writings, I have tried to show that the theory outlined by Adam Smith about feelings concerning the wages of different occupations was also sketched out by Max Weber in his widely discussed though controversial concept of *axiological rationality*. The most important in my view are the passages where he claims that social action always involves the two dimensions of instrumental and axiological rationality, and states that, although the two dimensions are always present in any actual social action, they should be considered as conceptually distinct from one another. While these works have given birth to a lasting flow of comments, with some going as far as to claim that the notion of axiological rationality is meaningless, I tried to make them analytical.

I must confess that I am surprised that theoretical notions such as these are seldom seriously discussed in the contemporary sociological literature, although they are crucial. My guess is that this state of the art results from the fact that the social and human sciences often accept the undesirable *naturalistic* principle that, as in the physical world, material and efficient causes — often called *structural* — are the only ones worth consideration in a genuinely scientific explanation. It is true that the natural sciences became scientific from the moment they substituted mechanical for final causes in their explanation of

natural phenomena. But people have intentions, desires and are able to evaluate. These features belong to their reality. Their intentions, reasons, values, preferences, goals are *facts*, even though they have to be indirectly observed or reconstructed. Human actions are not determined by social context, they are based on reasons whose parameters are set by context. Ignoring this hard fact is to doom oneself to *unrealism*. Now, how can an explanation be both scientific and unrealistic without being contradictory? It seems to me that the widely accepted failure of positivism in all its variants lies in this confusion between *realism* and *materialism*. The two notions are indistinct in the case of natural, but not of human phenomena. A failure to grasp this point is responsible for the decline of all the approaches which, like behaviourism, structuralism and the other variants of positivism, rest on the principle that human behaviour should be explained by some material causes or forces of cultural, social, psychological or biological origin rather than by reasons and motivations, as I have tried to show in my discussion with Jean-Pierre Changeux and Vincent Descombes (Bronner 2009) which for obscure reasons will remain unpublished.

Rationality

My theoretical interests naturally led me to reflect on the notion of rationality. Weber's *axiological rationality* is widely rejected by contemporary social scientists. Rationality is generally considered as exclusively instrumental (choosing the right means to reach one's goals). To Bertrand Russell (1954: viii) e.g., "*Reason* has a perfectly clear and precise meaning. It signifies the choice of the right means to an end that you wish to achieve. It has nothing whatever to do with the choice of ends". To Herbert Simon (1983: 7–8), "*Reason* is fully instrumental. It cannot tell us where to go; at best it can tell us how to get there. It is a gun for hire that can be employed in the service of any goals we have, good or bad". A consequence of this widely shared view is that the goals, ends, and values of social actors are either taken as mere facts that are worth being registered rather than explained, or explained by irrational causes, as the socialisation effects familiar to sociologists, the obscure psychological forces evoked by Freudians or the hypothetical biological forces referred to by sociobiologists. Being aware of the uncomfortable character of this situation, my friend James Coleman, another student of Lazarsfeld and Merton, proposed to apply the basic principles of economics to sociology, notably its instrumental view of rationality. This gave birth to so-called *Rational Choice Theory*.

The proposal was in part a wise one. Rational Choice Theory had been implicitly used with some success, long before it was given this name, to explain a number of problems concerned with politics, social movements, ideology and many others of interest to the field of sociology. As an obvious example I would simply refer to Mancur Olson's *Theory of collective action*. Nobody working in the field of social and political mobilization could ignore it, even if they propose to revise it in some fashion. But there are also many social facts that Rational Choice Theory is unable to explain for the obvious

reason that, as it has practically nothing to say on *normative*, *evaluative* and *representational beliefs* and on the *goals* of social actors, it has also practically nothing to say on social phenomena including normative, evaluative and representational beliefs and goals, whose explanation is not trivial matter. Now, a goal such as "staying alive" is trivial, but a goal such as "becoming a pianist" is not. Such a belief as "it is good to look to the right and left before crossing a street" is trivial; but not the belief that rain dances are an efficient means to help rain falling.

The success of RCT was understandable. It offered a solution to a widely recognized problem among contemporary sociologists: the problem of the identity — of the backbone — of sociology. At this point, I came to the conclusion that Rational Choice Theory employed a wrong theory of rationality. To be more precise it operates with an overly narrow conception of rationality. In other words, it is better thought of as a special case of a more general theory. I tried to show that this theory could be defined in an analytically acceptable fashion and applied to the explanation of a wide range of data. In doing this, I had the feeling I was merely elaborating on some implicit insights that were already present in a many great sociological works, past and present.

I must, somewhat immodestly, confess that I do not have the impression that the theory of rationality I have developed and proposed has yet received the attention it deserves. My Theory of Ordinary Rationality could, it seems to me, provide a backbone for the social sciences (Boudon 2009): a backbone with a basically cognitive orientation (Hamlin 2002). But I also have the feeling that this situation can be easily explained. The success of the social sciences, the fact that they are consulted on all kinds of questions today, has the consequence that it is much more rewarding for a social scientist, say, to produce reliable data on hot topics such as discrimination or poverty than to spend time on strategic but austere and difficult theoretical questions. Moreover, the same success has led to the development of more or less closed "corporations" among social scientists. These corporations are organized along a variety of dimensions. Some are defined by the paradigm they follow. Some by the goal they pursue: explanation of puzzling phenomena, collection of reliable data, but also political, cultural or social militancy. Many are concerned mainly with producing descriptive data on issues, such as elections or consumption, on which they aim at being recognized as experts. This heterogeneous character of the social sciences is widely recognized today. It explains why *sociological theory* and *general sociology* have practically disappeared. Not long ago, a German sociologist concluded from the present state of the social sciences that their true essence was revealed by this. They could not be a genuine science, despite Durkheim's or Weber's naive ambition, as he saw it. They are instead a type of *third culture*. But this culture is unfortunately of the *neither-nor* type: neither art nor science. It seems to me that, rather than be satisfied by the present state of the art and to christen and bless it, it is more fruitful to wonder whether this state of the art is really satisfactory, whether it optimises the production of new knowledge and finally whether it contributes to the enlightenment of

social and political actors and of citizens. Without a backbone or a grammar providing a discipline with a positive identity, it cannot be taught nor expect to be really cumulative.

As far as I am concerned, I deeply endorse Durkheim's statement that the main goal and social usefulness of sociology and the main service it can offer to society is to produce genuinely valid new knowledge on social phenomena. The tragedies which have covered and continue to cover the world and the persistence of strong inequalities in democratic societies ensure that social scientists often prefer militancy to the creation of new knowledge, while the complexity of the modern world inspires in others the idea that describing as honestly and reliably as possible the events occurring in some corner of the planet or in some dimension of the various economic, political and social activities is the only reasonable objective the social sciences can pursue.

I endorse Weber's view as well, one also shared by most classical philosophers, that men have in common a basic good sense. Albert Einstein (1936) maintained rightly that "Science is nothing more than a refinement of our everyday thinking". In the absence of this assumption many valuable ideas become empty. The idea of democracy has no meaning if it is not supposed, in agreement with the philosophers of the Enlightenment and the theorists of classical liberal democracy, that citizens forge their normative and representational beliefs on the basis of their good sense in every case where their opinions are not biased by their passions and interests (Boudon 2007). Now, many topics exclude such biases. If one prefers the irrational view of men to the rational view developed by the best philosophers and sociologists, politics becomes a mere confrontation between incompatible interests, democracy is an empty word and it is impossible to explain why, beyond the hocus-pocus of history, significant trends can be identified, such as the abolition of the death penalty in a growing number of countries. This trend is due to the fact that ideas tend in the long run to be rationally selected by the good sense of citizens, in other words by Adam Smith's *impartial spectator*. Contemporary impartial spectators no longer debate the death penalty because they recognize, in Europe at least, that it ought to be abolished everywhere for objective reasons: it is cruel, inefficient as a means of dissuasion and irreversible in case of wrong judiciary decisions. These remarks led me to propose, following in particular from Weber's and Durkheim's lead, a neo-Darwinian theory of social and political evolution where the role of mutations is fulfilled by mental innovations and the role of natural selection by rational selection.

Many political and social, as well as representational ideas, appear effectively in the long run as rationally selected. This selection supposes that men are guided in the long term by their good sense rather than by the hypothetical hidden forces so easily used by the contemporary social sciences. I see my own ideas in this respect as more beautifully expressed than I could by a quotation from Tocqueville's *Souvenirs*: "the future, enlightened and impartial judge, but who, alas, comes always too late" (*L'avenir, juge éclairé et impartial, mais qui arrive, hélas, toujours trop tard*).

A final note. *Zwei Seelen wohnen, Ach! in meiner Brust* (Goethe): "I have two souls, oh! in my breast." I worked hard because I wanted my writings to be as clear and uncontroversial as possible and for this reason rewrote many of my articles several times in order to achieve an illusory perfection. But I also appreciate not just the French *art de vivre*, as I mentioned at the beginning, but the *art de vivre* shortly. I must confess that I prefer many things to work: taking walks along the sea shore, fossil hunting, fishing, listening to my favourite composers, reading books and newspapers or sitting in *cafés*. So, the length of my list of publications results, not from my zeal at work, but from the fact that I have for many years been invited to many conferences taking place in fascinating places I wanted to experience, in Europe, America or Asia, and that I had to pay for this pleasure by writing a paper. In fact, none of my articles except the first one, a popularization of Lazarsfeld's *latent class analysis*, was written spontaneously. All are the products of these temptations. As to my books, they also testify to my basic laziness: for most of them are collections of articles drawn from these papers.

References

Boudon, R. (2007) "Modernity and the classical theory of democracy". *Tocqueville Review*, 28(1): 9–38.

— (2009) *Toward a General Theory of Rationality: A Defence of Common Sense*. Oxford, Bardwell Press.

— (2009) *Ordinary Rationality: The Backbone of the Social Sciences* (forthcoming). http://dx.doi.org/10.3917/anso.101.0019

Bronner G. (2009) *La pensée humaine en débat*: *Boudon, Changeux et Descombes*, unpublished.

Dubois M. (2000) *Premières leçons sur la sociologie de Raymond Boudon*. Paris, Presses Universitaires de France, Major.

Einstein A. (1936) "Physics and reality". *The Journal of the Franklin Institute,* 221(3), reprinted in S. Bargmann (ed.), *Ideas and opinions of Albert Einstein*. New York, Crown.

Feyerabend P. (1975) *Against Method*. London, N.L.B.

Hamlin C. (2002) *Beyond Relativism, Raymond Boudon Cognitive Sociology*. London/New York, Routledge.

Ingelhart R., Basañez M., Moreno A. (1998) *Human Values and Beliefs: A Cross-Cultural Sourcebook: Political, Religious, Sexual, and Economic Norms in 43 Societies: Findings from the 1990–1993 World Values Survey*. Ann Arbor, The University of Michigan Press.

Morin J.M. (2006) *Boudon: un sociologue classique*. Paris, L'Harmattan.

Russell B. (1954) *Human Society in Ethics and Politics*. London, Allen & Unwin.

Simon H. (1983) *Reason in Human Affairs*. Stanford, Stanford University Press.

Vautier M. (2002) *Raymond Boudon*, *Vie, Œuvres, Concepts*. Paris, Ellipses.

Wenner A. M. & Wells P. H. (1990) *Anatomy of a Controversy: the Question of a "Language" among Bees*. New York, Columbia University Press.

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2079

From positivism to generative mechanisms: Raymond Boudon's path-breaking research programme

Mohamed Cherkaoui
Emeritus Senior Research Fellow. Centre National de la Recherche Scientifique (CNRS)
cherkaoui@msh-paris.fr

Received: 18-03-2014
Accepted: 19-05-2014

Abstract

This paper briefly compares two research traditions: empiricism and generative mechanisms. Moreover, it briefly recalls the arguments of empiricist epistemology and its philosophical justification, whose paradigmatic expression is positivism. The paper then examines the criticisms that have been addressed against this trend and sets out the principles of the generative mechanisms strategy in sociology by focusing on several studies by Boudon which are the most representative of this methodological orientation.

Keywords: empiricism; epistemology; causal explanation; rationality.

Resumen. *Del positivismo a los mecanismos generativos: el programa de investigación pionero de Raymond Boudon*

Este artículo compara brevemente dos tradiciones de investigación: la del empirismo y la de los mecanismos generativos. Además, recuerda también brevemente los argumentos del empirismo epistemológico y su justificación filosófica, cuya expresión paradigmática es el positivismo. El artículo examina, pues, las críticas que se han dirigido contra esta tradición y expone los principios de la estrategia de los mecanismos generativos en sociología limitándose a las investigaciones de Boudon, que son las más representativas de esta orientación metodológica.

Palabras clave: empirismo; epistemología; explicación causal; racionalidad.

Summary

1. Introduction

The idea that a phenomenon is explained by constructing basic mechanisms that generate it is certainly not new. It can be traced back to Greek philosophers. However, it is only very recently that it has become a research strategy. By abandoning the classifications of natural history and by seeking to understand the causes of the invariance of certain characteristics of life or those viral or microbial causes responsible for particular diseases, biology was one of the early first sciences to implement the principles of the new methodology.

Breaking with the empiricist approach inherited from an interpretation of the natural philosophy of Newton, physics has been involved in the same way. Taxonomy remained nevertheless dominant but not exclusive in linguistics until recently. It is helpful to recall that formalist theoreticians, Chomsky (1964) for example, have played a critical role in the emergence of the generative approach. In sociology, systematisation and codification of generative mechanisms methodology took place later. The first attempt by Simon (1952, 1955) hardly had any influence on the sociological community. It is only with the seminal works of Boudon as well as Fararo (1969a, 1969b, 1989) that this research strategy began to be viewed as distinctive and its fruitfulness recognised.

In this contribution, I shall briefly compare the two research traditions. Moreover, without claiming to identify all the reasons that explain the almost exclusive dominance of empiricist methodology, I shall briefly recall the arguments of its epistemological and philosophical justification, whose strongest expression is positivism. I shall then examine the criticisms that have been directed at it and set out the principles of the strategy of generating models in sociology limiting myself to Boudon's research studies that are the most representative of this methodological orientation.

2. Why has empiricist epistemology been dominant?

According to the positivist programme, which had and still has a great influence on sociology, social phenomena are explained when correlations are identified by empirical analyses of functional relationships between variables. Why has such a methodological approach been dominant in social sciences? There are, I think, four reasons, namely the social demand for "cameral sciences" – following Schumpeter's phrase – that is, action-oriented sciences, the growing power of statistical techniques and computer technology, the institutional and social prestige of quantitative sociology, and finally, the intellectual comfort

that this approach provides. It is possible to identify the epistemological principles of such a mode of explaining social phenomena by studying the philosophical foundations upon which most current empirical research studies are based. However, one can reach the same objective by analysing the formal discourse of the sociological method, which has been historically the expression and justification of this strategy.

Empirical sociology today is the true forerunner of the nineteenth-century emerging sociology. It has been the result of a triple tradition, namely Comte's positivism, moral statistics and epistemological empiricism dominant in all sciences at the beginning of this century. There are surely some divergences between the protagonists of these three movements, but they actually share the same conception of this new science of society they aspired to create. If, for example, Quételet and moral statisticians on the one hand, and Comte and the positivist school on the other hand, did not agree on the use of probability theory in social sciences, they shared the same epistemological principles of empirical scientism that dominated science since the victory of Newtonian physics, or – to be more accurate – since the philosophical interpretation that most philosophers have given of this science. There are four principles:

I) According to the first principle, science must reject metaphysics and separate the empirical data from any ontology. Human beings and social phenomena are therefore "naturalised" in order to be subjected to scientific investigation, namely to observations, indirect experiments, and historical method according to Auguste Comte, and to measure according to Quételet and moral statisticians.

II) According to the second principle, the ultimate goal of scientific study is to demonstrate that phenomena comply with laws. Ideally, any explanation is supposed to be reduced to a mathematical function. Here, I would like to quote one of the most eloquent positivist texts borrowed from the *Course of Positive Philosophy*[1]:

> In the final, the positive state, the mind has given over the vain search after absolute notions, the origin and destination of the universe, and the coituses of phenomena, and applies itself to the study of their laws, that is, their invariable relations of succession and resemblance. Reasoning and observation, duly combined, are the means of this knowledge. What is now understood when we speak of an explanation of facts is simply the establishment of a connection between single phenomena and some general facts, the number of which continually diminishes with the progress of science.

This extract from the first lesson of the *Course of Positive Philosophy* meets the famous introduction of *Sur l'homme et le développement de ses facultés* (*On Man and the Development of his Faculties*) where Quételet explains his conception of naturalistic laws and macro-social regularities.

1. In this paper all translations are mine.

III) For the third principle, the rejection of any metaphysics banishes any scientific research into the realm of causes or generating mechanisms of phenomena. This is a leitmotif in the *Course of Positive Philosophy*.

> We shall admit, in physics, as a fundamental principle of the true theory of the institution of the assumptions, that any scientific hypothesis to be really tested (French text says "judgeable") should exclusively focus on the laws of phenomena and never on their modes of production.

Pierre Duhem (1906: 26), who influenced not only the Vienna Circle, but also all the twentieth-century positivists, is the heir of Comte's thought when he defines physical theory as follows: "a physical theory is not an explanation. It is a system of mathematical propositions, deduced from a small number of principles, which aims to represent as simply as completely, and as accurately as possible a set of experimental laws."

Later, Paul Samuelson (1964) says nothing else when he replies to Machlup (1964): "Scientists never 'explain' any behavior by theory or by any other hook. Every description that is superseded by a "deeper explanation" turns out upon careful examination to have been replaced by still another description."

For the father of positivism, Newton was right to limit his search to stating laws and not trying to explain, while Descartes was wrong in pretending to go further by offering accountability of laws by the swirls hypothesis. Who was really right, Descartes or Newton? For those who seek to predict, compute and act, specifically for Comte and the positivists, the English scientist approach is the right one. For those who want to understand the mechanism of movement, for example, Newton's theory is not satisfactory, and Descartes has at least the merit of having tried to decipher the puzzle in the second book of his *Principles of Philosophy* even if his explanation is laughable.

Quételet does not say anything else: he denies being a theorist, he rejects any system, and he voluntarily limits his search to the facts and to the study of their patterns.

In the eyes of positivist and social statisticians, as in the eyes of all those who were frightened by the rise of the "dangerous classes" and their "social harms," it is absurd to try to explain and understand phenomena with theories based on the rationality assumption as the Enlightenment philosophers wanted. The rational man of the eighteenth century belongs to the elite that are able to calculate probabilities, while the average man does not have this distinctive feature. During the Enlightenment, moral sciences endeavoured to reveal the rational foundations of action and belief. For them, society was also governed by laws insofar as it is an aggregation of rational individuals. In the nineteenth century on the other hand, social sciences sought to highlight social patterns and abandon the microscopic level of individual action. For them, society was governed by laws despite the irrationality of its members. It is well known that the theme of the irrationality of crowds and of the individual is commonplace in the nineteenth-century sociological literature.

More than that: even the use of assumptions, which are regarded merely as artifacts, is strictly prohibited. Auguste Comte distinguishes two kinds of assumptions. The first one, which is legitimate and authorised, concerns the analysis of phenomena in order to discover their laws. The second one, which is prohibited, is related to the nature, the cause or the mode of production of phenomena. The conclusion of the basic theory of assumptions developed in the twenty-eighth lesson of the *Course of Positive Philosophy* is that "any scientific hypothesis to be really subject to judgment should exclusively focus on the laws of phenomena and never on their modes of production". Quite obviously, the founder of positivism shares the empiricist interpretation of Newton's natural philosophy, which makes the English physicist a positivist *avant la lettre*. In fact, for the empiricists, the famous aphorism "*hypotheses non fingo*" is a profession of faith according to which Newton would have declared unlawful the search for explanations by hypothesis. Actually, the famous adage is taken from a passage of the *Mathematical Principles of Natural Philosophy* where Newton recognised that he was unable to find a hypothesis that could explain the properties of gravity. He therefore knew that gravitation requires an explanation, that is to say, something that goes beyond the law.

IV) According to the fourth principle, action is the ultimate goal of sciences, including social sciences. On this point, there is a consensus between Saint Simon, social reformers, Comte and Quételet who are the heirs of the French *Idéologues*. However, the action depends on the forecast that rests on the knowledge of the laws governing phenomena.

At the turn of the nineteenth century, some biologists and physicists broke with the traditional research programme corresponding to the macroscopic approach of Newtonian astronomy and physics. They developed a new paradigm consistent with new theories (e.g. viral, genetic, particle), and the problems posed by the micro-physical phenomena.

Claude Bernard is undoubtedly one of the first scientists to have understood the limits of empiricism and positivism, which were dominant in the late nineteenth century. In his *Introduction to Experimental Medicine* and in his *Principles of Experimental Medicine*, he opposed the empiricist doctor of the Hippocratic School to which the positivists belonged, and the experimentalist physician he represents himself. While the first observes and describes; the second leads experimentations, that is to say experiences brought about according to theoretical assumptions. If the empiricist relies on the statistical relationship between treatments and diseases, he is however unable to answer the question of why the administration of such pharmacopoeia has an effect on this particular disease.

"Empiricism stops science and dulls the mind, when one rests on it," Bernard (1947: 75) says. "The empiricist is satisfied when he manages to heal. The experimental doctor wants to go further and to penetrate with experimentation by explaining the vital mechanisms. The real goal of the researcher is to know and understand the generative mechanisms of the disease or, in the words of Claude Bernard (1947: 137), "the mechanism producing disease."

One perceives a similar attitude in physics at the beginning of the twentieth century that Emile Meyerson (1921: 62) admirably summed up:

> Just have a look at the collection published by the Council (the Brussels Congress of Physics in 1911) that reflects the communications that have been presented and the discussions they provoked, to find that the sole purpose of all this work was the search for a true physical theory, an assumption on the mode of production (so odious to Auguste Comte and inadmissible, indeed, according to his conception of science).

The strength of positivism is largely due to the belief according to which it is a philosophy that expresses, founds and justifies the new science. In its various forms, positivism was a philosophy of the nineteenth-century scholar. But it sometimes continues to play the same role today. As a research strategy seeking to explain nomologically and providing psychological satisfaction with a minimum intellectual investment, positivism is ineradicable because it is a lifeline to which clings the researcher who is often unable to speculate on the generative mechanisms of phenomena. At times, it happens we are positivist, and at times we reject positivism.

In the nineteenth century, few people understood the negative consequences of these epistemological principles and attempted to replace them with others. In his *Rules of the sociological method*, Durkheim (1895) remained a prisoner of the positivist dogma. However, in his research studies, the author of *Suicide* and *The Elementary Forms of Religious Life* tried to break with certain principles of positivist philosophy by demonstrating the inadequacy of nomological activity and the need to explain the laws themselves. It has been shown (Boudon, 1990, 1998b; Cherkaoui, 1997, 1998, 2008) that a sketch of a new strategy based on the search of the modes of production of social phenomena can be found in the work of the French sociologist that he neither developed nor systematised. It should be added that no one has been able to achieve this programme, even the most anti-positivist sociologist of his generation, namely Gabriel Tarde.

Let us now turn to contemporary sociology. Many research studies belong partly to the empiricist tradition. The methodology of multivariate analyses of large-scale quantitative data is based on the principles of the positivist model. This is the case of the work of Stouffer, Lazarsfeld and the Columbia school although some of them, including Lazarsfeld and Merton, expressed dissatisfaction here and there with the nomological explanation and tried to formulate hypotheses whose kinship with generative models is obvious. However, such assumptions remain *ad hoc* and *ex post*.

3. Boudon's new research strategy

In the 1960s, a new research programme in sociology appeared simultaneously and independently in Boudon's and Fararo's works, to mention only the most

representative work of this new trend which broke with the positivist model and its nomological explanation. The mode of explanation of this new strategy was to highlight the mechanisms that generate regularities of phenomena in a way identical to biology, physics or linguistics.

The merit of Boudon is to have clarified the principles of such a strategy, and to have tested them in such a vital domain as the sociology of mobility and social inequalities, proving their fruitfulness, and expanding their fields of application. He clearly saw their links with the paradigm of methodological individualism and the general theory of rationality, both of which he devoted the rest of his intellectual life to (Boudon 2009, 2010). It is noticeable that, at the same time, and obviously independently, Harré (1970) proposed a realistic philosophy advocating the same perspective.

The idea of explaining by social mechanisms runs through the work of Tocqueville, Durkheim, Simmel, Weber, or Merton (Cherkaoui, 2005), but in none of them was it recognised, developed and applied to such wide areas as did Boudon, Fararo, Simon (1952, 1955, 1968) or later Schelling (1978). Moreover, in the early 1970s, Boudon's approach was so unusual that even such acute minds as Hauser (1974) had trouble in understanding it. His critical review of Boudon's *Opportunity and Social Inequality* was evidence of a deep and symptomatic misunderstanding: there is indeed a dichotomy between the positivist empiricism that dominated research on mobility in particular and sociology in general, and the new research programme proposed by Boudon.

A quick examination of the development and directions displayed in Boudon's work offer a division of his oeuvre into three stages. The first period is dominated by contributions to the construction of statistical and mathematical models in the tradition of his friend Lazarsfeld (Boudon, 1967). A second period covers the late 1960s and 1970s up to the publication of *La place du désordre* in 1984 (translated in *Theories of social change*, 1986). It is characterised by the first version of the theory of rationality applied to social mobility, inequality and change. The third period has been marked by his efforts to generalise his theory and extend it to the most various topics and issues, such as ideological, scientific, and religious beliefs (Boudon, 1999, 2009, 2011).

It would not be difficult to point out the change of direction in Boudon's sociological work that first became evident at the end of the 1960s. He moved away from the studies whose main inspiration was the Lazarsfeldian tradition to a novel form of sociology whose primary concern was the construction of a new strategy based on the discovery of generative mechanisms; an approach systematised in Boudon (1973) *L'inégalité des chances* (translated in *Education, Opportunity and Social Inequality*, 1974), but actually visible in two articles written in the late sixties and published in the European journal *Quality and Quantity*, with the significant titles "Essai sur la mobilité sociale en Utopie" (Essay on social mobility in utopia) and "Éléments pour une théorie formelle de la mobilité sociale" (Elements for a formal theory of social mobility). In fact, this innovation was not an *ex nihilo* creation. Boudon had been for some time seeking a new methodology that could allow sociology to get out of the

rut in which it was stuck. In fact, in the article "La statistique psychologique de Tarde" which appeared in 1964, the premises of the new methodological orientation were revealed. Another article published in the same year by Boudon (1964b) with an even more suggestive title "Les mécanismes sociaux des abandons de poursuite" (Social mechanisms of prosecution dropouts) provides a still better example of how the new methodology could be applied to a specific area.

Boudon's methodology rests on five principles of analysis that will inform his later work. Firstly, he refuses to consider the correlation analysis or any standard statistical technique as an explanatory mode of phenomena. Secondly, he regards as impossible or at least highly difficult to draw any reliable conclusions from a direct reading of data tables since any table is only a provisional stage that is unable to express correctly the social process that generates it. Thirdly, for him sociological phenomenon is the result of complex social processes which require the use of longitudinal analyses. Fourthly, macrophenomena are the results of behaviours of partially autonomous social actors and their interaction structure (independence or interdependence). Finally, it seems necessary to build psychological and cognitive models to understand these behaviours.

Let us examine some significant aspects of Boudon's interpretation of Tarde's criminal studies. Tarde, he notes, is less interested in the explanation of crime but focuses on that of mechanisms of repression. For him, the statistical series to be explained are the joint product of an act, crime, and the judiciary institution. This judiciary may "class" without result, suppress or discharge. The statistical regularities are largely the product of judicial treatment. However, these macro-patterns are the result of interactions between actors with specific social roles and that mutually influence each other. To understand the functioning of judicial institutions, it is necessary to distinguish two sets of actors. The first, the magistrates, is characterised by interaction or reciprocal influences. The defendants and juries, who form the second set of actors, do not interact with each other. This distinction helps to explain the differences between some longitudinal data series: on the one hand, the proportion of prisoners who appeal the decision of criminal courts remain stable over time, while the proportion of appeals by the prosecution decreases during the same period. In the first case, the actors' independence explains the stability, while the interdependence between judges who monitor each other and anticipate their respective decisions accounts for the decline.

4. On phenomenalism and its critique

The positivist explains by establishing laws that he seeks to derive from other more general laws and so on. In its simplest form, the lawful relation is an expression, often mathematical, between at least two phenomena such as distance and time in Galileo-Descartes's law of falling bodies, or suicide and marital status in Durkheim's (1897) theory of integration. These relations are

empirical generalisations or laws with limited validity. The positivist explains the law of falling bodies by deriving it from Newton's laws of gravitation and motion. But sometimes he will face problems that cannot be solved in the context of his epistemology. To solve them, he will therefore be condemned to violate some of his principles. Let us limit ourselves to the following sociological examples.

Suppose we study the relationship between two variables X and Y. Assume that a correlation between the two variables is deduced from the analysis of empirical data from a cross-sectional survey. The empiricist will, however, face many problems if he wants to deduce certain consequences from this empirical fact.

Firstly, he cannot be sure that the relationship is a real one. Admittedly, he uses the basic rules of multivariate analysis to test whether or not the relationship is a true one by taking into account control variables. Actually, he is unsure that all relevant control variables have been taken into consideration in his analyses. Moreover, the assumption of a closing system of relations between variables is only a convenient fiction – the assumption allows us to suppose that exogenous factors (i.e. not controlled) are not simultaneously correlated with independent and dependent variables, but it does not tell us anything about the relevance of our choice of endogenous and exogenous variables. The *ceteris paribus* rule is itself no longer of any help.

Secondly, we should consider the case where data from a new survey show no correlation between the variables under study. If so, would we be allowed to conclude that these new data falsify this correlation? Certainly not! We therefore are in an undecidable situation.

Thirdly, the dependence between variables sometimes says nothing on the meaning of their relationship. Does the growth of education precede industrialisation? Or does the reverse hold? Should we rather suppose a retroaction effect between them? Here again, the empirical analysis does not allow us to decide.

Fourthly, if any modelling of the relationship between two phenomena (i.e. any identification of nomological invariance) is built on the basis of N observations, it is therefore possible to construct empirically not *one* but a very large number of mathematical "laws". Can we empirically make sure that a specific law is valid? Certainly not! All those laws are empirically valid. Actually, we often apply the principle of simplicity to choose a model among all those which are theoretically possible. However, this principle is frequently violated for theoretical reasons. Suppose an empiricist wishes to study the relationship between the size of the family and suicide rate on the basis of data provided by Durkheim in *Suicide*. The first order linear regression model between the dependent variable (S) and the independent variable (F) provides an excellent fit which gives a value of R^2=0.915 .

$$S = 1609.6 - 3.64\ F$$

The linear regression model with a negative slope expresses the Durkheimian proposition according to which when the density of the family increases,

the likelihood of suicide decreases. Such a result would satisfy the most demanding empiricist. A more complex model is, however, preferred because it is deduced from the theory of integration, according to which when integration increases, the suicide rate decreases and then increases. This theoretically based proposition is expressed by a parabolic model. From the same empirical data, the estimated parameters are

$$S = 7515\text{-}34{,}051\ F + 0{,}039\ F^2$$

In this case, the value of R^2=0.97 is greater than that given by the first order regression. I am not saying that the parabolic model is better than the first-order model on the basis of the value of R^2. If the parabolic model "explains" a greater percentage of the variance than the linear one does, it is because it contains an additional parameter. Choosing a model rests on the sole theory and not on the results of analyses of empirical data.

The simplicity principle cannot be therefore a selection criterion. If we base our choice on numerous examples borrowed from the history of science, we will be inclined to believe that frequently the more complex model is preferable to the simplest one. A classic example is the astronomical theory that describes the motion of planets around the sun. Kepler's ellipsoidal model is more complex than Copernicus's circular model.

To explain the correlation and make it understandable, the empiricist is usually forced to construct hypotheses on the mechanisms that generate empirical regularity. In doing so, he is no more an empiricist. But this is a *post hoc* strategy of generative mechanisms. In general, what he calls interpretative assumptions at the end of his research are mostly *ad hoc* and cannot be generalised.

Let us borrow from Boudon (1973) a suggestive example of an *ad hoc* mechanism. Different studies have demonstrated that social mobility changes according to various factors such as the increase in school enrolment, economic development, the nature of stratification such as the presence of legally established strata in the past as states in Europe. Lipset and Zetterberg (1956) noted that, contrary to what was thought, the mobility rate was only very slightly different in Europe and in the United States. They tried to improve the theory by introducing the hypothesis of the existence of a mechanism that would explain why the aspiration to climb the social ladder is small. According them, this aspiration is especially low when social barriers are less visible: where social distinctions are imperceptible, the income inequality is low and the standard of living is high, and the individual does not seek to improve his/her social status. If the United States has a mobility rate similar to that in Europe, this is due to factors whose conflicting effects nullify each other: the United States has no legally defined social hierarchy as in Europe; it has a less visible stratification than in the old continent. The combination of greater ease and less aspiration to climb the social ladder therefore makes intelligible the relative uniformity of mobility rates between societies whose systems of stratification are different. We have here a rudimentary but good example of the use of a *post hoc* mechanism.

As Boudon (1973) emphasises, the explanation being *ad hoc*, the only consequence that can be drawn is the proposition itself it seeks to explain. Boudon does not claim that what he calls factor analysis can be of no help despite its limitations and weaknesses. He disqualifies it as a scientific research strategy insofar as its objective is to explain and to understand phenomena. The aim of research in the empiricist approach is to find a relationship between the "*explicandum*" and the "*explanans*" as in statistical analyses. In the most favourable interpretation, this research strategy sometimes assumes the existence of a black box between the two variables that is either of no interest or difficult to observe, to decompose and to describe, as Bunge notes (1967). We will see later how the theory of rationality tries to avoid the existence of the black box in explaining phenomena.

Boudon goes further in his critique of factor analysis. He observes that whatever the degree of generality of the observed relations between variables, and assuming they are not spurious, they remain unintelligible and therefore offer no explanation insofar as we do not identify the generative mechanisms of phenomena and their co-occurrences, that is to say, essentially the reasons for actors' behaviours. There are reasons that convert the power of the actor to perform actions. Actions are therefore only observable effects of reasons. Reasons are not causes in Hume's meaning; they are yet intrinsic traits of actions. They are logically related to actions. Of course, the positivist can always protest against the fact that reasons are empirically inaccessible and that they remind him of metaphysical entities. Actually, this argument is unconvincing since, like physicists and biologists, we accept that unobserved entities may be requested to explain, providing that their traces are empirically controlled.

In fact, identifying dependency relationships between phenomena is a particular and feeble case of nomological activity. Suppose that the sociologist points out invariant relations between phenomena, and that he can therefore subsume under laws the phenomena he is studying. According to Hempel's (1965) deductive-nomological theory, a phenomenon is explained when one can deduce it from a general law and its particular antecedents. However, there are several general laws that do not explain the relationship between the phenomena they bind and do not make them comprehensible. One can recall many regularities in physics for instance. Consider the examples borrowed from Hempel (1965). Kepler's laws describe well the path of the planets around the sun, but do not explain it. The ideal gas law PV = a T that links the pressure P, volume V and temperature T, specifies how each factor varies depending on the other two, but tells us nothing about the mechanisms that account for these relationships. The explanation was possible later with the kinetic molecular theory. The law is indubitably derived from the molecular theory; nevertheless, the explanation does not reside in this deduction. It lies in the fact that the macro-phenomena are caused by microscopic phenomena according to theoretically built mechanisms. We will see later that what is valid in physics and biology is valid in sociology or social sciences.

It is clear that, according to the deductive-nomological model, to explain a phenomenon is solely to subsume it under laws and to deduce these laws from more general laws, exactly as required by Auguste Comte. It is true that this model does not justify the explanation by black boxes, but it certainly contributes to moving the scientist away from the search for generative mechanisms anathematised by Auguste Comte.

5. From generative mechanisms to a general theory of rationality

We are not always able to produce explanations without black boxes. When we use collective concepts or notions that we are unable to deduce from the structures of individual interactions, or when we ignore their mode of production, we can be assured that black boxes exist in the chain of reasons we give to make intelligible the phenomenon under study. How can we explain the behaviour of individuals in a crowd? How is it that "an assembly of harmless bourgeois can turn into a fearsome monster", according to Durkheim's (1897) famous phrase? For LeBon (1895), the individual undergoes a radical transformation in a crowd that makes the emergence of primitive and irrational elements possible. The person loses control of himself and behaves like an animal. The basic mechanism by which LeBon elucidates this behaviour is the "suggestion", that is, accepted orders given to that person. The person is literally hypnotised. Why does the "primitive" believe in magic? For Lévy-Bruhl (1922), their magical beliefs are explained by the fact that they have the "primitive mentality" different from the "civilised mentality". Why do students from different social origins formulate different educational demands although they are of the same age, educational level and scholastic attainment? Why do working class students prefer lower-level education than upper-middle class students? To explain this correlation, let us suppose that the sociologist argues that it is because children are socialised differently or that they do not have the same "habitus". Why do French social scientists produce more books than scientific articles compared to their American colleagues (Boudon, 1986; Cherkaoui, 2011 for a systematisation and a generalisation of the market theory based on empirical data)? A possible answer is that France has a "cultural specificity". "Suggestion", "primitive mentality", "socialisation", "habitus", and "cultural specificity" are black boxes insofar as we are unable to specify the mechanism that meets these vague notions.

How can we explain without black boxes, according to Boudon (1998a)? He distinguishes two types of explanation with mechanisms. The first one "gives the impression that the explanation is final" in the sense that our search ends. Being satisfied at least temporarily by the responses we give, we do not ask more questions. For the second one, the explanation does not seem to have exhausted all the questions that are likely to be asked. In some cases indeed, there are theories that use mechanisms but do not belong to the methodological individualist paradigm, as in the example of the explanation of the demand for education by socialisation (or some macro-structural theories).

Let us examine the example of crowd behaviour. Coleman (1990) shows that the structure of the game of this phenomenon is not comparable to that of the prisoner's dilemma since, in crowds, there is communication between the actors. We cannot therefore assert that there is a dominant strategy, which would be running to the fire exit in the event of a fire in a theatre. He notes that the concepts of suggestion, contagion, and social facilitation can be interpreted in terms of "transfer of power." In some situations indeed, the individual divests himself of the power or control he normally has of his own actions and transfers it to others. This interpretation has the advantage of making the action the result of an intentional act and allows us to ask the question why and in what circumstances the rational actor operates such a transfer. In addition, such a perspective can make the difference between groups; for example, between those who have potentially extremist behaviour and the others.

The behaviour of an assembly or a crowd in a closed space does not necessarily result in panic; it depends on the initial action of the actors who have not transferred their power to others: if it is quiet, the subsequent actions will also be so. Suppose that player A knows that the other players act independently of his behaviour. In his interest, he should try to quickly leave the room because he knows that his action has no effect on the actions of others. If, however, A knows that his action has an effect on others, he knows at the same time that fleeing can result in a traffic jam whose costs will be very high for him and for others. Escaping would be beneficial for him only if he is next to a door. It is more advantageous for him than for the others to leave the room in order. If the actions of others depend entirely on his action, it is in his interest to dictate orders so that the evacuation takes place peacefully. The situation is different if he assumes that each of the other group members does not act independently of him but has transferred to him and to others partial control. A can estimate that it is his interest either:

1) to run to the exit regardless of what others do,
2) to try to go out calmly regardless of what others do,
3) to try to go out quietly firstly, and secondly matching his action to that of others. All this depends on the degree of power that others have transferred to him, the probability that they will run towards the exit, and the benefits actor A can gain from each situation. We find ourselves in a situation characterised by the interdependence between actors, where an actor cannot define a rational strategy if he ignores the strategy of others.

Coleman's explanation is satisfactory and temporarily ultimate for the following reasons:

1) It answers all our why-questions,
2) It takes into account all possible scenarios (fleeing, go out in order, etc.). Unlike the theory that seeks to explain the consequences of only one behaviour, such as bottling,

3) It uses the same assumptions to account for all behaviours, contrary to the assumption according to which the actor is sometimes rational and sometimes irrational,
4) It generates the macro-phenomena from micro-social behaviours, that is to say behaviours of individuals that have the same properties but that are involved in different interaction structures.

The fact that the ultimate causes are related to individual decisions implies that the explanation by a social mechanism has been made in terms of the methodological individualist paradigm.

As a model that abstractly reproduces the phenomenon to be explained, the generative mechanism provides an interpretation in terms of individual behaviours. It considers the social actor as the sociological atom who is generally individual. It excludes structures and therefore any attempt to reify them, as Weber (1922b) taught us. The sociological explanation is intelligible because it refers to intentions, reasons or causes and consequences of the individual agent actions. But if the paradigm of methodological individualism is necessarily associated with the strategy of generative mechanisms, it is however not always easy to apply. In fact, according to the epistemological status of a particular research domain, it is sometimes difficult or even temporarily impossible to give an explanation at the individual level (see Cherkaoui, 2007).

Two options are open. In the first one, the most radical, the researcher refuses to take into consideration in any explanation macro-social phenomena or normative concepts that are not explained and reduced to their individual behaviours. In the second one, which is more flexible, one provisionally accepts these macro-phenomena. As elements of our explanation, norms, rules or institutions for example are provisionally accepted in this second option and rejected in the first one, even though both versions agree that these variables are an intended or unintended consequence of individual interactions or the result of the crystallisation of long historical processes that the researcher is not yet able to explain but should make intelligible later (Cherkaoui, 2007).

One can therefore understand why Boudon (1998a) draws attention to the fact that the methodological individualist paradigm offers no immunity against the presence of black boxes in a sociological explanation. It is a necessary but not sufficient condition. To be protected against any risk of infection, one should couple this paradigm with the assumption of rationality. If we want our explanation to be entirely satisfactory, we should express the ultimate causes in fully understandable individual actions, namely based on reasons. In other words, those social actors are not victim to any hidden and unconscious force (Boudon, 1979). Insofar as it is based on the assumption of rationality – such as the example of the explanation of collective behaviour borrowed from Coleman – the rational choice model guarantees against any explanation infected by black boxes. It leads to ultimate explanations that do not require any additional questions or whose contents cannot be substantially improved by adding new information produced outside its frame.

Boudon timely points out that in the rational choice model, "rationality" is defined by the cost-benefit operation or expected utility. It therefore cannot legitimately claim to generality, and is unable to be applied to the most different sociological phenomena. Boudon offers four critical examples. The first one concerns the demand for education mentioned above. The expected utility model is unable to give an account of the interaction effect between students' social background, their level of academic attainment and their demand for further education. In contrast, Boudon's (1973, 2007, 2010) cognitive theory is able to explain why the influence of students' social class is even higher on demand whereas the academic achievement is lower.

The second one is related to the voting paradox to which Boudon (1997) devotes a full analysis. According to the rational choice model prediction, citizens should not vote to the extent that the costs of voting largely outweigh the benefits. Yet they vote. The third example focuses on what he calls "the overreaction paradox". In many circumstances, the violent individual reaction to insignificant daily life events is disproportionate to the costs. Sometimes costs are immeasurable compared to profit, which is zero or negative.

Finally, the fourth example, which is the most important, is related to the explanation of beliefs. In the last three cases, we are not forced to accept the heavy assumption of actors' irrationality. We should simply question the claim to generality of rationality as supposed by the expected utility model. In addition, cognitive and normative statements as well as beliefs concerning ends and not means are beyond the domain of validity of the rational choice model. This model can account for beliefs formally translated by the statement "X is good because X removes unwanted consequences" since in this proposition X refers to its consequences. But it does not apply to non-sequential normative beliefs that have nothing to do with the consequences of X. When we say "X is good," X does not refer to its possible consequences. Such statements refer to what Weber meant by "axiological" rationality or "Wertrationalität" opposed to "Zweckrationalität" or instrumental rationality that is mean-end rationality.

Boudon's cognitive theory extends the notion of rationality and generalises it in order to integrate instrumental rationality, cognitive rationality and axiological rationality. It is based on four assumptions. For the first one, actions, beliefs and attitudes are perceived by the actor as meaningful, that is to say, based on reasons. The other three specify the conditions under which the three types of rationality are applied. It should be noted that this theory does not apply to all actions and accepts the existence of a-or irrational actions in the meaning Weber (1913, 1922) gives to traditional and *affectual* behaviours (particularly emotional), two of the four types of actions with Zweckrationalität and Wertrationalität actions. But a good research strategy must first begin by identifying the reasons behind the action or belief and not by using the a-or irrationality hypothesis that could lead to an aporia.

By reconstructing the chain of the actor's supposed reasons, social science proposes a theory that generates empirically testable propositions.

The rational choice model becomes a special case of Boudon's cognitive theory for different reasons. While offering the same epistemological benefits as the first, by for example allowing the construction of hypothetico-deductive and predictive models and explaining mean-end actions, Boudon's theory is more general than the expected utility model that is unable to make intelligible non mean-end phenomena. If cognitive theory gives the impression that it is not as unified as the rational choice model, it is due to the diversity of the types of reasons it takes into account and which depends on the nature of the circumstances in which actions, beliefs and attitudes are taken in or adopted.

Bibliographic references

Bernard, Claude (1865). *Introduction à l'étude de la médecine expérimentale,* 3rd edition. Paris: Delagrave, 1912.

— (1947). *Principes de médecine expérimentale.* Paris: Presses Universitaires de France.

Boudon, Raymond (1964a). "La statistique psychologique de Tarde". *Annales internationales de criminologie*. In: Boudon (1971).

— (1964b). "Les mécanismes sociaux des abandons de poursuite". *Année sociologique,* 111-244.

— (1967). *L'analyse mathématique des faits sociaux*, Paris, Plon.

— (1970). "Essai sur la mobilité sociale en Utopie". *Quality and Quantity,* IV, 2, 212-242.

— (1971). "Éléments pour une théorie formelle de la mobilité". *Quality and Quantity,* V, 2, 39-86.

— (1973). *L'inégalité des chances. La mobilité sociale dans les sociétés industrielles*. Paris: A. Colin.

— (1979). *La logique du social*. Paris: Hachette.

— (1984). *La place du désordre*. Paris: Presses Universitaires de France; tr. *Theories of social change*, 1986.

— (1986). "L'intellectuel et ses publics". In: Reynaud, J.D. and Grafmeyer, Y. (eds.) *Français, qui êtes-vous ? Des essais et des chiffres*. Paris: La Documentation française.

— (1990). *L'art de se persuader.* Paris: Fayard. Tr. *The Art of Self-Persuasion. The Social Explanation of False Beliefs*. Polity, 1997.

— (1997). "Peut-on être positiviste aujourd'hui". In: Cuin, C.-H. (ed. 1997). *Durkheim d'un siècle à l'autre. Lectures actuelles des Règles de la méthode sociologique,* Paris: Presses Universitaires de France.

— (1998a). "Social mechanisms without black bloxes". In: Hedstrom, P. and Swedberg, R. (eds.). *Social mechanisms. An analytical approach to social theory*. Cambridge: Cambridge University Press.

— (1998b). "Les formes élémentaires de la vie religieuse: une théorie toujours vivante". *Année Sociologique*, vol. 49, 149-198.

— (1999). *Le sens des valeurs*. Paris: Presses Universitaires de France. *The Origin of Values*, 2013.

— (2007). *Essai sur la théorie de la rationalité*. Paris: Presses Universitaires de France.

— (2009). "Why I became a sociologist". In: Cherkaoui, M. and Hamilton, P. (eds.) *Raymond Boudon. A Life in Sociology*, vol. I, 53-75. Oxford: Bardwell Press..

— (2010). *La sociologie comme science*. Paris: La Découverte, Eng.tr. *Sociology as Science*. Oxford: Bardwell Press, 2013.

— (2011). "La nature de la religion selon Durkheim". In: Boudon, R. (ed.). *Durkheim fut-il durkheimien*. Paris: A. Colin.
Bunge, Mario (1967). "Scientific Research". In: *The studies of the foundations, methodology, and philosophy of science*. Berlin: Springer Verlag.
Cherkaoui, Mohamed (1997). "La théorie de l'explication chez Durkheim". In: Cuin, C.H. (ed.). *Durkheim d'un siècle à l'autre. Lectures actuelles des Règles de la méthode sociologique*. Paris: Presses Universitaires de France.
— (1998). *Naissance d'une science sociale. La sociologie selon Durkheim*. Geneva: Droz.
— (2005). *Invisible codes*. Oxford: Bardwell Press.
— (2007). *Good intentions*. Oxford: Bardwell Press.
— (2008). *Durkheim and the puzzle of social complexity*. Oxford: Bardwell Press.
— (2011). *Crise de l'Université. Le nouvel esprit académique et la sécularisation de la production intellectuelle*. Geneva-Paris: Droz.
Chomsky, Norman (1964). "Current issues in linguistic theory". In: Fodor, J. A. and Katz, J. J. (eds.). The structure of Language. Readings in the philosophy *of language*. Prentice Hall: Englewood Cliffs.
Coleman, James S. (1990). *Foundations of social theory.* Cambridge: The Belknap Press of Harvard University Press.
Comte, Auguste (1830-1842). *Cours de philosophie positive.* New ed. Paris: Hermann, 1975.
Descartes, René (1644). *Principes de la philosophie.* New ed. Adam and Tannery, vol. VIII-1. Paris: Vrin, 1964.
Duhem, Pierre (1906). *La théorie physique. Son objet, sa structure*. New ed. Paris: Vrin 2007
Durkheim, Émile (1895). *Les règles de la méthode sociologique.* Paris: Alcan. New ed. Paris: Presses Universitaires de France, 1963.
— (1897). *Le suicide. Étude de sociologie.* Paris: Alcan (2nd ed. 1912).
— (1912) *Les formes élémentaires de la vie religieuse,* Paris: Alcan. New ed. Paris: Presses Universitaires de France, 1968.
Fararo, Thomas (1969a). "The nature of mathematical sociology". *Social Research,* XXXVI, 75-92.
— (1969b). "Stochastic processes". In: Borgatta, Edgar F. (ed). *Sociological methodology.* San Francisco: Josey Bass.
— (1989). *The meaning of General Theoretical Sociology. Tradition and formalization*. Cambridge: Cambridge University Press.
Harré, Rom (1970). *Principles of Scientific Thinking.* Chicago: University of Chicago Press.
Hauser, Robert M. (1974). "Review essay: On Boudon's model of social mobility". *American Journal of Sociology,* LXXXI, 911-928.
Hempel, Carl (1965). *Aspects of scientific explanation and other essays in the philosophy of science.* New York: The Free Press.
LeBon, Gustave (1895). *Psychologie des foules.* New ed. Retz: Paris, CEPL, 1975.
Lévy-Bruhl, Lucien (1922). *La mentalité primitive.* Paris: Alcan. New ed. Paris: Presses Universitaires de France, 1960.
Lipset, Seymour M. and Zetterberg, Hans (1956). "A theory of social mobility". In: *The transactions of the Third World Congress.* London: International Sociological Association, 155-177.
Machlup, Fritz (1964). "Professor Samuelson on theory and realism". *American Economic Review*, 55, 733-36.

Meyerson, Emile (1921). *De l'explication dans les sciences*. París: Fayard.
Newton, Isaac (1687). ***Philosophiae Naturalis Principia Mathematica***, tr. *Mathematical Principles of Natural Philosophy.*
Quételet, Adolphe (1835). *Sur l'homme et le développement de ses facultés ou essai de physique sociale.* Paris: Bachelier. New ed. Paris: Fayard, 1991.
Samuelson, Paul (1964). "Theory and realism: A reply", *American Economic Review*, vol. 54, 736-39.
Schelling, Thomas (1978). *Micromotives and macrobehavior.* New York: Norton and Co.
Simon, Herbert (1952). "A formal theory of interaction in social groups". *American Sociological Review,* XVII. In: Simon (1957) *Models of man. Social and rational.* New York: John Wiley.
— (1955). "Mechanisms involved in pressures toward uniformity in groups". *Psychological Review,* LXII. In: Simon (1957) *Models of man. Social and rational.* New York: John Wiley.
— (1968). "On judging the plausibility of theories". In: Simon, Herbert (1977). *Models of discovery.* Dordrecht: Reidel Publishing Co.
Tarde, Gabriel (1895). *Les lois de l'imitation.* New ed. Geneva: Slatkine, 1979.
Weber, Max (1913). "Über einige Kategorien der verstehenden Soziologie". In: Weber, Max (1922). *Aufsätze zur Wissenschaftslehre.* Tübingen: Mohr.
— (1922). *Wirtschaft und Gesellschaft.* Tübingen: Mohr.

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2068

Raymond Boudon: A review*

Ángeles Lizón Ramón
Universidad Nacional Autónoma de México. Centro de Estudios Básicos en Teoría Social
Universitat Autònoma de Barcelona. Department of Sociology. Currently retired.
angeles.lizon@uab.cat

Received: 18-02-2014
Accepted: 19-05-2014

Abstract

A brief retrospective on Boudon's academic contribution through the three major stages of his academic career (path regression models, game theoretical mechanisms and subjective rationality) allows one to retain as his major contribution the proposal of theory construction by means of causal intentional mechanisms based on models of strategic decision-making. Recurrent structures of cogent systems of preferences formalized in game theoretical language seemed to offer him a suitable way to characterize some of the main or more significant macro-social effects in terms of micro-decisions taken by individual intentional agents.

Keywords: game theory and strategic analysis; rational choice theory; decision-making mechanisms; game theory analytical tools; methodological individualism; explanation by mechanisms and theory construction; subjective or cognitive rationality.

Resumen. *Raymond Boudon: una retrospectiva*

Atendiendo a los tres estadios centrales de su carrera (modelos de regresión y análisis de sendero, modelos de juegos de estrategia y racionalidad subjetiva), esta breve retrospectiva sobre Raymond Boudon pretende señalar como su mayor contribución a la sociología de la segunda mitad del siglo xx la propuesta de construcción de teorías de mecanismo intencional. Formalizadas en lenguaje de juegos de estrategia, un significativo número de estructuras recurrentes de preferencias le proporcionaron una forma innovadora de caracterizar efectos macrosociales significativos en términos de microdecisiones de agentes intencionales.

Palabras clave: teoría de juegos y análisis estratégico; teoría de la elección racional; mecanismos de toma de decisiones; herramientas analíticas de teoría de juegos; individualismo metodológico; explicación por mecanismos y construcción de teoría; racionalidad subjetiva o cognitiva.

* A previous version of this paper was published in Mohamed Cherkaoui and Peter Hamilton (eds.). *Raymond Boudon: a life in sociology*. Oxford: Bardwell Press, 2009.

Summary

1. Introduction

If generating social mechanisms as a strategy for theory building is a useful tool, then there is no doubt that the work of the French sociologist Raymond Boudon has been fundamental for the development of modern sociological theory. He was, in fact, a prime example of applying this strategy to complex social phenomena and at the time represented a major chapter in the history of contemporary analytical sociology.

As expressed in his constant arguments supporting these possibilities of explanation through mechanisms, one outstanding feature of Boudon's academic career was his persistent determination to actively maintain both the applied and theoretical requirements of sociology, and his tenacious pursuit of an empirical and scientific dimension for the discipline. Since his early training in the ***École*** *Normale*'s tradition, he always showed a strong aversion to merely rhetorical, demagogic or manipulative discourse, instead, "admiring and emulating the objectivity and clarity of men of science" (Boudon, 1996a: 77). It was probably this same critical spirit which led him to Columbia University where, for a short time, he worked with Paul Lazarsfeld who he already knew through the publication of *The Language of Social Research* (1955). From this point on, regression analysis, secondary analysis and text analysis became a part of his academic and professional training, and continued to be a permanent feature of his work.

Fluent both in English and German thanks to his work at the universities of Columbia and Freiburg, from very early on in his academic life he was in direct contact with these two traditions and assimilated their most salient features. From the Anglo-Saxon empirical school he learned methodological rigour and became interested in regression techniques, statistical control and path analysis. Mindful of the limited or scarce possibilities of these statistical procedures as a means to serious causal analysis, he promptly embraced Lazarsfeld's claim in favour of non-statistic or theoretical assumptions. Following then Herbert Simon's leading work on asymmetry and causal concepts, he saw in the introduction of intentional mechanisms a secure way to overcome shortcomings widely spread in the standard statistical sociological research (Lizón, 2006). From the German sociological school, in particular from Max Weber, he took the basic notion of an intentional explanation that elicits the motives or reasons for an action as principles of meaning as well as causes of human agency. Accordingly then, throughout the different stages of his own intellectual and academic career, Boudon shaped his own

research programme that came to us as a curious intersection between these two major traditions.

One of the most impressive aspects of his work was the tenacity with which he came to illustrate and exemplify this strategy of explanation through social mechanisms in a wide range of areas. Attending either to simple mechanisms of cost-benefit (Boudon, 1973), systems of interaction with interdependent decisions (Boudon, 1979a), or a more lax notion of subjective or cognitive rationality (1992, 1996b) which he promoted as more in accordance with the principles of Weberian sociology of action (Boudon, 2003, 2006), the most outstanding result of his work was his capacity to integrate and articulate numerous macro-effects or social results within the theoretical framework of rational action.

Regarding his contributions to analytical sociology in general, one could discern at least three clearly defined moments or stages (Hamlin, 2002). In the first, Boudon (1965, 1967, 1968) mainly focused on questions regarding empirical methodology, factorial analysis, and path models using regression. From here onwards, the generation of models as a strategy for interpreting statistical structures (Boudon, 1979b) placed him in the wake of Lazarsfeld-Merton and Simon-Duncan's thesis of theoretical models to interpret statistical structures. This particular course was to bring him in contact with the pioneering studies of unconventional economists such as Albert Hirschman, Thomas Schelling and Mancur Olson, all of them deeply influenced by Herbert Simon's "*Models of Men*" and his notion of 'satisficing' or bounded rationality.

By the end of the 1970s, Boudon had arrived at a conception of social science in general, and of sociology in particular, which, to a large degree, he shared with the then young Norwegian philosopher Jon Elster (1978). Although working along independent lines, they both presented similar programmes, mainly built on an intentional behaviour model linked to the methodological individualism of rational choice explanatory thesis, and supported by game theory analytical tools. In this central stage of his work, Boudon (1977, 1979a) focused on the analysis of interactive social patterns, producing an indisputable accomplishment regarding his strategy for explaining macro or social phenomena by means of intentional interdependent decision-making mechanisms.

Even though he was familiar with the analytical tools of rational decision and game theories, from the very outset of his work one can appreciate his critical distance from the prevailing economic model. In this respect, he had particularly strong reservations regarding the position of the Chicago School (Gary Becker and James Coleman) or any claims of 'economic imperialism'. In *La Logique du social*, Boudon (1979a)[1] explicitly set out an alternative version of *homo sociologicus* that, by necessity, required the inclusion of "more complex" assumptions and a conception of an autonomous agent "better adapted to sociological thinking". With this, he clearly aimed to go far beyond the

1. Quotations are from the Spanish translation, Madrid, Rialp, 1980.

standard *homo economicus*' monist set of motivations, expressively centred on the agent's self interest and utility expectations. His analytical model, a peculiar form of 'methodological intentionalism' and 'methodological rationalism' (Van Parijs, 1990: 48), gained theoretical credibility thanks to his major contributions and achievements, reconstructing major examples from the sociological tradition in a continued effort to interpret unintentional lateral macro-effects in terms of micro-processes or intentional mechanisms (Boudon 1979a). In some way then, his work became one of the most effective and useful responses to Merton's long unattended claim concerning the possibilities of linking research and theory in sociological work.

2. Boudon's analytical proposal

By arguing for a change of paradigm, Boudon attempted to achieve an enriched version of intentional explanation as the ultimate aim of any sound sociological analysis. His programme was based on the thesis that, in order to explain any social phenomenon, one has to start from the actions-decisions of intentional agents (a clear preference for methodological individualism) confined to interactive systems (his predilection for strategic interdependent decision making). With this programme in mind, he tried to distance himself from the sociology of social causes (Durheim's 'social facts'), while leaning towards models of intentional interdependent decision (a neo-Weberian approach).

While distancing himself from the passive *homo sociologicus* as the result of social determinisms (Parsons, 1937) or social roles (Dahrendorff, 1958), his essentially interactionist notion of sociology also brought him into conflict with the standard notion of 'parametric rationality'. Thus, his new proposal ended up being conceived as a kind of intermediate position between diametrically opposed options, and would culminate in his own vision of an active *homo sociologicus* as the focal point of his analytical and theoretical proposal (*op.cit.*: 223-241).

Within this paradigm, social phenomena are no longer seen as mere reflections of society and culture, but rather as the result of the human capacity to reason and decide. Consequently, the social facts are to be interpreted as the manifest or latent result of actions and decisions of intentional agents that interact in socially indexed contexts. Faithful to the Weberian tradition closely linked to the notion of autonomous social agents, the proximity to the theory of decision making and games seemed particularly suited to his aims. This did not necessarily imply that he accepted the assumptions of the economic model as is, and its view of human emotions, motivation and behaviour. Far from adopting the view of social agents as mere utility maximizers, Boudon attempted to describe them by means of a complex set of preferences activated throughout the decision-making process within specific strategic, normative and cultural contexts. This was how his version of the *homo sociologicus* came to anticipate important corrections and additions to the standard economic model. According to him, perhaps the main differences between both concep-

tions are that, in the former, the agent is taken as being rational in the strict sense of 'objective' rationality,[2] while in the latter this is only the case in a peculiar meaning of limited or 'subjective' rationality (Boudon, 2003).

In fact, the rational behaviour of individuals is a far cry from being the behaviour of a utility maximizer and cannot be automatically reduced to the standards of economic rationality. Leaving aside questions related to the natural limitations of human cognitive and perceptual abilities (Boudon, 1986a, 1990), within a given interactive context the individual may also find his/her elective preferences influenced by the decisions of others; and, more definitively so, by some social basic norms and the inertia of beliefs and values that provide a sort of 'background' that, by nature, influence them (Searle, 1995, Bourdieu, 1980). Thus, Boudon's final proposal came to be a hypothesis for rational behaviour that tended to be a good deal more lax and wide ranging than in the prevailing economic model.

Starting from the premise that the motives (or reasons) behind actions make them 'intelligible' and 'explain' them, in a true Weberian sense, he incorporated into his programme the unavoidable interpretative or subjective moment of the "motives that lead to action", yet not renouncing the also unavoidable need to go beyond merely understanding and describing the evidence in question. His approach becomes particularly interesting if one takes into account that, in certain areas, particularly in post-modern sociology, there has been a tendency to substitute any attempt to explain for merely interpreting social phenomena[3]. Clearly, for Boudon the art of interpreting did not grant the sociologist of the observance of the scientific requirements regarding explanation. The peculiarity here is the fact that action constitutes its own category of facts (Boudon, 1979a: 239) and, consequently, he made use of the empathic assumption not only as a criterion applied to the individual to indicate the reasons that 'justify' his/her actions, but, also, as a principle for explaining them. Thus, motives or reasons contribute to an inevitable interpretative moment, and at the same time endow the action with its own explanatory principle (1979a: 237-241, 1992: 31).

It is perhaps this demand for explanation which constitutes the most distinctive feature of Boudon's 'neo-Weberian interactionism', which he used to definitively distance himself from the old prejudices of operationalists and behaviourists and their systematic exclusion of mental states. However, the same demands also allowed him to take distance from the non-causalist hermeneutic versions, interpretations equally founded on understanding or *Verstehen*, but only in a narrow Diltheian sense, as regarding only subjective intentions or meanings. Rather than the motives which specific actors attribute to equally

2. In an ideal competitive market, individuals are believed to be fully informed and all-knowing in the sense that they know the only pertinent information about the prices. Since the agent knows all he needs to know, their knowledge is taken to be "objective".
3. To a large degree, "due to the predominance of this kind of interpretive discourse, sociological theory appears to carry less and less weight in empirical research." (Boudon, 1998: 127).

specific actions (as found in some merely descriptive ethno-methodological exercises), Boudon's concern was clearly both actionalist and indisputably explanatory and theoretical.

3. The appeal of game theory

While game theory hardly had any immediate impact on the empirical theory of action followed by the Columbia school, Boudon found himself decisively influenced by the logic of games. Paul Lazarsfeld was who introduced these mathematical tools as part of the on-going training of the members of his *Bureau of Applied Social Research*. Though his central aim was to improve their formal skills, judging by the results, von Neuman and Morgenstern's theses did not seem to have any substantial influence on most of them; although one must underline that the lucid essay by Luce and Raiffa (1957) on strategic games came out of these seminars. Merton did not mention this possibility (in his successive editions of *Theory and Social Structure*) nor did Coleman (the one member of this group who was closest to the rational choice models) embrace matrix games in any significant way. Perhaps the main achievement of Lazarsfeld's efforts was to familiarize those fellow sociologists, completely alien to the Weberian tradition, with the idea of an explanation based on human agency. Following the logic of interdependent decision frameworks, they were exposed to alternative hypotheses on human behaviour which could be used to substitute their empirical programme focused on *ad hoc* psychological variables (as in Stouffer's epigones), or based on rudimentary functional mechanisms (following Merton). In stark contrast, a very young Boudon clearly saw the possibilities offered by the mathematical language of games to provide a causal dimension hitherto unknown in sociological statistical practice. In fact, its interactive models seemed designed to offer social research the possibility of integrating causal reasoning into the analysis of the social interaction of autonomous decision-makers.

Lacking a language able to express cause in an efficient way, it is only natural that statistics did not encourage the idea of a strictly explicative empirical sociology (Lizón, 2006). Nevertheless, this would all change radically with the introduction of the language of games. Leaving aside the constitutive problems that have prevented this applied mathematics from being fully developed, matrix games decisively broke down the obstacles to causal thinking imposed by the symmetry of statistical correlation. It finally allowed for formally posing a basic causal reasoning in matters of social theory and social research (Simon, 1964).

Given that Boudon always stressed methodological issues, it comes as no surprise that he placed emphasis on the use of different matrix games, as they provided him with excellent tools to establish causal grounds for some general structures of human interaction. These mathematical structures allowed him to do so within a logic that presented the social facts as lateral results, often not intended, or even contradictory to the intentional actions of intentional

interacting agents. Hence, in stark contrast to many of his colleagues, he was fully convinced that, by learning to identify relevant structures of preferences in crucial matrix games, and using them properly in tasks of description and analysis, sociologists could make a solid contribution to the understanding of interesting social processes.

In this way, game theory offered him a formal and analytical framework in which the 'rational', 'consistent', 'non-contradictory, etc. came to be defined in terms of interdependent decisions within strategic interaction systems. This approach would finally lead him to integrate Popper's former analysis of the logic of the situation and the interdependent decision model into a basic scheme (Boudon, 1979a); a reformulation that also allowed him to associate this decisional stance with the idea of composition effects or lateral social results, now understood in terms of intentional reasons or causes. In merging these ideas, he found a general logic that could shed light on the behaviour of social actors and, furthermore, do so from a formal framework closely tailored to real forms of social interaction.

An immediate consequence of rational strategic behaviour in game theoretical terms is its essential interactive or social character. In fact, in this particular mathematical language, rationality of choice comes to be described in relation to a context of interaction and in strict dependence on the type of interactive system in which the decision is made. Effectively, this is a formal modality of interdependent decisions that accounts for maximization problems and strategic equilibrium within systems in which the decision of some individuals has a decisive influence on the decision of others. It is probably because of this that Boudon ended up stating that "since it is unrealistic to try and explain a social phenomenon outside an interactionist model […] these systems of interdependence are of particular importance for sociological analysis" (1979a: 129).

Once one adopts the framework of rational decision and game analysis and begins to experiment with the various matrix games, one of the most surprising outcomes is the potential wealth of its formal tools. In fact, they can simulate a wide range of interdependent relations, in which central notions such as 'risk', 'agreement', 'mediation', 'coalition' or 'social contract', etc. come to identify different critical structures of preferences, which describe a wide frame of empirical arrangements of great interest to social research. The basic utility of these formal tools is commonly associated with the fact that, "rather than a theory in commonly understood terms", the logic of interdependent decision appears to refer to "an indispensable natural system for understanding human interaction" (Elster, 1989: 36)[4]. Although it has infinite mathematical possibilities (Schelling, 1984), social scientists should only be concerned with identifying a finite number of game matrices that seem to be particularly pertinent to questions that directly concern them.

4. Quotation is from the Spanish translation, Barcelona, Gedisa, 1991.

4. The meaning of methodological individualism in Boudon's sociology

Along with the rationality assumption, the explanatory principle of methodological individualism constitutes for Boudon (1992: 26-31) the other mainstay of the sociology of action. In a strictly Weberian sense, methodological individualism for him is equivalent to the assumption that social phenomena must be explained as the result of individual actions, actions which, in turn, have to be explained in terms of those intentional stances which guide the individual agents and cause them to undertake them.

In such a case, human actions are to be understood as the set of behaviours motivated by 'significant' mental states responding to 'expressed' motives or reasons. Its central relevance is naturally tied to humans' basic capacity for empathic understanding; an innate ability that allows us to 'look inside' and understand the reasons for our actions, and relate to others by understanding the motives behind theirs. It is because of this sort of human endowment as natural psychologists (Humphrey, 1986) that the action comes to constitute the core of any sound analysis and explanation of human affairs. One must not forget that this insight was precisely that which provided Weber with the definitive argument to award action a privileged position in scientific social explanations. With this it was also implied the consequent need to endow the social sciences with a character that was not only intentional, but also, intrinsically interpretive.

Closely linked to the Weberian tradition, Boudon continuously underlined the primacy of intentional regularities. The analytical focus of sociology must be on autonomous individuals capable of non-regulated decision making: "The causality relation that is observed between the parameters of the interaction system and the behaviour of the actors is only intelligible if seen in terms of the behaviour of actors endowed with autonomy" (1979a: 35-36). So, as a methodological rule, the sociologist must adopt the consideration of individuals or agents as no longer "left to their fate in a social vacuum" (Boudon, 1992: 28), but instead included in interactive systems, where the intentional actions come to be "the [true] logical atoms of analysis" (Boudon, 1979a: 63).

All of this is in clear contrast to Durkheim's collectivist tradition and his proposal for structural states as being responsible for social aggregates or 'social facts'. Adopting a completely opposed view, Boudon systematically argued in favour of the idea that any correctly established social regularity is to be understood as the result of intentional facts, that is, in terms of human actions and human interactions (Boudon, 1986b). Whether the suicide rate remains stable or not in the face of different statistical controls – as argued by Durkheim (1897) – in the end it is individuals who in fact commit suicide and they do so in accordance with 'their reasons' (Douglas, 1967). This, if any, is the exact meaning Boudon gives to Coleman's idea of the intentional explanation as a 'final explanation' or explanation with proximate mechanisms and distal causes.

To define social facts from the perspective of game analysis does not only mean presenting them as the result of individual intentional actions, but also

assuming that these actions are shaped within authentic systems of interdependent decision making (Bunge, 1999: 30)[5]. On this matter, his proposal also ended up distancing itself from the standard methodological individualism of parametric rationality, since he clearly opted for a notion of strategic rationality that he believed adapted better to sociological analysis. In this way, more than a mere analytical tool, games end up making a kind of ontological claim. In the end, one would have a theory that prescribes how rational agents behave in contexts of interdependence, which is the hypothesis behind Boudon's (1979a) own interactionist paradigm. Thereafter, interactive systems came to be seen as the basic molecule of social analysis; in fact, they are the interdependent decision systems that incorporate the intentional actions of decision makers.

5. The question of ontological atomism

When it comes to methodological individualism and ontological atomism, Boudon appears to have fallen victim to a certain degree of confusion. On the one hand, he readily acknowledged the intrinsically interactive or social nature of agents as strategic decision makers. On the other, however. he did not consider it necessary to revise the ontological assumptions which have traditionally gone hand in hand with the explanatory thesis of the methodological individualism of rational choice. In the end, it is not clear what he really meant when he stated that the principle of methodological individualism, "does not imply […] a perception of society as a juxtaposition of '*solitudes calculatrices*'. *It does not convey an atomist image, but rather an interactionist* image of society, which is clearly very different" (1992: 28, emphasis added). In effect, we do not really know if, according to Boudon, we are social because we interact, or, from a more profound and essential perspective, we tend to interact precisely because *we are social*. What is argued in this last case is that human distinctive capacities – centrally our capacity for thought and decision – depend in a "non-causal but constitutive way" on the enjoyment of social relationships (Pettit, 1993),[6] and perhaps require being more attentive to our evolutionary make up (Lizón and Masjuan in press).

This matter appears to merit attention, even if only to rule it out in a solvent way. Since Boudon (1986a) considers methodological individualism the "fruit of a rationalist epistemology" that "only has a methodological status", he believes that the ontological questions linked to this explanatory principle must be postponed or ignored as "naïve" and "redundant" issues that "lead nowhere". He is so decidedly in favour of the explanatory principle of methodological individualism that he overlooks the fact that this explanatory thesis

5. Quotation is from the Spanish translation, México, Ediciones Edaf, 2000.
6. This final interpretation would effectively avoid the Hobbesian idea of a pre-social mind; a topic that has now taken on interest not only in scientific circles that sustain a biological and evolutionary approach, but also in metaphysics and philosophy of the mind-intention where it concerns how significant people's relationships are in their essential constitution as subjects and agents.

is not bound *a priori* by any specific ontological claim regarding the nature of the human mind and human subjects. In effect, methodological individualism does not constitute a uniform doctrine and, certainly, has no given prior commitment to any claims concerning the nature of the mind or the content of the mental or intentional states that motivate individuals (Udehn, 2002).

As can be easily found in any dictionary of philosophy, the term 'individualism' designates a doctrine according to which the individual constitutes the basis of all structural regularities or social law. Nevertheless, given the elementary meaning of the 'individual' as an 'atom' or indivisible unit, individualism has always been concerned with at least two distinct and different conceptions. On the one hand, there is the definition of the individual in negative terms, that is, simply in opposition to any other composed reality (society, community, the state, etc.). In contrast, the other tradition has opted to define the individual in positive terms as a 'human individual', that is, someone in possession of certain impregnable characteristics and essential capabilities that confers the individual a basic ontological possibility to interact socially with others.

Although in both cases the explanatory principle of individualism is opposed to methodological collectivism, the two approaches imply completely different ontological conceptions. One way of highlighting the difference between ontological atomism and methodological individualism would be then to clarify that, while the former definition assumes a complete reduction of sociology to pre-social Hobbesian psychology, the explanatory thesis of methodological individualism is also compatible with a richer idea of constitutive social – albeit non-causal – human individuals.

Given the significant differences in interpretation, one has to adopt a clear position on this matter, regardless of how open the issue may be. In any case, just resting on the assumption that the explanatory principle of methodological individualism "has no more basis than its efficacy" (Boudon, 1979a: 65) implies a considerable degree of oversight regarding deeply rooted philosophical and even biological questions. Therefore, by ignoring this issue, Boudon succeeded only in obscuring the meaning of his own proposal. Unless one is sensitive to the atomistic ontology underlying the standard thesis of the methodological individualism of rational election, it is not possible to avoid some of the pitfalls that he himself attributed to Hayek-Popper's version (Boudon 1992: 28). Neither can one take for granted that the explanatory principle does not imply "conceiving society in terms of a juxtaposition of *solitudes calculatrices*" (*Ibidem.*) and, at the same time, attempt to offer an alternative image of a socially constitutive human mind by the mere fact of undersigning an interactionist scheme. At least not, as is the case here, if one wants to criticize in any depth or even go beyond the limitations of the economic model (Boudon, 1979a: 224). What is lacking is a more clearly thought out and well-informed reconsideration of the ontological assumptions underlying the explanatory principle of methodological individualism, which is essential to obtain a final fit of the central pieces of his theoretical bet. It is only in this way that he could offer more secure and solid arguments in favour of his *homo*

sociologicus, or at least more sound than by making loose statements against ontological atomism.

In fact, one of the immediate consequences of the axiomatic basis of neoclassical economics was precisely the exclusion of social and political motivations. Central notions to sociological analysis such as 'altruism', 'solidarity', 'legitimacy', and 'social commitment' are not included in its formalized model (Eisenstadt and Roniger, 1984). Yet it is precisely this set of economic monist motivations, and not anything else, which Boudon seems to criticize responding with the idea of a more complex social agent. If this is the case, he certainly cannot limit himself to adopting methodological individualism as a mere explanatory thesis with the idea of undermining or transcending the explanations based on social facts. He also needs to provide some ontological counterview that would be relevant for his idea of intentional active actors, basically autonomous individuals and yet ruled by, and integrated into, the historical structures or institutions of their time. By taking distance from Hayek and Popper's economic atomism, he urgently needs a social ontology that, though still unsolved, was somehow prefigured in his early proposal of a neo-Weberian paradigm.

Boudon was a long way from dealing with this, but it is very much to his merit that he at least posed a question that sociology will have to resolve if his active *homo sociologicus* is to prove to be "alive and well" (Boudon, 1979a: 224).

Bibliographic references

Boudon, Raymond (1965). "A method of linear causal analysis: Dependence analysis". *American Sociological Review*, 30, 365 – 374.

— (1967). *L' analyse Mathematique des Faits Sociaux*. Paris: Plon.

— (1968). "A new look at correlation analysis". In: Blalock, Hubert and Blalock, Ann B. (eds.). *Methodology in Social Research*. N.Y.: McGraw-Hill.

— (1973). *L'inégalité des chances: la mobilité sociale dans les sociétés industrielles.* Paris: Armand Colin. [(1974). *Education, Opportunity and Social Inequality.* N. Y.: Wiley.]

— (1977). *Effets pervers et ordre social*. Paris: PUF. [(1982). *The unintended consequences of social action.* London/Basingstoke: Macmillan.]

— (1979a). *La logique du social*. Paris: Hachette. [(1981). *The logic of social action.* London/Boston: Routledge & Kegan Paul.]

— (1979b). "Generating models as a research strategy". In: Merton, Robert K., Coleman, James S. and Rossi, Peter H. (eds.). *Qualitative and Quantitative Social Research. Papers in Honor of Paul F. Lazarsfeld*. N.Y.: The Free Press.

— (1986a). *L'ideologie ou l'origine des idées reçues*. Paris: Fayard. [(1989). *The analysis of ideology*. Cambridge: Polity Press, cop.]

— (1986b). "Individualisme et holisme dans les Sciences Sociales". In: Birnimbaum, P. and Leca, J. (eds.). *Individualism, Theories and Methods*, Oxford: Claredom Press.

— (1990). *L'art de se persuader des idées douteuses, fragiles ou fausses*. Paris : Fayard. [(1994). *The art of self-persuasion: the social explanation of false beliefs.* Cambridge, UK; Cambridge, MA: Polity.]

— (1992). "Action". In: Boudon, Raymond *et al.* (eds.). *Traité de Sociologie*. París: P.U.F.
— (1996a). "Pouquoi devenir sociologue? Réflexions et évocations". *Revue Française de Science Politique*, 46 (1), 52-79. <http://dx.doi.org/10.4000/ress.654>
— (1996b) "'The Cognitivist Model'. A generalized 'Rational-Choice Model'". *Rationality and Society,* 3, 123-150. <http://dx.doi.org/10.1177/104346396008002001>
— (1998). "Social mechanisms without black boxes". In: Hedström, Peter and Swedberg, Richard (eds.). *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
— (2003). *Raison, bonnes raisons*. Paris: P. U. F.
— (2006). "*Homo sociologicus*: neither a rational not an irrational idiot". *Papers. Revista de Sociologia*, 80, 149-169.
Bourdieu, Pierre (1980). *Le sens practique.* Paris: Ed. De Minuit.
Bunge, Mario (1999). *The Sociology-Philosophy Connection.* Transaction.
Dahrendorf, Ralph (1958). *Homo Sociologicus*. London: Routledge & Kegan Paul. 1973.
Douglas, Jack D. (1967). *The Social Meanings of Suicide*. Princeton: Princeton University Press.
Durkheim, Émile (1897). *Le suicide.* Paris: P.U.F. 1973.
Eisendstadt, Shmuel N. and Roniger, Louis (1984). *Patrons. Clients and friends: Interpersonal relations and the structure of trust in society*. Cambridge: Cambridge University Press.
Elster, Jon (1978). *Logic and Society.* Chichester: John Wiley and Sons.
— (1983). *Explaining technical change*. Cambridge [etc.]: Cambridge University Press.
— (1989). *Nuts and bolts for the social sciences.* Cambridge [etc.]: Cambridge University Press.
Hamlin, Cynthia L. (2002). *Beyond Relativism. Raymond Boudon, cognitive rationality and critical realism*. London: Routledge Studies in Critical Realism, 2002.
Humphrey, Nicholas (1986). *The Inner Eye*. Faber & Faber.
Lizón, Ángeles (2006). "Estadística, causalidad y sociología empírica del siglo XX", *Papers*. *Revista de Sociologia*, 80, 223-255.
Lizón, Ángeles and Masjuan, Josep Maria (2014). *Instintos, emociones e inteligencia social. Lecturas evolucionarias para científicos sociales,* In press.
Luce, R. Duncan and Howard Raiffa (1957). *Games and Decisions: Introduction and Critical Survey*. N.Y: John Wiley.
Parsons, Talcott (1937). *The structure of social action*. N.Y.: Free Press [etc.], cop, 1968.
Pettit, Philip (1993). *The Common Mind: An Essay on Psychology, Society and Politics*. N. Y., Oxford University Press. <http://dx.doi.org/10.1093/0195106458.001.0001>
Schelling, Thomas (1984). *Choice and Consequence*. Cambridge, Mass.: Harvard University Press.
Searle, John R. (1995). *The Construction of social reality.* N. Y. [etc.]: Free Press, cop.
Simon, Herbert S. (1964). *Models of Man: Social and Rational. Mathematical Essays on Rational Human Behavior in a Social Setting*. N. Y.: John Wiley & Sons, Inc.
Udehn, Lars (2002). "The changing face of methodological individualism", *Annual Review of Sociology*, 28, 479-507. <http://dx.doi.org/10.1146/annurev.soc.28.110601.140938>
Van Parijs, Philip (1990). *Le Modèle économique et ses rivaux.* Genève: Libraire Doz.

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2076

The Explanation of Everything. A Critical Assessment of Raymond Boudon's Theory Explaining Descriptive and Normative Beliefs, Attitudes, Preferences and Behavior

Karl-Dieter Opp
University of Leipzig / University of Washington
opp@sozio.uni-leipzig.de

Received: 11-03-2014
Accepted: 05-06-2014

Abstract

Raymond Boudon proposes a theory that explains attitudes, descriptive and normative beliefs, preferences and behavior, in other words: everything – or at least almost everything – social scientists are interested in. The basic idea is that *reasons* are a major causal factor, but there are also *irrational factors* (Boudon's term) such as affective causes. This is the first paper that provides a detailed critical analysis of this theory. We first identify the major problems of the theory. One is its relatively low explanatory power: it is largely left open how to select the causally relevant reasons and irrational factors for a given explanandum. A second problem is the validity of the theory: is it plausible that a single theory can explain the wide range of phenomena Boudon focuses on? A final question is whether Boudon's rejection of utility maximization is acceptable.

To answer these questions two social psychological theories are applied to each of the explananda of Boudon's theory: value expectancy and balance theory. It is shown that the two theories are capable of improving the explanatory power of Boudon's theory. They further confirm that a single theory can explain the explananda of Boudon's theory. Finally, both theories imply that Boudon's rejection of utility maximization is not tenable.

Keywords: Rational choice theory; Raymond Boudon; explanation by reasons; utility maximization; value expectancy theory; balance theory; explanatory power.

Resumen. *La explicación de todo. Una evaluación crítica de la teoría de Raymond Boudon que explica las creencias descriptivas y normativas, las actitudes, las preferencias y la conducta*

Raymond Boudon propone una teoría que explica las actitudes, las creencias descriptivas y normativas, las preferencias y la conducta, en otras palabras: todo —o como mínimo casi todo— en lo que los científicos sociales están interesados. La idea básica es que las *razones* son un factor causal fundamental, pero que existen también, en términos de Boudon, *factores irracionales*, como las causas afectivas. Este es el primer artículo que ofrece un análisis crítico detallado de esta teoría. En primer lugar se identifican los principales problemas de la teoría. Uno es su relativamente bajo poder explicativo: la cuestión de cómo seleccionar las razones y los factores irracionales causalmente relevantes para un determinado *explanandum* se deja abierta en una medida importante. Un segundo problema consiste en la validez de la teoría: ¿resulta plausible que una única teoría pueda explicar un rango de fenómenos tan amplio como el que aborda Boudon? Una última cuestión es si resulta aceptable el rechazo de Boudon de la maximización de utilidad.

Para responder a estas preguntas se aplican dos teorías socio-psicológicas a cada uno de los *explananda* de la teoría de Boudon: la teoría del valor esperado y la teoría del equilibrio. Se muestra que ambas teorías son capaces de mejorar el poder explicativo de la teoría de Boudon. Adicionalmente, dichas teorías confirman que una única teoría puede dar cuenta de los *explananda* de la teoría de Boudon. Finalmente, ambas teorías implican que el rechazo de la maximización de utilidad por parte de Boudon no se sostiene.

Palabras clave: teoría de la elección racional; Raymond Boudon; explicación por razones; maximización de utilidad; teoría del valor esperado; teoría del equilibrio; poder explicativo.

Summary

1. Introduction

Raymond Boudon is the only social scientist who has proposed a theory that is supposed to explain every phenomenon, or at least most of the phenomena, social scientists are interested in: descriptive and normative beliefs, preferences, attitudes and behavior. In contrast to social psychological theories such as learning theories, Boudon's theory is strikingly simple, as will be seen below. Furthermore, because Boudon subscribes to methodological individualism, his theory is supposed to contribute to the explanation of macro phenomena as well. Thus, compared with "grand" theorists, such as Karl Marx or Talcott

Parsons, Boudon's theory refers to micro as well as macro phenomena. And, it seems, it is a testable theory. Another attractive feature of Boudon's theory is that he illustrates it with numerous examples, mostly taken from classical writers such as Émile Durkheim, Alexis de Tocqueville and Max Weber. This relates the theory to the classical core of sociology and is thus rooted in the sociological tradition. Perhaps Boudon's theory is the overarching theoretical system social scientists have dreamt of?

The generality of Boudon's theory, its simple structure and its close relation to the work of major classical writers deserve a detailed discussion. It is surprising that this is lacking so far. The present paper attempts to close this gap.

In what follows, we first present Boudon's theory and discuss its major problems. Two social psychological theories are then applied and it is examined whether they are capable of contributing to the solution of the problems of Boudon's theory: balance theory and value expectancy theory. We apply the theories successively to the different explananda of Boudon's theory: descriptive beliefs, normative beliefs, attitudes, preferences and behaviors. After these detailed analyses, the question is addressed whether Boudon – who is a vehement opponent of rational choice theory – does not implicitly apply a wide version of this theory. It is argued that this is actually the case.

Before we address the issues mentioned, it is useful to define some of the basic concepts. *Descriptive beliefs* – also called representational beliefs (e.g., 2012a: 8)[1], positive or cognitive beliefs – refer to empirical statements such as "X is true" or "the wages of women in Europe are lower than of men." The explanatory question is under what conditions individuals more or less accept such statements as valid.

Another dependent variable of Boudon's theory are *normative beliefs*, that is, beliefs of the kind "X is good." Examples are "the state should support the poor" or "it is not allowed to kill somebody." These statements cannot be validated by confrontation with the real world (1996: 125-126). The question to be answered is when individuals accept such statements.

Attitudes are evaluations of objects, but without an oughtness component. For example, the statements "I like to spend money" or "I like others to spend money" refer to a positive feeling, but do not imply any normative claim that one should spend money.

A discussion of Boudon's work is burdened with the problem that he addresses the same questions in numerous writings and that it sometimes seems that there are incompatibilities. The concern of this paper is not an exegesis or interpretation of Boudon's work. The aim is a discussion of substantive issues. For each claim discussed, citations or quotations are provided. Thus, these theses and claims are held by Boudon. I leave it open whether in other work other claims are made.

1. Citations of years with page numbers refer to papers or books by Boudon.

2. Boudon's Theory

Boudon formulates the "basic principle" of his theory that he calls the *cognitive equilibrium principle* (2012b: 18) or the *cognitivist model* (e.g., 1996) – henceforth abbreviated as CM – in the following way: "… the fact that subject X subscribes to idea Y, that the subject believes in Y, can be explained … by the *reasons* that the subject has for believing in it" (1994: 3, italics in the text). Reasons are thus *causes* (1994: 4) of beliefs. Another formulation of this theory is: "[P]eople believe that X is true, acceptable, good, legitimate, etc. as soon as they have the feeling that X rests upon a set of acceptable reasons." This hypothesis explains beliefs. But the central variable – "acceptable", "strong" or "good" reasons – is also a condition for behavior: "[P]eople have strong reasons to believe what they believe, to do what they do" (1996: 140).

The "reasons" need not be idiosyncratic but may be "transsubjective" (1996: 130) or "collective" in the sense that they are accepted by many other individuals. Reasons may be objectively wrong, but can nonetheless be "good reasons" (1989: 174) for a belief. Beliefs, then, may be wrong. An example is the false belief in the causal impact of rain dances on the generation of rain (see the discussion below). This false belief is based on "good reasons" which are false as well. Important for explaining beliefs and action is thus which reasons an actor accepts. Actors who act on the basis of good reasons are, as Boudon called it, "subjectively rational" or simply "rational." "Good" means that the reasons are plausible in the situation of the actors.[2] To illustrate (1989: 188-189), employers often believe that replacing human work by machines increases unemployment. This is based on their experience because when they buy new machines they release workers. In the economy, however, new machines must be produced and maintained and may therefore increase employment.

What is the meaning of "reasons"? Boudon gives the following general definition: When actors are confronted with a question they try to answer it "by making a guess, a conjecture, or by applying a theory or a general principle valid in many cases" (1989: 176). Reasons are thus also beliefs, viz. descriptive as well as normative beliefs. For example, a good reason for believing in the effectiveness of rain dances is the belief that there are gods that can fulfil the desires of the tribe members. Another belief – a reason – may emerge due to the observation that it often (or most of the time) happens that it rains after the ritual has been performed. In this example, there might also exist a general belief, a sort of everyday induction principle, that A causes B if B occurs relatively often some time after A. Thus, *reasons are a certain type of beliefs* that are

2. There are more detailed descriptions in Boudon's writings about what *good* reasons are. For example: "good reasons have the status of conjectures, principles, or theories that most people with the same level of information and/or interest in the question they are confronted with would endorse" (1989: 180, see also 175; further 1994: 34). We will not go further into what "good" (or sometimes Boudon speaks of "strong") reasons are. It suffices for what follows that reasons are beliefs that are subjectively considered valid by individual actors.

"relevant" for accepting other beliefs – we will return to the relevance criterion later. One could call the beliefs that explain other beliefs *second-order beliefs*. Reasons are thus second-order beliefs. The *first-order beliefs* are those beliefs that are to be explained.

Not only reasons are causes. There are other causes that are called *irrational factors* (e.g., 1996:126). These are, among other things, "affective causes" or "passions" (1994: 4). An illustration is Othello's belief that his wife is unfaithful. The reasons for accepting this belief is the evidence Jago presents to him, but everybody else knows "that the credence he gives to these reasons is itself an effect of his jealousy" (1994: 4). There are further "non-affective causes" that are not beliefs either (1994: 5). These are, for example, "psychic causes located beyond any control of the subject" such as a "primitive mentality" (1989: 180). Biological factors or "consumption of some chemical substance" (cocaine) are causal factors as well. Further examples are "absent-mindedness", "deficiency of cognitive capacities" (1996: 125), "internalization of collective beliefs through socialization" or "effects of cultural or of biological evolutionary processes" (1996: 126). We may add factors such as global warming or a natural catastrophe such as a Tsunami. "Sentiments of justice or injustice, legitimacy or illegitimacy … include an affective dimension: nothing is more painful than injustice" (1996: 145). The norms themselves that are adduced are reasons.

These definitions imply that *preferences* are not reasons, they fall under the irrational factors. But sometimes the terminology is not clear. For example, Boudon notes that reasons can be "cognitive" as well as "utilitarian." In one of his examples, Boudon states that there were "'cognitive' reasons to be anticlerical, but also 'utilitarian' reasons not to oppose the anticlerical actions and declarations" (1996: 140). Thus, preferences seem to be reasons. *Constraints* such as available income are not reasons, only the perceptions of constraints which are then beliefs are reasons.

The "irrational" causes – henceforth we will omit the quotation marks of "irrational" – may give rise to beliefs that are based on "solid reasons" (1996: 128). For example, a person has been taught that "2 + 2 = 4," but the person holds this belief because it is based on "solid reasons" (1996: 128). These solid reasons are taught as well.

There are thus two types of causes – reasons and other causes – that may be conditions for beliefs or actions. "Undoubtedly, irrational factors, notably affective ones, can in many circumstances legitimately be evoked to the effect of explaining beliefs" (1996: 126). Thus, "I do not in any way draw the conclusion that all beliefs have to be explained by reasons" (1994: 20). But in many cases, Boudon argues, his CM (i.e., explanations by reasons) is superior, and often explanations with irrational factors can be legitimately replaced by explanations with reasons (see in particular 1994).

How can the reasons be identified? Boudon mentions "linguistic expressions" (1989: 174) that are normally used. This suggests that the actors themselves provide information about their reasons.

Having a "reason" to believe something or to act in a certain way means that there is a belief that is "relevant" or "meaningful" (1996: 136) or "good" (1996: 136) to the actor. A reason is a belief *for* accepting some other belief or for performing an action. It seems that "relevant" means that the actor perceives some beliefs as a justification or a supporting argument for another belief. What these second-order beliefs are must be empirically ascertained. In his numerous examples Boudon tries to find out what the "relevant" reasons might have been. But, again, he does not provide any criterion of how to measure the relevant reasons.

The previous quotations refer mainly to beliefs. This is the major explanandum Boudon is concerned with. But he also claims (see, for example, 2009b: 192 – the quotation is given below) that reasons explain attitudes, preferences and action (see the respective sections below).

We summarize the theory in the form of an if-then statement:

> If individuals have good reasons for accepting a (descriptive or normative) belief, an attitude, a goal or an action, or if there are irrational factors, then individuals accept the belief, hold the attitude or goal or perform the behavior.

One part could be called the *reason proposition* – if we drop the part "or if there are irrational factors." This is the proposition Boudon focuses on. A summary of the theory must also include irrational factors. Boudon does not specify the kind of irrational factors that determine the explananda. Therefore, the partial sentence mentioned before was added.

3. The Major Problems of the Theory

One criterion for evaluating a theory is its *explanatory power* (or, equivalently, explanatory content or information content). The basic idea is (Popper, 1959): the more a theory excludes or forbids, the higher is its explanatory power. This implies, among other things, that a theory has a high explanatory power if it explains a *large* class of relatively *specific* phenomena – for details see below. If this is the case the theory is incompatible with a relatively large number of predictions. It thus forbids much and, therefore, has a high explanatory power. What is the explanatory power of the CM? This is the first question that is discussed in this section.

A second criterion for judging the quality of a theory is its *validity*. This problem is discussed in this section as well.

3.1. The Explanatory Power of the Theory

The explanatory power of a theory depends, among other things, on the degree to which a theory can explain a large class of very specific phenomena. For example, the hypothesis "if people are frustrated, they act aggressively" can only explain that *some* kind of aggression will occur when people are frustrated.

The explanatory power of this hypothesis would be higher if it could explain for which kind of frustration which kind of aggression occurs. What is the explanatory power of the CM? We discuss this issue only for the explanation of beliefs. The argument also holds for the other explananda of the theory. Let us first assume that *only reasons* are causes. As an illustration, assume that Swiss tennis fans accept the following belief:

> Belief b: Roger Federer will win the next Wimbledon championships in July.

What may be good reasons for this belief? Let the Swiss tennis fans accept the following set of beliefs:

> Belief 1: Federer was among the top five players several months before Wimbledon.
> Belief 2: Federer is a better player on lawn than the other players.
> Belief 3: The unemployment rate is lower in Switzerland than in Great Britain.
> Belief 4: Federer is married.

If the theory has a high explanatory power we would expect that we can predict for any *given* possible reasons (i.e., beliefs) what exactly the ensuing belief is. This would be the case if there is a *selection criterion* specifying *which possible reasons lead to which beliefs*. The theory would then state:

> (1) Given a set of possible reasons r: reasons of type i cause the person's belief of type i.

This would allow us to predict, for example, that beliefs (1) and (2) and not beliefs (3) and (4) are causes for accepting b.

It may be argued that it is difficult to imagine how a theory in the social sciences could be so specific. The theories applied below show that such theories already exist. To add an example from learning theory: it hypothesizes, for example, that a reward for an activity a has the effect that the frequency of a increases. Thus, very specific instances of a large class of phenomena can be explained.

Now assume that we do not know which reasons are relevant for which beliefs, that is, no selection criterion is specified. The theory thus only asserts:

> (2a) Given a set of possible reasons as causes for a belief: the reasons are causes for **some** belief b_1 or b_2 … or b_n.

For example, if an actor accepts beliefs 1 to 4 in our example, it could not be predicted which belief b_i will ensue.

Furthermore, the theory without a selection criterion would imply:

> (2b) Given a belief b as a dependent variable: **one or several** possible reasons (i.e., beliefs) may be causes for b.

Thus, given the previous belief b, it is not specified which of the four possible reasons in our example are causes for b.

The lack of a selection criterion has further the consequence that *falsification is difficult or impossible*. When we don't find a reason we might always assume that we had bad luck or were not intelligent enough to find the right reason or reasons. The lack of a clear selection criterion also opens the door for *ad hoc explanations*. When one wants to explain a belief one always finds other beliefs that actors accept, and these beliefs can then be claimed to be causal factors. This is certainly an unsatisfactory situation.

We have assumed so far that only reasons are relevant for beliefs. The previous analysis can be expanded to *irrational factors*. In the previous cases (1), (2a) and (2b) we only need to replace "reasons" with "irrational factors."

Now let us look at the full theory that assumes that *reasons as well as irrational factors* are relevant. Let us introduce a new *term f* that refers to reasons (i.e., rational factors) or irrational factors. The previous analysis holds for f as well: we only need to replace "reasons" with "factors."

When the previous argument is expanded so that reasons *and* irrational factors are included simultaneously, the explanatory power of the theory decreases dramatically, compared to the separate analysis of reasons and irrational factors. This expanded theory tells us neither which reasons nor which irrational factors from a set of possible reasons and irrational factors lead to a specific belief. This corresponds to case (2a) above. The expanded theory does not tell us either which belief is the consequence if a given set of possible beliefs and irrational factors is given. This corresponds to case (2b) above.

Which of these cases applies to the CM? To be sure, Boudon speaks of reasons "for" a belief, as was said before: "to account for a belief, or an action, always try to find the reasons for it" (1994: 18). But this selection criterion is relatively vague. In none of his examples does it become clear what exactly the general hypothesis (or "rule") is that Boudon applies to select the relevant reasons as explanatory variables. Just stating that the selected reasons are those which are relevant *for* certain explananda is not sufficient advice about how to ascertain the right reasons.

Reasons refer to subjective states of mind. So why not ask persons which reasons were relevant? Boudon is aware of the problems of asking subjects for their reasons. He mentions rationalizations (by citing Freud). We may add that there are spontaneous behaviors where people are not aware why they have done something. Wishful thinking or social desirability effects in surveys show the problems of measuring reasons by simply asking people. Boudon further notes that the reasons are generally "metaconscious in the mind of people" (2014), that is, people are not or need not be aware of the reasons. This makes it particularly difficult to find the right reasons.

So our conclusion is that the explanatory power of the theory is rather low. But assume the reader is very tolerant and argues that Boudon's selection criterion that reasons "for" beliefs or other explananda are relevant is precise enough. But for the irrational factors there is clearly no selection

criterion at all: it is not even insinuated which irrational factors explain which phenomenon.

Nonetheless, the theory is not completely without content. It suggests that certain *kinds of factors* are relevant for explaining social phenomena. Boudon's claim is that it is reasons that are of major explanatory importance. This is an *orienting hypothesis* in Merton's sense (Merton, 1957: 88; for a discussion see Opp, 2014b: 174-177). Perhaps the following quotation shows with particular clarity the orienting character of the CM and, thus, its low explanatory power:

> Firstly, social action in the general case *depends on* beliefs. Secondly, beliefs, actions, attitudes should *as far as possible* be treated as rational, more precisely as the effect of reasons perceived by social actors as valid. Thirdly, reasons of the "cost-benefit" type should not be given *more attention than they deserve.* Rationality is one thing, expected utility another. (2009b: 192, italics not in the original text)

The text printed in italics shows that the theory points in a general way to *kinds of factors* that might be causally relevant. For example, the kind of belief that explains action is not specified. The phenomena to be explained should be explained "as far as possible" as the effect of reasons – what kinds of reasons and what does "as far as possible" mean? It is not denied that actors sometimes maximize utility, but sometimes they do not. If so, how do they decide in which situations?

3.2. The Validity of the Theory

Does it make sense to analyze the validity of a theory if it has such a low explanatory power? The answer is that the theory makes some empirical assumptions that can indeed be empirically tested. There are two assumptions that might compromise the validity of the theory: one is the wide range of phenomena the theory is supposed to explain (i.e., the generality assumption), the other is Boudon's rejection of utility maximization.

3.2.1. The Generality of the Explananda

Is it really plausible that a single theory can explain such diverse phenomena as normative beliefs, descriptive beliefs, attitudes, preferences and behavior? This is the claim of the CM. Such a theory does not exist so far, and there will certainly be many social scientists who are skeptical towards the validity of such a claim. We will provide a plausibility test of this claim: we will examine whether two widely used social psychological theories can be applied or expanded to substantiate this claim.

3.2.2. The Rejection of Utility Maximization

Boudon claims that costs and benefits are often not relevant as explanatory variables, in particular for explaining beliefs. "Reasons," Boudon argues, "can-

not be reduced to mere considerations of costs and benefits" (1996: 124). The CM "is drawn from the 'rational-choice model' by lifting the restriction that the reasons of social actors should always be of the cost-benefit type" (1996: 124). Thus, in some circumstances reasons are not of this type (1996: 147). This claim is inconsistent with major existing theories. So the question arises what arguments Boudon submits for his claim. These are discussed below.

But assume Boudon is right: utility maximization often does not apply. The question then arises as to what the alternative hypothesis is. For example, if the decision to prefer theory A to theory B is made, how do actors decide if they do not in some way maximize utility? Boudon does not answer this question. This has the consequence that the theory has an additional severe problem that *diminishes its explanatory power considerably*: it cannot be explained why people choose certain options, be they beliefs, attitudes, preferences or behaviors. We will return to these questions later in this paper.

3.3. Summary

Let us summarize the major problems of the CM. First of all, a *selection (or relevance) criterion* for the kind of reasons and irrational factors that are causes for the explananda is lacking. Secondly, it is not clear what the *joint effects of reasons and irrational factors* are. These problems refer to the explanatory power of the theory. Another problem is the *validity of the theory:* is it possible to explain the wide range of phenomena, the CM tries to explain, with a single theory? This problem refers to the *generality of the theory*. Another validity problem is the *rejection of utility maximization*.

Despite these problems, the theory can be seen as a general *orienting hypothesis* claiming that in explaining beliefs etc. one should in any event consider beliefs as major causes. In the 21th century this is hardly a very exciting advice, it is rather a truism in the social sciences.

4. How to Select the Causal Factors: Applying Social Psychological Theories to Evaluate Boudon's Theory

How can the CM be improved? One possibility is to compare the CM with social psychological theories that are widely applied in social psychology. This is possible because these theories address at least some of the explananda of the CM. If this is the case it can be examined to what extent the theories give more specific guidelines to select the relevant factors in order to explain relatively specific explananda. Thus, we examine to what extent there are theories that could improve the explanatory power of the CM.

Applying social psychological theories may further shed light on the validity of the CM. To what extent do the theories encompass the wide range of explananda Boudon tries to explain? If they do not, is it plausible to expand their range of explananda? This is a plausibility test of Boudon's generality assumption that holds that only one theory is needed to explain everything

from beliefs to actions. Secondly, applying the theories is a test of Boudon's rejection of utility maximization. Do the theories make this assumption, or is there an alternative hypothesis?

It is striking that Boudon never systematically compared his CM with existing social psychological theories. We don't know whether he had "good reasons" for this. It seems that there are *no* good reasons not to apply social psychological theories. On the contrary, the good reasons for applying social psychological theories are that they could improve the CM or confirm some of its assumptions. Furthermore, the social psychological theories could be improved by expanding their explananda.

There are numerous social psychological theories. We selected two theories that are widely applied in social psychology and that might solve the selection problem: one is balance theory (e.g., Heider, 1958, see also van de Rijt, 2011 with further references), the other value expectancy theory (e.g., Feather, 1982, 1990; for an overview see Wigfield, Tonks and Klauda, 2009).

The two theories are applied to each of the explananda of the CM: descriptive beliefs, normative beliefs, attitudes, preferences and action. For each explanandum the focus is thus on the following questions. (1) Do the theories include reasons as major variables? (2) Do the theories show how reasons as well as irrational factors influence explananda? (3) Are the theories capable of explaining the wide range of phenomena Boudon wants to explain with his theory? (4) Do the theories assume some version of utility maximization?

5. Explaining Descriptive Beliefs

We will begin with one of Boudon's major examples and then examine to which extent two major theories can be applied to solve the problems of the CM, based on the example. These theories are balance theory and value expectancy theory.

5.1. An Example: Explaining the Belief in the Effectiveness of Rain Rituals

One of Boudon's examples to illustrate his theory is taken from Émile Durkheim's "Les formes élémentaires de la vie religieuse" (first 1912, see Book III, chapter II).[3] Durkheim explains why tribes of central Australia believed that rain rituals generated rain. To be sure, the tribes had empirical knowledge about how plants grow and die. Thus, there are correct beliefs that include, among other things, information about the importance of water for the growing of plants. But the tribes did not have at their disposal the results of modern scien-

3. This subsection is largely based on Boudon (2014) which was submitted to a special issue of the *Kölner Zeitschrift für Soziologie und Sozialpsychologie* shortly before his death. This paper thus contains the last version of the CM. Therefore, I use this paper. Page numbers are not included in quotations because the paper has not yet been published. For the explanation of magical beliefs see also Boudon (1989).

ce implying that rituals do not generate rain. According to Durkheim (1915: 25-26), for the "primitive man … there is nothing strange in the fact that by a mere word or gesture one is able to command the elements, … bring rain or cause it to cease." The rites "do not appear more irrational … to his eyes than the technical processes of which agriculturists make use." In other words, the "primitive men" have good reasons for engaging in the rain rituals.

Shouldn't the tribe members learn over time that there is no causal effect of their ritual on rain? There are several "good reasons" for keeping the false belief. First of all, the rituals are practiced at a time when rain is likely to fall anyway. There is thus a relatively close temporal association between the ritual and the rain. The belief of the effectiveness of the ritual is thus, as Boudon writes, "rational": scientists would use the same rule of inference. Second, if it turns out that sometimes rituals do not work tribes use auxiliary hypotheses. One might be that the rituals were not performed in the right way.

A third reason for the persistence of the belief about the effectiveness of the rain ritual is the existence of a general belief that is backed by numerous everyday experiences: if an action is only sometimes successful it is unlikely that it will be completely unsuccessful in the future. For example, if phone calls are sometimes not answered that does not mean that phone calls remain always unanswered. However, if an action is always successful for a relatively long period of time and then suddenly the outcome does not occur anymore, the likelihood is very low that the outcome will occur later. This is consistent with hypotheses about intermittent reinforcement in learning theory (see, for example, Ferster and Skinner, 1957 and any textbook on learning theory like Schwartz and Reisberg, 1991). Applied to the belief in the effectiveness of the rain rituals, tribe members will trust the success of the ritual even if it sometimes does not work.

A fourth reason for the stability of the effectiveness belief is that ineffectiveness is explained by some action of other groups:

> The efficacy of these rites is never doubted by the native: he is convinced that they must produce the results he expects, with a sort of necessity. If events deceive his hopes, he merely concludes that they were counteracted by the sorcery of some hostile group. In any case, it never enters his mind that a favourable result could be obtained by any other means. If by chance the vegetation grows or the animals produce before he has performed his Intichiuma, he supposes that another Intichiuma has been celebrated under the ground by the ancestors and that the living reap the benefits of this subterranean ceremony. (1915: 333)

This is a strategy to immunize the belief of the effectiveness of rain rituals against falsification. Another belief makes the ineffectiveness assumption plausible: the ritual leads to rain because god or the gods make the rain, and the dance is supposed to prompt the gods to let it rain. There is thus a whole set of "good" reasons that the Australians had for their false belief about the effectiveness of rain rituals.

The example clearly shows the problems of the CM: why are the "reasons" Boudon mentions the real causes? What is the role of irrational factors? There is no systematic analysis of these factors. What could they be? Beliefs are often transsubjective – as Boudon puts it –, that is, shared by others, and there is joint action as in the example of the rain rituals. An individual may accept a shared belief because he or she is afraid of sanctions: the tribe members put pressure on each other to accept beliefs. This would be an affective cause and not a reason for accepting a belief. Are there other irrational factors and how important are they, compared to reasons? These questions are not answered.

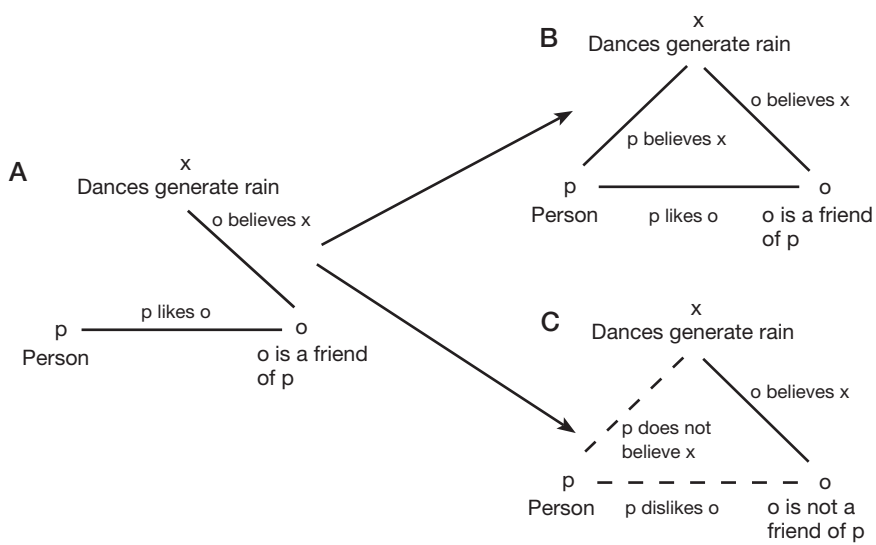
5.2. The Application of Balance Theory

We first provide a short introduction to balance theory (BT). The reader who is familiar with this theory might skip this part. We will then apply the theory to the example. The question is to what extent the theory can solve the problems of the CM.

5.2.1. A Short Introduction to and Application of Balance Theory

We begin with modeling irrational causes. Assume a tribe member, person p, is a friend of another person o (or a set of other persons o). For some reason p does not yet accept that rain dances generate rain. Let p learn that o has this belief x (i.e., that rain dances lead to rain). These three objects – p, o and x – can be depicted in a plane, as figure 1 shows. Graph A shows two relationships:

Figure 1. Application of Balance Theory to Explain the False Belief about the Effectiveness of the Rain Dance.



p likes o and p perceives that o believes x. There are thus a liking relationship L (such as friendship) and a unit relationship U (such as a belief x of a person p).

The relationships in such a pox system may be of different *kinds*: they may be positive or negative (e.g., p may like or dislike o) or they may be present or absent. For example, p might not yet have an opinion about x (graph A of figure 1). Relationships may further have different *intensities*, but we will only assume that relationships are present or absent, and, if present, can be positive or negative. This suffices in this context.

Certain distributions of lines are defined as *balanced* or *unbalanced*. For example, if all three connections in a pox system are positive there is balance. If two relationships are negative and one positive there is balance as well. However, if one is negative and two are positive there is imbalance. Balance is a psychologically pleasant state, imbalance an unpleasant one. To illustrate, assume p has the same beliefs as o and o is p's friend. This means that all three relationships are positive (see graph B of figure 1). Apparently, it is a pleasant feeling when I share my friend's beliefs. However, if the beliefs of my friend and me differ this is unpleasant. For example, if o approves of terrorist activities x and I disapprove of these activities this is certainly unpleasant. It is also unpleasant when I realize that my friend believes x but that I have not yet have formed a belief about x, that is, there is no relationship between p and x (see part A of figure 1). Thus, the pox system is incomplete. It would be more pleasant if I would believe x as well. Balance theory (BT) hypothesizes, among other things, that a lacking relationship in a pox system yields imbalance and is thus unpleasant.

So far we have defined balance and imbalance. A *proposition* is that individuals try to change unbalanced states. In our example, balance would exist if p adopts the same belief as o (see the upper right part B of figure 1).

If a pox system is in an unbalanced state, balance can be brought about by several changes. For example, the imbalance of graph A in figure 1 has been removed by adding a line between p and x (i.e., p adopted belief x; see graph B). There could have been another reaction by p (see graph C of figure 1): p could reject the belief x (i.e., px would become negative) and at the same time terminate friendship with o (po would thus become negative). This would result in a balanced state as well. It is not unpleasant if my beliefs differ from the beliefs of people that I don't like.

Would p prefer the situation depicted in graph B or C? BT assumes that those balanced graphs which require a relatively low number of changes are preferred. The underlying idea is that changing a relationship is costly. One thus prefers a balanced state that requires a relatively low number of changes. Had we introduced intensities, an additional assumption is that the lines with the lowest intensities are most likely to be changed. For example, assume p is quite sure that o believes x and that p could not find any evidence that this belief is wrong. So ox is strong. In addition, let po be strong as well. Thus, for p the least costly change that leads to balance is to add px (graph B).

The pox system can be extended. For example, there could be *other beliefs* (other x's) that are consistent or inconsistent with x (i.e., there may be positive or negative unit relationships with x). For example, x may follow from several religious beliefs. Furthermore, there may be other persons with different relationships to p. We will not discuss more details because this is not necessary for the following analysis.

5.2.2. Some Implications
This very short introduction to BT suffices to illustrate the following points.

(1) BT explains, among other things, beliefs. It is thus possible to compare BT with the CM.
(2) BT specifies a *relevance criterion*. For example, px originates because it makes p better off. Graph C is not chosen because this is costlier than graph B. What about the belief that apples are healthy or that capital punishment does not deter crime? Assume we add these beliefs as elements u and v in graph A. They would be irrelevant because there would not be a relationship of these elements to x and o. But if p perceives that o believes in the deterrence effect of capital punishment then this would be relevant for p's psychic well-being.
(3) BT includes *irrational factors*. A liking relation is not a second-order belief, it is an "affective" relationship. Nonetheless, BT explicitly includes this kind of factor, together with reasons in Boudon's sense. Among the irrational factors are *preferences* as well. We could extend the pox system by assuming that p has a strong motivation to adhere to the *norm* to participate in the rain dances. The norm could be added as an additional object z in the graphs. A preference for adhering to the norm is a positive line between p and z. Thus, irrational factors are included in the theory. Furthermore, BT *integrates* both factors. There is no need to distinguish rational and irrational factors.
(4) An underlying assumption of BT is that there is cognitive optimization or, put differently, *cognitive utility maximization.* In the example, actors do not choose the cognitive structure C but B. The reason is that B is more pleasant or less costly than C. In other words, actors are better off when they choose B instead of C. This is clearly not in line with Boudon's claim that acquiring or changing beliefs has nothing to do with costs and benefits and utility maximization.
(5) BT provides some evidence for Boudon's *generality assumption* (a single theory can explain everything from beliefs to action). Lines between cognitive elements may refer to beliefs (including norms), attitudes, preferences and behaviors.

The previous example suggests *extending Boudon's use of the term "reason."* Reasons are, by definition, beliefs but not feelings. It makes sense to say that p's friendship with o is a "good reason" to adopt belief x because otherwise o

would terminate the friendship relationship. In everyday language, it is certainly a "good reason" to do something in order to achieve a goal and to avoid an unpleasant state of affairs. This suggests that perhaps the meaning of "reasons" should be changed. It could refer to beliefs as well as motivational states.

5.3. Applying Value Expectancy Theory

The theory is usually applied to explain behavior and not beliefs. According to Boudon's generality assumption it seems plausible to expand the range of application of value expectance theory (VET) to explain beliefs as well. We will first provide a short introduction to VET and then explore its capability to explain beliefs.[4]

5.3.1. A Short Introduction to Value Expectancy Theory

The theory asserts that among at least two perceived behavioral alternatives the action with the highest subjective expected utility (SEU) is chosen. This overall utility for a behavior is computed in the following way. A first step is to find the perceived *behavioral alternative*s to a behavior that is to be explained. In a second step the *behavioral consequences* for each perceived alternative must be ascertained. For each consequence, the expected *subjective probability* and *utility* (valuation) must be determined. The sum of the product terms of each behavioral consequence for a *given* behavioral alternative is, by definition, equal to the SEU of the respective behavior. This is its overall utility. The *empirical proposition* is: a person chooses the behavioral alternative that has the highest SEU. The theory becomes more understandable when we apply it to our example.

5.3.2. An Application of VET to Explain the Belief about the Effectiveness of the Rain Dance

VET explains behavior. If its range of application is expanded to explain beliefs, the subjective expected utility should refer to holding a *belief* (instead of performing an *action*). The SEU should depend on the likelihood and utility of the consequences of holding a belief, from a set of alternative beliefs. Is this a plausible expansion of VET?

In order to answer this question we apply VET to our example: can VET explain the adoption of the false belief B about the likelihood that rain dances lead to rain? A first consequence of holding B is that the individual i's belief matches the beliefs of friends. It is assumed that having the same belief as one's friends is beneficial (i.e., it has a positive utility for the actor). This is the same assumption that was made when we applied balance theory. However, VET introduces the subjective probability p that the beliefs of i and his or her friends match. For the members of the Australian tribe this probability is probably

4. There is little work that applies this theory to the explanation of beliefs. See in particular Breen (1999), Matsueda et al. (2006), Breen and Goldthorpe (1997), and Becker (2013).

1 because everybody knows that everybody else accepts this belief. In other situations, however, p may be smaller than 1.

A further consequence of holding B is that it is consistent with the religion (or a set of religious beliefs) that i accepts. If this is the case, this is certainly pleasant for i. If i observes that rain regularly follows the rain dance it would be unpleasant for i not to believe in the causality of the rain dance. Finally, i may be relatively sure that his or her dance influences the intention of the gods to let it rain. Such action is beneficial for i because i believes that this intention will lead to the respective action.

Let us formulate this argument more precisely. We write an equation that consists of the single product terms and thus defines the SEU of holding a belief. We introduce the following abbreviations:

B_i = individual i's belief that rain rituals generate rain
SEU = subjective expected utility of holding belief B_i[5]
p = subjective probability that the consequence occurs
U = utility

The equation for the SEU of B is as follows:

(1) SEU_i (Descriptive belief B_i) = p_{BF} · U(Consistency with friends' B) + p_{BR} · U(Religion is consistent with belief B_i) + p_{BO} · U(Observation that rain follows the dance is consistent with B_i) + p_{BR} · U(Ritual activates Gods' intention to make rain)

The right-hand side of the equation consists of product terms. Each of the product terms consists of a probability that the consequence occurs, given the belief B. The first subscript refers to the dependent variable B, the second to the utility term. This probability could be different if the belief is not held.

A second equation should be added for the SEU for *not* accepting belief B or for accepting perceived alternatives to B. In this equation (or in such equations), the probabilities on the right hand side could be lower than those in the first equation, perhaps even zero. This means that not believing B would in all likelihood not have the consequences mentioned before. The utilities would be the same as in the first equation. The lower probabilities imply that the SEU of the first equation is higher than the SEU of the second equation and that, according to VET, B will be accepted.

5.3.3. Some Implications

The implications are very similar to those for BT.

(1) The previous analysis suggests that VET can explain beliefs. This supports Boudon's *generality assumption*: it is plausible that VET can explain descrip-

5. We assume that in the example "belief" is a dichotomous variable: one may or may not have the belief in the effectiveness of the rain ritual. Were we to distinguish between degrees of beliefs, we would need an equation for each degree.

tive beliefs. Whether the other explananda can be explained remains to be seen.

(2) VET specifies the *kind* of belief that is to be explained, and thus includes a relevance or *selection criterion*. It holds that only those consequences are relevant that are related to the SEU of the belief B. This implies, for example, that the belief that apples are healthy would not be included.

(3) The example includes utilities and thus *irrational factors*. Furthermore, VET *integrates* reasons (in this case subjective probabilities) and irrational factors in the product terms. Again, there is no need to distinguish between rational and irrational factors. It is shown how both factors determine the origin of beliefs.

(4) VET clearly assumes a kind of *utility maximization.* The hypothesis that the action with the highest SEU is chosen means that the actor chooses among the given options the one that is best for him, given his or her subjective beliefs and utilities. This is thus not consistent with Boudon's rejection of utility maximization.

The CM does not include the terms "utilities" and "subjective probabilities." However, the previous argument can easily reformulated by using Boudon's terminology: one could say that each of the consequences provides a "good reason" – referring to the values of the subjective probabilities – for adopting the false belief. The utilities are not to be regarded as reasons. The suggested change of the terminology seems useful here as well: why not say that high utilities are good reasons as well?

The conclusions of applying VET are the same as the conclusions of applying BT: VET remedies some deficiencies of the CM. Our reformulation confirms Boudon's idea that one theory can explain the different phenomena the CM comprises.

6. The Explanation of Normative Beliefs

In discussing Boudon's explanation of normative beliefs we will proceed in the same way as before: we will first describe an example that illustrates the CM and then discuss the theory, based on the example, by applying BT and VET.

6.1. An Example: Why Should Miners Get a Higher Salary than Soldiers?

Boudon's example is taken from Adam Smith's "An Inquiry into the Nature and Causes of the Wealth of Nations" (1776, Book I, Chapter X, Part I.).[6] The question is why there is a strong feeling among 18th-century Englishmen that miners should be paid higher wages than soldiers. The issue is thus to explain a norm. Boudon's explanation is as follows. A salary is a reward. There is a

6. Boudon describes this example in several publications, for example in 1996: 146; 2009a: 36-43. The following is based on Boudon (2014).

general norm that rewards should correspond to contributions to the society. Thus, miners and soldiers should get the same salary if their contributions are equal, as they are valued by the population. Contributions consist of *investments* that are necessary to acquire the competence needed to accomplish the contributions, and the *risk* involved in bringing about the contributions. The investment and risks are similar for soldiers and miners.

However, the *social meaning of the activities* of soldiers and miners is different. Soldiers preserve the existence of the nation, whereas miners perform only economic activities. Furthermore, a soldier's death is considered a sacrifice, whereas the death of a miner is an accident. The soldier thus gets *symbolic rewards* "in terms of moral prestige, symbolic distinctions, glory notably when he has won a battle." These rewards do not accrue to the miner. If the salaries of the miners were not higher, "an unjustifiable *disequilibrium* between the contributions and the rewards of the two categories would appear."

6.2. Applying Balance Theory to Explain Normative Beliefs

The explanandum in the previous section referred to a *descriptive* belief, and it was assumed that important others such as friends accept the belief. As long as p has not formed a belief about the effectiveness of the rain ceremony the situation was unbalanced (see figure 1). Two changes in p's cognitive system were discussed.

Now assume that x is not a descriptive but a *normative* belief (viz., the belief that miners should get higher wages than soldiers). We call this the *wage norm*. As in figure 1, there is a unit relationship ox (p perceives that o accepts norm x) and a liking relationship between p and o. Assume further that p has not yet decided whether the norm about the different wages of miners and soldiers is acceptable (see figure 2, graph A). This is an unbalanced state, as described before. Accepting the normative belief yields balance (see graph B in figure 2). A balanced state would also be achieved if p rejects the norm and terminates friendship with o (graph C). However, balance theory assumes that B is preferable to C, as was said before.

Note that the type of x is irrelevant in BT: x may be a descriptive or normative belief or any other object (such as a third person q). Only the structure of the graph (i.e., the distribution of positive or negative relationships between cognitive elements), and thus balance or imbalance, is of importance.

The example from Adam Smith also refers to *relationships between norms*: there is a general norm about fairness of rewards and a special norm about the fairness of wages. The latter norm is implied by the general norm. This situation is depicted in figure 3. Assume p accepts a general norm g and believes that a special norm s follows from g. It would be a balanced state when p also accepts s. A negative line from p to s would yield imbalance and would thus be costly.

Note that the *perceived* implication of norm s is important, not the actual implication. It may thus be the case that p does not think that s follows from g. For

Figure 2. Application of Balance Theory to Explain the Norm about Fair Wages.

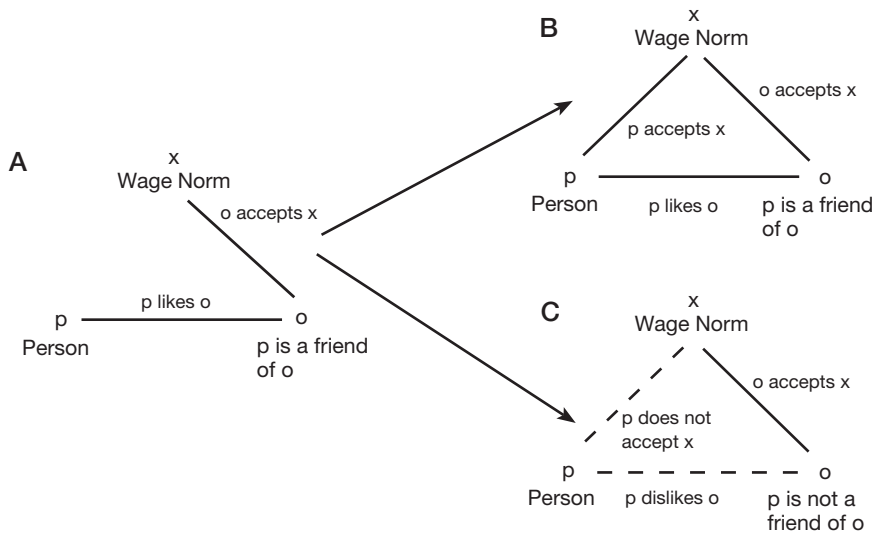
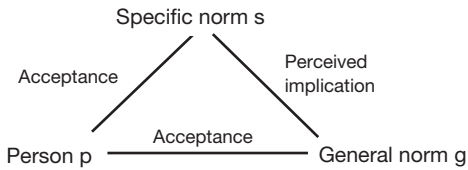


Figure 3. Acceptance of a General Norm and of Specific Norms.



example, p may think that symbolic rewards must not be counted in determining overall rewards. Thus, it is not the logical relationships between norms but the psycho-logical relationships which are relevant. Note further that even in a logical argument where a norm follows from a set of statements that include a norm utilities are involved.

In the previous section about descriptive beliefs some *implications* of the application of BT to explain descriptive beliefs and its compatibility with the CM were discussed. The points that were made there fully apply to the explanation of normative beliefs as well. So we need not repeat these implications.

6.3. Value expectancy theory

In applying VET, the dependent variable would not be a descriptive belief that something *is* the case but a normative belief that something *should*

be the case. If one holds the specific norm s that soldiers should get lower wages than miners, a perceived behavioral consequence is that one follows a general norm g which is beneficial (i.e., provides utility). VET allows us to introduce subjective probabilities. In the extreme case, an actor may be certain that norm s is implied by norm g. However, the probability need not be 1, that is, there may be some doubt about the implication. The equation would thus read:

(2) SEU (Belief that norm s holds) = $p_{SG} \cdot U$(s is implied by the general Norm g)

This equation includes only one behavioral consequence. We could add others. For example, the belief in norm s could be compatible with the normative expectations of friends. We will not add further components because our goal is only to show that VET can be applied.

In the discussion of descriptive beliefs, several *implications* of VET were discussed. The points made there hold for the application of VET to explain normative beliefs as well.

7. Explaining Attitudes and Preferences

In this section, we will first present a version of VET to explain attitudes because this version – the Fishbein-Ajzen theory – is well confirmed and widely used in social psychology. Based on this analysis we will apply BT. This is an easy exercise because we can build on our previous analyses.

7.1. Applying Value Expectancy Theory to Explain Attitudes

Boudon does not explicitly deal with the explanation of attitudes, although he sometimes mentions them (e.g., 2001: 200). Furthermore, it seems that he does not clearly distinguish between attitudes and norms. An attitude is, by definition, a positive valuation without any implication of oughtness. The lack of distinction between attitudes and norms is apparent when Boudon discusses the example that people may "prefer" to drive relatively fast in the city and for this reason regard traffic lights as a "good" (yet unpleasant) thing. This is because "traffic is more fluid with traffic lights than without." Therefore, the "value statement" that "traffic lights are a *good* thing" is accepted (Boudon, 2001: 150-151, italics in Boudon's text). It can be doubted that the previous statement – "traffic lights are a *good* thing" – is a norm. The term "good" may mean "effective" and, thus, may express the *fact* that traffic lights prevent accidents or make traffic faster. Furthermore, "good" may refer to a positive attitude toward the consequences of traffic lights. This implies that oughtness or morality is not involved. In general, traffic rules are just conventions that serve some purpose, and one does not have a bad conscience when they are violated, and there is no intrinsic valuation.

When discussing his example, Boudon explains the valuation of traffic lights by the fact that existing traffic lights have consequences that people like (or consider "good"). This argument is consistent with the well-confirmed attitude theory by Fishbein and Ajzen (see, e.g., 2009). The theory hypothesizes that there will be a positive attitude towards an object if an individual associates with this object positive features with high subjective probability. This is also held by VET, but the dependent variable there is attitudes and not behaviors.

Let us apply this version of VET to Boudon's example of the rain dances. One would predict that the Australians do not only engage in the rain dances as an instrumental activity (i.e., to bring about rain), but that they also like the dances or participation in the dances. Thus, a positive attitude toward the rain dance develops. What might be the "good reasons"? The explanation of the Fishbein-Ajzen theory would be that the rain dances are associated with very positive features. One is the rain. Furthermore, the members of the tribe might enjoy the social gatherings for their own sake.

A simplified example of an equation of VET where the dependent variable is an attitude (i.e., the SEU of having a positive attitude toward the rain dance) is the following:

(3) SEU(Attitude toward the rain dance) = $p_{AR} \cdot U(\text{Rain}) + p_{AF} \cdot U(\text{Friends present})$

Thus, the tribe members will like participating in the rain dance (i.e., they have a positive attitude toward the attitude object "rain dance") if the subjective probability p_{AR} that it rains and the utility U of rain is high and, furthermore, if it is likely that friends are present and if this has a high utility as well.

7.2. Explaining Preferences with Value Expectancy Theory

Attitudes differ from goals or preferences, but it often happens that goals develop if objects are valued positively. For example, if someone values computers very positively he or she will often have the goal of buying or possessing one. Boudon briefly addresses the explanation of goals and illustrates it with an example (2014). He assumes that the educational and social goals individuals acquire are due to taking as a reference people "they are mainly in relation with." This would yield the following equation:

(4) SEU (Educational goals) = $p_{EI} \cdot U(\text{Consistency with goals of important others})$

One could speculate that having a positive relationship to important others and not sharing their goals is psychologically unpleasant.

In general, it seems plausible to expand the previous equation by including other consequences. We will not explore this possibility further. For the

purpose of this paper it is important to note that VET is apparently capable of explaining preferences.

7.3. Applying Balance Theory

There is no question that BT can be applied to explain attitudes as well. To illustrate, it would be unpleasant if p perceives many *single* features of an object (z_1 to z_n of an object x – there are thus unit relations between x and the features z of an object) that are positively valued and ascribed with high probability, but if a negative evaluation is attributed to the *entire* object.

In regard to the explanation of preferences BT would imply, po would be the positive relationship of p and the important others. If p perceives that o (the important others) want, for example, to attend university (x), it would be unpleasant not to have the same preferences.

We will not explore the details of the applications sketched before. For the purposes of this paper it is only of importance that BT can apparently explain attitudes as well as preferences.

7.4. Conclusion

It is plausible that Boudon's explanation of attitudes and preferences could also be improved by applying the two social psychological theories: the selection problems can be solved, and irrational factors (such as affective factors) are included and integrated. Furthermore, the two theories clearly imply utility maximization. Finally, Boudon's general claim that there needs to be only one theory that explains a wide range of phenomena is confirmed by the previous analyses.

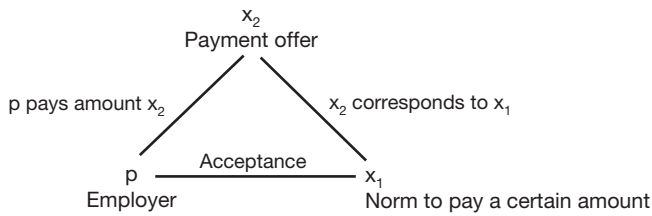
8. The Explanation of Behavior

As was said before, the major application of VET was the explanation of behavior. BT can explain behavior as well. We will show this by sketching the application of BT and VET with the previous Adam Smith example. This time, the explanandum is not the wage *norm.* The question we address is: will employers *pay* workers the wage they deserve according to the fairness norm outlined before? Assume an employer knows that there is a norm (say a law) that miners have to get a certain wage, and assume a miner applies for a job and is interviewed. The employer first makes an offer.

BT would be applied in the following way. Let the offer of p (the employer) be x in a pox system, where o refers to the important others. The employer knows what the important others (o) pay. There would thus be positive relationships between p and o on the one hand and o and x on the other. The expectation is thus that the wage offer would be x.

Assume instead that the cognitive system consists of p, x_1 (norm to pay an amount x) and x_2 (payment offer x) – see figure 4. It would be dis-

Figure 4. Explaining Payment Offers



sonant not to pay the amount that is demanded by the norm. In this case, px_2 would be negative. The system could be extended by including o and assuming that o pays x_2 and accepts the norm x_1. The reader might write o next to x_2 and connect these elements with a line. This would result in a balanced system.

In order to apply VET, let us change our example to explain participation in the rain dance. The following consequences could be included in a behavioral equation. Let there be a norm to participate in the ceremony, and let non-compliance be sanctioned negatively. The tribe members might value the company of others and they might have developed a positive attitude toward the ceremony itself. This yields equation 5:

(5) SEU(Participate in rain ceremony) = $p_{PN} \cdot$ U(Follow the norm to participate) $-$ $p_{PS} \cdot$ U(Negative Sanctions for not participating) $+$ $p_{PC} \cdot$ U(Company of others) $+$ $p_{PI} \cdot$ U(Intrinsic Motivation of participation in the rain dance)

So far we have formulated a total of five equations. They are a *hierarchical explanatory argument*. Equation (5) includes utilities which can be explained by previous equations. For example, probabilities are descriptive beliefs which could be explained by equation (1). The norm in equation (5) can be explained by equation (2).

Our conclusions for the explanation of behavior are the same as those for the explanation of descriptive and normative beliefs, attitudes and goals (or preferences): the two theories solve the problems of the CM; they are inconsistent with Boudon's rejection of utility maximization; and they confirm the generality assumption.

9. Boudon's Implicit Background Theory: The Wide Version of Rational Choice Theory

VET and BT are versions of RCT: they assume utility maximization and that preferences and constraints are determinants of the explananda (for details see below). The constraints are, for example, the perceived properties of the social network. Boudon is an emphatic opponent of RCT: he wrote three articles that attack this theory (1998, 2003, 2009b), and there are critiques in passing in

numerous publications. If Boudon's critique of RCT is correct then perhaps BT, VET and other social psychological theories have major weaknesses so that it is not meaningful to apply them in order to remedy problems of the CM. So perhaps one replaces one evil with another. It is thus important to examine to what extent Boudon's critique of RCT is tenable.

A major problem of Boudon's critique of RCT is that he does not distinguish between different versions. This is of utmost importance because these versions face different problems. RCT is not a single theory but a family of theories. In general, a critique of "the" theory of rational action is completely mistaken.

In analyzing Boudon's critique of RCT, the following questions are discussed. (1) It is first necessary to clarify which version of RCT is Boudon's target: is it an outdated narrow neo-classical version or an increasingly accepted social psychological wide version? (2) Does Boudon's critique hold for a wide version of RCT as well? (3) What are the differences between the CM and the wide version of RCT? (4) In regard to Boudon's rejection of utility maximization we examine whether his arguments are acceptable. The general conclusion is that the CM is consistent with a wide version of RCT.

9.1. A Brief Outline of the Narrow and the Wide Version of Rational Choice Theory

There are three hypotheses that characterize every version of RCT.[7] It is held that preferences and constraints determine behavior, and that individuals choose the behavior with the highest utility. The different versions impose restrictions on the kinds of variables and on the kind of utility that is maximized. It is useful to distinguish a narrow and a wide version. In a nutshell, the wide version admits a wide range of preferences and constraints and assumes that utility maximization occurs from the perspective of the actor. A narrow version does not accept these assumptions. In particular, the differences are as follows.

(1) It is often held that only egoistic preferences matter. In contrast, a wide version includes all kinds of preferences. In particular, altruistic preferences are admitted and goals to follow internalized norms.
(2) There are no restrictions on the kinds of constraints in the wide version either: not only material constraints, but also, for example, social sanctions or expectations of others are admitted (which may be constraints if they affect goal attainment of the actors).
(3) A narrow version assumes that reality is perceived as it is. According to a wide version, perceptions, or equivalently beliefs (which may be wrong), are explanatory variables.

7. For a detailed discussion of the different versions see Opp (1999) and Kroneberg and Kalter (2012); for the role of norms in RCT see Opp (2013a, 2014a). See further Braun and Gautschi, (2014) for an innovative new formal model of a wide RCT version.

(4) Utility maximization in the narrow version means that the actor chooses the behavior that is objectively (i.e., from the perspective of an omniscient observer) best for him or her. In a wide version, the actor chooses the alternative that is best from his or her perspective.

It goes without saying that the specific kinds of preferences and constraints that are assumed to be relevant for a behavior must be determined empirically. Obviously, circular reasoning or tautologies are excluded.

The wide version is increasingly accepted. To social psychologists, it is obvious that all kinds of motivation and perceptions are relevant. This is shown in numerous applications of VET. Unexpected empirical findings in economics have led to increasing skepticism about a narrow model that only admits egoism. For example, in the ultimatum and dictator game people typically apply fairness norms. The assumption of pure egoism would suggest that a subject in an experiment who can decide to keep a given amount of money or share part of it with an unknown subject would keep the whole amount. Actually, this happens rarely (see, for example, Henrich et al., 2004). In the same vein, the phenomenon of altruistic punishment (e.g., Fehr and Gächter, 2002) and much work in behavioral economics that points to subtle incentives (e.g., Thaler and Sunstein, 2009 and Ariely, 2009) are not compatible with a narrow version of RCT.

This short discussion indicates that it is absolutely necessary in critiques of RCT to distinguish different versions. It is a major problem of Boudon's critique of RCT that he attacks RCT as if it is one theory and not a whole family of theories.

9.2. Boudon's Critique of Rational Choice Theory and His Alternative

To be sure, Boudon grants that rational choice theory "is a family of theories with many versions" (2009b: 180). But he does not address their differences in detail. Instead, he presents six postulates (2003: 3-4, see also 2009b: 180) that describe RCT "in a general way" (2009b: 180) – see the summary in Table 1. But he does not show which of these postulates pertain to which version. As a matter of fact, the postulates address a mixture of the wide and the narrow version. The former seems to be Boudon's CM. Let us look at the postulates in detail.

Boudon's first postulate P1 with which he characterizes RCT is methodological individualism, that is, the claim to explain macro phenomena by processes on the micro level. This is shared by every version of RCT. P1 is thus consistent with the CM and Boudon's methodological orientation. P1 as well as P2 refer to the interpretive sociology ("Verstehende Soziologie") of Max Weber (Boudon, 2009b: 186). P2 contends that the "meaning" of an action to an individual is an explanatory variable, and this meaning consists of the reasons that an actor regards as valid (2009b: 192). P3 is an equivalent formulation of P2. P1 to P3 characterize what Boudon calls "the general theory of rationality" (2009b: 186). It is equivalent to the CM.

Table 1. Boudon's Characterization of Rational Choice Theory, the Narrow and Wide Version of Rational Choice Theory and Boudon's "Cognitivist Model" (CM)

Postulates	Rational Choice Theory: Boudon's characterization	Consistency of the postulates with the narrow, the wide version and the CM
P1	"[A]ny phenomenon is the effect of individual decisions, actions, attitudes etc." (methodological individualism).	This refers to micro-macro explanations which is a goal of any version of RCT and of the CM.
P2	"[I]n principle, an action can be understood."	The action's meaning (i.e., the reasons of an action) to the actor is important. P2 is held by the wide version and the CM.
P3	"[A]ny action is caused by reasons in the mind of individuals" (rationality postulate).	P3 is identical with the wide version and the CM.
P4	"[T]hese reasons [see P3] derive from consideration by the actor of the *consequences* of his or her actions as he sees them" (consequentialism, instrumentalism).	Restrictions to "instrumentalism" in a narrow sense (see the text) are an assumption only in the narrow version of RCT. The wide version and the CM address norm compliance and other motivations.
P5	"[A]ctors are concerned mainly with the consequences to themselves of their own action" (egoism).	Restrictions to egoistic motivations are an assumption only in the narrow version. All kinds of motivations are admitted in the wide version and the CM.
P6	"[A]ctors choose the line of action with the most favorable balance" (maximization, optimization).	*Boudon rejects P6.* Plausible: Boudon assumes that actors maximize *subjective* utility (see text). This is consistent with the wide version of RCT.

There can be no doubt that these postulates are in line with the wide version of RCT. Micro-macro explanations are a major goal of proponents of this version, and "reasons" are included in this version as well.

P4 to P6 are, Boudon argues, only sometimes true. They are the distinguishing features of RCT and the CM. This can only mean that these are postulates of the narrow version. Let us look at these postulates in more detail.

P4 clearly differs for the two versions of RCT. Boudon asserts that RCT can only deal with "instrumental rationality." This excludes internalized norms. Norm following in this sense is not "instrumental," in contrast to pursuing goals like earning more money. However, those who act according to internalized norms also pursue goals, namely following a norm or avoiding a bad conscience (for details see Opp, 2013a). One might distinguish several kinds of goals. Some goals refer to external states (such as earning money), others to internal states (such as doing one's duty). A wide version of RCT includes all kinds of goals. Pursuing norms is explicitly a motivation in the wide version *and* in the CM.

P5 (see also Boudon, 2012a: 17) is certainly correct for applications of RCT in many fields such as economics. Egoistic preferences are the only driving force of actors. But a wide version of RCT holds that *any* preferences may be explanatory factors. Accordingly, preferences may vary: people may more

or less take into account others' welfare (i.e., may be altruistic). At least this is explicitly considered as a possible motive in the wide version. In commenting on P5, Boudon criticizes that RCT does not explain normative phenomena. However, as has been argued before, RCT can actually be applied for explaining these phenomena as well. Anyway, Boudon's critique does not hold for the wide version. Furthermore, the claim not to restrict the theory to egoism is endorsed in the CM as well.

P6 means, Boudon asserts, that actors maximize utility from the perspective of an impartial and omniscient observer. This is held by proponents of a narrow version, but definitely not by those who advance a wide version. Here the hypothesis is that actors engage in *subjective* utility maximization. This means that actors do what they think is best for them in a given situation. Boudon rejects P6 without making any distinction about different versions of this assumption. We will return to P6 in the next sub-section.

In other writings where Boudon criticizes RCT without mentioning the previous postulates explicitly, he clearly refers to the narrow version. For example, he asserts that RCT "introduces the fiction of a solipsistic *homo sociologicus,* whereas the CM recognizes the *homo sociologicus* as a social being" (2012b: 18). This critique is clearly directed towards the narrow version. Obviously, the social environment imposes various costs and benefits on actors and is taken account of in a wide version and in the CM as well (see in particular Boudon, 2014: XX). In his preface to a collection of his essays in German (2013), he asserts that rational choice theory "postulates that the reasons which stimulate individuals are egoistic and instrumental," whereas in his CM there can be "supraindividual and cognitive reasons" (translation by KDO). The former are clearly a characterization of the narrow version, whereas the latter – if it means that reasons might be shared by a group of individuals – refer to the wide version.

Boudon's critique of the narrow version can be illustrated with his discussion of so-called "paradoxes," in particular with the paradox of voting (see, for example, Boudon, 2003: 6-7; 2012a: 7-8) that, in his opinion, RCT is not able to solve. From a narrow RCT it follows that nobody will participate in an election because a single voter has no influence on the outcome of an election. In reality, however, we find that election participation is far from zero. From the perspective of a wide version, various kinds of costs and benefits (in M. Olson's terms: selective incentives) may have an impact on voting. Which ones are important has to be tested empirically.[8] Boudon strongly criticizes this procedure. His argument is that introducing these other incentives is ad hoc and unacceptable. The exact reasons for this critique are not clear. RCT is a general theory that says that preferences and constraints influence individual behavior and that people do what they think is best for them. The theory imposes no limitations on the kinds of preferences and constraints. These

8. For a more detailed discussion of Boudon's analysis of the paradox of voting, see Opp (2014a) and also Opp (2001).

limitations – such as only considering egoistic preferences or ignoring internalized norms – are introduced ad hoc. The social psychological version of VET shows this clearly: no social psychologist sees any problem in introducing all empirically relevant consequences and testing their influence. This procedure is by no means ad hoc in order to "salvage" the theory (Boudon, 1998: 821), it is embodied in the theory. This critique is strange because factors such as (false) beliefs are also ingredients of the CM.

As was said before, the CM lacks an explicit reference to utility maximization. We will therefore discuss this assumption in more detail in the following section.

9.3. A Reconstruction of Boudon's Implicit Use of the Assumption of Utility Maximization

It happens that scientists explicitly reject certain hypotheses or methodological rules but actually apply them. For example, many scholars are against rational choice theory but actually apply it in their work (which holds, incidentally, for Analytical Sociology, see Opp, 2013b). This might be the case with Boudon's work as well. It is therefore instructive to look at some of Boudon's examples in order to examine the extent to which he implicitly assumes utility maximization.

Boudon tries to explain why clerks in a firm had violent conflicts on minor issues such as "being seated closer to a source of heat or light" (1996: 144 – the example is based on C.W. Mills's *White Collar* from 1951). A "cognitivist interpretation" of this "overreaction paradox" (2009b: 183) is the following. All workers get equal pay and their work is similar. There is further a norm that contributions to the production should match rewards. Any unequal reward (such as sitting closer to a window) is perceived and intolerable. As soon as the advantage of sitting close to a window is due to the decision of a supervisor it is an injustice. From a "utilitarian viewpoint" (as Boudon puts it), sitting close to a window matters little, but it is regarded as an injustice and therefore instigates conflicts.

Boudon claims that this is contrary to RCT. This is wrong when a wide version is used: norms such as equal pay for equal work are included as possible factors, and it is also possible to include environmental factors such as the workplace in the explanation. They are perceived constraints. So far, then, the explanation of the CM and the wide version of RCT do not differ.

However, the wide version would further argue that eliminating the injustice makes individuals better off. Thus, if this elimination is brought about by the conflicts, individuals prefer the new situation to the previous one. This is exactly what Boudon implicitly assumes. The conflicts are solved when no one has a privilege such as sitting close to a window. And this is in the interest of the workers – which is implied in the example. This is equivalent to arguing that solving the conflicts makes the workers better off or maximizes their subjective utility.

Let us look at the Adam Smith example again. Boudon's argument lacks a central explanatory step if an assumption about utility maximization is not included. On the one hand, Boudon writes that violation of a justice norm results in a "disequilibrium" (see the quotation above), which is obviously unpleasant. In other words, this situation is costly. Boudon's implicit assumption is that individuals want to avoid the disequilibrium. If this is not assumed, why is the disequilibrium regarded as relevant for the explanation? Isn't the assumption also that avoiding the disequilibrium is better for individuals than staying with the disequilibrium? Thus, individuals choose what is best for them. This is the assumption of subjective utility maximization.

Another assumption in this example is that the general fairness norm implies the acceptance of the norm that miners should get a higher salary. Why do individuals accept the implication of the general norm? Apparently, individuals would feel uneasy if they would not. This would be costly. Furthermore, accepting the implication is best from the viewpoint of the individual.

Utility maximization is further plausible when we imagine that the Australian tribes come into contact with modern science and learn that the traditional belief that rain dances generate rain is wrong. They are taught which processes lead to rain. Assume further that the new beliefs are acquired because they are convincing. The latter term means that they are regarded as true or superior to the competing traditional belief.

One explanatory step is missing: if modern science is convincing: *why will the traditional beliefs be given up*? Why not hold both types of beliefs? Apparently, holding both beliefs (i.e., beliefs that are inconsistent such as "p and non-p") is highly unpleasant, and therefore costly. It is not only cognitively unpleasant but will also be punished by important others. If you tell your friends, for example, that you believe "apples are healthy" and "apples are not healthy," you will be considered feeble-minded.

Holding a belief that is regarded as wrong further violates goals that most people subscribe to, namely, knowing the truth (at least for some matters). Reaching such a goal is beneficial. That is to say, not accepting a true statement is a cost.

In addition, accepting the belief supported by modern science has concrete advantages. The time and other resources invested in performing the rain dances could be used for setting up irrigating systems. Thus, giving up the false belief is clearly a benefit, and holding it is a cost.

The confrontation with modern science will probably lead to rejecting an entire *system* of traditional beliefs and replacing it with a new belief system. This is not only a question of just checking what is acceptable (see the previous quotation from Boudon, 1998: 824). It is a painful psychological process where some form of subjective utility maximization is involved.

Thus, in general, arguing that an actor prefers a (descriptive or normative) belief A to another belief B (or an action A to another action B) because he or she has good reasons for A and not B, means that he or she has *better* reasons for A than for B. This means that there is some choice, and that an individual

is *better off* when a certain belief is chosen and not another. In other words, the actor chooses what is subjectively better for him or her. This holds for the choice between descriptive beliefs, normative beliefs (see the discussion in Opp, 2013a, see also Opp, 2001), attitudes, preferences and behaviors.

These examples illustrate that Boudon seems to implicitly assume that actors do what they think is best for them. This is actually subjective utility maximization. If this is granted, then the CM is consistent with the wide version of RCT.

Many formulations in Boudon's work come very close to or are even identical to the hypothesis of utility maximization. For example, the CM states that "social actors try to act in congruence with reasons they perceive as valid" (2009b: 192). Why do they try to reach congruence if that doesn't make the actors better off? Another statement by Boudon is that an actor prefers the theory that accounts "for given phenomena in the most satisfying possible way (in accordance with given criteria)" (2009b: 184). "Satisfying" points to the fact that acceptance is "more satisfying" than non-acceptance – a clear case of subjective utility maximization. Given various goals, Boudon argues (2009b: 193), actors "are rational in the sense that they look for the best or at least for a satisfactory system of reasons able to provide a ground to their answer." Can there be a clearer description of subjective utility maximization?

But assume we accept Boudon's claim. What is the alternative to subjective utility maximization? In regard to beliefs, the answer that an actor accepts a belief that is convincing begs the question of why this is done. Why not keep the old belief?

What are Boudon's *explicit arguments* against utility maximization? Let us look at some quotations. The CM "is drawn from the 'rational-choice model' by lifting the restriction that the reasons of social actors should always be of the cost-benefit type" (1996: 124). In another article, Boudon argues (1998: 824): "endorsing a theory is a noninstrumental action … the question the actor is confronted with here is not to maximize a cost-benefit balance, but to check whether, to the best of his knowledge, an idea is acceptable." It is striking that no empirical evidence is provided for his rejection of utility maximization. He just asserts that it is false. Furthermore, Boudon does not put forward an alternative proposition.

The most important argument against Boudon's claim is that it is inconsistent with major social psychological theories. This holds true for the two theories applied before, but also for other theories such as learning theories. Their assumption which is usually not formulated in an explicit way is that individuals choose a situation that they think is best for them, in a given situation.

10. General Conclusions

Boudon's basic idea that everything social scientists are interested in can be explained by a single theory is new and has been confirmed in the previous

analyses. This should encourage social scientists to try to expand the range of application of their theories to encompass not only behavior. It would be an important research program to work on an encompassing theory that Boudon had in mind.

The basic idea of Boudon's theory that beliefs (i.e., "reasons") are explanatory variables of a wide range of phenomena is nowadays widely accepted in quite different schools of the social sciences. However, alternative approaches focusing on objectivist propositions such as a narrow RCT and on methodological collectivism are still strong. So discussing Boudon's work and expanding it may in general strengthen a subjective approach that is advocated by this author as well.

The title of the paper is "The Explanation of Everything." Are there questions social scientists are interested in that are not addressed by the CM? The answer is yes. Among the questions not answered by Boudon's theory are at least the following. (1) The question of when a behavior is planned and when enacted spontaneously has become a focus of social sciences since the rising interest in dual-process theories. This question has not yet been addressed in the present paper and I have not found any detailed analysis in Boudon's work either. (2) Another issue is the role of objective factors in the formation of beliefs. We know at least since the work of Kahneman and Tversky that misperceptions are ubiquitous. But we still know little about the conditions of when people believe what. (3) When does a negative attitude develop toward a behavior and when does a norm come into being? For example: when do people not *like* it when somebody smokes in their presence, and when does a *norm* emerge that one should not smoke in one's presence? This question is rarely addressed in the literature and is not dealt with in the CM either. But perhaps Boudon would answer: a norm and not only an attitude will emerge if the actor has good reasons for demanding something. This raises the question of what these reasons might be.

For this writer, the most fascinating part of reading Boudon's presentations of his theory was his creative use of the work of classical writers, in particular Max Weber, Émile Durkheim and Alexis de Tocqueville. Boudon's reconstructions of these explanations are fascinating in their own right.

Bibliographic references

Ariely, Dan (2009). *Predictably Irrational. The Hidden Forces That Shape Our Decisions*. New York: HarperCollins.

Becker, Dominik (2013). "The Impact of Teachers' Expectations on Students' Educational Opportunities in the Life Course: An Empirical Test of a Subjective Expected Utility Explanation". *Rationality and Society*, 25 (4), 422–469. http://dx.doi.org/10.1177/1043463113504448

Boudon, Raymond (1989). "Subjective Rationality and the Explanation of Social Behavior". *Rationality and Society*, 1 (2), 173-196. http://dx.doi.org/10.1177/1043463189001002002

— (1994). *The Art of Self-Persuasion. The Social Explanation of False Beliefs*. Cambridge: Polity Press.

— (1996). "The 'Cognitivist Model': A Generalized 'Rational-Choice-Model'". *Rationality and Society*, 8 (2), 123–150. http://dx.doi.org/10.1177/104346396008002001
— (1998). "Limitations of Rational Choice Theory". *American Journal of Sociology*, 104 (3), 817–828.
— (2001). *The Origin of Values. Sociology and Philosophy of Beliefs*. New Brunswick: Transaction.
— (2003). "Beyond Rational Choice Theory". *Annual Review of Sociology*, 29, 1–21. http://dx.doi.org/10.1146/annurev.soc.29.010202.100213
— (2009). "On Axiological Rationality". In: Hill, Paul; Kalter, Frank; Kopp, Johannes; Kroneberg, Clemens and Schnell, Rainer (eds.). *Hartmut Essers Erklärende Soziologie. Kontroversen und Perspektiven*. Frankfurt and New York: Campus.
— (2009b). "Rational Choice Theory". In: Turner, Bryan S. (ed.). *The New Blackwell Companion to Social Theory*. Oxford: Blackwell. http://dx.doi.org/10.1002/9781444304992.ch9
— (2012a). "The Micro-Macro Link. Notes on a planned course on 'truth, beliefs, interpretation'". Unpublished manuscript.
— (2012b). "'Analytical Sociology' and the Explanation of Beliefs". *European Journal of Social Sciences*, 50 (2), 7–34. http://dx.doi.org/10.4000/ress.2165
— (2013). *Beiträge zur allgemeinen Theorie der Rationalität*. Translated by Felix Wolter. Tübingen: Mohr Siebeck.
— (2014). "On Context". In: Friedrichs, Jürgen, and Nonnenmacher, Alexandra (eds.). *Soziale Kontexte und soziale Mechanismen*. Special issue 54 of the *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (*Cologne Journal of Sociology and Social Psychology*). Wiesbaden: Springer VS. In press.
Braun, Norman; Gautschi, Thomas (2014). "'Zwei Seelen wohnen, ach! in meiner Brust'. Ein Rational-Choice-Modell innerer Konflikte". *Zeitschrift für Soziologie*, 43 (1), 5–30.
Breen, Richard (1999). "Beliefs, Rational Choice and Bayesian Learning". *Rationality and Society*, 11 (4), 463–479.
Breen, Richard; Goldthorpe, John H. (1997). "Explaining Educational Differences: Towards a Formal Rational Choice Theory". Rationality and Society, 9 (3), 275–305. http://dx.doi.org/10.1177/104346397009003002
Durkheim, Émile (1915, first 1912). *The Elementary Forms of the Religious Life* (translated by Joseph Ward Swain). London: George Allen & Unwin.
Feather, Norman T. (1982). *Expectations and Actions: Expectancy-Value Models in Psychology*. Hillsdale, N. J.: Lawrence Erlbaum.
— (1990). "Bridging the Gap between Values and Actions. Recent Applications of the Expectancy-Value Model". In: Higgins, E. T. and Sorrentino, R. M. (eds.). *Handbook of Motivation and Cognition. Foundations of Social Behavior*. Vol. 2. New York: Guilford Press.
Fehr, Ernst; Gächter, Simon (2002). "Altruistic Punishment in Humans". *Nature*, 415 (March), 137–140. http://dx.doi.org/10.1038/415137a
Ferster, Charles B.; Skinner, Burrhus F. (1957). *Schedules of Reinforcement*. New York: Appleton-Century-Croft.

Fishbein, Martin; Ajzen, Icek (2009). *Predicting and Changing Behavior. The Reasoned Action Approach*. New York and Hove: Psychology Press.
Heider, Fritz (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.
Henrich, Joseph; Boyd, Robert; Bowles, Samuel; Camerer, Colin; Fehr, Ernst; Gintis, Herbert (eds.) (2004). *Foundations of Human Sociality. Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
Kroneberg, Clemens; Kalter, Frank (2012). "Rational Choice Theory and Empirical Research: Methodological and Theoretical Contributions in Europe". *Annual Review of Sociology*, 38, 73–92.
http://dx.doi.org/10.1146/annurev-soc-071811-145441
Matsueda, Ross L.; Kreager, Derek A.; Huizinga, Davis (2006). "Deterring Delinquents: A Rational Choice Model of Theft and Violence". *American Sociological Review*, 71 (1), 95–122.
http://dx.doi.org/10.1177/000312240607100105
Merton, Robert K. (1957). Social Theory and Social Structure. Glencoe, Ill.: Free Press.
Opp, Karl-Dieter (1999). "Contending Conceptions of the Theory of Rational Action". *Journal of Theoretical Politics*, 11 (2), 171–202.
http://dx.doi.org/10.1177/0951692899011002002
— (2001). "Warum denken normale Leute, sie seien politisch einflußreich? Die Erklärung einer kognitiven Illusion". *Jahrbuch für Handlungs- und Entscheidungstheorie* 1, 9–47.
http://dx.doi.org/10.1007/978-3-663-10417-9_1
— (2013a). "Norms and Rationality. Is Moral Behavior a Form of Rational Action?" *Theory & Decision* 74 (3), 383–409.
http://dx.doi.org/10.1007/s11238-012-9315-6
— (2013b). "What is Analytical Sociology? Strengths and Weaknesses of a New Sociological Research Program". *Social Science Information* 52 (3), 329–360.
http://dx.doi.org/10.1177/0539018413483939
— (2014a). "Instrumental, Axiological Rationality and the Explanation of Norms. Cherkaoui's (and Boudon's) Critique of Rational Choice Theory and Its Ability to Explain Norm Emergence". In Manzo, Gianluca (ed.). *Paradoxes, Mechanisms, Consequences: Essays in honor of Mohamed Cherkaoui*. Oxford: Bardwell Press. In press.
— (2014b). (7th ed.). *Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theorienbildung und praktischen Anwendung*. Wiesbaden: Springer VS.
Popper, Karl R. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.
Schwartz, Barry; Reisberg, Daniel (1991). *Learning and Memory*. New York: W.W. Norton.
Thaler, Richard; Sunstein, Cass R. (2009). *Nudge. Improving Decisions About Health, Wealth, and Happiness*. London: Penguin Books.
Van de Rijt, Arnout (2011). "The Micro-Macro Link for the Theory of Structural Balance". *Journal of Mathematical Sociology*, 35 (1 to 3), 94–113.
http://dx.doi.org/10.1080/0022250X.2010.532262
Wigfield, Allan; Tonks, Stephen; Klauda, Susan Lutz (2009). In: Wentzel, Kathryn R. and Wigfield, Allan (eds). *Expectancy-Value Theory. Handbook of Motivation at School*. New York: Routledge.

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2114

Are there many types of rationality?

Pierre Demeulenaere
Université Paris-Sorbonne
pierre.demeulenaere@gmail.com

Received: 16-06-2014
Accepted: 10-07-2014

Abstract

This paper has three basic objectives. First, it summarizes the general theories of rationality that are currently available in the social sciences. Secondly, it describes the specificity of Boudon's approach to rationality in contrast to the previous ones. Finally, it suggests a few ways in which this approach can be enhanced and complemented in order to achieve a solid theory on the interpretation of human actions in a social world.

Keywords: Boudon; axiological rationality; cognitive rationality; ordinary rationality; explanation.

Resumen. *¿Existen muchos tipos de racionalidad?*

Este artículo tiene tres objetivos principales. En primer lugar, se ofrece una panorámica de las teorías de la racionalidad actualmente disponibles en ciencias sociales. En segundo lugar, se discute la especificidad de la aproximación de Boudon en contraste con las anteriores. Finalmente, se sugieren algunas vías a través de las cuales la aproximación puede ser mejorada y completada con la finalidad de alcanzar una teoría sólida de la interpretación de las acciones humanas en el mundo social.

Palabras clave: Boudon; racionalidad axiológica; racionalidad cognitiva; racionalidad ordinaria; explicación.

Summary

1. Introduction

Raymond Boudon's contribution to social theory has two main dimensions. First, he has been a pioneer in the theorization of mechanisms that govern social reality and allow social explanations (Boudon, 1974). Second, he has been responsible for a broadening of the notion of rationality, beyond the one involved in the standard rational choice model (Boudon, 2007, 2009). The link between the two aspects is that mechanisms in social life necessarily involve individual actions, and that these actions can be to a certain extent interpreted in terms of rationality. This implies a reflection on the notion of rationality itself, and its characterization. It is part of the indispensable more general reflection on the process of understanding action in a social context. This is a central issue for social sciences: collecting or simulating data is only the first step; the second step is to explain social phenomena, and this always requires procedures for understanding behaviour.

Boudon's contribution to the theory of rationality has itself two main dimensions (Boudon, 2011): on one side, he has insisted that rationality should not be understood only as instrumental or utilitarian, but should also involve what he has named "cognitive" and "axiological" dimensions. The idea is that the standard rational choice model does not satisfactorily take into account these two essential aspects of action that are necessary to its interpretation. Therefore, the grammar of the interpretive work in social sciences requires the introduction of cognitive and axiological features I will describe further on in this paper. Second, although rationality has various dimensions (mainly, in his words, instrumental, utilitarian, cognitive and axiological), it is unified as a general structure of human behaviour and named as such as "ordinary rationality". He has described his theory of rationality as referring to this "ordinary rationality" beyond the various dimensions he had emphasized. This poses the problem of the link between the general idea of rationality and its various dimensions. What is the common element of rationality that allows us to speak of rationality when we introduce several types of rationality?

In this paper, I will try to:

1) Summarize the general theories of rationality that are currently available in the social sciences.
2) Describe the specificity of Boudon's approach to rationality in contrast to the previous ones.
3) Suggest a few ways by which this approach can be enhanced and completed in order to achieve a sound theory of the interpretation of human actions in a social world.

I will concentrate on a general approach to the idea of rationality which is rarely followed, oriented by the question of the justification of the very idea of rationality: when are we entitled to speak of rationality and why should we consider an action as a rational one? What is, at a meta level, the justification

of such an idea? In particular does it involve a normative dimension and, if this is the case, how is it articulated to the positive investigation of social science?

2. Various theories of rationality

There are five basic widespread notions of rationality in the social sciences (including Boudon's approach). They are classically linked to methodological individualism, since theories of rationality are theories of rational actions. All of them can be related, more or less clearly, to Popper's "problem-solving" notion (Popper, 1967).

The first one is to consider that individuals are rational whenever they act in an intentional way. The intentional decision to act corresponds to the reason someone has to act, for instance smoking for her pleasure, or not smoking in order to avoid health problems. This classical idea can be found in Weber.

An irrational behaviour would be then to act on the basis of non-intentional causes. For instance, Kahneman (2011) reports a study about parole judges observed in their decisions. In this example, individual judges tend to be more severe when they are hungrier: clearly in this case non-intentional and unconscious causes affect intentional decisions. The judges are not aware of the influence of those factors on their deliberate decisions. This means that intentional motives do not have the sole influence on individual decisions; however, in this example it is still intentional decisions that are at stake, partly determined by unconscious trends. It can be also the case that people will act in a manner that is opposed to their intentions, for instance when they smoke although they would prefer not to smoke.

Intentional action can clearly be linked to and dependent on psychological features or social norms. When we speak of reasons to act, these reasons can obviously include the acceptance of given social or cultural norms. Therefore, in this first meaning of intentionality, rationality does not give clear criteria for decisions that would lead towards specific kinds of choices, which would, for instance, be different from and maybe opposed to the acceptance of social norms. But it stresses the fact that in order to be rational, an intention of action must be the effective source of action: I decide to act according to my intention to act, whereas, when I am not rational, I act on the basis of causes that do not correspond to my intentions, either because they are opposed to my intentions, or because they overcome my intentions, or because I have no intentions at all. This has a problem-solving dimension since I must decide how to act given my situation (Shall I Smoke? Shall I not smoke given all the consequences I know about smoke?). Ideally, an individual decision is the "correct" decision one should take, according to one's situation, although the criteria of such correctness are not one-sided and do not lead to interpersonally valid norms of decision. The emphasis is not so much on the criteria of the decision, but on the fact that the intention corresponds to a deliberate choice. Therefore, a choice is often based on subjective motives. This is a very weak sense of rationality, although an important preliminary one.

The second notion of rationality is linked to the ideal norm of consistency or transitivity of choices. The rationality does not stem from the contents of individual choices as such, but just from the fact that individuals are consistent in the ordering of their subjective preferences. Preferences are clearly not considered to be rational in this case. It is the official position of neo-classical economics, where rationality does not involve any kind of "right" decision, but only the fact that people do not take inconsistent decisions (Sen, 1977). Irrationality here would correspond to inconsistent choices of the kind, again, described by Kahneman and Tversky (2000), or Elster (2010). Here again, the transitivity of choices implies that the contents of the choice are not part of the rational dimension, and therefore depend on other dimensions (either subjective, psychological or social). Preferences are considered to be outside the realm of rationality, sometimes characterized as irrational, and sometimes as "a-rational". The criterion of rationality is here only consistency. Although limited in scope, it clearly has a normative aspect, which poses two questions: Why should consistency be assimilated to rationality, that is, why does rationality involve consistency? Alternatively, is it conceivable for someone to have reasons not to be consistent in her preferences? This poses the problem of the reversal of preferences, and of the localization of consistency in time: should an apparently inconsistent choice be interpreted as irrational, or as an adaptation of choices given the possible evolution in time of given preferences? This is the position defended by Elster, who does not consider that reversal of preferences in time should be considered as irrational. It is a difficult topic: on the one hand, it is clear that in sociology the idea of stable preferences relatively independent of time change is useless, and should not be the default solution that would be highlighted only for modelling reasons. We have to take into account, in a realistic way, all the changes of preferences that can be observed, and there is moreover room for a rational interpretation of change of preferences. On the other hand, we cannot escape the issue of consistency of choices (for example in polls analysis), although successive choices always intervene in definite periods of time.

The third notion of rationality links it to an instrumental dimension of action, where an actor chooses the adequate means to reach an end. The idea of rationality derives here from an ability to select the *correct* means to achieve an end, which requires that the normative notion of correctness is here implied, whereas in the previous notion of rationality, the normativity of rationality was located in consistency. It should be noted that, although instrumental rationality is widely associated with standard economics, it can be separated from it from two opposed points of view: first, as was noted by Weber, instrumental action does not necessarily imply any kind of economic motives. Second, the transitivity of preferences, which is the core of neo-classical formalization of economic action, does not in fact involve a notion of instrumental action. Here again, instrumental rationality presupposes and involves motives that can be related to a social dimension. However, the criteria of rationality are clear and rather univocal. Although this notion of instrumental rationality has

been defined by western philosophy and sociology in specific contexts, and has its own scientific history, it can in retrospect be applied to many human conducts in the most diverse cultural settings. An important question in this respect is whether animal behaviour, which is often clearly instrumental, can be defined in terms of rationality on this basis (Turner and Maryanski, 2008). A more common challenge for such a theory of instrumental rationality is to interpret the behaviours that seemingly depart from its requirements: for instance magic rituals.

The fourth notion of rationality is clearly the most widespread in today's social sciences, although it suffers major ambiguities (Demeulenaere, 1996) and can be the subject of deep criticism. It is the idea that people act rationally whenever they efficiently pursue their self-interest (Coleman, 1990). This is a much different and narrower conception of rationality: it does not insist only on the deliberate choice of an end, nor on the means to achieve an end, nor on the consistency of the ends, but on the ends of action themselves, namely self-interest.

In this scenario there is, however, a deep hesitation between two alternatives (Demeulenaere, 2011): either the preferences are supposed to be the same for all the actors (like a typical preference for an increase in wealth), or they can be dissimilar (like smokers' or non-smokers' preferences). Consequently, it is not easy to interpret the notion of costs and benefits, since they can correspond to the same choices for a set of actors, or, conversely, to opposed choices; for instance, in the case of smokers and non-smokers who clearly do not have the same appreciation of costs and benefits regarding smoking. On the one hand, it seems clear that people do often have the same sense of costs and benefits and that it is relevant to interpret their actions in terms of maximizing benefits and reducing costs on the common basis of shared conceptions of advantages and disadvantages. On the other hand, they often do not agree on what is a cost, and what is a benefit, and in this case any reference to costs and benefits has no specific contents and only designates different choices and different subjective (or socially prescribed) evaluations of a situation. Saying that people maximize their benefit tends to be tautological, since any choice is by definition an attempt to maximize satisfaction. This problem is a fundamental one and is unresolved by the standard rational choice model which tends to assert at the same time that preferences are diverse, and to reason in terms of implicitly or explicitly unified costs and benefits: the two contentions are not consistent.

This theory is strongly normative since it reduces rationality to the choice of self-interest, that is, it considers that it is rational to select one's interest instead of, for instance, to adhere to a value or a social norm (and values or social norms should be interpreted in terms of self-interest). However, it can be argued that the theory is not normative at all (Homans, 1987), since it is just a description of a typical behaviour and an attempt to model the fact that people tend to maximize their self-interest, without any evaluative appreciation of that attitude. The rational choice model would be just a matter of action

modelling, devoid of any action evaluation. This position can be defended with good arguments. It would have as a consequence that the notion of rationality is here rather useless, and that its concept (with all its background) is not really part of the theory. The idea would only be that people typically tend to maximize their interest, without any consideration about the rationality of this conduct. The notion of rationality would be taken in a sense that is devoid of all its usual requirements. Therefore, it should be in fact better to avoid it because of its normative connotations.

However, this is implausible. Although the rational choice model can be reduced to a self-interest maximization modelling of action, it does in fact involve specific rationality considerations for two different and unequally acceptable reasons. One of them is indeed very disputable.

The first reason is the instrumental dimension that is involved in this characterization of action: people, when they pursue their self-interest, tend to select the appropriate means to reach their ends. This is the core idea that led Weber and Pareto to interpret economic actions in terms of rationality, and precisely in terms of instrumental rationality, whereas before them, for instance in Mill's theorization of economic action, no explicit reference to rationality was made. This assimilation of economic action to instrumental rationality can be challenged in return on two bases: first, it can be noticed that most often there is no availability of definite means to reach economic ends; second, there is no reason to exclude the fact that in pursuing economic ends people can sometimes use non-rational or irrational methods (like, for instance, feng shui beliefs). The connection of economic motives to adequate instrumental choices is not an obvious one. It has a normative side, since it is considered that, having economic motives, people *should* select the adequate means to reach them.

The second reason is more controversial: it has its roots in some type of positivist belief that asserts that it is rational for someone to pursue her interest. This derives from a negative proposition: belief in values cannot be rational, since they are not based on fact description. The positive proposition would be to assert that, by nature, people tend to pursue their interest (as opposed to cultural values) and this is why it is rational to adopt this conduct. Pareto was ambiguous about the interpretation of the status of interests: on the one side he maintained that any interest, as much as a value, does not belong to the logical side of action, since it cannot be based on the description of a fact. On the other hand, he suggests on several occasions that people who pursue their interests have so-called logical attitudes. At any rate, it has been since then commonly admitted that rationality involves the pursuit of interest, not just as an instrumental choice of the means to reach them, but also as the choice of interest itself as a rational end. For instance, in Parsons' theorization of action, when he refers to values, he refers to "non-rational" ends, as opposed to rational interests. Similarly, Elster (2009, 2010) refers to irrationality when people adopt attitudes that do not correspond to self-interest.

In sum, we have four different criteria for rationality:

— Intentionality, the idea of rationality being here that people have reasons to decide what they decide.
— Consistency, since it appears intuitively that people should not contradict themselves in formulating their choices.
— Adequacy of the choice of the means to reach an end, when a correct choice of means is in fact available, and since it would be ineffective to select the wrong means.
— The choice of interest, since it is a sound choice regarding human nature.

Clearly, the fourth notion does not provide the same self-evidence as the three previous ones, as I have just pointed out. A fifth notion of rationality is much rarer in social sciences, as a consequence of the influence of positivism: it extends the scope of rationality to the determination of somehow right choices that include not only correct beliefs but also normative deliberations about ends. Boudon has advocated the reintroduction in sociology of this classical meaning of rationality in philosophy. I will outline now his conception.

3. Boudon's theory of rationality

Boudon's main idea regarding the theory of rationality is that people have *good reasons* to do what they do, that is, not only reasons, but precisely good ones. Given their situation and given the information they have, they tend to do the somehow *right* choice that is available from their point of view. To introduce the notion of point of view is to consider that a good choice is often, in principle, attainable. However, people do not necessarily make this ideally good choice, but the one they do is the best from the point of view of their situation and information. When the notion of right choice is introduced (through the notion of "good reasons" to make a given choice) this can involve at the same time truly right choices, *per se* (given our capacities to justify a choice), and sometimes erroneous choices, but the latter appear to be the best ones given the situation of the actor and her point of view. This theory is rather similar to Popper's one, or Dray's one (1964), but Boudon's attempt was more explicitly intended to limit the scope of the predominant rational choice theory and to reintroduce a more specific treatment of the rationality of normative issues. It is different from Simon's theory of bounded rationality in his emphasis on *good* reasons.

Boudon's theory of rationality is based on two main critiques of other conceptions of rationality. One is addressed to the classical Humean theory. There is a debate about the exact significance of Hume's contention about rationality I will not discuss here. I will just mention that Hume's theorization is more complex and subtler than its caricature. Boudon develops his comments on the basis of two common interpretations of Hume's legacy:

— The fact/value dichotomy
— The restriction, given this dichotomy, of rationality to the choice of means, not to the choice of ends.

Hume's legacy has been influential in the social sciences, especially in economics, and it can be said to have partly led to the current rational choice model which associates an instrumental dimension regarding the choice of the means, and a so-called utilitarian one regarding the ends: people rationally pursue their self-interest. I have just indicated that those two dimensions are fundamentally different and that they should be separated in the conceptualization of the issue of rationality.

Boudon's critique regarding this tradition and this model has three main aspects, one about beliefs, one about the instrumental dimension, and one about the choice of norms. Regarding beliefs, Boudon has three main contentions. First, he introduces the idea that a cognitive dimension (and rationality) is necessary to interpret positive beliefs (for instance about magic) that are shared by people in given situations. Although Boudon says that cognitive rationality does not belong to the Humean theorization of rationality, we can reply that this is not the case: Hume clearly asserts that reason can select correct beliefs; in particular the choice of means to reach an end presupposes such beliefs. Naturally the word "cognitive" is a modern one, but the idea of a relevant positive belief based on an appropriate use of reason clearly belongs to Hume's legacy.

Regarding errors, Boudon has been critical of Kahneman's and Tversky's perspectives when they insist on the impact of biases in the formation of beliefs and decisions. Boudon criticizes their trend (although not an overwhelming one) to interpret such biases as causal ones, and consequently to characterize the behaviours that are dependent on such biases as irrational. Boudon has instead insisted on the reasons for which individuals endorse false beliefs, given their situation and information. He has pledged, following Weber, for a rationalistic interpretation of error instead of a causal and irrationalistic one. This is a major discussion for the philosophy of social sciences: the debate must clearly continue to go on; it is related to the link between psychology and the social sciences.

The most important topic Boudon has insisted on is that beliefs are also implied in matters about norms. This means that normative beliefs are implied in the adoption of norms as well as positive beliefs. Both depend on reasons: reasoning about norms and values does exist in everyday life. This line of argument is a clear rupture with Hume's or Pareto's tradition. Boudon has tried to convince that this theory of reasoning about values was on the contrary defended by Weber or Durkheim (Boudon, 1998, 2000), despite the common thinking that they did not do so. This particular historical point will not be discussed here. The main idea here is that people have systems of reasons about the values and norms they endorse, as much as they have systems of reasons about the positive beliefs they endorse. Therefore, it is irrelevant to separate the positive and the normative aspects, since they are both unified by the exist-

ence of beliefs and by the fact that those beliefs are supported by reasons people have, in their situation, to accept them. For instance, people support the death penalty (normative position) because they believe it is efficient. They can renounce it when they find out that it is not in fact efficient (positive belief).

Instrumental rationality appears in this perspective as only one aspect of having reasons to take a decision or to make a choice. Boudon, however, is critical of three dimensions commonly associated with the reduction of rationality to instrumental action. He does not so much insist, as Hume did, on the empirical evidence for the selection of the right means as a base for the idea of a rational selection. Instead, he describes instrumental action as the choice made by an actor of a definite action because of its consequences. In Weber's definition of *Zweckrationalität*, both aspects are present, and even a third one: the right choice of the means (supported by empirical evidence), the anticipation of consequences, and the comparison of the ends given those predictable consequences.

What Boudon has claimed is that people often behave not only on the basis of the consequences of their action (which he envisages as utilitarian ones), but on behalf of principles they endorse. Moreover, they do not behave only to fulfil their self-interest, but they obey to moral principles.

Two examples in this case are central: first the paradox of voting, second the prisoner's dilemma. Both have no solution on an instrumental rationality principle. In a large poll, people should not vote on an instrumental basis because their vote makes no difference. But they do vote. People should not cooperate in a prisoner's dilemma situation, because their interest leads them not to cooperate. However, they do cooperate. This is why the notion of axiological rationality is introduced.

Boudon's reference to axiological rationality is his most innovative contribution. The main idea is that it is possible to reason about values and norms, and that values and norms problems can have good solutions. Boudon distances himself from the rational choice model that would treat the adoption of norms either as a part of given irrational preferences, or as the result of a calculation of self-interest. Since there are reasons that are given by actors for their choices of values, it cannot be said to be just the manifestation of an irrational trend. Since people often endorse principles that do not correspond to their self-interest, it cannot be said that they adopt values and norms only when they suit their interest. They do not stem either from natural unconscious trends.

On the other hand, Boudon rejects a culturalist perspective that would reduce the adoption of norms and values to a consequence of various social and cultural settings, without any intrinsic justification. The idea is again that people believe in the values they adopt and are often able to give reasons for their beliefs. This is why Boudon contrasts his interpretation of Weber's axiological rationality with the one which is commonly held: axiological rationality does not just correspond to coherence with values, whatever their roots are; axiological rationality is instead a rational ability to reason about norms, and to adopt a behaviour that stems from this rational acceptance of norms.

For instance, people can understand that from a consequential point of view, they have no reason to vote in a large-scale poll, because one vote makes no difference. Despite this knowledge, they will vote because they understand, on the basis of rationality, that participation is important in order to respect the democratic principles they adhere to from the perspective of axiological rationality.

What then are the criteria that allow us to determine, rationally, the validity of normative principles? Boudon has mostly developed his theory as a comment on crucial examples. The good reasons are not unified in his presentation as a deductive body of propositions: they mostly arise from particular discussions of particular issues, which are not grounded in a few general assumptions. Instead, every single case is envisioned from its own properties.

However, it can be said that his reflection about the rationality of norms has led Boudon to stress three general principles. They clearly do not exhaust all the necessary elements that are present in reasoning about norms, but they constitute recurrent major ingredients (although very different in nature) of this reasoning.

The first one is the principle of human dignity, often referred to as a substantial overall source of normative beliefs. It seems to Boudon that this principle does not have to be justified as such, it has a kind of self-obviousness.

The second is the principle of impartiality theorized in Smith's impartial spectator. People have a more appropriate sense of the correct normative attitude when their own interest is not at stake.

The third is an evolutionary stance Boudon has increasingly introduced in his reflection about norms: new principles tend to arise, and are selected because of their strength and intrinsic appeal; for instance, the principles of the separation of powers. Once they are defined and applied, it becomes extremely difficult to challenge them. This leads to a rather optimistic and evolutionary theory of norms where better norms, in the long run, tend first to appear and then to replace inadequate ones. Boudon has repeatedly, for instance, described the death penalty as a more and more inacceptable sanction: because it is inefficient, because it is cruel, and because it makes any judicial error irreparable. Boudon thought that the death penalty would be progressively abolished, in particular in the United States of America. In his view, the theory of axiological rationality could allow us to make some specific predictions.

4. Towards an integrated theory of rationality

Boudon's theory of good reasons has major theoretical advantages. It is an attempt to bridge values and facts, positive beliefs and normative beliefs. It is an effort to link sophisticated intellectual investigations and ordinary thinking about norms and values. It is therefore a combination of philosophical investigation and of sociological explanation. It puts together normative considerations and empirical data. It relates correct beliefs and right norms to false beliefs and unjust practices in a common interpretation. It combines a

diversity of rationality aspects (cognitive, instrumental, utilitarian, axiological) in a unified perspective. It departs from a narrow naturalism that would not be sensitive to social variations and cultural diversity of norms, but it similarly rejects a simple culturalism that would not succeed in explaining the variation of norms on the basis of the change of situations and of the rational dimension of any belief. It integrates more diverse social particular situations to its framework than other theories of rationality. It leads us to abandon a naïve and intuitive sociological commentary on data by providing a more systematic guide to interpretation that takes into account the logic of formation of beliefs and decisions in various social situations and individual positions. In particular it gives all its strength to the idea that people do not hold beliefs or decide actions on the basis of only cultural or subjective grounds: their rationality is what allows us to interpret their formation.

However, Boudon's theory can be completed and enhanced in several dimensions. First, the advantage of the classical Humean criteria for rationality is that they are clear and undisputable: it is experience, or empirical evidence, that provides the elements for a correct belief about a fact (namely facts about the proper means to reach an end). We do not have the corresponding evidence to interpret as rational the choice of the various and opposing norms that govern social life. Boudon's theory tends first to assert that some norms are better than others. Their advantage stems from particular reasoning, which refers to various principles. He then tries to demonstrate that the better norms tend to prevail and are widely shared by people. But he does not really display clear criteria to demonstrate that the choice of those norms is rational. The problem of the reference to good reasons is that in some sense everyone has good reasons to do what he does, but that does not necessarily provide a framework for a unified theory of behaviour. Boudon tries to reach this unified position, but it is not sure he always succeeds in doing that.

We should therefore reconsider the elements that can justify the idea of rationality. The strength of Boudon's theory's can be seen partly as a weakness. The strength is the pluralistic, supple, and open dimension of rationality, sensitive to the variation of individual and social situations. The consequence of this, however, is the risk of having little difference between reasons (which depend on various subjective and cultural motives) and good reasons (truly rational), and therefore no clear criterion for discriminating right decisions among competing ones.

To partly overcome this problem, I have suggested (Demeulenaere, 2003) that we should introduce a reflection that provides us with the general basis for defining something as rational. Something is said to be rational when it obeys a norm that drives the decision; a norm that is not subjective or cultural. For instance, following Weber (1978), to say that two plus two equals four corresponds to a *correct* choice that is neither subjective nor cultural. It corresponds to a norm of correctness that "forces us " to admit that two plus two equals four. If everything is subjective, or cultural, there is no room for the idea of rationality, and no idea of the correctness of this sort. Similarly, an adequate

belief regarding a fact is something, based on a norm of correctness (Putnam, 2002), which should not be seen, if it is to be rational, as a subjective belief or a cultural attitude. We are complied by the facts, in one manner or another, to adopt beliefs that are congruent with them (even if errors are common). Here again, any correct belief is dependent on a norm of correctness, which is not subjective or cultural. Naturally, it is possible to introduce the notion of subjective rationality the way Boudon (1989) has done. It allows us to interpret, the way Weber had previously conceived them, typical errors that are rational in the sense that given the information someone has, and given the norms of rationality, she could not achieve the exact correct solution to a problem. The one reached was, however, the best option in her situation. It has a subjective dimension, although based on non-subjective norms.

But how should we characterize these norms and where do they come from? What is the common element between intentionality, consistency of choices, choice of adequate means, and axiological rationality that entitle us to speak of rationality? I will not develop all this in this paper, but I will suggest two elements of reflection I have developed elsewhere (Demeulenaere, 2003):

- Rational norms derive from intrinsic features of action and reflection. Thinking and acting have their own norms. Consistency of choices, for instance, derives from the very notion of making a choice, or of ordering preferences. The idea of rationality derives from the features of action. Similarly, when we describe a fact, the very notion of description implies that the description should be correct, and therefore poses its conditions for correctness.

- Because of these intrinsic features of action and reflection, which are constitutive of human behaviour, rationality norms are not subjective nor cultural, although they can be variously defined and more or less enhanced in various social contexts. The scope of the idea of rationality is to define interpersonally valid norms that are not reducible to cultural various settings. Their existence allows us to interpret particular behaviours that are inevitably concerned by these norms.

- On the basis of those rationality norms, particular social norms, although normative, can correspond to rational solutions to problems. For instance, I cannot convince someone to accept a norm that is not in her interest, just by saying that it is in my own interest. It typically does not work from the point of view of the meaning of the process of justification of a social norm. The other person has to find an interest in the norm, or to find a more general reason to accept it, which cannot be only *my* own interest. This typically will not work (except if the person has a particular interest of her own in my interest). This is a point that had been made by Durkheim and Pareto. Therefore any social norm that is intended to be legitimate must inevitably take into account this constraint on any process of justification. So, any attempt to find legitimate social norms is constrained by this feature of any process of legitimization. The solution to this problem evolves given other social parameters. The rational choice is here to find solutions to this problem that are consistent with the general features of the process of legitimization.

Once they are conceived in particular social settings, the mechanisms by which these norms are diffused should be described more precisely. Boudon was inclined to put the emphasis on the correspondence between correct norms, or better norms (in his view) and a wide acceptance of those norms. Situations where unjust norms are widely accepted or where just norms are not diffused do naturally exist. For sure, he has many elements in his theory to explain such situations. But he did not really pay attention to the diffusion mechanisms that lead from one situation to another.

Neither has Boudon specifically considered the influence of social groups on the formation of individual beliefs. People do not form their beliefs alone: when, for instance, they believe that there are massive destruction weapons in Iraq, they believe such a thing because they are part of a group where dissent is difficult and where everyone's certitude is reinforced by that of the others. This is not incompatible with a theory of rational behaviour, but reasons are shared, and the fact that they are has a major influence on the formation of a singular belief. It is particularly true through educational processes.

Finally, the role of emotions in the formation of beliefs, that is, the fact that beliefs are sensitive to desire (Elster, 2000, 2010), should be more precisely described. For sure, this does not mean that there are no rational beliefs. When someone adopts a belief she generally has more than just a desire to accept it; there must be some cognitive elements that give credit to the belief. But it is important to note that rationality by itself is supported by emotions (Damasio, 1999), and that other emotions tend to interfere with its verdicts.

Bibliographic references

Boudon, Raymond (1974). *Education, opportunity and social inequality*. New York: Wiley.

— (1989). "Subjective Rationality and the Explanation of Social Behavior". *Rationality and Society*, (1) (2), 173-196. <http://dx.doi.org/ 10.1177/1043463189001002002>

— (1998). *Etudes sur les sociologues classiques*. Paris: Presses Universitaires de France.

— (2000). *Etudes sur les sociologues classiques*, *II*. Paris: Presses Universitaires de France. Quadrige.

— (2007). *Essais sur la théorie générale de la rationalité*. Paris: Presses Universitaires de France.

— (2009). *La rationalité*. Paris: Presses Universitaires de France. Collection Que sais-je?

— (2011). "Ordinary rationality: the core of analytical sociology". In: Demeulenaere, P. *Analytical Sociology and Social Mechanisms*. Cambridge: Cambridge University Press, 33-49.

Coleman, James S. (1990). *Foundations of Social Theory*. Cambridge and London: The Belknap Press of Harvard University Press.

Damasio, Antonio (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

Dray, William H. (1964, 1993). *Philosophy of History*. Upper Saddle River (NJ): Prentice Hall.
Demeulenaere, Pierre (1996). *Homo oeconomicus, Enquête sur la constitution d'un paradigme*. Paris: Presses Universitaires de France. Réédition Quadrige, 2003.
— (2003). *Les normes sociales. Entre accords et désaccords*. Paris: Presses Universitaires de France.
— (2011). "Introduction". In: Demeulenaere, Pierre (ed.). *Analytical Sociology and Social Mechanisms*. Cambridge: Cambridge University Press. 1-30.
Elster, Jon (2000). *Strong Feelings: Emotion, Addiction and Human Behavior*. MIT Press.
— (2009). *Le désintéressement. Traité critique de l'homme économique*, I. Paris: Seuil.
— (2010). *L'irrationalité. Traité critique de l'homme économique*, II. Paris: Seuil.
Homans, George C. (1987). "Behaviourism and after". In: Giddens, Anthony and Turner, Jonathan (eds.). *Social Theory Today*. California: Stanford University Press, 58-81.
Kahneman, Daniel (2011). *Thinking fast and slow*. New York: Farar, Straus and Giroux
Kahneman, Daniel and Tversky, Amos (eds.) (2000) *Choices, Values, and Frames*. New York: Cambridge University Press and Russell Sage Foundation.
Popper, Karl (1967). "The Rationality Principle". In: Miller, David (ed.) (1985). *Popper Selections*. Princeton (NJ): Princeton University Press.
Putnam, Hilary (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, Massachussets and London, England: Harvard University Press.
Sen, Amartya (1977). "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory". *Philosophy and Public Affairs*, 6 (4), 317-344. <http://dx.doi.org/ 10.2307/2264946>
Turner, Jonathan H. and Maryanski, Alexandra (2008). *On the origin of societies by natural selection*. London: Paradigm Publishers.
Weber, Max (1922, 1978). *Economy and Society*. Berkeley: University of California Press.

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2125

Reason-based explanations and analytical sociology. A rejoinder to Boudon*

Gianluca Manzo
GEMASS – CNRS / Université de Paris-Sorbonne
gianluca.manzo@cnrs.fr

Abstract

The paper discusses Raymond Boudon's theory of ordinary rationality as well as his assessment of the so-called "analytical sociology". On the first point, the paper argues that, in order to combine the realism of an unconstrained conception of rationality with the *ex-ante facto* predictive power of the narrow version of rational choice theory, we should better understand the relation between potentially triggering events and the actor's "reasons". Heuristics, social identity, and emotions are regarded as good candidates to advance in that direction. On the second point, the paper recalls some factual elements suggesting that Boudon's assessment of analytical sociology is excessively severe and explains why some analytical sociologists are right in devoting especial attention to computational modelling.

Keywords: rationality; heuristics; social identity; emotions; analytical sociology; agent-based simulation.

Resumen. *Explicaciones basadas en razones y sociología analítica. Una réplica a Boudon*

El trabajo discute la teoría de la racionalidad ordinaria de Raymond Boudon, así como su valoración de la denominada «sociología analítica». Respecto a la primera cuestión, el artículo argumenta que, para combinar el realismo de una concepción de la racionalidad sin constreñimientos con el poder predictivo *ex-ante facto* de la versión estándar de la teoría de la elección racional, debemos comprender mejor la relación entre eventos potencialmente desencadenantes y las «razones» del actor. Heurísticas, identidad social y emociones son vistas como buenas candidatas para avanzar en esa dirección. En lo que se refiere a la segunda cuestión, el artículo recuerda algunos elementos factuales, sugiriendo que la valoración de Boudon de la sociología analítica es excesivamente severa y explica por qué algunos sociólogos analíticos dedican especial atención a los modelos computacionales.

Palabras clave: racionalidad; heurísticas; identidad social; emociones; sociología analítica; simulación multiagente.

* This article was first published in *Revue européenne des sciences sociales* 2/2012 (50-2), p. 36-65. It was a rejoinder to a Raymond Boudon's paper also published in that issue. Re-printed with permission from Librairie Droz.

Summary

Introduction

One of the most important findings in the contemporary philosophy of social sciences is that the materially constrained pursuit of single-minded self-interests is far from being the only way in which human rationality, hence rational action, can be conceived (see Searle, 2001). In social sciences, this fact is well reflected by the plurality of forms that, despite its apparently unifying label, the theory of rational choice has assumed over the years (Goldthorpe, 1998). This heterogeneity expresses a progressive shift from the narrow to the broad conceptions of rational action. According to the latter, all kinds of desires and constraints are admitted, full information is no longer assumed, and actors' subjectivity in constraint perception is crucial (Opp, 1999). This move away from the narrow versions of rational action theory is partly due to the lack of empirical support for its core micro-level assumptions – where experimental psychology (Shafir and LeBoeuf, 2002) and behavioral economics (Camerer and Loewenstein, 2004) played the crucial role – and partly to its explanatory failures (Ostrom, 1998).

However, the dichotomous distinction between narrow and broad versions of rational action theories tends to hide an important fact: that broad versions are not all equally broad. In this regard, the most noticeable change over the last two decades has been the diffusion of an extreme variant of broad rational choice theory in which actors' rationality is equated to having subjectively well-founded "reasons". Rational action thus basically amounts to "reasoned action" (more than to "reasonable action": on the distinction between "rationality" and "reasonableness" see Beaney [2001]).

In economics, this point of view is endorsed by Sen (2009, p. 180), who expresses it as follows: "Rationality of choice, in this view, is primarily a matter of basing our choices – explicitly or by implication – on reasoning that we can reflectively *sustain* if we subject them to critical scrutiny". Reason sustainability is not only a matter of "self-scrutiny", Sen (2009, p. 196) adds, but also of "defensibility in reasoning with others" – here, Sen (2009, p. 44-46) refers back to Smith's metaphor of the "impartial spectator". Similar views can be found in political philosophy (see Rawls, 2003; Young, 2005; and Pizzorno, 2007, p. 123).

In sociology, over the last two decades or so, Raymond Boudon has been among the scholars who have taken this open conception of rationality as a set of well-grounded subjective reasons to its furthest extent (see Boudon 1989,

1993, 1996, 1998a, 2003, and 2011). His plea for a "reason with a small *r*" is based on the conviction that every other definition of rationality would be insufficient fully to account for the variety of ways in which human behaviour can be called "rational" (see Boudon, 1994, ch. 9). Similarly to Sen (2009, p. 191-193), Boudon's extreme extension of the concept of rationality is strictly related to his intention rationally to explain not only individuals' choices of means but also their choices of ends (see Boudon, 2001).

As I read it, Boudon's "'Analytical sociology' and the explanation of beliefs" (see, in this issue, p. 7-34) is the most remarkable synthesis of his theory of rational action. To the best of my knowledge, this paper also contains his first explicit assessment of an approach to sociological theory that, over the last ten years or so, has received considerable attention around the world, *i.e.* so-called "analytical sociology".

I have two purposes in what follows. First, I intend to discuss what seems to me the crucial difficulty with a conception of rational action that equates "rationality" with "reasons". On this view, indeed, one can no longer benefit from the "predictive device" (Sen, 2009, p. 175, p. 183) contained in the narrower conception that frames rationality in terms of instrumental rationality. The line of reasoning that I shall seek to defend is that, in order to combine the realism of an unconstrained conception of rationality with the deductive power of the narrow version of rational choice theory, a possible solution is to look for regularities in the relation between potentially triggering events and the actor's "reasons". In particular, I shall consider heuristics, social identity, and emotions as explanatory factors which should be more systematically analysed to find such regularities. I regard this line of reasoning as one possible starting point, but not as the full answer, which would be beyond the scope of this paper. My second goal is even more modest. I intend only to recall some factual elements suggesting that Boudon's assessment of analytical sociology is excessively severe. In particular, I shall explain why computational modeling, to which analytical sociologists devote especial attention, is a central resource with which to test theoretical explanations referring to complex systems of individuals' reasons.

The paper is organized as follows. First, I discuss the main difficulty with an unconstrained conception of rationality *à la* Boudon; second, I analyze this conception of rational action in the light of important contributions on "heuristics", on "social identity, and on "emotions" in psychology, in economics, and in political science. Finally, I discuss Boudon's assessment of analytical sociology.

1. What We Can (not) Do with Boudon's Theory of Ordinary Rationality

"'Analytical sociology' and the explanation of beliefs" (see p. 7-34, in this issue) contains a clear exposition of what Boudon now calls a "theory of ordinary rationality" (Boudon [1989] first adopted the label of "subjective rationality", and then that of "cognitive rationality": see, for instance, Boudon, 1996). The theory relies on the "cognitive equilibrium principle" according to which one

must examine actors' reasons to find the proximate causes of their choices – in Boudon's words, "people believe that X is true, acceptable, good, legitimate, etc. as soon as they have the feeling that X rests upon a set of acceptable reasons" ("'Analytical sociology'...", p. 18 – note the similarity between this statement and Sen's definition of rational action quoted in the introduction).

Thus, the distinctiveness of Boudon's theory of rational action is that actors' reasons are completely unrestricted in the sense that no specific class of reasons is given explanatory priority. The rhetoric of cost-benefit calculation disappears behind the variety of arguments that actors can endorse and publicly defend to sustain the beliefs that motivate them to act. In the paper under discussion, Boudon demonstrates that this open conception of actors' rationality makes it possible to explain much more social regularity than allowed by the narrow version of the rational choice theory. In particular, he argues, by extending the set of acceptable reasons, it is possible to explain not only the choice of means but also the choice of ends. In this way, sociology can explain the genesis of complex sets of human values like individuals' feelings of justice (on this point, see, in particular, Boudon and Betton, 1999).

This achievement comes with a cost, however. To restrict actors' motivations to a specific class of reasons – instrumental reasons, in the case of the narrow variants of the rational choice theory – makes it possible to form expectations on the micro-behaviours and their macroscopic consequences that are most likely to appear, given a certain set of constraints. A given empirical observation can thus be compared with a clear benchmark that is formulated before the observation is made. As acknowledged by Sen (2009, p. 175 and p. 183), who, as we have seen, endorses a very open conception of rational action, the capacity to figure out a single outcome *ex-ante facto* is lost when actors' rationality is given a completely unconstrained form. On this view, the expression "rational action theory" itself is inappropriate. Strictly speaking, indeed, there is no theory, but rather a single framework in which every sort of reason-based explanation – *i.e.* an explanation focusing on the system of arguments that a given set of actors endorse to act in the way that they act – can be conceived.

The loss of the "predictive device", to use Sen's expression, affecting the conception of human rationality as a complex set of subjectively well-grounded reasons should be carefully distinguished from two related but analytically distinct objections usually brought against the broad versions of rational choice theory: namely their lack of deductive power, and the danger of *adhocness* to which they are exposed.

As correctly pointed out by Boudon (1998b, p. 195) himself, the deductive power of a theory involving individuals' reasons does not depend "on the nature of reasons mobilized in a model". Given a set of postulated "reasons", in fact, it is possible to form expectations on what behaviour is likely to appear. As no explanatory primacy is given to a specific set of "reasons", however, the deductive power only exists once the set of reasons have been postulated. But how could one achieve such a system of reasons? This question leads directly to the *adhocness* objection that an unrestricted conception of human rational-

ity increases the probability that the set of hypothesized reasons will continue to be re-adjusted until it is possible to demonstrate that a given macroscopic regularity in fact derives from a population of actors acting in a rational way (Pizzorno, 2007, p. 65-66). As discussed at length by Opp (1999), however, this danger can be powerfully counteracted by the use of empirical data to assess the extent to which the hypothesized set of reasons is tenable. More specifically, the empirical solution says that, given a certain individual behaviour and a set of macroscopic consequences, once the set of individual reasons leading to this behaviour has been formulated, one must test whether the reasons postulated are empirically tenable. If not, one should revise the protocol of data collection and / or modify the set of hypothesized reasons.

The "empirical argument" thus amounts to an iterative procedure of *post hoc* theory testing which tends to be case-oriented. It only suggests that theories built on a very open conception of rational action can be falsified, but it does not help to remedy the problem of the reduced predictive power of these theories. In other words, Opp's empirical approach does not solve the problem of the variety of predictions than can be made on the basis of a conception of rational action that equates "rationality" and "reasonableness" – a problem that Boudon himself has acknowledged elsewhere (1998b, p. 195)[1].

Would it be possible to go beyond the empirical approach? In particular, is there any way to combine an open conception of rationality with the predictive power of the narrow versions of the rational choice theory? My answer is that we may eventually achieve this result by identifying some mechanisms that tend systematically to trigger certain set of reasons. In this case, we would have access to a set of regularities that enable us to formulate *ex-ante facto* clear expectations on what micro- and macro-level outcomes are more likely to be observed[2].

2. Three Classes Of "Reason Triggers"

I propose to define a "reason trigger" as a structural or individual-level factor that increases the probability that a specific set of reasons will arise in the

1. Boudon's reply is that "the reconstruction of reasons is a *theory* and that, as soon as the elements of a theory are acceptable, because they consist of empirical data and acceptable laws, the challenge is to find another theory that would be better in some respect […] If you do not like the latter, the challenge is this: find a better theory. We can never prove that a theory is true, only that one theory is better than another" (personal communication from Boudon's after he had read a first draft of the present paper). But this does not help solve the problem. Like Opp's empirical argument, this one only suggests that reason-based theories can be falsified. It does not address the issue of their capacity to point out *ex-ante facto* what system of reasons is likely to appear under certain circumstances, which is what one needs to produce *ex-ante facto* testable propositions.
2. The "empirical approach" can, of course, contribute to this result. In the long run, indeed, if it is applied systematically, one of its by-products may be the discovery of regularities in the connection between certain incentive structures and specific sets of actors' reasons. These regularities may then be incorporated into reason-based theories *à la* Boudon, thereby increasing their predictive capacity.

actor's mind. I focus here on three broad class of reason triggers that may help in finding regularity in the individuals' belief formation process, namely (i) "cognitive and social heuristics"; (ii) human invariants; (iii) "emotions". For each of these factors, I will also suggest that, while Boudon has never explicitly analyzed them in depth, none of them is in contradiction with his analysis of how reasons form in actors' minds.

2.1. Reason-based explanations and "heuristics"

Building on Simon's concepts of "bounded rationality" and of "heuristic search" (see, respectively, Simon, 1979; and Newell *et. al.* 1958), a large body of psychological literature has developed on "heuristics", *i.e.* cognitive shortcuts adopted by actors when they have to decide and solve problems (see Goldstein, 2009). In economics, this concept was used by Kahneman and Tversky to demonstrate that actors' reasoning is variously "biased" and that it does not conform with the way in which the narrow version of rational choice theory frames human decision-making (for an overview see Kahneman, 2003).

Readers who are familiar with Boudon's work may be surprised by my reference to this literature. In effect, Boudon has repeatedly criticized Kahneman and Tversky for treating "cognitive biases" as black boxes (see, for instance, Boudon, 1998b, p.180; 2004, p.186). Whilst one may agree that many of the "frames", "scripts", or "biases" mobilized in cognitive psychology and in economics are nothing more than labels – which is sometimes admitted even in the behavioral economics (see, for instance, Frederick *et al.*, 2002, on the concept of "inter-temporal discount rate") – the so-called "fast-and-frugal heuristic" research program (Gigerenzer, 2008) studies "heuristics" in a far less black-box fashion and explicitly aims to go beyond the concept of "cognitive biases" as conceived by Kahneman and Tversky.

Whilst, to the best of my knowledge, Boudon and Gigerenzer do not cite each other, their conceptions of human rationality are in fact strikingly similar. First of all, Gigerenzer (2008, p. 7) makes exactly the same criticism as Boudon of Kahneman's "heuristic-and-biases program". He maintains that "heuristics" are not directly and explicitly modeled in this research tradition, so that they end up with "mere verbal labels". Moreover, like Boudon, Gigerenzer argues that Kahneman and Tversky in fact adhere to the standard rational choice approach because they still assume that actors' rationality should comply with logic- and probability-based rules. According to Gigerenzer, it is instead the structure of the information contained in the environment in which actors are embedded that generates what is rational for them. He calls this conception of rationality "ecological rationality" (in economics, see Smith, 2008, p. 36, p. 151 and p. 168).

This largely overlaps with Boudon's frame of rational action as "reason-grounded action". In the paper under discussion, for instance, Boudon states that, in order to understand actors' reasons, "the observer should be aware of the relevant features of the social and cognitive context in which the individual

is embedded." ("'Analytical sociology'...", p. 16). Thus, according to him, magical rituals should be interpreted as perfectly rational from the point of view of those individuals who are ignorant of biological, chemical and physical mechanisms that are instead known by many contemporary observers. In the two cases, the structure of the information is simply different.

Given the strong similarity between Boudon's "ordinary rationality" and Gigerenzer's "ecological rationality", what should we expect to gain from a more systematic integration between the two theoretical perspectives? My argument is that the experimental evidence accumulated within the "fast-and-frugal heuristic" (for an overview see Gigerenzer *et al.*, 2011) provides useful empirical material with which to find regularities between the environment in which actors are embedded and the system of reasons that they tend to develop. We may discover, for instance, that, when information is highly skewed so that certain outcomes are rare, actors have good reasons to believe that they perform (or can avoid risk) better than they do in reality. Systematic links of this kind among given informational structures, beliefs, and certain mental shortcuts may help in building reason-based explanations that do not rely on *a posteriori* operations of reason reconstruction, thereby helping us to specify in advance the micro-level behaviour and its macroscopic consequences that are most likely to be observed.

2.2. Reason-based explanations and "social identity"

The second class of reason triggers that might be fruitfully incorporated into reason-based explanations are "human invariants", *i.e.*, as I conceive them here, behavioral, cognitive or emotional patterns common to all human beings regardless of the culture in which they live.

At first sight, similarly to the concepts of "heuristics" and biases", that of "human invariant" may seem incompatible with Boudon's theoretical framework. In effect, the French sociologist has never concealed his distaste for "dispositional variables", which, in the paper under discussion, he defines as "conjectural causes operating in the backs of [the actors'] mind" ("'Analytical sociology'...", p. 17). On the other hand, however, in his final comment on Durkheim's analysis of magical rituals, he judges Durkheim's explanation convincing because it introduces "either empirical statements [...] or psychological uncontroversial laws, as 'in general people want to survive'" (*ibid.*, p. 22). Boudon is thus implicitly admitting that "dispositional variables" are legitimate explanatory factors as long as we can demonstrate that these dispositions correspond to well-defined basic individual needs or desires (see also Elster, 2011, p. 61).

This is an extremely important point because it opens rational action theory in sociology to research in anthropology (see, for instance, Brown, 1991, 1999, 2004), in evolutionary psychology (see Pinker, 2002) and, partly, in behavioral economics (see, for instance, Henrich *et al.*, 2001; Gächter and Herrmann, 2009) intended to establish the existence of individual invariants

empirically. This literature can be of benefit to sociologists defending a conception of human rationality which equates rationality and "reason-grounded action" in that it can provide insights on universal psychological motivations that tend regularly to activate specific set of reasons.

Among basic individual needs, an individual's desire to possess and to defend a well-defined social identity seems especially important for sociological analysis. In social psychology, it is commonplace to consider the need for social belonging as a fundamental motivation (see, for instance, Fiske, 2011, p. 116). Pizzorno's and Akerlof's conceptualizations of social identity are two interesting starting points from which to assess the fruitfulness of a closer integration between reason-based explanation *à la* Boudon and identity-based explanations. In effect, Pizzorno overtly builds his theoretical proposal against the theory of rational choice and methodological individualism, whereas Akerlof aims to extend the standard model of the actor in mainstream economics[3].

Pizzorno's (1986, p. 366-372) thesis is that social action can be framed in terms of rational action provided we accept that rationality needs identity. He explains the source of this link as follows. When an actor is choosing between, say, X and Y, at time t, according to rational action theory, he should be able to evaluate the expected benefits of the two alternatives at time t+n. However, Pizzorno argues, in order for this evaluation to be possible, the actor's identity should be stable over time. If not, the expected future benefits of X and of Y cannot be really evaluated and compared because the actor does not know what his point of view on X and on Y will be at time t+n. Thus, without inter-temporal identity stability, he argues, rational action is impossible. Pizzorno's proposal is that social recognition, hence the social circles that provide it, should be considered as identity "stabilizers". That is why the actor's logic, according to Pizzorno, is driven more by the quest for sources of social recognition than by self-interest (see Aguiar and de Francisco [2002] for a criticism of Pizzorno's argument).

Pizzorno's and Boudon's analytical framework are less distant that might seem. In the paper under discussion, Boudon states that actors can believe that the reasons that they endorse to act in the way that they act are strong only if they also believe that these reasons can in principle be shared by other actors. This is the concept of reason trans-subjectivity (see "'Analytical sociology'...",

3. Within the micro-foundationist tradition, Little (1998, ch. 6) has argued that there is no incompatibility in principle between identity- and reason-based theories. Broad rational-choice theorists have explicitly attempted to incorporate identity into the rational-choice framework. Some have done so by introducing the idea that actors are animated by a multiplicity of selves, the objective of this hypothesis being to account for some violations of the predictions that would ensue from the standard rational-theory, like the so-called weakness of will phenomenon (see, for instance, Elster, 1985, and Coleman, 1990, ch. 19). Others have tried to incorporate identity-based preferences into rational choice theory by conceiving "social identity" as a complex set of beliefs about oneself and about the group to which one thinks/wants to belong (see Aguiar and Francisco, 2009). Here my point of view is different, in that, on the one hand, I am not mobilizing social identity to remedy some explanatory failures of the narrow version of the rational choice theory, and, on the other hand, I consider the desire-component of social identity more than its belief-component.

p. 18). Pizzorno's idea of "recognition circles" implies a similar notion: the actor needs others with similar views in order to be reassured about his identity, which is what makes belief and reason formation possible. In both cases, it is postulated that actors need to search for potential sources of (more or less) local social consensus.

A similar concern for social conformism is behind Akerlof's endeavour to devise a new approach in economics called "identity economics" (see Akerlof and Kranton, 2001). The basic assumption here is that beliefs about oneself (identity), beliefs about expected behaviors (normative beliefs, in Boudon's terminology) as a function of this identity, and choices are closely interrelated. On this basis, Akerlof and Kranton build an analytical framework that can be summarized as follows: a) actors belong to social categories; b) social categories convey beliefs about the self (actors' identity); c) social categories are also associated with norms about the prevailing behavior within the category; d) actors' utility increases/decreases if they conform with / violate these norms (*ibid.*, p. 14 and chap. 3). Conformism assures social belonging (*ibid.*, p. 22).

Despite the different pathways that Pizzorno and Akerlof follow to plead for an integration of identity concerns into rational action theory, both of them establishes a link among social belonging, social identity and actors' beliefs. Again, my argument is that this link can help in building reason-based explanations that lead to fine-grained predictions *ex-ante facto*. For instance, when actors are deeply concerned to secure their social identity, one may expect that, in order to reinforce their participation in a social group, they will be more likely to accept material and psychological costs so that collective action (see Willer, 2009) or even extreme choices like terrorist attacks (see Tosini, 2011) become possible.

When such concerns for social identity and status drive actor's behaviour, we can also expect them to express different convictions in different social circles, to change their minds as the networks to which they belong change, or publicly to express opinions that they do not endorse in private – the so-called "unpopular norms" (see Bicchieri, 2006, ch. 5; Centola, Willer, and Macy, 2005; Willer, Kuwabara, and Macy, 2009). Within this analytical framework that conceives the need for social identity as a fundamental human motivation, preference inconsistency can thus be anticipated and explained without introducing dubious hypotheses like that of a fragmentation of the actor's self.

2.3. Reason-based explanations and "emotions"

Emotions are the last reason trigger that I shall briefly discuss as a basic micro-level element that might help to increase the predictive power of theories based on a conception of rational action as "reason-grounded action"[4].

4. Stets and Turner (2006) provide a thorough overview of the variety of theoretical perspectives on emotions in sociology. By contrast, the empirical description of emotions is only in its early stages in sociology. As remarked by Golder and Macy (2011), large-scale web-based data might help improve this situation.

In the paper under discussion, Boudon does not address the place of emotions in the explanation of social action. Elsewhere, however, he explicitly admits that actors' emotions and actors' reasons are closely related, with the causality going in both directions, i.e. from actors' beliefs to actors' emotions and the reverse (see, for instance, Boudon, 2003b, p. 150-151).

This is important because, as pointed out by Frijda *et al.* (2000), much research has been conducted to understand how emotions are triggered by the way in which actors perceive external events (see, for instance, Scherer, 2011). By contrast, "oddly", say Frijda *et al.* (2000, p. 1), the causal direction going from actors' emotions to actors' beliefs "has received scant attention". When this is the case, they remark, "the emphasis has been on the assumption that the former [emotions] distort the latter [beliefs]" (*ibid.*, p. 2).

Elster is a good example of this asymmetry within the philosophy of social science and sociology. He has focused closely on the effect that actors' cognitions (or perceptions) have on actors' emotions (for a recent overview see Elster, 2011). In particular, he has carefully studied, for each emotion, the action tendency that this emotion is likely to trigger once it has been activated by a given belief. As regards the causal link from emotion to beliefs, however, Elster's (2009) analysis only focuses on the negative consequences that emotions can have on beliefs by triggering under-investment in information (urgency) or under-estimation of the long-term consequences of action (impatience). The results are, respectively, low-quality and biased beliefs.

To deepen our understanding of emotions as "reason triggers", the "positive" role that emotion plays in belief formation should be analyzed more systematically. That actors' emotions do not necessarily distort cognitions is stressed by Scherer, for instance, who suggests that emotions can be rational in the sense that they can help actors to reach their goals (functional rationality), to make correct inferences (intellectual rationality), and to be accepted by others as persons that react in the right way (reasonable or consensual rationality). An extreme empirical example of the not-necessarily-biasing effect of actors' emotions on their beliefs is that of depressive people, who tend to assess the reality more realistically than optimistic persons (see Scherer, 2011, p. 340). Similarly, it is often observed that fear or anxiety may induce actors to invest in information search in order to clarify their perception of what political parties offer, thereby acquiring beliefs more accurate than would be possible in the absence of such emotions (see, Jaspers, 2011)[5].

Among the variety of emotions that can act as "reason triggers", I regard interaction-comparison-based emotions as especially important. These are emotions, like envy, jealousy, indignation, humiliation, shame or resentment,

5. The emotion-to-cognition and the cognition-to-emotion patterns can co-exist. The simplest example is a dissonance-reduction-based mechanism of belief change where two (or more) beliefs that are discrepant generate negative feelings which induce the actor to change one (or more) of his initial beliefs (see, for instance, Harmon-Jones, 2000).

that tend to be triggered by the comparisons that actors make within the dyadic interactions in which they are embedded. Social networks are a fundamental part of social life, and actors perform every sort of social comparison within these networks. Interaction-comparison-based emotions are thus likely to be ubiquitous, and they may lie at the origin of many of the beliefs that populate actors' minds[6].

Recent research in social psychology suggests that this is a fruitful idea. Fiske (2011) builds on an impressive amount of empirical studies to demonstrate that social comparisons tend to generate specific emotions that in turn tend to generate specific beliefs. In particular, she shows that two basic patterns are especially frequent. On the one hand, downward comparisons tend to trigger disgust and scorn, emotions that tend to induce actors to believe that people below them are less warm, less familiar, less competent, less articulate, less intelligent, and, in short, less typically human. On the other hand, upward comparisons tend to trigger envy and resentment, emotions that tend to induce actors to believe that people above them forgo their humanity to get ahead, that they are cold and calculating, even though they are competent. Coldness and competence tend to trigger the belief that wealthy people are engaged in a conspiracy and, ultimately, a threat to "us".

This is precisely the kind of evidence that we need to increase the predictive capacity of an approach equating rational action with "reason-grounded action". If we know that upward social comparison, for instance, tends to trigger specific emotions (like envy) with specific objects (like wealthy people), then we may expect to find that specific sets of reasons are also triggered in actors' minds. Members of lower and middle social groups may be more likely to think that the members of upper groups do not deserve what they have and / or that they have obtained what they have by unfair or corrupt means. As a consequence, they may also be more likely to believe that the economic organization that supports those groups should be changed. A reason-based theory incorporating such emotion-belief linkages might thus be better able to predict *ex-ante facto* single outcomes at macroscopic level, like waves of anti- or pro-capitalist attitudes (see Jaspers 2011, for a plea for emotions to be included in the analysis of social movements).

2.4. The argument in a nutshell and Boudon's expected reply

The argument that I have outlined so far is that an extremely open rational action theory conceiving actors' rationality as a bundle of subjectively well-founded reasons can profit from stronger connections with three research areas:

6. Elster (1999, p. 141-142; 2007, p. 58) distinguishes between comparison-based emotions, like envy, and interaction-based emotions like resentment. In order to draw attention to the fact that these emotions are often a by-product of the social comparisons driven by dyadic links between actors, I propose the hybrid conceptual category of "comparison- interaction-based emotions".

1) the "new" literature in cognitive psychology about mental and social shortcuts, *i.e.* "heuristics"; 2) the literature in anthropology and in evolutionary psychology about basic and culturally invariant psychological needs; 3) the literature in cognitive psychology on emotions.

When "heuristics", "social identity", and "emotions" are considered as "reason triggers", the benefit that one may expect from integrating them into a reason-based theoretical perspective in which actors' rationality is unconstrained concerns the increase in the predictive capacity of this perspective. The more we know about the regular linkages between the structures of information in which the actors are embedded and their beliefs, among actors' networks, their social identities and their beliefs, and between actors' emotions and actors' beliefs, the more, it seems to me, we should be able to figure out *a priori* the micro- and macro-level outcomes observable under given social circumstances. In this way, we may eventually combine the realism of an open conception of rational action with the "predictive device", to use Sen's expression, contained in the narrower versions of the rational choice theory.

Boudon's reaction to the line of reasoning proposed would probably be that integrating cognitive-, identity-, and emotion-based mechanisms into a subjective conception of human rationality would expose sociologists to the risk of paying attention to unnecessary psychological details. Over the years, indeed, he has consistently defended the thesis that sociology should be based on a "conventional psychology", that is to say, a highly abstract depiction of how actors think and feel (see, for instance, Boudon, 2003c, p. 169-170; 2007, p. 44, footnote 1). Within the micro-foundationist tradition, among others, Coleman (1990, ch. 1) and Goldthorpe (1998, p. 181-182) have also made a strong case against the introduction of an elaborate individual psychology into an appropriate rational action theory for sociology. Their argument is that we do not need to go into much detail about actors' psychology because the explanatory focus of sociology is the macroscopic consequences of individuals' actions. The implicit assumption behind this argument is that, where large populations of actors are concerned, psychological differences across actors cancel each other out, so that we are entitled to focus only on ideal-typical actors.

In my view, the analysis of "heuristics", "social identity", and "emotions" as "reason triggers" does not necessarily lead to a psychological-based theory of social action. My proposal certainly presupposes a stronger interaction with social and cognitive psychology; but what one should look for in this literature is a set of regularities between specific individual-level factors and the genesis of certain set of reasons, rather than details about actors' personalities and idiosyncrasies. That said, on a methodological level, we today have access to techniques that enable study of the macroscopic consequences of models of actors as complex as we want, so that we have fewer good reasons than in the past for omitting too many details at the micro-level. This is a point that I address in the next section.

3. Boudon and Analytical Sociology

As announced in the introduction, apart from arguing for a general theory of rational action, "'Analytical sociology' and the explanation of beliefs" contains Boudon's first explicit assessment of the so-called "analytical sociology". Here it is:

> I have the impression, though, that the handbooks on "analytical sociology" insist on secondary technical details and fail to identify clearly the common paradigm that underlies many illuminating sociological works, *i.e.* the paradigm that I have tried to identify as grounded on three principles: *methodological singularism*, *methodological individualism* and the *cognitive equilibrium principle* ("'Analytical sociology'...", p. 31).

Analytical sociology is a complex intellectual movement with respect to both its distant and proximate historical roots and its position within contemporary sociology. In other words, the diachronic and the synchronic frontiers of analytical sociology are still not well-defined (see Manzo, 2010). Since Boudon explicitly limits his objection to the "handbooks of analytical sociology", I shall restrict myself to programmatic books on analytical sociology in assessing this objection (namely, Hedström, 2005, and Hedström and Bearman, 2009a).

To this end it is important to give more precise definition to what analytical sociology is. Hedström and Bearman's (2009b, p. 16) propose the following:

> The explanatory strategy can be described as follows (see also Epstein, 2006): 1. We start with a clearly delineated social fact that is to be explained; 2. We formulate different hypotheses about relevant micro-level mechanisms; 3. We translate the theoretical hypotheses into computational models; 4. We simulate the models to derive the type of social facts that each micro-level mechanism brings about; 5. We compare the social facts generated by each model with the actually observed outcomes.

Although restrictive – more qualitative-oriented, yet analytically rigorous, scholars would consider steps 3 and 4 unnecessary (see, for instance, Elster, 2007, p. 455) – this definition is useful for discussing the two main points addressed by Boudon's critical assessment of analytical sociology: the excessive importance attributed by analytical sociologists to techniques, and their myopia with respect to the most important principles that animate all scientific sociological works.

3.1. Are analytical sociologists really myopic?

Let me start with the second point. Although Hedström and Bearman do not use the term "methodological singularism", it seems to me that their step 1 clearly follows this principle. In effect, Hedström and Bearman's advice here is to focus on *explananda* whose temporal and spatial contours are clearly specified.

Boudon's second principle, *i.e.* "methodological individualism", seems to me outlined in Hedström and Bearman's step 2, *i.e.* "we formulate different hypotheses about relevant micro-level mechanisms." Even more explicitly, they claim: "[...] all social facts, their structure and change, are in principle explicable in terms of individuals, their properties, actions, and relations to one another" (Hedström and Bearman, 2009b, p. 8). Moreover, similarly to Boudon, Hedström and Bearman (*idem*) also attempt to avoid the reductionist interpretation of this principle by remarking that "As we define the term, structural individualism is a methodological doctrine [that] differs from traditional notions of methodological individualism (*e.g.* Elster, 1982) by emphasizing the explanatory importance of relations and relational structures".

Boudon's last principle, *i.e.* the "cognitive equilibrium principle", which basically states that human actions must be conceived as reason- based, also seems to be at the core of analytical sociology. In its manifesto, Hedström (2005, p. 38-39) posits: "the desires, beliefs and opportunities of an actor are here seen as the proximate causes of the actor's action and, he continues, "beliefs and desires are mental events that can be said to cause an action in the sense of providing reasons for the action". As I stressed earlier, Boudon's fundamental contribution is to demonstrate that there is no compelling justification for restricting actors' reasons to instrumental ones. Contrary to what some critics of analytical sociology maintain (see Gross, 2009), a similar open conception of rationality as "reason-based actions" is also at the heart of analytical sociology – "DBO theory makes no assumption that actors act rationally, however; it only assumes that they act *reasonably* and with intention", states Hedström (2005, p. 61, emphasis added).

Thus, the analytical sociology manifestos suggest that the three methodological principles which Boudon recognizes in every scientific sociological analysis do not animate this perspective "implicitly", as he claims. On the contrary, they are explicitly, consciously and programmatically put at the core of the approach – could it be otherwise, one may ask, given that Boudon is one of the main inspiring intellectual sources of the analytical sociology movement (see Hedström, 2005, p. 6-9)[7]?

I would expect Boudon's counter-objection to be that, if analytical sociology consists explicitly in the three-rule paradigm that he has identified, then he is fully entitled to claim that "although 'analytical sociology' is a new expression, it is actually old wine in new bottles, since it essentially revitalizes the principles more or less implicitly used by classical sociologists, notably by Weber and Durkheim" ("'Analytical sociology'…", p. 19).

7. In this respect, the following coincidence is also significant. In the paper under discussion, Boudon cites a recent article by Pawson who, according to Boudon, has "convincingly shown that the paradigm described by those three principles disentangle the meaning of "middle range theory" ("'Analytical sociology'...", p. 29). The second chapter of *The Oxford Handbook of Analytical Sociology* argues that "the theories found in this book are contemporary incarnations of Robert K. Merton's notion of middle-range theory" (see Hedström and Udhen, 2009, p. 25).

That mechanism-based theorizing is at the core of classical sociology is a well-established historical fact (see, for instance, Cherkaoui, 2005, ch. 1 and 4). It is clearly evident to, and explicitly acknowledged by, analytical sociologists (see Hedström, 2005, p. 6; Hedström and Edling, 2009). It is also clear that contemporary analytical sociology can be traced back to research in mathematical sociology, in sociological theory, and in philosophy of social sciences in the 1960s and 1970s (see Manzo, 2010). Moreover, it would also be easy to demonstrate that specific pieces of analytical sociology are at the heart of several strands of the theoretical and empirical literature in contemporary sociology. But does this suffice to deny the novelty of analytical sociology? I have argued elsewhere that the novelty of analytical sociology consists in its integration of epistemological, theoretical, and methodological proposals that only exist separately in the rest of the discipline (see Manzo, 2011).

An example of this federative power of analytical sociology is the theory of action that it tries to set up. Its most distinctive feature is its attempt to endogenize the proximate causes of individuals' action, *i.e.* desires, beliefs, and opportunities, by taking social interactions into account (see Hedström, 2005, p. 42-59). Theoretically, this has a notable consequence: social interactions open the theory of action to ego- and alter- centered mechanisms that are usually not linked to reason-based explanations within the micro-foundationist tradition.

Social interactions may induce actors to imitate each other, imitation being a heuristic-based mechanism responsible for belief and desire changes (see Gigerenzer, 2008, p. 31). The theory of action is thus open to the literature on heuristics that I mentioned earlier (see Goldstein, 2009). Social interactions also trigger social comparisons which tend in turn to activate specific emotions. The theory of action is thus open to the literature on emotions. We saw earlier that individual and social identity is strongly related to comparison processes that take place within dyadic interactions. Social interactions thus indirectly open the analysis of belief and desire formation and change to theoretical models of action that stress identity and social recognition more than individual reasons, such as Pizzorno's framework or the recent developments in economics to which I referred earlier. Finally, taking interactions seriously into account to explain the genesis of beliefs, desires and opportunities makes it possible to establish theoretical and methodological bridges between action and network theory – something that, in the paper under discussion here, Boudon himself considers a desirable development.

One may retort that this ambition of analytical sociology to integrate different strands of the literature in order to develop a more realistic theory of action is excessive because sociology does not have methodological tools with which to study the macroscopic consequences of complex sets of micro-level mechanisms. As steps 3 and 4 of Hedström and Bearman's research strategy show, many analytical sociologists think that simulation is a promising solution.

3.2. Do analytical sociologists have good reasons to be technique-addicted?

This brings me to the second component of Boudon's caustic comment on analytical sociology, *i.e.* "that the handbooks on 'analytical sociology' insist on secondary technical details". My final remarks aim to explain why a specific type of simulation method, namely agent-based modeling, can in fact be regarded as a crucial resource with which to move sociology towards being a deeper and more rigorous discipline (I note in passing that Boudon was a strong advocate of formal modeling and simulation in the early stages of his intellectual career: see, for instance, Boudon, 1965, 1979).

What are agent-based models? The British computer scientist Michael Wooldridge (2009, p. 5) defines an agent as "a computer system that is capable of independent action on behalf of its user or owner". A single agent is thus nothing more than a computational entity. "A multiagent system", Wooldridge continues, "is one that consists of a number of agents, which interact with one another, typically by exchanging messages through some computer network infrastructure".

This class of formal models is so important for social sciences because it is infinitely flexible. Virtually any substantive mechanisms can be represented and studied within the framework of agent-based modelling. At the deepest level, this flexibility relies on the specific type of programming language adopted to build this model, namely the so-called "object-oriented programming" which allows specification of each computational entity as a set of attributes and rules and their arrangement into different relational topologies and across several levels of organization (see Hummon and Fararo, 1995).

This flexibility is especially attractive for sociologists for the following reasons. First, agent-based modeling can represent entities and have them interact at any level of analysis. A computational agent need not necessarily represent an individual. Whatever entity we wish to represent can be programmed. According to the attributes and the activities associated with the entities, agents can represent cells, atoms, molecules, individuals, organizations, groups, nations, and so forth.

Second, agent-based modeling makes it possible to introduce as much heterogeneity as believed necessary for the problem at hand. Agents can be heterogeneous in terms of attributes and/or in terms of the values they get on these attributes. More radically, agents can be heterogeneous in terms of activities, tasks, or the behavior rules by which they are driven. This is a fundamental point. As Gallegati and Kirman (1999) pointed out in their critique of mainstream economics, agent-based modeling constitutes a robust formal tool that indeed enables us to go beyond the metaphor of the "representative agent". In the paper under discussion, Boudon constantly refers to ideal-typical actors. In reality, actors are heterogeneous in terms of beliefs and desires, and heterogeneity matters in explaining macroscopic outcomes because heterogeneity spreads across social networks. Agent-based modeling allows us to represent heterogeneity and study its macroscopic effects.

Third, agent-based modeling is entirely agnostic about the logic of action by which agents are driven. This means that we are no longer obliged to represent actors who maximize or optimize some quantity; nor are we obliged to suppose that actors possess the very large amount of information needed to compute the future consequences of alternative choices. With respect to the theory of action, agent-based modeling is the domain of heuristics. Whatever mental or social shortcuts are assumed to be at work in the real world, we can design and study them by means of an agent-based computational model. Since the model is solved by simulation, that is to say by iterating the constitutive rules of the model several times, mathematical tractability is no longer a constraint for the kind of actors that we want to represent. This is why some have argued that agent-based modeling is the right mathematics for social sciences (see Borrill and Tesfatsion, 2010).

Finally, agent-based modeling is all about social interactions. Every network structure that we are able to imagine can be designed, and agents can be embedded within it so that their beliefs, desires, and opportunity can be seen as locally constrained and influenced by the choices of other agents and by the network's topology.

For these reasons, agent-based modeling cannot be considered a "secondary technical detail". This class of formal models makes it possible to design theoretical models as complex as we need them to be with respect to both action logic and structural / relational constraints affecting social actions. If we are really interested in the macroscopic consequences of a given set of hypothesized micro- and interaction-based mechanisms, agent-based modeling is the most powerful method available today for the rigorous study of every substantive problem in which aggregation matters. In the late 1980s, Coleman (1986, p. 1316) complained about the existence of "extraordinarily elaborated methods for analysis of the behavior of a set of independent entities (most often individuals), with little development of methods for characterizing systemic action resulting from the interdependent actions of members of the system". This lack of methods with which to study the micro-macro transition is in principle solved by the use of agent-based modeling.

The flexibility of this class of models may also have important consequences in resolving a fundamental difficulty with the analysis of social mechanisms. As remarked by Elster (2011), one constantly has to deal with the problem of indeterminacy: on the one hand, the indeterminacy of the conditions that trigger a given (set of) mechanism(s); on the other, the indeterminacy of the resulting effect of mechanisms operating at the same time but in opposite directions. In both cases, agent-based modeling constitutes a powerful virtual laboratory in which to design triggering conditions and to determine the resulting microscopic and macroscopic effects of concatenations of mechanisms.

Despite these objective advantages of agent-based modeling, to what extent might the importance that many analytical sociologists attribute to this method give rise to another form of "hard obscurantism" generating essentially only "science fiction" (I borrow the two terms from Elster's (2007, p. 458-465)

criticism of quantitative social sciences)? My answer is that this danger can be limited by giving priority to the so-called empirically-calibrated agent-based models (see Hedström, 2005, ch. 6), that is to say, artificial societies in which agents' attributes and behavior rules rely on empirical information provided by ethnographic studies, experiments, or survey data. Although not easy, this combination is technically possible. I do believe that analytical sociology should be given the chance to prove that this is a research pathway that is worth exploring.

Conclusion

I have discussed an article by Boudon which I regard as important for two reasons. On the one hand, it is a synthetic exposition of his conception of rational action as "reason-based action"; on the other, it is also Boudon's first direct assessment of the growing intellectual movement labeled "analytical sociology".

As regards the former aspect, I first pointed out that Boudon's conception of rational action is part of a more extensive shift from a narrow to a broad version of rational action theory. To some extent, Boudon's theory represents the extreme version of this trend, in that it equates rationality to the subjectively perceived reasons that an actor endorses to believe/do what he believes / does. In this respect, my argument has concerned the main problem that must be tackled when we assume that rational action amounts to "reason- grounded action", *i.e.* the reduced capacity to figure out *ex-ante facto* a single micro- or macro-level outcome that should be observed under certain social circumstances. As Boudon honestly admits, the larger the set of acceptable reasons, the less unique are the theory's predictions.

To solve this problem, I have suggested that, instead of going back to a narrower conception of rational action (as Abell [1992] suggested, for instance), we may try to accumulate regularities on "reason triggers": that is to say, micro- or structural-level facts that increase the probability that specific sets of reasons will appear to actors' minds. Among possible "reason triggers", I have focused on "heuristics", "social identity", and "emotions". In particular, I have stressed that the experimental evidence accumulated within the "fast-and-frugal heuristic" research program in cognitive psychology shows systematic links among given informational structures, given beliefs, and certain mental shortcuts. Recent research in social psychology, political science, and economics on social identity conceived as a fundamental psychological need is of help in establishing regularities in the connection among social belonging, social identity and actors' beliefs that increase our capacity to predict actors' preference inconsistencies across social circles and over time. Social and cognitive psychology research on the positive, not necessarily distorting, role of emotions in belief formation can help in detecting regularities in actors' systems of reason, in particular when interaction-based social comparisons are at the origin of emotions like envy, jealousy, indignation, shame, or resentment.

In regard to Boudon's critical stance on analytical sociology, I have sought to show that the basic principles of this approach are identical with Boudon's conception of what scientific sociology should be. A minor disagreement concerns the stress that some analytical sociologists put on the technical side of the enterprise. On this point, my argument has been that a specific class of formal models, namely agent-based computational simulations, is in fact one of the best resources available today for the study of theoretical models based on a complex form of methodological individualism. I have suggested that this conviction does not arise from a naïve love of technicalities, but rather from the close match between the theoretical requirements of this form of methodological individualism and the structural features of the computational methodology.

Acknowledgments

The first draft of this paper was prepared for a PhD course on "Norms, Values, and Normative Argumentation" organized by the department of sociology and human geography of the University of Oslo and held at the *Centre de Coopération Franco-Norvégien* in Paris in January 2012. I am very grateful to participants at this seminar, in particular Ragnvald Kalleberg and Hans Joas, for their comments and suggestions. I have also greatly benefited from discussing the article with Raymond Boudon, Karl-Dieter Opp, and Mohamed Cherkaoui, who carefully commented on a previous version of the manuscript. I have finally profited from dense discussion with James Jaspers on emotions and rationality: I thank him for kindly agreeing to meet me during his busy Paris fellowship in spring 2012. Last but not least, I wish to express my gratitude to Adrian Belton for revising my English. The usual disclaimers apply.

Bibliographic references

Abell, P. (1992). "Is rational choice theory a rational choice of theory?". In James S. Coleman; Thomas J. Fararo (eds.). *Rational Choice Theory. Advocacy and Critique*. Newbury Park, California: Sage, 183-207.

Aguiar, F. and De Francisco, A. (2002). "Rationality and Identity: A critique of Alessandro Pizzorno". *European Journal of Sociology / Archives Européennes de Sociologie*, 43-1, 119-131. <http://dx.doi.org/10.1017/S0003975602001042>

— (2009). "Rational choice, social identity, and beliefs about oneself". *Philosophy of the Social Sciences*, 39-4, 547-571. <http://dx.doi.org/10.1177/0048393109333631>

Akerlof, G. A. and Kranton, R. (2001). *Identity economics. How our identities shape our work, wages, and well-being*. Princeton: Princeton University Press.

Beaney, M. (2001). "Rationality and Reasonableness". In: Berit Brogaard, Barry Smith (eds.). *Rationality and Irrationality: Proceedings of the 23rd International Wittgenstein Symposium*. Vienna: öbv & hpt, 18-24.

Bicchieri, C. (2006). *The Grammar of Society. The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Borrill, P. and Tesfatsion, L. (2010). "Agent-based modeling: The right mathematics for the social sciences?". Working Paper no 10023. Iowa State University, Department of Economics.
Boudon, R. (1965). "Réflexion sur la logique des modèles simulés". *Archives Européennes de Sociologie*, VI-1, 3-20.
— (1979). "Generating Models as a Research Strategy". In: Robert K. Merton, James S. Coleman, Peter H. Rossi (eds.). Qualitative and quantitative social research. New York: The Free Press, 51-64.
— (1989). "Subjective Rationality and the Explanation of Social Behavior". *Rationality and Society*, 1-2, 171-196.
<http://dx.doi.org/10.1177/1043463189001002002>
— (1993). "Toward a Synthetic Theory of Rationality". International Studies in the Philosophy of Science, 7-1, 5-19.
<http://dx.doi.org/10.1080/02698599308573439>
— (1994). The Art of Self-Persuasion. Cambridge: Polity Press.
— (1996). "The cognitivist model. A generalized rational-choice model". *Rationality and Society*, 8-2, 123-150.
<http://dx.doi.org/10.1177/104346396008002001>
— (1998a). "Limitations of Rational Choice Theory". *American Journal of Sociology*, 104-3, 817-828.
<http://dx.doi.org/10.1086/210087>
— (1998b). "Social Mechanisms without black boxes". In: Peter Hedström and Richard Swedberg (eds.). *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press, 172-203.
— (2001). *The Origin of Values*. New Brunswick and London: Transaction Publishers.
— (2003a). "Beyond Rational Choice Theory". *Annual Review of Sociology*, 29, 1-21.
<http://dx.doi.org/10.1146/annurev.soc.29.010202.100213>
— (2003b). *Raison, bonnes raison*. Paris: PUF.
— (2003c). *Y a-t-il encore une sociologie?* Paris: Odile Jacob.
— (2004). *The Poverty of Relativism*. Oxford: Bardwell Press.
— (2007). Essais sur la théorie générale de la rationalité. Paris: PUF.
— (2011). "The cognitive approach to morality". In: Steven Hitlin and Stephen Vaisey (eds). *Handbook of the Sociology of Morality*. New York and Berlin: Springer, 15-33.
— Betton, E. (1999). "Explaining the Feelings of Justice". *Ethical Theory and Moral Practice*, 2, 365-398.
<http://dx.doi.org/10.1023/A:1009918510993>
Brown, D. (1991). *Human Universals*. New York: McGraw-Hill.
— (1999). "Human Nature and History". *History and Theory*, 38, 138-157.
<http://dx.doi.org/10.1111/0018-2656.00108>
— (2004). "Human Universals, Human Nature & Human Culture". *Daedalus*, 133-4, 47-54.
<http://dx.doi.org/10.1162/0011526042365645>
Camerer, C. F. and Loewenstein, G. (2004). "Behavioral Economics: Past, Present, Future". In: Colin F. Camerer, George Loewenstein, Matthew Rabin (eds.). *Advances in Behavioral Economics*. Princeton: Princeton University Press, 3-51.
Centola, D.; Willer, R. and Macy, M. (2005). "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms". *American Journal of Sociology*, 110-4, 1009-1040.
<http://dx.doi.org/10.1086/427321>

Cherkaoui, M. (2005). *Invisible Codes: Essays on Generative Mechanisms*. Oxford: The Bardwell Press.
Coleman, J. S. (1990). *The Foundations of Social Theory*. Cambridge and London: Belknap Press of Harvard University Press.
— (1986). "Social Theory, Social Research and a Theory of Action". *American Journal of Sociology*, 91-6, 1309-1335.
<http://dx.doi.org/10.2307/2779798>
Elster, J. (ed.). (1985). *The Multiple Self*. Cambridge: Cambridge University Press.
— (1999). *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.
— (2007). *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*. New York: Cambridge University Press.
— (2009). "Emotions". In: Peter Hedström, Peter Bearman (eds.). *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press, 51-71.
— (2011). "Indeterminacy of emotional mechanisms". In: Pierre Demeleunaere (ed.), *Analytical Sociology and Social Mechanisms*. Cambridge: Cambridge University Press, 50-63.
Gross, N. (2009). "A Pragmatist Theory of Social Mechanisms". *American Sociological Review*, 74, 358-379.
<http://dx.doi.org/10.1177/000312240907400302>
Fiske, S. (2011). *Envy up, Scorn down. How Status Divides Us*. New York: Russell Sage.
Frederick, S.; Loewenstein, G. and O'donoghue, T. (2002). "Time Discounting and Time Preference: A Critical Review". *Journal of Economic Literature*, 40-2, 351-401.
<http://dx.doi.org/10.1257/002205102320161311>
Frijda, N.; Manstead, A. S. R. and Sacha, B. (2002). *Emotions and Beliefs: How Feelings Influence Thoughts*. Cambridge: Cambridge University Press.
Gächter, S. and Herrmann, B. (2009). "Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment". *Phil. Trans. R. Soc. B*, 364, 791-806.
<http://dx.doi.org10.1098/rstb.2008.0275.
Gallegati, M. and Kirman, A. P. (1999). *Beyond the Representative Agent*. Aldershot and Lyme, NH: Edward Elgar.
Gigerenzer, G. (2008). *Rationality for Mortals. How People Cope with Uncertainty*. Oxford: Oxford University Press.
Gigerenzer, G.; Hertwig, R. and Pachur, T. (2011). *Heuristics. The Foundations of Adaptive Behavior*. Oxford: Oxford University Press.
Golder, S. and Macy, M. W. (2011). "Diurnal and Seasonal Mood Vary with Work, Sleep and Daylength Across Diverse Cultures». *Science*, 333, 1878-1881.
<http://dx.doi.org/10.1126/science.1202775>
Goldstein, D. (2009). "Heuristics". In: Hedström Peter, Bearman Peter (eds.). *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press, 140-167.
Goldthorpe, J. (1998). "Rational Action Theory for Sociology". *The British Journal of Sociology*, 49-2, 167-192.
<http://dx.doi.org/10.2307/591308>
Harmon-Jones, E. (2000). "A cognitive dissonance theory perspective on the role of emotion in the maintenance and change of beliefs and attitudes". In: Nico H. Frijda, Antony S. R. Manstead and Sacha Berm. (eds.). *Emotions and Beliefs: How Feelings Influence Thoughts*. Cambridge: Cambridge University Press, 185-212.

Hedström, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
Hedström, P. and Bearman, P. (eds.). (2009a). *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press.
— (2009b). "What is Analytical Sociology all about?". In: Peter Hedström, Peter Bearman (eds.). *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press, 3-24.
Hedström, P. and Edling, C. (2009). "Tocqueville and analytical sociology". In: Mohamed Cherkaoui and Peter Hamilton (eds.). *Raymond Boudon: A Life in Sociology*. Oxford: The Bardwell Press, 153-172.
Hedström, P. and Udéhn, L. (2009). "Analytical sociology and theories of middle range". In: Peter Hedström, Peter Bearman (eds.). *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press, 25-47.
Henrich, J.; Boyd, R.; Bowles, S.; Camerer, C.; Fehr, E.; Gintis, H. and Mcelreath, R. (2001). "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies". *American Economic Review*, 91, 73-78.
Hummon, N. P. and Fararo, T. J. (1995). "Actors and networks as objects". *Social Networks*, 17-1, 1-26. <http://dx.doi.org/10.1016/0378-8733(94)00245-6>
Jasper J. M. (2011). "Emotions and Social Movements: Twenty Years of Theory and Research". *Annual Review of Sociology*, 37, 285-303. <http://dx.doi.org10.1146/annurev-soc-081309-150015.
Kahneman, D. (2003). "Maps of Bounded Rationality: Psychology for Behavioral Economics". *The American Economic Review*, 93-5, 1149-1475. <http://dx.doi.org/10.1257/000282803322655392>
Little, D. (1998). *Microfoundations, Methods, and Causation: Essays in the Philosophy of Social Sciences*. New Brunswick and London: Transaction Publishers.
Manzo, G. (2010). "Analytical Sociology and Its Critics". *European Journal of Sociology / Archives Européennes de Sociologie*, 51-1, 129-170. <http://dx.doi.org/10.1017/s0003975610000056>
— (2011). "The Whole is Greater than the Sum of its Parts: Some Remarks on The Oxford Handbook of Analytical Sociology". *European Sociological Review*, 27-6, 829-835.
Newell, A.; Shaw, J. C. and Simon, H. A. (1958). "Elements of a Theory of Human Problem Solving". *Psychological Review*, 65-3, 151-166.
Opp, K.-D. (1999). "Contending Conceptions of the Theory of Rational Action". *Journal of Theoretical Politics*, 11-2, 171-202. <http://dx.doi.org/10.1177/0951692899011002002>
Ostrom, E. (1998). "A Behavioral Approach to the Rational Choice Theory of Collective Action". *The American Political Science Review*, 92-1, 1-22.
Pinker, S. (2002). *The Blank Slate*. New York: Viking.
Pizzorno, A. (1986). "Some other kind of otherness: A critique of rational choice theory". In: Alejandro Foxley, Michael S. McPherson, Guillermo A. O'Donnell, Albert O. Hirschman (eds.). *Development, Democracy and the Art of Trespassing: Essays in Honor of Albert O. Hirschmann*. Notre Dame (Ill.): University of Notre Dame Press, 355-373.
— (2007). *Il velo della diversità. Studi su razionalità e riconoscimento*. Milano: Feltrinelli.
Rawls, J. (1993). *Political Liberalism*. New York: Columbia University Press.

Scherer, K. (2011). "On the rationality of emotions: or, When are emotions rational?". *Social Science Information*, 50-(3-4), 330-350. <http://dx.doi.org/10.1177/0539018411411018>

Searle, J. (2001). "The Classical Model of Rationality and its Weaknesses". In: Berit Brogaard and Barry Smith (eds.). *Rationality and Irrationality: Proceedings of the 23rd International Wittgenstein Symposium*. Vienna: öbv & hpt, 311-324.

Sen A. (2009). *The Idea of Justice*. London: Penguin Books.

Shafir, E. and Leboeuf, R. A. (2002). "Rationality". *Annual Review of Psychology*, 53, 491-517. <http://dx.doi.org/10.1146/annurev.psych.53.100901.135213>

Simon, H. (1979). "Rational decision making in business organizations". *American Economic Review*, 69-4, 493-513.

Smith, V. (2008). *Rationality in Economics. Constructivist and Ecological Forms*. Cambridge: Cambridge University Press.

Stets, J. E. and Turner, J. H. (2006). *Handbook of the Sociology of Emotions*. Springer.

Tosini, D. (2011). "Agents and Mechanisms of Terrorist Campaigns: A Contribution to a General Theory of Rationality". *Revue Européenne des Sciences Sociales*, 49-2, 43-70. <http://dx.doi.org/10.4000/ress.973>

Willer, R.; Kuwabara, K. and Macy, M. (2009). "The False Enforcement of Unpopular Norms". *American Journal of Sociology*, 115, 451-490. <http://dx.doi.org/10.1086/599250>

Willer, R. (2009). "Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem". *American Sociological Review*, 74, 23-43. <http://dx.doi.org/10.1177/000312240907400102>

Wooldridge, M. (2009). *An Introduction to Multi-Agent Systems*. Glasgow: John Wiley & Sons.

Young, Y. (2005). "The (Un)Reasonableness of Rawlsian Rationality". *South African Journal of Philosophy*, 24-4, 308-320. <http://dx.doi.org/10.4314/sajpem.v24i4.31431>

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2083

Social contagion and homophily within romantic networks: A simulation analysis

Francisco Linares
Universidad de La Laguna
flinares@ull.es

Received: 24-03-2014
Accepted: 13-06-2014

Abstract

La logique du social anticipated, 35 years ago, the current conception of complex adaptive systems, where agents behave following some kind of rational behavior, interact among themselves and, as a result of those interactions, produce an unintended outcome. The aim of this paper is to illustrate Raymond Boudon's approach by means of an agent based simulation model, where agents represent teenagers who attempt to find a romantic relationship looking for a partner within their friendship ties. Partner choice is ruled by a homophilic principle which is, in the case of this artificial society, assumed to operate in a very simple way: agents look for someone who is similar to them in a given socio-cultural trait. At the same time, the value of this trait (which is assumed to be measured on a quantitative scale) for every agent is influenced by the values of other agents in the immediate environment. It is shown how these two social processes are interdependent, thus producing outcomes which are causally related.

Keywords: homophily; partner choice; social contagion; social networks; social simulation.

Resumen. *Contagio social y homofilia en una red romántica: un análisis de simulación*

La lógica de lo social avanzó, hace treinta y cinco años, la actual concepción de los sistemas adaptativos complejos, en los que los agentes actúan siguiendo alguna forma de comportamiento racional, interactúan entre ellos y, como resultado de tales interacciones, producen un efecto inintencionado. El objeto de este trabajo es ilustrar la aproximación de Raymond Boudon a través de un modelo de simulación basado en agentes, donde los agentes representan a adolescentes que intentan mantener relaciones románticas buscando a su pareja entre sus lazos de amistad. La elección de pareja se rige por un principio de homofilia que, en el caso de esta sociedad artificial, se asume que opera de una forma muy sencilla: los agentes buscan a alguien que sea similar a ellos en un cierto rasgo sociocultural. Al mismo tiempo, el valor de ese rasgo para cada agente (que, se asume, puede medirse en una escala cuantitativa) se ve influenciado por los valores de otros agentes en su entorno inmediato. Se muestra cómo estos dos procesos sociales son interdependientes, por lo que producen resultados que están causalmente relacionados.

Palabras clave: homofilia; elección de pareja; contagio social; redes sociales; simulación social.

Summary

1. Introduction

The aim of this paper is to illustrate Raymond Boudon's sociology by means of the analysis of a population of artificial agents. These agents represent teenagers who attempt to find a romantic relationship looking for a partner within their friendship ties. Partner choice is ruled by a homophilic principle which is, in the case of this artificial society, assumed to operate in a very simple way: agents look for someone who is similar to them in a given socio-cultural trait. At the same time, the value of this trait (which is assumed to be measured on a quantitative scale) for every agent is influenced by the values of other agents in its immediate environment. By building an agent-based model, I attempt to show how these two social processes (contagion of a social trait and partner choice) are interdependent, thus producing outcomes (i.e., number of relationships and variation of the trait among agents) which are causally related, even if statistical analysis would suggest the contrary.

The analysis briefly summarized here shares, not accidentally at all, a number of characteristics with Boudon's approach to social explanation. It is assumed that agents act in a certain social environment which, in this case, is fixed. It is also assumed that agents pursue certain goals (having a romantic relationship) and that, in pursuing these goals, they adjust their behavior to each other. It is also assumed that as an effect of this mutual adjustment a feedback process between agents' actions and their results happens, thus producing a number of social outcomes (properties of the system of action) which were not intended by the agents themselves.

The paper will proceed as follows: first a brief exposition of the characteristics of Boudon's sociology is provided. Then follows a small account of two simple simulation models of contagious and partner choice in which every social dynamic is considered to be independent of the other, thus making them useful as baseline models. Thirdly the social-contagion-and-partner-choice-model (SCPCM) is briefly described; details of the model can be found in the appendix. The most important results of the analysis of the model are then shown. Finally, the paper ends with a discussion of the results and a brief valuation of Boudon's sociology.

2. The sociology of Boudon

"Complex Adaptive Systems" is the term which has been coined to refer to systems – whether social, biological, or of another kind – whose aggregate behavior is the result of actions performed by different individuals who continuously adapt to an environment, which is itself constituted by other individuals also engaged in an adaptive behavior (Miller and Page, 2007). These multiple behaviors, may combine in complex ways to produce self-reinforcing dynamics that are difficult to understand. This is so because the behavior that emerges from individuals' mutual adaptations, aggregate behavior, may in fact be the product of a long chain of interactions and show patterns that are quite distant from individuals' initial motivations to act.

The conception of social systems as complex adaptive systems was advanced, 35 years ago in Raymond Boudon's pathbreaking books: *La logique du social* (1975) and *Le place du désordre* (1984). Summarizing Boudon's theorization:

1) The social explanandum, whether the development of a certain agricultural society, the spreading of the use of a new drug among doctors, or the rate of anomie in a population, is the product of the combination of individual actions.
2) These individual actions are to be explained within a Weberian paradigm, which is to say that reasons must be found to account for those actions. These "reasons" needn't be all of the same type.
3) Individuals themselves act in a given environment, which they usually cannot control and which influence their actions. Thus, as an example, doctors working in a hospital will not be influenced by each other in the same way as private practitioners. So, even if they have the same motivation to treat ill patients, the influence process will be different in each case.
4) These basic principles lead, depending on the characteristic of the system, to reproductive processes, cumulative processes or transformational processes. Cumulative and transformational processes imply feedback dynamics between the system of action and its outcomes.

Besides these basic guidelines for theory construction, it is interesting to remark on two further features of Boudon's work: the use of well-established sociological knowledge, often from the classic period, to develop his arguments, and the use of simple formal models in order to better present his argument and make his conclusions easier to understand (and difficult to refute). A clear example is his examination of the relative-deprivation effect found in Stouffer's *American Soldier* by means of a simple lottery model where rewards depend on the number of people buying a ticket. This allows him to show in a quite straightforward way how higher opportunities of upward mobility (environmental conditions) make it reasonable for a larger amount of people to "buy a lottery ticket" (Weberian paradigm), thus producing widespread frustration (social fact to be accounted for). In this way, Boudon shows how a

very simple mechanism operating at the motivational level of action, together with certain structural conditions, is enough to account for a perverse unintended social effect which has previously appeared as a sort of contradiction: the higher the opportunities the higher the frustration.

Although the potential of this paradigm is successfully demonstrated in his writings, one may wonder whether this method would work with more complex systems, where less stylized facts are to be explained. What if different individuals have different attitudes towards risky options, hold different assets, etc. And, even more, what if the decisions in realm A, let's say working opportunities, is somehow interconnected by decisions in realm B, let's say friendship relationships, which is itself ruled by different mechanisms.

Of course these questions are not easy to answer in a rigorous, deductive way. Fortunately, there is nowadays a tool that may help: agent-based models (ABMs). This tool was not available when Boudon wrote his masterpieces, although he was able to use a different type of simulation (micro-simulations) in addressing the issue of educational choices. An ABM is a formal and simple representation of the reality which, unlike other formal and simple representations (such as differential equations), can easily deal with heterogeneity in a population of individuals (that is, individuals may differ in many traits), as well as with decision rules other than rationality. ABMs are thus nicely suited for analyzing complex adaptive systems.

These models have several applications. By means of empirically calibrating their parameters, they have been used to explain real data, such as fertility trends in France (González-Bailón and Murphy, 2013), local youth unemployment rates in Stockholm (Hedström, 2005), or educational achievement in France (Manzo, 2013). Whatever the empirical applications, in the realm of sociology ABMs are also a formal tool for developing and exploring the implications of middle-range theories (Gilbert, 2008). According to this aim, which is also the aim of this paper, ABMs are mainly used to explore the logical consequences following from a set of assumptions about the characteristics of agents, their rules of interaction, and the characteristics of the environment. The main theoretical and methodological implications of these kinds of models, related to the sociology of Boudon, are: bottom-up explanations, the analysis of cumulative systems and the production of artificial experiments.

2.1. Explanations, generative social science and mechanisms

As Epstein and Axtell (1996) nicely put it at the end of their pathbreaking *Growing Artificial Societies*, where the now well-known Sugarscape model is analyzed:

> From an epistemological stand point, what "sort of science" are we doing when we build artificial societies like Sugarscape? Clearly, agent-based social science does not seem to be either deductive or inductive in the usual senses. But then what is it? We think generative is an appropriate term. The aim is

> to provide initial micro-specifications (initial agents, environments and rules) that are sufficient to generate the macrostructure of interest. We consider a given macrostructure to be "explained" by a given micro-specification when the latter's generative sufficiency has been established. (p. 177)

This "generative" approach, which implies that a given social pattern is explained from the bottom-up, provides a sensible answer to the micro-macro problem masterfully traced by Coleman (1986) ten years earlier. It is also very close to Elster (1989) and others' (e.g., Hedström and Swedberg, 1998) defense of "mechanisms" as the building blocks of sociological explanations. According to Hedström and Bearman (2009: 5), a mechanism "refers to a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome." Although a common unit of analysis in sociology is the individual, nothing in the concept of "mechanism" precludes the unit either being a "supra-individual" entity, such as a collective, or a "sub-individual" entity, such as the components of an individual decision-making process (e.g., attitudes, values, emotions, etc.). The concept of "mechanism" does not exclude a rational conception of action either. What the concept of "mechanism" does imply is that whatever the entities and their rules of behavior, it has to be shown that they must regularly produce the outcome that is to be explained. This conception of social explanation is essentially the same as that proposed in points 1 to 4 summarized above.

2.2. Emergence and cumulative systems

One of the most intriguing characteristics of society is the strong stability of many social patterns. Despite the fact that we all have the experience of living in an era of change, certain characteristics of society seem either to change very slowly or not to change at all. The distribution of wealth among different social classes, rules of domestic labor assignment and school achievement rates of students of different backgrounds are just a few examples. Assuming that all that happens in society is a result of individual actions, the question to answer is, how is it that individuals act in ways that produce such aggregate patterns, which are often unintended, undesired, and even detrimental to many of them?

The answer lies in the fact that the relations individuals produce when interacting with one other often produce a "new reality" that, so to speak, "traps" individuals. This "new reality" is called an *emergent* outcome of the system. As in the case of undergraduate students living in a residence hall who, in a few days, develop a stable system of informal rules concerning the use of the common kitchen, once a given distribution of rights and resources is established in any realm of society it will likely show a self-perpetuating trend, since agents are now forced to mutually adjust their behavior under the new conditions, eventually reaching an equilibrium (though possibly "unfair"). These complex adaptive systems, where the emergent outcome feeds back on

the original system of action, are known as "cumulative systems" (Boudon, 1979) and are a common object of sociological analysis.

2.3. Research methods and simulation experiments

Both quantitative and qualitative data are essential to produce simulation models insofar as these models are not built to reflect someone's fantasies about society. The representation of reality – the model – has to be grounded in empirical knowledge of the world if it is to provide an explanation of that world at all. However, in any given theory (whatever the theoretical style), we will find concepts (e.g., properties of individuals such as "sensitivity to the influence of others") with difficult, or even impossible, empirical measure. When agent-based modelers are faced with this problem, some solution must be found in order to make the simulation run. The solution consists of substituting unknown empirical data for random numbers which are extracted from a theoretical distribution (this is why the results of ABMs must be presented as averages of a sufficiently large number of simulation runs). While this procedure may be considered an artifact, notice that it is quite honest: the modeler explicitly recognizes the lack of knowledge that in a narrative style of theorizing often goes unnoticed (and sometimes hidden under a prose whose eventual literary beauty is not an essential element of a proper explanation).

This "artificial" way of proceeding has a further advantage, which is key for the analytical agenda in social sciences: the possibility of carrying out more complex "thought experiments" than those performed in the absence of this tool (such as Boudon's lottery thought experiment). When conducting field social research, it is almost impossible to answer "what if" questions that may be relevant for increasing the understanding of a social phenomenon. When there are competing theoretical understandings of an issue, relevant questions arise, such as, "What if the topology of the social network were different?" "What if people were not sensitive to others' expectations?", etc. Nevertheless, by artificially manipulating parameters, it is possible to show whether a given prerequisite (e.g., network closure) is actually a necessary condition to "grow up" the social pattern.

3. Contagion and homophily in a social network

3.1. A simple model of contagion

"Contagion" is a fairly well known social phenomenon. Since the celebrated study of Coleman, Katz and Menzel (1957) on the diffusion of the use of "gammanym" among doctors, it is widely accepted that the influence of peers on individuals' decision to accept or refuse a given socio-cultural trait produces a kind of "snow-ball process" that can usually be represented with a typical S-shaped diffusion curve, where the speed of the process depends on certain

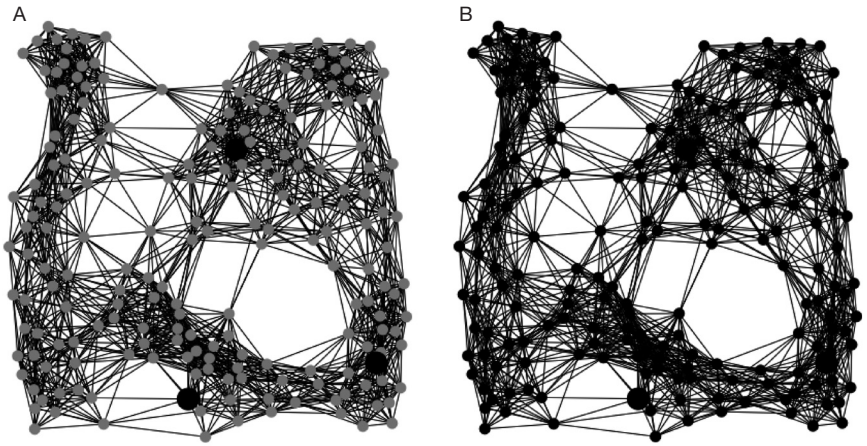
characteristics such as the "critical mass" or the network topology. This process usually ends with a fairly large proportion of the population adopting the new trait (Rolfe, 2009).

An example of this is Stonedhal and Wilensky's (2008) computational model of "virus on a network". The model represents a network of agents, a few of which are "infected by a virus" or, using a more general example, hold a certain trait with a dichotomic value. A number of parameters, such as the probability of virus-spreading or the probability of recovering after being infected, control the diffusion process in this model. However, the result is almost always the same. Even under the worst conditions for the spreading of infection allowed by the model, a very small number of initially infected nodes will foster the spreading of the virus through a large part of the population, as soon as there is a small probability of infection (5%) and the network is dense enough.

In Stonedhal and Wilensky's model, the artificial network is produced by a very simple algorithm which asks the potential nodes to create links with randomly selected partners until a certain number of connections are created. This final number of links is the result of multiplying the number of nodes by the average node degree (the average number of links per node), which is a parameter that can be manipulated in the model. This algorithm almost grants that, in every simulation, the topology of the network fits a "small world" topology, that is, a network with an average low distance between any two random nodes.

Figures 3.1.A and 3.1.B show a typical network generated by this model with 200 nodes and an average node degree equal to 20. Parameters that refrain diffusion are set to the minimum, and the virus spread chance is set to 5%. In figure 3.1.A, the large black circles represent the initial "infected" nodes.

Figure 3.1. A) Virus on a network (initial state). B) Virus on a network (final state)

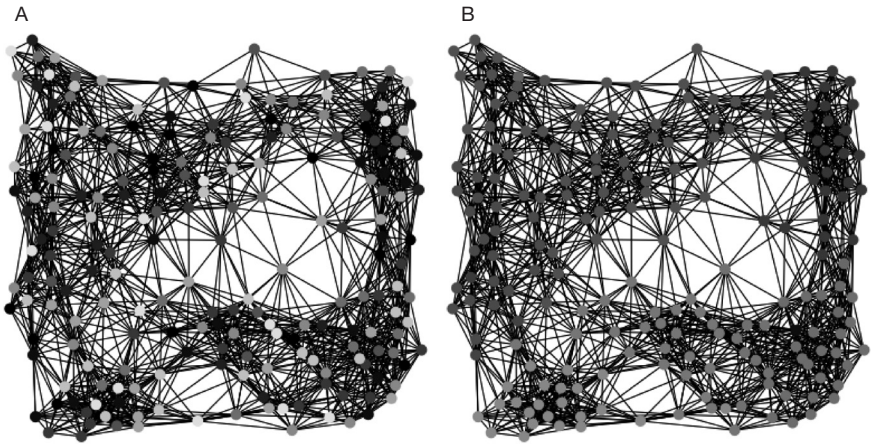


But what if the trait to be diffused is not a discrete-dichotomous variable, but a continuous one? The question can be explored by modifying the initial "virus on a network" model in the following way:

a. All nodes have a state variable (representing a social trait) which can adopt values between 0 and 9; with the values being randomly assigned.
b. Nodes are sensitive to contagion from others with a probability which is controlled by a "social influence" parameter.
c. With probability fixed at point b), the nodes change the value of their traits to the median (the use of the median instead of the mean is recommended since individuals are not likely to be influenced by extreme options, but the main results of the analysis do not depend on the use of this statistical measure) of the nodes with which they have a directed link (i.e., nodes within a path distance of 1).

Figures 3.2.A and 3.2.B show the same network before and after the contagion process is finished. The color of the nodes indicates the "intensity" of their trait, and is darker the lower the values of the trait. It is important to note two main differences between these networks: first, a clustered distribution substitutes the initial random distribution, so nodes with the same color are close to each other. And secondly, the variability of the initial distribution has been lowered; which is shown by the fact that, in this network, the initial coefficient of variation of the trait decreases from 0.6 to 0.3. This process has little sensitivity to the "social influence" parameter (i.e., the end of the process is the same even for low chances of social influence).

Figure 3.2. A) Virus on a network with a continuous trait (initial state). B) Virus on a network with a continuous trait (final state)



3.2. A simple model of homophily in partner choice

Homophily is a basic principle of the structuring of social relations, meaning that similar individuals are connected among themselves more often than dissimilar ones. This tendency may be the product of the distribution of population over relevant social attributes (Blau, 1977), the structuring impact of social foci of interaction on individuals' networks (Feld, 1981), or the preferences of individuals for similar others (Kossinets and Watts, 2009). Whatever its cause, homophilic patterns imply larger homogeneity in social relations than would otherwise be expected. It is a well documented pattern in many realms of social life (see McPherson et al., 2001 and Cruz, 2013). In the realm of partner choice, Bearman et al.'s (2004) analysis of the structure of the romantic and sexual relations of 832 students at "Jefferson High" (a high school located in "Jefferson City") provides an excellent illustration. By conducting a series of simulation experiments, the authors conclude that homophily is a necessary mechanism (even if not sufficient) to account for the topology of this romantic network.

We may replicate a simplified version of that model with the same artificial network as in the previous section in the following way:

For every node:

a) Look for a partner among your directly linked neighbors, following the rules:
   a. The partner must be single (at the initial state of the simulation all of them are).
   b. The partner must be of the opposite sex (at the initial stage of the simulation sex is assigned with 50% chance).
   c. If your sex is male, the partner must be younger than you. If your sex is female, the partner must be older than you (age of the nodes ranges from 14 to 17).
   d. The difference between the values of the trait must be equal to, or lower than, 10%
b) With a certain likelihood, which is controlled by a parameter, a relationship may be broken.
c) The decisions iterate until no new relationships emerge.

The main results of this model are shown in Table 3.1. The figures (averages and standard deviations over 50 simulation runs) show that when partner choices are not homophilic, the number of relationships created through the simulation is higher the higher the value of the probability that a relationship will be broken, β. However, when partner choices are homophilic, assuming a tolerance to partner's difference of 10%, the number of relationships created is not only lower but also less dependent on the likelihood of breaking relationships (β).

Table 3.1. Number of relationships

	No homophily		homophily	
	μ	σ	μ	σ
β = 0.25	68.3	5.4	25.8	4.0
β = 0.50	92.2	9.0	30.4	5.6
β = 0.75	158.2	18.6	30.3	6.0

4. A social contagion and partner choice model

In order to analyze how these two basic and simple dynamics interact, I have built an agent-based simulation model (which I call SCPCM) where contagious and partner choice co-evolve at the same time within a population of 200 agents embedded in the same network as figures 3.2.A and 3.2.B. The model is fully described in the appendix following the ODD protocol designed by Railsback and Grimm (2012). A brief description of SCPCM is provided in this section. Afterwards some hypotheses on the expected relationships among variables are suggested.

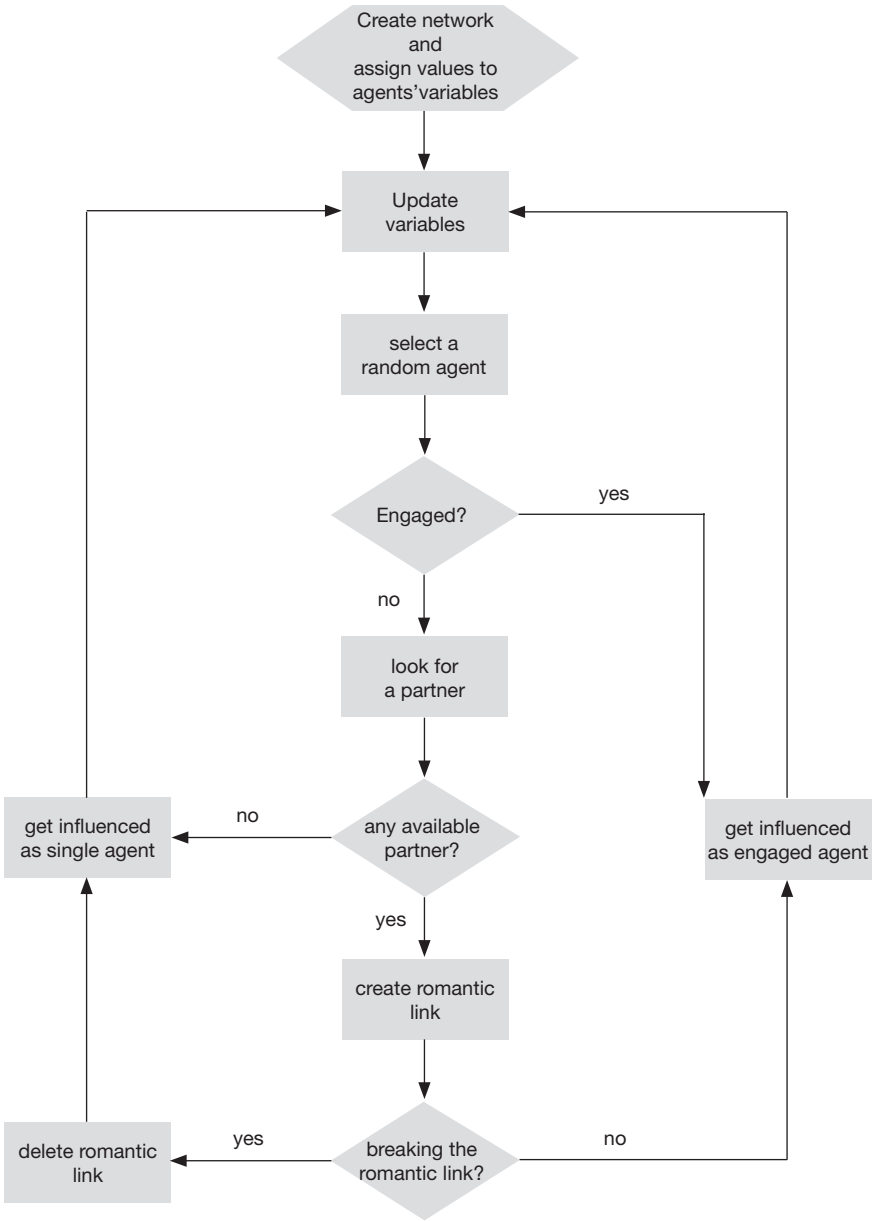
4.1. Brief description of SCPCM

The program was implemented in the Netlogo platform (Wilenski, 1998) and reproduces the following steps:

a) One of the agents is randomly chosen.
b) If the agent does not have a partner, the agent is asked to look for someone according to the following rules:
   a. The partner must be found among linked neighbors
   b. The partner must be of a different sex.
   c. If the agent is male, the partner must be younger; and the other way around if the agent is female.
   d. The difference between the values of the trait must be within a range of tolerance, which is set by a tolerance-parameter (τ).
c) If a partner is found, both agents engage. This relationship may be broken with a probability which is set by a breaking-probability-parameter (β).
d) Whether a partner is found or not, the agent is influenced by its linked neighbors according to the following rules:
   a. If the agent is not engaged, the value of the agent's trait becomes the median of its linked neighbors.
   b. If the agent is engaged, the value of the agent's trait is determined by both the value of the trait of the agent's partner, weighted by a weight-parameter (ω), and the median of the agent's linked neighbors, weighted by 1 - ω .

In summary, the model contains two different mechanisms of social interaction: on the one hand, agents select their partners following a homophilic

Figure 4.1. Flow chart



rule. The homophilic strength of the choice is determined by parameter τ, which ranges from 0 to 1.

On the other hand, agents are influenced by the other agents they are tied to, so the values of their traits converge to a central value of the local environment. This contagious process is, nevertheless, affected by the previous partner selection, since the value of the partner's trait has a special weight. The strength of the partner's influence, relative to other agents' influence, is determined by parameter ω, which also ranges from 0 to 1.

In the end there is a feedback process between partner choice and trait contagion: the distribution of trait values influences the agents' partners' pool; and, at the same time, agents' choices of partner influence the distribution of trait values. These dynamics are summarized in the flow chart above. The model attempts to show the outcomes of these reinforcing flows, paying special attention to the fact that the variation among agents' trait values within the network is determined by parameters τ, β and ω.

4.2. Hypothesis

Concerning the process of contagion, the variation in the trait distribution should be positively associated with tolerance, since tolerant individuals will be "comfortable" in a world with high diversity. It should also be negatively associated with the weight of the partner's influence, since if my partner has a strong influence on me, overall diversity is reduced. Nevertheless, there is not an obvious way to relate it to the probability of breaking a relationship. Therefore it can be expected that:

— The higher τ, the higher the coefficient of variation of the trait (H1).
— The coefficient of variation of the trait will not be sensitive to β (H2).
— The higher ω, the lower the coefficient of variation of the trait (H3).

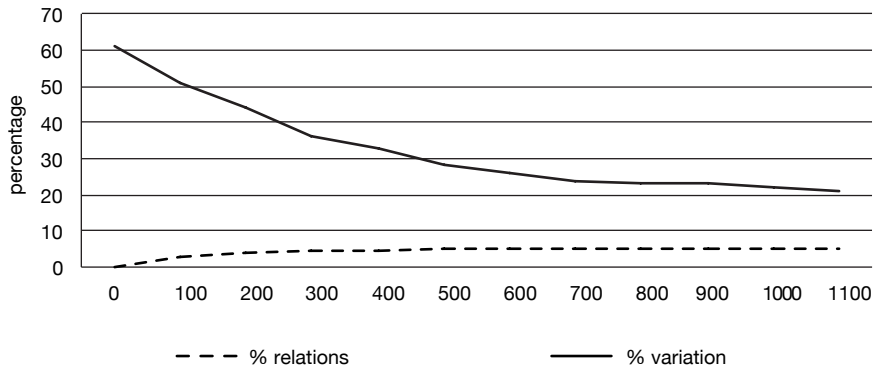
On the other hand, concerning the partner choice process, it is straightforward that as the probability of breaking romantic relationships increases, the final number of relationships must also increase. It would also seem quite obvious that the higher the tolerance to difference, the number of relationships should also increase. Therefore it can be expected that:

— The higher τ, the higher the number of relations (H4).
— The higher β, the higher the number of relations (H5).
— The number of relations will not be sensitive to ω (H6).

5. Results

A series of simulation experiments exploring the parameter space of τ, ω and β were conducted; reiterating the simulation 50 times for every experimental condition, which amounts to 66,500 simulation runs. The results of these experiments show the emergence of patterns which are quite different from the

Graph 5.0. Typical simulation run ($\beta = \omega = \tau = 0.5$)



simple models, where every dynamic operated independently. Graph 5.0 shows the evolution of the coefficient of variation (%) and number of relations (as a percentage of total friendship links) in a typical simulation run. The trends are quite clear: the trait variation continuously decreases from roughly 60% to roughly 20% as the simulation progresses, while a number of romantic relationships emerge in the early stages of the simulation. Although some of them disappear and new ones appear, the rate to total relations remains practically constant throughout the simulation run at a value of roughly 5%.

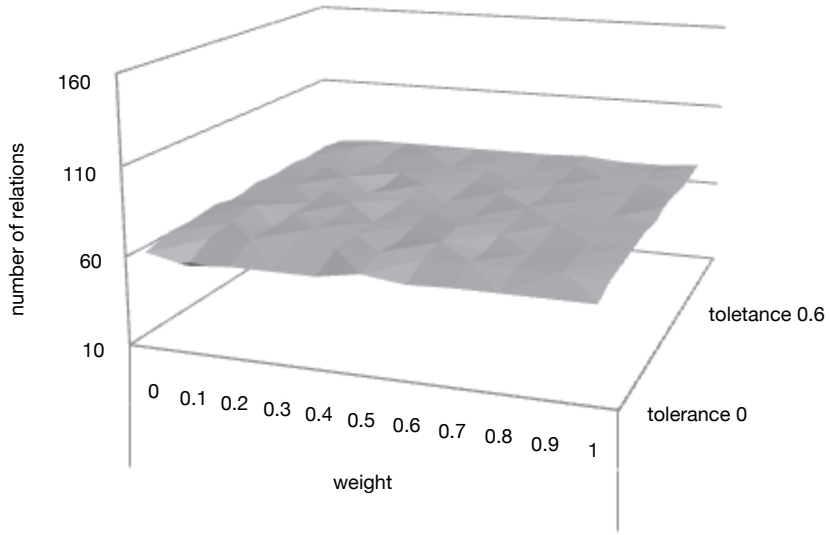
On the one hand, the number of relations is higher than in the simple homophilic partner choice (the data are actually similar to the simple partner choice model when homophilic choice is *not* allowed). On the other hand, the standard result in the simple model of contagion is just a special case of SCPCM. Thus, the almost deterministic result found in the simple model (above), in which the coefficient of variation always decreases from 0.6 to 0.3, is no longer valid. In the new model, the coefficient of variation drops below that level (as low as 0.13) for most of the combination of values of the parameter space, but increases above that value (as high as 0.43) when the values of ω are very high.

5.1. Number of romantic relationships

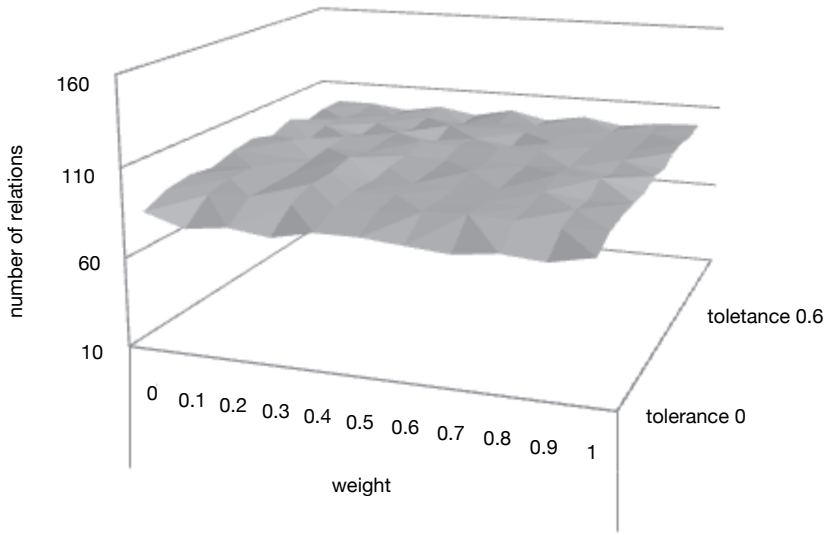
Concerning the number of romantic relationships, the simulation provides clear support for hypothesis H5 and H6, as can be easily seen in graphs 5.1, 5.2 and 5.3, which represent the number of final relationships for every combination of the spectrum parameter of τ and ω, when β equals 0.25, 0.50 and 0.75, respectively[1]. It is quite obvious that the variation in parameter β

1. The results shown in graphs are the mean values of 50 repetitions for every combination of parameters.

Graph 5.1. (β = 0.25)



Graph 5.2. (β = 0.50)



Graph 5.3. ($\beta = 0.75$)

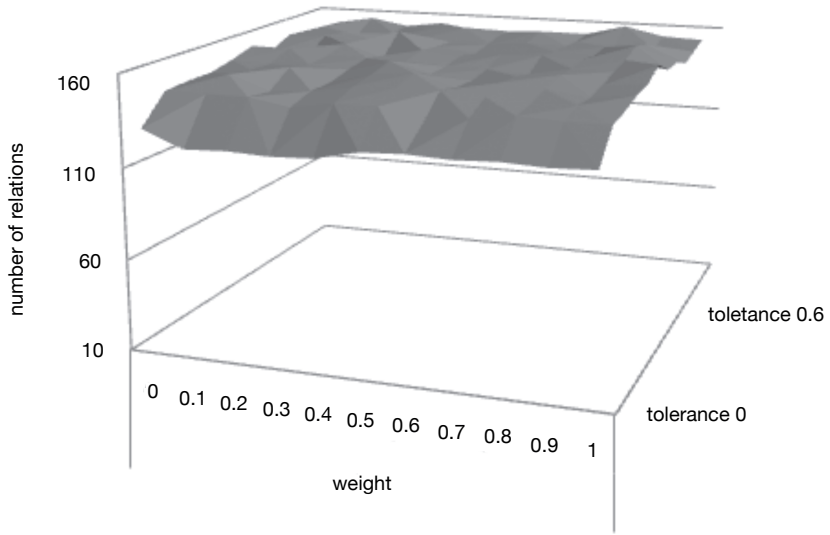


Table 5.1. Dependent variable: Number of relations

Model		Unstandardized Coefficients		Standardized Coefficients			99% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	30.075	.171		176.289	.000	29.635	30.514		
	probability-of-breaking-up	109.753	.203	.738	540.454	.000	109.230	110.276	1.000	1.000
	tolerance	8.801	.185	.065	47.606	.000	8.324	9.277	1.000	1.000
	weight	-.248	.185	-.002	-1.338	.181	-.725	.229	1.000	1.000

has the expected effect: the higher the probability of breaking a relationship, the higher the number of final relationships. It is not only the expected effect but also the greatest effect, since parameters ω and τ do not seem to have any influence. This result is clearly counterintuitive, since one would expect the number of relations to increase with tolerance to the partner's trait, as suggested by H4.

The linear multivariable regression model estimated for this dependent variable clearly confirms the impression produced by the graphs. "Probability of breaking a relationship" has the strongest significant effect on the dependent variable, while "weight" has no significant effect at all and "tolerance" has a

very weak (although significant) effect as shown in Table 5.1, which displays the results of the model. The value of R squared for the model is 0.547[2].

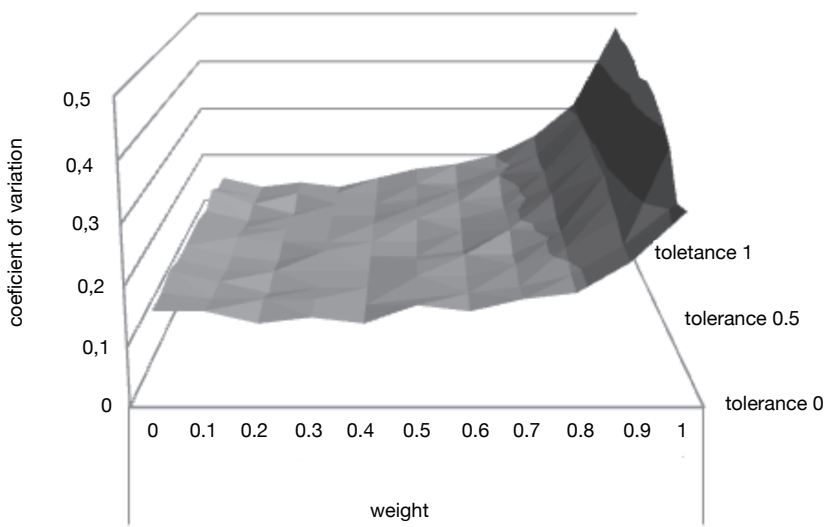
5.2. Coefficient of variation

The results concerning the coefficient of variation of the trait are even more counterintuitive. The trait variation among agents increases as β decreases, but only with high values of both ω and τ. If the weight of the partner's trait is high but tolerance to partner's difference is low or the other way round (i.e., tolerance is high but weight is low), the probability of breaking a relationship does not seem to have an effect on the coefficient of variation.

Tolerance to partner's trait and weight of partner's influence have very different effects. On the one hand, the coefficient of variation *does not seem to be very sensitive to the values of parameter* τ. On the other hand, *parameter ω seems to have a strong influence*, as the coefficient of variation of the trait clearly increases the higher the values of ω.

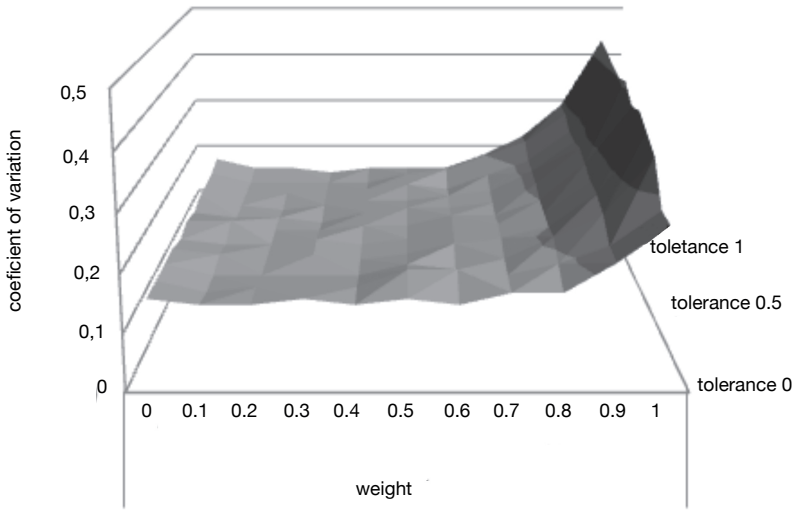
The linear multivariable regression model estimated for this dependent variable shows (see Table 5.2) that "weight" has the strongest significant effect on the dependent variable, but it is positive rather than negative. The effect of "tolerance" has the expected direction, although it is rather small. Contrary

Graph 5.4. (β = 0.25)



2. It cannot be taken for granted that the distribution of residuals is fairly homoscedastic in this statistical model or in the models presented below.

Graph 5.5. (β = 0.50)



Graph 5.6. (β = 0.75)

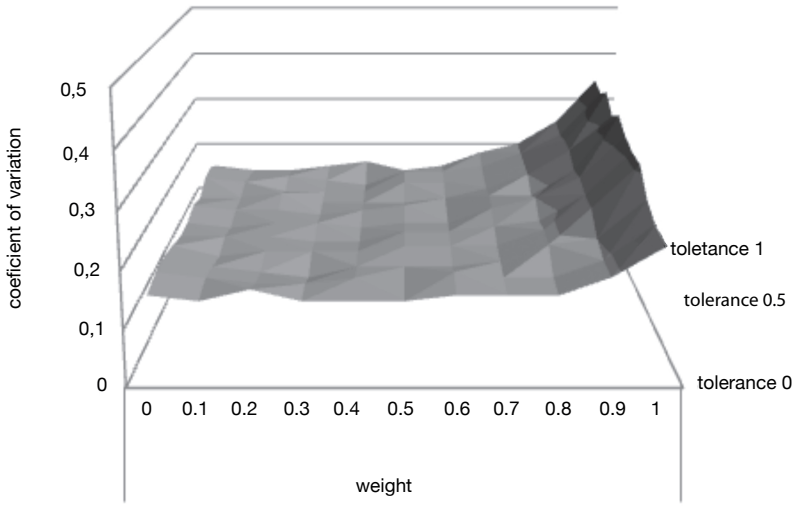


Table 5.2. Dependent variable: Coefficient of variation of trait

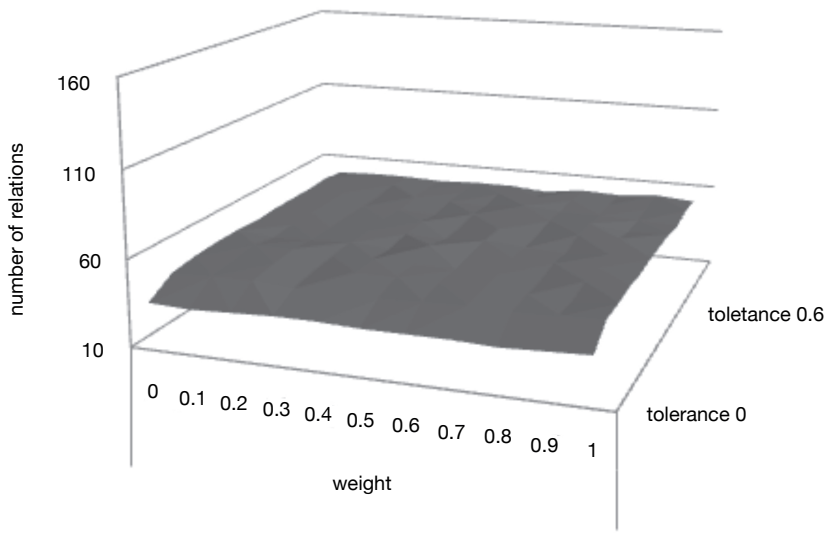
Model		Unstandardized Coefficients		Standardized Coefficients			99% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	.181	.001		353.862	.000	.180	.183		
	probability-of-breaking-up	-.030	.001	-.090	-49.124	.000	-.032	-.028	1.000	1.000
	tolerance	.018	.001	.059	32.170	.000	.016	.019	1.000	1.000
	weight	.129	.001	.424	232.622	.000	.128	.131	1.000	1.000

to our expectations, "probability of breaking a relationship" has a significant, although small, negative effect. The value of R squared for the model is only 0.192, implying that the model poorly captures the logic behind the variation of the dependent variable.

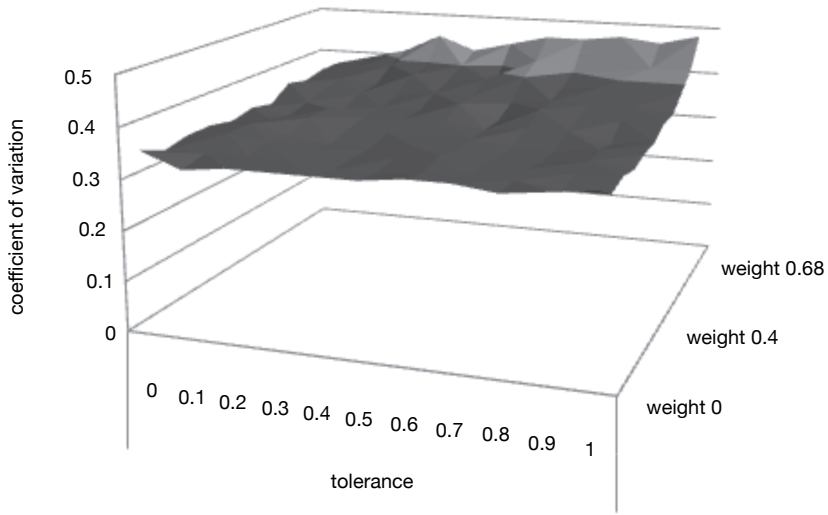
5.3. Sensitivity to average number of links

A straightforward question regarding these results is whether they are dependent (and if so, to what extent) on the topology of the network. As explained above (see the appendix for a more detailed account), the network is created by means of an algorithm which randomly assigns links to agents until the number of links per agent fits a certain average node degree parameter, which has been set to 20 throughout the whole range of simulations.

Graph 5.7.



Graph 5.8.



In order to answer this question, new simulations were conducted varying the average node degree. Graphs 5.6 and 5.7 show the number of relations and coefficient of variation when the average node degree equals 4 (i.e., agents have 4 links on average) and the probability of breaking a relationship equals 0.5 (so they can be compared to graphs 5.2 and 5.5). The influence of average node degree on the number of relations is obviously a deterministic outcome of the model: since agents choose their partners from among their linked neighbors, the lower the number of ties, the lower the number of romantic relationships. The estimated regression model (see Table 5.3) shows a strong significant effect for this variable. The R squared for this model increases to 0.743.

Table 5.3. Dependent variable: number of relations

Model		Unstandardized Coefficients		Standardized Coefficients			99% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-12.346	.162		-76.092	.000	-12.764	-11.928		
	probability-of-breaking-up	109.790	.153	.738	716.837	.000	109.396	110.185	1.000	1.000
	tolerance	8.773	.139	.065	62.921	.000	8.413	9.132	1.000	1.000
	weight	-.253	.140	-.002	-1.811	.070	-.613	.107	1.000	1.000
	average-node-degree	3.394	.008	.442	429.131	.000	3.374	3.415	1.000	1.000

Table 5.4. Dependent variable: coeficient of variation of trait

Model		Unstandardized Coefficients		Standardized Coefficients			99% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	.302	.001		599.056	.000	.301	.304		
	probability-of-breaking-up	-.030	.000	-.090	-63.098	.000	-.031	-.029	1.000	1.000
	tolerance	.018	.000	.059	41.360	.000	.017	.019	1.000	1.000
	weight	.129	.000	.424	297.775	.000	.128	.131	1.000	1.000
	average-node-degree	-.010	.000	-.561	-393.762	.000	-.010	-.010	1.000	1.000

On the other hand, the influence of average node degree on the coefficient of variation is less straightforward. The linear multivariable regression model estimated for the coefficient of variation (see Table 5.4) shows that "average node degree" has the strongest significant effect, which is negative. The effects of "weight", "probability of breaking a relationship" and "tolerance" are similar to the model shown in Table 5.2 above. The value of R squared increases to 0.507 in this model.

6. Discussion

Summarizing, in SCPCM the probability of breaking a relationship has a very strong positive effect on the number of relations, and a weak (but still significant) negative effect on trait variation. Tolerance to others has a positive significant effect on both variables, but in every case this effect is rather small. The strongest effects on trait variation comes from the average node degree and from the weight of partner influence (which has no effect at all on the number of relations).

There are a number of counterintuitive results that should be stressed. One would expect higher levels of trait variation the higher the tolerance; however, *tolerance to others has no strong effect on trait variation*. In the model, the coefficient of variation invariably falls, mainly driven by the number of available ties per agent and the weight of partners' influence. These effects are also counterintuitive insofar as one would expect the influence of these variables to operate in the opposite direction than they actually do. The trait variation should be expected to increase as the number of different neighbors also increases, and to decrease as the weight of partner's influence increases. But the statistical tests show that it actually happens the other way round. Why is this so?

The dynamic of the simulation model allows us to understand this puzzle. Because the process of contagion is necessarily stronger the denser the network of ties, the diversity among agents is reduced (and, at a network level, the

emergence of a number of trait-clusters is fostered); a result which has already been observed in previous models (Schelling, 1978; Axelrod, 1997). Furthermore, when the influence of the (similar) partner has a higher weight than the influence of other linked neighbors, homophilic choices of partners seem to reinforce the homogenization effect of social contagion. To the best of my knowledge there is no previous research accounting for this effect nor a clear explanation of why tolerance does not have a strong, statistically significant influence on trait variation.

Finally, an important result of the analysis is that since the effect of "weight" on the trait variation necessarily depends on the number of romantic relationships created through the simulations, and because this number is only a small proportion of the total amount of relations (unlikely to be higher than 5%), it follows that the behavior of a small number of agents has a strong impact on the evolution of the whole system, which is an usual feature of complex adaptive systems.

7. Conclusion

One of the most interesting features of Raymond Boudon's sociological theory is the use of simple "thought experiments" in order to ascertain the basic logic underlying the emergence of "social facts". This methodology can be widely extended nowadays by means of agent-based models, a computational tool for analyzing complex adaptive systems, such as those which are usually the object of sociological analysis. Following Boudon's approach, in this paper we have built a "thought experiment" whose goal is to ascertain the joint consequences of a contagious process and a homophilic partner choice in a population of artificial agents.

Agent-based models clearly reproduce the basic principles of Boudon's sociological analysis. In our model, agents are programmed to pursue certain goals (i.e., finding a partner) while being sensitive to other agents' characteristics. Due to the effect of agents' interactions in a given social environment (a social network), some properties of the system (i.e., total number of romantic relations and the variation of a trait among agents) change. It is worth noting that these system-level changes are driven by very simple mechanisms which operate at the level of agent's behavior (i.e., peer influence and homophilic choice).

Following this generative approach, system properties are shown to emerge from a cumulative process: when an agent finds a partner, it has a consequence for the agent and for the agent's partner; but it also has consequences for agents to which they are both linked. These consequences may drive other agents to change their behavior, and so forth. Of course these long chains of reactions cannot be disentangled in the absence of simulation experiments. Since the value of partner trait has a special weight, the influence of an engaged agent on others will not be the same as that of a "single" agent. Even if engaged agents are a minority in every simulation run, their presence has an effect

which spreads over the whole network. In the end, it is possible to observe the emergence of patterns which were not "intended" by any of agents since they are the product of local responses that have global consequences.

Acknowledgments

I am grateful to Beatriz Abdul-Jalbar and to two anonymous reviewers from *Papers*, whose comments were useful to improve a previous version of the paper. I am also grateful to my excellent students of social theory at Universidad de La Laguna (year 2013-2014) whose faithful attendance to my lectures encouraged this work. Financial support is acknowledged from research projects CSO20012-31401 (MINECO) and CSD2010-0034 CONSOLIDER-INGENIO (MICINN).

Bibliographic references

Axelrod, Robert (1997). "The dissemination of culture: a model with local convergence and global polarization". *Journal of Conflict Resolution*, 41 (2), 203-226.

Bearman, Peter S.; Moody, James and Stovel, Katherine (2004). "Chains of Affection. The Structure of Adolescent Romantic and Sexual Networks". *American Journal of Sociology*, 110 (1), 44-91.

Blau, Peter M. (1977). "A Macrosociological Theory of Social Structure". *American Journal of Sociology*, 83 (1), 26-54.

Boudon, Raymond (1979). *La logique du social. Introduction à l'analyse sociologique*. Paris: Hachete.

— (1984). *Le place du désordre*. Paris: Presses Universitaires de France.

Coleman, James S. (1986). "Social Theory, Social Research and a Theory of Action." *American Journal of Sociology*, 91 (6), 1309-35.

Coleman, James S., Katz, Elihu and Menzel, Herbert (1957). "The Diffusion of an Innovation among Physicians". *Sociometry*, 20: 253-270.

Cruz Gómez, Irene (2013). "La homofilia como forma de estructuración de las relaciones sociales en Cataluña". Universitat Autònoma de Barcelona. PhD. Diss.

Elster, Jon (1989). *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Epstein, Joshua M. and Axtell, Robert L. (1996). *Growing Artificial Societies: Social Science from the Bottom Up.* Washington (DC): The Brookings Institution.

Feld, Scott L. (1981). "The Focused Organization of Social Ties". *American Journal of Sociology*, 86 (5), 1015-1035.

Gilbert, Nigel (2008). *Agent-based Models*. London: Sage Publications.

González-Bailón, Sandra and Murphy, Tommy E. (2013). "The effects of social interactions on fertility decline in nineteenth-century France: An agent-based simulation experiment". *Population Studies: A Journal of Demography*. http://dx.doi.org/10.1080/00324728.2013.774435.

Hedström, Peter (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

Hedström, Peter and Bearman, Peter S. (2009). "What is analytical sociology all about? An introductory essay". In: Hedström, Peter and Bearman, Peter (eds.). *The Oxford Handbook of Analytical Sociology.* Oxford: Oxford University Press.

Hedström, Peter and Swedberg, Richard (1998). *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
Kossinets, Gueorgi and Watts, Duncan J. (2009). "Origins of Homophily in an Evolving Social Network". *American Journal of Sociology*, 115 (2), 405-450.
Manzo, Gianluca (2013). "Educational Choices and Social Interactions: A Formal Model and a Computational Test". *Comparative Social Research*, 30, 47-100.
McPherson, Miller; Smith-Lovin, Lynn and Cook, James M. (2001). "Birds of a Feather: Homophily in Social Networks". *Annual Review of Sociology*, 27, 415-444.
Miller, John H. and Page, Scott E. (2007). *Complex Adaptive Systems*. Princeton, NY: Princeton University Press.
Railsback, Steven F. and Grimm, Volker (2012). *Agent-based and Individual-based Modeling*. Princeton, NY: Princeton University Press.
Rolfe, Meredith (2009). "Conditional Choice" In: Hedström, Peter and Bearman, Peter (eds.). *The Oxford Handbook of Analytical Sociology.* Oxford: Oxford University Press.
Schelling, Thomas (1978). *Micromotives and macrobehavior*. New York: Norton.
Stonedhal, Forrest and Wilensky, Uri (2008). Netlogo Virus on a Network model. http://ccl.northwestern.edu/netlogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
Wilenski, Uri (1998). NetLogo. http://ccl.northwestern.edu/netlogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Appendix: Odd protocol

##Overview

1) *Purpose*: The model aims to explore how two different social dynamics, diffusion of a cultural trait and romantic matching, influence each other. The specific problem the model addresses is how both of these processes are dependent on interaction based on three individual characteristics: sensitivity to others' similarity, influence of partner on one's own decision, and likelihood of breaking a romantic relationship. The model explores these dynamics in a fixed network of 200 agents which are intended to represent teenagers who have friendship relationships which may evolve, if the right partner is found, to romantic relationships.
2) *Entities, state variables, and scales*: The model has three kinds of entities: boys, girls and links. The environment consists of a torus of 81x81 patches which have no state variable. All agents, whether boys or girls, have the following state variables: sex (boolean), age (numerical), engaged? (boolean), partners-memory (list), trait (numerical), and influence-threshold (numerical). Links represent the type of relationship between two agents by means of a color code (see below).

Global variables are: number-of-romantic-relationships (numerical) and trait-variability (numerical), which are the main outputs of the model. Other global variables are set as parameters: likelihood-of-breaking-a-relationship (numerical) tolerance-to-cultural-difference (numerical) and weight-of-partner-influence (numerical). All three of these variables are key parameters to explore in the

model. Besides these variables, average-node-degree (numerical) and mean-influence-threshold (numerical) are parameters that control the average number of ties per agent, and the average sensitivity of agents to social influence.

There are no temporal or spatial scales, since real time and/or real environment are not simulated.

3) *Process overview and scheduling*: The model includes the following actions executed every time-step in the same order:
   1. One agent is randomly chosen.
   2. If the agent is not engaged, the agent is asked to look for a partner.
   3. If a partner is found, the agent is asked to engage.
   4. Whether engaged or not, the agent is always asked to be culturally influenced (i.e., change the value of its cultural trait).
   5. Variables are updated.
   6. The simulation stops after 1200 time steps, which is enough for the model to reach an equilibrium point.

##Design concepts

4) *Design concepts*

Basic Principles: The model attempts to capture the interaction of two different mechanisms: homophily and contagion. Homophily is the principle by which people tend to engage in relations with other people similar to them in certain traits. For the sake of simplification, only one trait is represented. Contagion is a process which produces the spread of a certain trait among a population by means of social influence. In the model, agents look for a romantic partner similar to them in a certain cultural trait, which is measured on a quantitative scale. At the same time, agents are also influenced by their relationships, whether romantic or friendship, although these two different sources of influence do not have the same weight.

Emergence: The model shows how the dynamics of romantic-matching and social influence are interdependent so the rate of variation of the cultural trait among the population and the number of romantic relationships both differ from the scenario where these two processes are independent.

Adaptation: Agents perform two kinds of adaptive behavior. They become engaged if there is an agent in their local environments who meets the conditions to be chosen as a partner (details below). Second, agents change the value of their trait by means of a social influence process (details below).

Objectives: There is not a fitness or utility measure in the model to be optimized. However, agents behave as if they had the goal of finding a romantic partner.

Learning: Agents do not learn from past experience.

Prediction: Agents do not predict future conditions.

Sensing: All agents occupy a position in a network, which is assumed to not evolve as time progresses. The network represents the web of friendship

relationships among teenagers. When searching for a partner and when updating the value of its cultural trait, every agent has access to state variables of its local environment (i.e., other agents it has a direct tie with).

Interaction: Boys and girls in the same local environment interact by making (and breaking) romantic relationships (see details below). All agents in the same local environment interact by influencing one another on the value of their cultural traits (see details below).

Stochasticity: Stochastic processes are used in the initialization in different ways. The social network is seeded with random number 1111 in order not to confound the effect of variation in the network topology with the effect of agents' behavior. State variables of agents are randomly initialized in every simulation run. The agent behaving in every simulation is also randomly chosen. Since there are 200 agents and the simulation lasts for 1200 ticks, every agent has on average 6 chances of engaging in a relationship and being influenced. Random numbers are also used in some sub-models (see details).

Collectives: There are two agent sets: boys, who may match with girls younger than them; and girls who may match with boys older than them. Both boys and girls are the subject of social influence in the same way.

Observation: At the end of every simulation run, the required outputs are: a) number of social relationships engaged in through the simulation; b) actual coefficient of variation of the cultural trait. Plots show the evolution of these indicators through time steps. In addition, it is also shown in the interface whether a certain link represents friendship (black links), a current romantic relationship (green links) or a past relationship (grey links). Agents are represented by means of circles whose color shows the value of the cultural trait (from light gray for low values to dark gray for high values).

##Details

5) *Initialization*: The simulation is initialized with 200 agents, whose state variables are randomly assigned. Sex is assigned with a 50% chance. The age of agents is picked from a uniform distribution within the range 14 to 17. The trait of agents is picked from a uniform distribution within the range 0 to 9. Influence-threshold is set by a parameter between 0 and 1 (currently set to 1, i.e., maximum sensitivity to influence). The variable engaged? is set to false for all agents. Memory of past partners is initially empty.

Links are then created with a random seed. The random assignment of links to agents ends when the condition of 20 links per agent on average is met. This produces a small-world type of network. The procedure is copied from Stonedhal and Wilensky (2008).

6) *Input data*: No input data are required.
7) *Submodels*:

Look for a partner:

If an agent is selected to look for a partner it will randomly pick, if any, one of its linked neighbors which meets the following three conditions:

a) opposite sex
b) if the agent is male, the partner must be younger. If the agent is female, the partner must be older.
c) the absolute difference between the two trait values divided by ten must be less than the value set by the parameter tolerance. This grants that agents will engage with agents with a very similar trait value when tolerance is low, but the pool of possible partners will be larger when tolerance is high.

Get engaged:

If a partner has been selected, the agent checks that the partner is not a member on the list of previous partners. Then it includes the partner in this list, changes the state of engaged? to true, and asks the partner to do both actions. However, if a random number extracted from a uniform distribution between 0 and 1 is below the value set for the parameter probability of breaking the relationship, the variable engaged? is again set to false for both agents.

Get influenced:

Regardless of whether the agent is engaged or not, it will be the subject of social influence. If the agent is engaged, the agent's trait will become equal to the value of the trait of its partner, weighted by the value of parameter weight, plus the median of the values of its local neighbors, weighted by one minus weight. When the agent is not engaged, the value of the agent's trait becomes the median of the value of the agent's local relationships.

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2080

The art of self-beliefs. A Boudonian approach to social identity

Fernando Aguiar
IESA-CSIC
faguiar@iesa.csic.es

Received: 19-03-2014
Accepted: 05-06-2014

Abstract

In spite of recognizing that the concept of beliefs is a basic ingredient of our identity, Raymond Boudon neither developed a belief-based theory of social identity nor paid special attention to the concept. This article will attempt to shed some light on why Boudon excluded identity from his work and inquire if his social theory can be used to address the issue. We suggest that it is possible to build the concept of identity on Boudon's model of action by defining identity as a special kind of belief – a belief about oneself or self-belief.

Keywords: cognitive rationality; identity beliefs; *homo economicus*; *homo sociologicus;* Raymond Boudon.

Resumen. *El arte de las autocreencias. Una aproximación boudoniana a la identidad social*

A pesar de que Raymond Boudon reconoce que el concepto de creencia es uno de los ingredientes básicos de nuestra identidad, no desarrolló una teoría de la identidad social basada en las creencias ni prestó especial atención al concepto. En este artículo se intentará arrojar luz sobre los motivos por los que Boudon excluyó de su obra la identidad y se analizará si su teoría social se puede emplear para abordar el asunto. Lo que proponemos es que cabe la posibilidad de construir el concepto de identidad apoyándose en el modelo boudoniano de acción, lo que lleva a definir la identidad como una clase especial de creencia: una creencia sobre uno mismo o autocreencia.

Palabras clave: racionalidad cognitiva; creencias de identidad; *homo economicus*; *homo sociologicus*; Raymond Boudon.

Summary

1. Introduction

The fertile work of Raymond Boudon spans the most diverse fields of sociology, from his empirical work on unequal opportunities to his theoretical work on the nature of social action. Following in the footsteps of the great classics such as Tocqueville, Marx, Weber, and Pareto – whose works he knew well – there are few sociological issues that Boudon did not address. There is, however, one exception: social identity. It is not that Boudon did not dedicate any of his work to this topic (there was no need for him to do so), but the term hardly appears in his most important works, which is surprising given its importance. For example, the concept of social identity is not dealt with in the analytical indexes of *Theories of Social Change* (1986), *The Analysis of Ideology* (1989), *The Art of Self-Persuasion* (1994) or *The Origin of Values* (2000) nor can it be found in most of his articles or in the *Dictionnaire critique de la sociologie* (1982, 2004, written with Françoise Bourricaud).

Not surprisingly, the issue of social identity is also absent in those works that analyze Boudon's contribution. For instance, in her interesting book, Cynthia Lins Hamlin (2002) does not discuss social identity and in the monumental tribute to Boudon edited by Mohamed Cherkaoui and Peter Hamilton (2009), comprising four volumes, only the chapter by Max Haller explores identity, but in a way that has nothing to do with Boudon's work.

Why did the author of *The Analysis of Ideology* ignore the issue of social identity despite it being central to sociology? What are the reasons for this decision? This article will attempt to shed some light on why Boudon excluded this key concept from his work and inquire if his theoretical approach can be used to address the issue of social identity that he chose to overlook. To this end, the following section will present some of the criticisms of the often vague and confusing concept of identity, as well as two attempts to overcome these criticisms – a sociological approach influenced by interactionist theory and the rational-choice reductionist project. After presenting Boudon's model of social action in the third section, we will try to understand why he did not address the question of identity. In the fourth section we make a proposal for what could be a Boudonian approach to identity. Falling back on our previous work (Aguiar and de Francisco, 2009) we defend that social identity is a set of positive and normative collective beliefs individuals have about themselves; beliefs that give

social actors reasons for action. As we are going to see, Boudon's social theory and his anti-Humean theory of action fit well in with this belief-based conception of identity (section five). Conclusions are drawn in the final section.

2. *Homo sociologicus*, *homo economicus* and identity

The literature on social identity is simply overwhelming. The topic of identity, which became "one of the unifying frameworks of the intellectual debate of the nineties" (Jenkins, 1996: 7), has not decreased in importance in the twenty-first century. However, when delving further into the literature we soon come up against numerous conceptual obstacles. To begin with, we must tackle the enormous amount of adjectives that accompany the noun *identity* (personal, sexual, racial, ethnic, political, civic, cultural, social, collective, national, religious, local, and so on) (Somers and Gibson, 1994: 66). To the best of my knowledge, no other relevant category in the social sciences, perhaps with the exception of *role*, is bedecked with such a long cloak of adjectives. Take, for example, the habitual use of categories such as class, status, power, capital, group, action, structure, authority or organization. It seems as if the semantic content of the concept of identity were plunged into obscurity and light can only be shed upon it by resorting to other concepts that serve to define it. Starting from this conceptual inflation, the main criticisms regarding the concept of social identity have been grouped into three categories (Brubaker and Cooper, 2000; Davis, 2006).

a) Conceptual ambiguity

The meanings of "identity" abound in the sociological, political science and psychological literature. Identity can be understand as people's concept of who they are and how they relate to others; biological aspects grounded in a socially constructed meaning of identity (race, gender); identification with national, cultural or linguistic symbols; role-specific understanding and expectations about self; cognitive schemata by which the individual knows the world; the prescriptive representation of political actors; expressivist behavior or non-instrumental modes of action; or the unstable, multiple and fragmented contemporary self, among others (Giddens, 1991). These myriad definitions pose reasonable doubts as to the usefulness of a concept which is at best vague and encompasses multiple definitions, as well as the causal tie between identity and action (MacInnes, 2004).

b) Categories of practice vs. categories of analysis

It is often difficult to distinguish if identity refers to the way in which people understand or see themselves in everyday life and in their social, political or economic practice, or if identity refers to an analytical concept embedded in a theory of social action. In other words, identity is not distinguished so much as a category of social practice but rather as a category of analysis (Brubaker and Cooper, 2000: 4).

c) Social identity does not exhaust personal identity

When identity is defined as a process of social identification, people are considered to be "embodiments of group prototypes rather than as independent individuals" (Davis, 2006: 6). Personal identity thus vanishes in a hypersocialized conception of the individual. But given the notably widespread consensus that people have multiple identities, it is necessary to appeal to individual or personal identity as a reference for the individual who reflexively chooses what he or she wants to be. That is, if social identity involves fitting individual action to a social category that others attribute to us, personal identity would imply fitting the action to the image that one has or wishes to have about oneself. This is a complex process full of ambiguities and incoherencies that can affect identification with others. For this reason, "without some account of the 'bearer' of a set of social identities, saying that an individual identifies with others is largely an empty claim" (Davis, 2006: 9).

The criticisms of hypersocialized versions of identity have, however, been a catalyst for the development of new theories that try to strike a balance between elements that are the product of social context, those that are a product of interaction, and those that are irreducibly personal. In what follows, we will see two prominent examples in this regard, which will be useful later to approach identity in Boudonian terms.

2.1. Homo sociologicus *and identity*

Luis is a 35-year-old black male, the father of one child, tall, married, a member of a football club, Spanish, a trade unionist, a high school teacher, of Guinean descent, an anti-bullfighting activist, a fervent defender of gender equality and outgoing. Some of Luis' traits are biological, others are social; some are a product of chance, others of need; some are the result of deliberate decisions, others of the consequences (foreseen or not) of those decisions; and yet others are culturally determined. Surely all these traits do not encompass all of what Luis is; but we *can* say that Luis is, at least, all that. In reality, any subset of traits that we might select would accurately identify Luis to a greater or lesser degree, although none would encompass all his properties. It should go without saying that what interests us here are Luis' socially shared traits; those that shape his social identity.

However, Luis' identity would not permit us to affirm, for example, that he takes part in a pro human rights movement *because* he is black, or that he participates in a demonstration on May 1st *because* he is a trade unionist, or that he defends his homeland *because* he is Spanish. In fact, there are many black people who do not participate in anti-racism movements, many Spaniards who are not willing to defend their homeland, and many trade unionists that go for a picnic on the 1st of May. Hence, it appears that something more must be added to these identifying traits in order to explain Luis' social action, either as an individual or as a member of a group.

Indeed, this "something more" is composed of three elements whose purpose is to integrate individual and social aspects in a general conception of identity. What we call "self" is a complex mixture of individual and social elements in which we must first distinguish the self-concept: "In general, the self-concept is the set of *meanings* we hold for ourselves when we look at ourselves" (Stets and Burke, 2003: 130). This set of meanings is constructed by observing ourselves, through inferences about ourselves, the behavior of others towards us, and our desires. Part of what we call "self", then, emerges through a process of evaluative self-reflection (Mead's "looking-glass self").

The self-concept, however, is not a passive aspect of people, but is instead subject to self-assessment, the second component of "self". Self-assessment has received much attention in recent years (Cast and Burke, 2002). It is understood as self-esteem in a twofold manner as being competent and able (efficacy-based self-esteem) and feeling valued and accepted (worth-based self-esteem).

Finally, social identity is the third element of a "self" understood as a synthesis of individual and social aspects. Identity is the organization of that "self" into "multiple parts (identities), each of which is tied to aspects of the social structures" (Stets and Burke, 2003: 132). What does it mean, for example, to be a parent, a trade unionist, a colleague or friend? From the perspective of agency, it is, firstly, the ability of people to give content (meaning) to these roles of parent, trade unionist, colleague or friend. Social roles are not cast in molds, but provide a margin of freedom: social agents can interpret and implement them in different ways. This, in turn, implies the possibility of creating new roles, as in the case, for example, of what is known as "new masculinities", that is, new ways of interpreting the social role of "man" – men's engagement in active fatherhood, masculinities embodied in managerial practices, and so on (Anderson, 2009).

From a structural perspective, adopting roles means we are governed by certain norms and rules, since ability to recreate social roles does not mean that we should act with our children as a trade unionist or with union members as a parent. Thus, the answer to the question of what roles mean (both in their agential and structural aspects) is the "content of identities" (Stets and Burke, 2003).

The expression of the social "self" through social roles that acquire their meaning through interaction and contribute to the development of a reflective self-concept is therefore far from the old idea of a rigid, hypersocialized *homo sociologicus*. Our individual Luis, for example, holds egalitarian beliefs about what it means to be a father or a husband, and these beliefs are shared with other parents and husbands, but not all of them. To put it another way, Luis plays the role of father/husband on egalitarian terms. Moreover, very complex elements have intervened in his self-esteem, such as his profession or the fact that his parents were Guinean immigrants. These elements mean that Luis is embedded in a system of roles whose meanings provide him an identity that far from being rigid can be interpreted in strategic and normative terms, among others.

So, is social identity from this perspective a category of analysis or a category of practice? It is both because identities are not only attributed to individuals in the form of labels, but the categorization is related to the meaning the roles have for the individuals themselves (Turner, 1999).

2.2. Homo economicus *and identity*

The most orthodox rational choice has always ignored the question of social identity. For the classical model of rational choice theory, the goal of individuals is to maximize their welfare or utility given their beliefs and preferences. Beliefs are rational if they are based on all the available information. In turn, preferences are rational when they are logically consistent, that is, if they do not contradict each other, if one thing and its opposite are not preferred at the same time, if they are transitive, complete, etc. Under these conditions, it can be assumed that individuals always act *as if* they would try to maximize their interest, and that is the only reason needed to explain social action.

According to this approach, theories of social identity are said to needlessly multiply the reasons for action, that is, they are not parsimonious. The best thing therefore is to adopt a reductionist strategy concerning social identity. In its more orthodox versions of rational choice, that strategy has been radical because it assumes that speaking about identity is equivalent to speaking about the specific interests of a person or a group of people who attempt to maximize their utility. It doesn't matter that Luis is a Spaniard of Guinean origin and a trade unionist; what essentially matters are the preferences he reveals through the action and not a supposedly objective identity (Hardin, 1995: 7).

A second, strategy, which is less orthodox and radical, also holds that identity is reducible to interest, but not only instrumental interest, as there are also expressive interests. According to Morris Fiorina, for example, "expressive factors probably dominate instrumental factors as an explanation of turnout" (1976: 410). What is at issue, then, is to consider the expression of identity – what one is – as another argument of the utility function (Akerlof and Kranton, 2010). For instance, we do not vote for one or another party just to maximize our self-interest, but to express our political identity, as the cost of voting is greater than the benefit to be gained. To continue with our example, Luis does not vote for the Socialist Party because it will benefit him, but because he *is* a socialist. Hence, we must include among Luis' preferences the fact that what he wishes to express is his identity. Only then can the fact that he votes be explained (Schuessler, 2000).

Without the conceptual richness of the interactionist-based sociological perspective, rational choice has managed to include identity in its research program (Calvert, 2000), and although it may have become less parsimonious, it is now more realistic. Both currents, that of sociological tradition and that of economic tradition, have left behind the old evils of the structural-functionalist concept of identity (conceptual ambiguity, the quasi-objective conception of

identity, hypersocialization, and so on). Let us now see if we are capable of understanding why Boudon did not partake in this revival.

3. Boudon's theory of social action and the identity oblivion

Boudon's theory of action may provide us with some clues to understand, firstly, why the French sociologist almost entirely ignored social identity in his work, although it has remained one of the key issues in the social sciences, as we have just seen. Furthermore, this will allow us to ask which conception of identity follows from his theory. Although the answer to these questions is speculative, since Boudon himself never raised these issues systematically, it is worth inquiring into them as they shed light on his theoretical position as well as the concept of identity.

As is known, for Boudon a social phenomenon M is the aggregate product of a set of actions *m* taking place in the situation S (Boudon, 1989):

$$M = M \{m [S (P)]\}$$

Let us look at each of the elements of this equation, which we will subsequently tie in with social identity. The phenomenon M we are interested in is the result of actions –explained in terms of attitudes and beliefs – of a set of individuals *m*. Following in the Weberian tradition, Boudon denied the explanatory value of collective concepts, which he instead understood as the result of individual interaction. For example, we cannot say that the Protestant ethic – to once again use this classic example – is responsible for the rise of capitalism without explaining the actions, attitudes and beliefs (interaction in short) of Protestants. In other words, the phenomenon M is a function of or a variable that depends on the set of individuals that are responsible for the phenomenon, that is, $M=M(m_i, m_j)$, where the subscripts denote the *categories* to which these individuals belong. To avoid being accused of inconsistent social atomism, Boudon ignored individuals "who are responsible for M in their concrete individuality, but we shall classify them in general categories" (Boudon, 1989: 243); categories that may have to do with the sex, gender or social roles of individuals. That is important to understand that the type of methodological individualism Boudon defended is close to what has been termed structural individualism (Hedström and Bearman, 2009: Chapter 1; Boudon and Fillieule, 2004). Indeed, the interaction of a set of individuals *m* gives rise to the phenomenon M in a given situation S that allows these individuals to be categorized. On the one hand, explaining or understanding (*verstehen*) a social phenomenon implies understanding the logic of the situation S, which is determined by a structural factor P –the labor market, for instance, or a set of social norms. On the other hand, however, it is important to examine individual attitudes and beliefs *as understood by the social actors themselves involved in the structural situation S(P).*

In this way, Boudon avoided what in his opinion were two of the evils of sociology: structuralist hyperfunctionalism and the rationalism of the *homo economicus* (Boudon, 2006; Lins Hamlin 2002: 12). In the first case, the social actor disappears under the weight of omnipresent and omnipotent objective structures that leave no room for explaining action in terms of intentions. In the second case, the social structure loses its importance to the sovereign action of rational utility-maximizing individuals. In contrast to the rationalism of rational choice, Boudon defended a Simonian model of subjective rationality (cognitive rationality) (Boudon, 1994; Boudon and Viale, 2000) in which people are not assumed to act as if they were rational and selfish, but that they have and give good, satisficing subjective reasons for behaving the way they do (Boudon, 1994, 2001, 2006).

This brief summary of Boudon's model will help us understand the feasible reasons why the French author ignored the concept of identity and how we can rescue it using his model. First, Boudon explicitly rejected identity as an essentialist theoretical concept that is a product of a hyper-socialized conception of social actors. Thus conceived, people would have rigid identities that would determine their social behavior, which in turn would prevent them from interpreting their action on their own terms, that is, sticking to their own reasons. Identity would therefore be a structural concept that corsets action completely. Consider, for example, the concept of role. The role identity as a parent or high school teacher of our character, Luis, does not explain his action in a corseted way. Rather, as Boudon pointed out, individuals interact within systems of roles that do not override their intentionality, because they retain their ability to make decisions (Boudon, 1981). The beliefs and preferences of individuals are influenced by the role they play (the logic of the situation), but they do not define closed and absolute identities.

Let us think about cultural identity. Luis is a Spaniard of Guinean origin. What is his supposed cultural identity? What values have more weight when defining his social "self", those that his parents gave him or those of the country of his birth? The fact that Luis is Spanish does not mean that he has a set of values that cannot be judged, evaluated, compared and even rejected without Luis ceasing to be who he is. Cultural identity does not "lock" Luis into a set of incommensurable community values:

> "Les notions indéfiniment déclinées aujourd'hui d'«identité culturelle» ou de «cultural embeddedness», qui invitent à voir les sociétés comme communautés ou des réseaux de «communautés» assises sur des systèmes de valeurs incommensurables, tendent à écarter par principe toute possibilité d'appréciation par l'observateur extérieur des pratiques ou des institutions en vigueur ici ou là. Pourtant, nous sentons bien que nous avons un droit à en juger. De façon générale, l'interprétation «postmoderniste» des valeurs apparaît comme en contradiction avec une multitude de faits facilement observables". (Boudon, 2000: 2).

The essentialism of postmodern cultural identity prevents the outside observer from understanding why individuals act, which leads to relativism and violates the neo-Weberian postulate of explanation as understanding: the individual who does not belong to a culture or social category, it is said, cannot understand and therefore cannot explain the social action of members of that culture or category.

Thus, rather than the closed and incommensurable identity of the hypersocialized *homo sociologicus* (functionalist or postmodern), Boudon is closer to the reductionist strategy of orthodox rational choice, which is perhaps one of the reasons why he ignored the concept of identity. However, Boudon's reductionism was very different from that of rational choice. We have seen that the most orthodox rational choice rejects the explanatory utility of the concept of identity and reduces it to individual interests. Boudon, however, would be unable to accept this rationalist reductionism. He could not accept that identity is one more argument in utility functions. This would turn all individuals into irrational seekers of the best expression of their "best" self, just as an employer would try to maximize her profit or a party would try to maximize its number of votes. Arguably, then, Boudon's reductionism has nothing to do with rational choice.

The fact that Boudon did not pay as much attention to identity as to other key issues of sociology therefore has to do, on the one hand, with his outright rejection of functionalism and structuralism and their postmodern heirs. On the other hand, he could not be convinced of reducing the concept as rational choice theory does because it meant not rejecting the rationalist model of the utility-maximizing *homo economicus*. This gives us good reason to think that maybe for him the concept was useless.

However, we think it is not useless. In fact, what we think is that Boudon did not pay attention to the identity revival because his criticism focused on the orthodox versions of functionalism/postmodernism and rational choice, as we have just seen. Had he been interested in those sociological and economic identity theories that tried to overcome orthodox approaches, he might have developed his own theory. Unfortunately, he did not do that in spite of remaining alert to the evolution of sociological theory – but not regarding identity.

In the next section, then, we are going to tentatively see if his formula could permit us to find a way of interpreting identity in Boudonian terms.

4. The possibility of a Boudonian approach to social identity

Social identity can be the dependent variable to be explained or the independent variable which helps to explain other phenomena. In the first case, when the social phenomenon to be explained is identity, I (identity) would be equal to M:

$$I = M = M \{m [S (P)]\}$$

Or what is the same,

$$I = I \{m [S (P)]\}$$

Social identity is explained through the actions *m* of a set of individuals in a situation S: the social action of individuals is what explains their collective identity. This would be the case, for example, of rites of passage, where a series of actions must be performed in order to be admitted to the group to which we want to belong. Here the intentionality of the action is unquestionable: people seek an identity because they want that identity. This version of Boudon's formula has to do, then, with the agential aspect of identity, that is, the aspect in which individuals make or create their own role (identity role).

Let us consider again the case of Luis. Since we are not interested in his concrete individuality but the degree to which he belongs to a general category, we must explain (understand) his actions, which together with the actions of other individuals, shape his social identity. As a trade unionist and school teacher, Luis has a set of positive beliefs about his profession. Luis believes, for example, that teachers earn a low salary, work long hours, lack the means to improve the quality of education and that their profession is not valued highly by society. These beliefs may be true or false, but Luis has good reason to take them into account given the information available to him, the opinion of his peers, his assessment of the professional career of his friends who are not teachers, etc.

In addition to positive beliefs, Luis has normative beliefs about what is just and what is unjust, what is right and what is wrong in the world of work, what is fair and what is not. Positive and normative beliefs, which are not the product of a simple cost-benefit analysis, are of a "trans-subjective" nature (Boudon, 2001: 123), as Luis believes that his reasons are well grounded and he can therefore share them with others. This allows us to say that Luis does X (go on strike, for example) because he *believes* that Y is true, fair, good, and so on (Boudon, 1994: 255).

What then would be the identity, the "I" in the above formula? In the reductionist program we have ascribed to Boudon, beliefs are sufficient to explain social action and identity would therefore be a redundant concept. However, we are not forcing Boudon's model if we add, on an expressivist base, that social action is understood insofar as it reveals or expresses the positive and normative beliefs of social actors, both about the world and themselves.[1] When Luis

1. Boudon's use of "belief" is not always clear, but we think it is implicitly close to that of the analytical philosophy of mind: "Analytic philosophers of mind generally use the term 'belief' to refer to the attitude we have, roughly, whenever we take something to be the case or regard it as true. To believe something, in this sense, needn't involve actively reflecting on it […] Nor does the term "belief", in standard philosophical usage, imply any uncertainty or any extended reflection about the matter in question (as it sometimes does in ordinary English usage). Many of the things we believe, in the relevant sense, are quite mundane" (Schwitzgebel, 2014). On the contrary, normative beliefs cannot be true or false. But both are mental states (Elster, 2007: Ch. 7).

performs action X, he is expressing what he *is*, he is expressing his social identity: a trade unionist and teacher in our example. Harboring the positive belief that situation X is true ("school teachers are paid little") and unfair ("it's not fair that some school teachers are paid so little") and acting in accordance with those beliefs involves developing a positive belief about oneself as a fair person: I (think that I) am a fair person. The notion then that identity is a set of positive and normative beliefs about ourselves that give us good reasons to act can be supported in Boudon's model. The set of actions (m_i, m_j), based on the positive and normative beliefs X, Y, and Z, give rise to a collective identity I.

This interpretation of Boudon's formula in terms of social identity, which he did not do but can be deduced from his model of social action without forcing it, allows us to address the other side of the matter, namely, those situations in which identity is the independent variable:

$$M = M \{i [S (P)]\}$$

Here it is not a question of explaining how identity is formed from a set of actions *m,* but understanding how certain aggregate phenomena M occur (go on strike, make war, vote, make family decisions, etc.) from the expression of identity *i*, which is given. Once identity is understood as a set of positive and normative beliefs, the most varied of social phenomena can be explained. Thus, for example, the joint action of Catholics who, with good subjective reasons, believe in God and believe that they should go to church and go because they think they are the type of people who fulfill their religious duties make the Church stronger.

This tentative Boudonian approach to identity seems to lack the defects that the French author attributed to hypersocialized theories or rational choice. It is neither a deterministic approach that establishes direct causal links between identity and action without attending to the microfoundations of identity, nor an empty maximizing exercise. On the contrary, social identity is a set of collective beliefs (positive and normative) individuals have about themselves; beliefs that give social actors reasons for action (Aguiar and de Francisco, 2009).

5. Boudon's anti-Humean theory of action

We have seen that the identity of the *homo sociologicus* is complex. It is shaped by the social roles of a "self" that conceives itself through its own observation and that of others, and well as its desires, beliefs and self-assessment. The most recent theories of identity do not rely on the classical *homo sociologicus* of functionalism or the orthodox *homo economicus* of rational choice. What is the point then of reducing this complex identity to mere beliefs? Beliefs are a part of the self-concept, but not the only one; desires should also be a prominent part of identity. However, does not a theory that focuses on beliefs impoverish the concept of identity?

As Cynthia Lins Hamlin points out, Boudon's theory of cognitive rationality is not only a theory of action, but "a theory of beliefs (both positive and normative) which draws on the strong reasons that agents have for doing what they do or believing what they believe" (Lins Hamlin, 2002: 2-3). Or to put it another way, beliefs provide reasons for action and that is what helps explain the action from the perspective of the subjects themselves. Beliefs that are positive (which may be true or false) and normative always respond to situations of interaction: "they are instituted at the meeting point of a personal history, personal projects and the situation of the actor" (Boudon and Bourricaud, 1986: 46). That meeting point could be seen as the social identity of an individual, and if the identity can explain action, we have to appeal to the beliefs that explain (permit understanding) that action. This does not mean that affective and volitional aspects are not important to understanding action; Boudon often stresses the importance of emotions and desires or preferences (Boudon, 2001). The implication is that even if they are sufficient, they are not necessary, unlike beliefs, which are necessary and often sufficient. Boudon's theory is thus clearly anti-Humean: beliefs have their own motivational force and the action does not *always* need the desire-belief pair to be explained. If I have a beer in a pub, my action to pay for it is not explained by my desire to pay, but by the normative belief that I must pay (Searle, 2001). But the fact that the action does not *always* need the desire-belief pair to be explained does not mean that it *never* needs it. What is important here to understand Boudon is that these desires are often the dependent variable of an explanation in which beliefs are independent variables: "beliefs can play the role of independent variables, that is to say, they appear as cause rather than effect, not only in individual development but in social change" (Boudon and Bourricaud, 1986: 46). Beliefs are, then, the cause of action and when the action relates to what a person believes he is or believes he should be these beliefs about one's self or identity beliefs explain the action (I vote socialist because *I am* a socialist, for example).[2] On the other hand, when identity beliefs are the dependent variable we have to explain their genesis based on "what we sometimes call, in rather a vague term, social structures" (Boudon and Bourricaud, 1986: 44). In any case, far from impoverishing the concept of identity, Boudon's action theory allows us to reconstruct it on grounds that go beyond the Humean model inasmuch as he gives motivational force to beliefs.

It is doubtful, however, that Boudon would support the interpretation of identity we have presented here. Given Boudon's rejection of all forms of psychologism, he might well consider that the concept of identity involves an unnecessary psychologization of social beliefs. However, it is a path that is worth pursuing to at least overcome two of the problems of his theory of action. The first is the absence of a clear definition of "belief" throughout his

2. Having the desire to vote socialist is explained by the fact of being socialist, not the other way round – I am not socialist because I desire to vote socialist. Being socialist is to share with other persons a set of beliefs about the world and about ourselves.

work. Although here we have implicitly assumed that beliefs are mental states (see footnote 2), this "mentalism" seems to be foreign to the French author's work, although he sometimes succumbs to it. A Boudonian theory of identity beliefs would make it necessary to define the very concept of belief in more detail.

The second problem is that of the "danger of adhocness" (Manzo, 2012: 39), which affects Boudon's version of good, subjective reasons for action. Indeed, one can always resort to a subjective reason to understand social action (workers have good reason to accept exploitation, the members of tribe X have good reason to believe in magic, etc.), meaning that his theory runs the risk of being uninformative: there will always be ad hoc reasons to explain any action. However, the expression of identity can be one of the "human invariants" that allow defining the scope of subjective reasons for action. As Gianluca Manzo argues with regard to the work of Boudon, "the link among social belonging, social identity and actor´s beliefs…can help in building reason-based explanations that lead to fine-grained predictions ex-ante facto" (Manzo, 2012: 45).[3] Identity gives us good reasons to act based mainly on positive and normative beliefs about ourselves. This also gives us good reason to think that perhaps Boudon would have accepted a Boudonian approach to identity.

6. Conclusion

In this paper we have tried to show that it is possible to draw a belief-based definition of social identity from Boudon's theory of social action. We think this is an interesting speculative task to perform because the French sociologist did not address the issue. He did not give us many clues to understand the reason why he rejected one of the key sociological concepts. However, in a first speculative exercise it can be established that Boudon was not interested in the identity question due to the ambiguity of a concept either hypersocialized by orthodox functionalists and postmodernists or reduced to preference maximization by the orthodox rational choice theory. Rejecting both theoretical trends entailed rejecting identity as an analytically useful concept. However, why did Boudon not pay attention to the new developments coming from interactionists and heterodox rational choice theorists, among others? We can only conjecture that it may have had to do with the logic of Boudon´s situation: in France, the cradle of postmodern functionalism, Boudon felt the need to incessantly quarrel with postmodern thinkers. In fact, the only quotation on identity we have found is a criticism of postmodernism.

The second speculative exercise has consisted in posing the following question: Does Boudon's social action formula permit us to address the identity issue? Of course it does. At the same time it helps us to understand some interesting features of his work. First, we have seen that Boudon gives beliefs causal power: Beliefs are causes, not only effects, of social action. Inasmuch as beliefs

3. Manzo, however, does not accept a belief-based definition of social identity (Manzo, 2012).

have motivational force – give us reasons for action – we do not always need a desire to understand individual and social actions. In breaking the Humean belief-desire pair that way we can defend that Boudon´s theory is clearly anti-Humean. Thus, it can be stated that a belief-based conception of identity fits in with his theory of action.

Of course these two speculative exercises are not free from serious problems. On the one hand, there are hermeneutic problems, that is, problems of interpreting Boudon correctly. Taking into account, for instance, that his belief definition is far from clear, it can be contested that we are right in saying the French sociologist is an anti-Humean thinker that gives beliefs motivational force. On the other hand, there are theoretical problems. If we put aside interpretation matters, the very proposal of understanding identity as a set of beliefs about oneself can be still questioned. However, tentative works are not useless if they open new research paths.

Bibliographic references

Aguiar, Fernando; de Francisco, Andrés (2009). "Rational choice, social identity and beliefs about one self". *Philosophy of the Social Sciences,* 39 (4), 547–571. http://dx.doi.org/10.1177/0048393109333631

Akerlof, George; Kranton, Rachel (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton: Princeton University Press.

Anderson, Eric (2009). *Inclusive masculinity: The changing nature of masculinities*. London: Routledge.

Boudon, Raymond (1981). *La lógica de lo social. Introducción al análisis sociológico*. Madrid: Rialp.

— (1989). *The analysis of ideology*. Chicago: The University of Chicago Press.

— (1990). "Subjective rationality and the explanation of social behavior". *Rationality and Society,* 2 (2), 224–228.

— (1994*). The art of self-persuasion*. Cambridge: Polity Press.

— (2000). "Pluralité culturelle et relativisme". *Comprendre*, 1, 311–339.

— (2001). *The origin of values. Sociology and philosophy of beliefs.* London: Transaction Publishers.

— (2006). "*Homo Sociologicus:* Neither a rational nor an irrational idiot". *Papers*, 80, 149–169.

Boudon, Raymond; Bourricaud, Francoise. (1986). *A critical dictionary of sociology*. Chicago: Chicago University Press.

Boudon, Raymond; Viale, Riccardo (2000). "Reasons, cognition and society". *Mind and Society*, 1, 45–56.

Boudon, Raymond; Fillieule, Renaud (2004). *Le méthodes en sociologie.* Paris: PUF.

Brubaker, Rogers; Cooper, Frederick (2000). "Beyond identity". *Theory and Society*, 29, 1–47. http://dx.doi.org/10.1023/A:1007068714468

Calvert, Richard (2000). "Rationality, identity and expression". In: Ira Katznelson and Helen Milner (eds). *Political Science: The State of the Discipline*. New York: Norton.

Cast, Alice; Burke, Peter (2002). "A theory of self-esteem". *Social Forces* 80, 1041–1068.
http://dx.doi.org/10.1353/sof.2002.0003
Davis, John (2006). "Social identity strategies in recent economics". *Journal of Economic Methodology*, 13, 371–390.
http://dx.doi.org/10.1080/13501780600908168
Elster, Jon (2007). *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.
http://dx.doi.org/10.1017/CBO9780511806421
Fiorina, Morris (1976). "The voting decision: Instrumental and expressive aspects". *Journal of Politics*, 38, 390–413.
http://dx.doi.org/10.2307/2129541
Giddens, Anthony (1991). *Modernity and self-identity: Self and society in the late modern age*. Cambridge: Polity Press.
Haller, Max (2009). "Language and identity in the age of globalization". In: Cherkaoui, Mohamed and Hamilton, Peter (eds). *Raymond Boudon, a life in sociology. Essays in Honour of Raymond Boudon.* Oxford: The Bardwell Press.
Hardin, Russell (1995). *One for All*. Princeton: Princeton University Press.
Hedström, Peter; Bearman, Peter (2009). "What is analytical sociology all about? An introductory essay". In: Hedström, Peter and Bearman, Peter (eds). *The Oxford handbook of analytical sociology.* Oxford: Oxford University Press.
Jenkins, Richard (1996). *Social Identity.* London: Routledge.
Lins Hamlin, Cynthia (2002). *Beyond relativism: Raymond Boudon, cognitive rationality and critical realism.* London: Routlege.
MacInnes, John (2004). "The sociology of identity: Social science or social comment?". *British Journal of Sociology*, 55, 531–543.
http://dx.doi.org/10.1111/j.1468-4446.2004.00036.x
Manzo, Gianluca. (2012). "Reason-based explanations and analytical sociology". *European Journal of Social Sciences* 50 (2), 35–66.
Searle, John (2001). "Rationality and action". In: Branquinho, João (ed). *The foundations of cognitive science*. Oxford: Oxford University Press.
Schuessler, Alexander (2000). "Expressive voting". *Rationality and Society*, 12, 87–119.
http://dx.doi.org/10.1177/104346300012001005
Schwitzgebel, Eric (2014). "Belief". *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). Edward N. Zalta (ed.).
http://plato.stanford.edu/archives/spr2014/entries/belief. Retrieved 3 June 2014.
Somers, Margaret; Gibson, Gloria (1994). "Reclaiming the epistemological 'Other': Narrative and the Social Constitution of Identity". In: Calhoun, Craig (ed.). *Social Theory and the Politics of Identity*. Oxford: Blackwell.
Stest, Jan E.; Burke, Peter J. (2003). "A sociological approach to self and identity". In: Leary, Mark R. and Tangney, June (eds.). *Handbook of self and identity*. New York: The Guilford Press.
Turner, J. C. (1999). "Some current issues in research on social identity and self-categorization theories". In: Ellemers, Naomi; Spears, Russell and Doosje, Bertjan (eds). *Social identity: Context, commitment, content*. Oxford: Blackwell.

ISSN 2013-9004 (digital); ISSN 0210-2862 (paper)
http://dx.doi.org/10.5565/rev/papers.2091

Reasons and biological causes. Some reflections on Boudon's Theory of Ordinary Rationality*

Francisco José León Medina
Universitat de Girona
francisco.leon@udg.edu

Received: 22-04-2014
Accepted: 28-05-2014

Abstract

In his effort to provide sociology with a theory of behavior for the microfoundation of social phenomena, Raymond Boudon searched for a theory that could be presented as *general* (i.e., a theory that, given its strength, can be used "by default" in explanations). In this search, Boudon disregarded biological causes and stated that his Theory of Ordinary Rationality was the best choice, since it offers final explanations: when a behavior is explained as a result of beliefs that are grounded on good reasons, we are offering a black-box-free explanation. In this paper, I shall state that there are serious problems in the arguments that Boudon used to dismiss the explanatory strategy based on "biological causes". Secondly, I shall point out that some recent findings of several behavioral sciences constitute a radical questioning of the value of his Theory of Ordinary Rationality, as well as a positive revaluation of Evolutionary Psychology. In light of these findings, we can state that on many occasions, either reasons are systematically biased by biological causes, or these causes cause behavior, thus reasons are mere rationalizations. Therefore, neither the reason-based explanatory strategy nor the biological causes-based explanatory strategy can be used "by default". Given the current state of our knowledge, Evolutionary Psychology cannot stand as a *general* theory of behavior but is better placed to do the job in the future: it will depend on its ability to build models that integrate reasons and biological causes.

Keywords: rationality; rationalism; reasons; evolutionary psychology; adapted mechanisms; biological causes; Raymond Boudon.

* This work has benefited from a MINECO R&D project grant (reference number CSO2012-31401) and from a CONSOLIDER-INGENIO 2010 project grant (reference number CSD 2010-00034).

Resumen. *Razones y causas biológicas. Algunas reflexiones sobre la teoría de la racionalidad ordinaria de Boudon.*

En su esfuerzo por proveer a la sociología de una teoría del comportamiento con la que microfundamentar los fenómenos sociales, Raymond Boudon buscó una teoría que pudiese presentarse como *general* (es decir, una teoría que, dada su fortaleza, pudiera usarse «por defecto» en las explicaciones). En esta búsqueda, Boudon desconsideró las causas biológicas y defendió que su teoría de la racionalidad ordinaria era la mejor elección, pues ofrecía explicaciones finales: cuando una conducta se explica como resultado de creencias que están bien fundamentadas en buenas razones, estamos ofreciendo una explicación libre de «cajas negras». En este artículo, sostendré que existen serios problemas en los argumentos que usó Boudon para descartar la estrategia explicativa basada en las causas biológicas. En segundo lugar, señalaré que algunos hallazgos recientes de varias ciencias del comportamiento suponen un cuestionamiento radical del valor de su teoría de la racionalidad ordinaria así como una reevaluación positiva de la psicología evolucionista. A la luz de estos hallazgos, podemos sostener que, en muchas ocasiones, o bien las razones están sistemáticamente sesgadas por causas biológicas, o bien esas causas causan directamente la conducta, por lo que entonces las razones son meras racionalizaciones. Por lo tanto, ni la estrategia explicativa basada en razones ni la basada en causas biológicas pueden usarse «por defecto». Dado el estado actual de nuestro conocimiento, la psicología evolucionista no puede proponerse como una teoría *general* del comportamiento, pero está mejor situada para hacerlo en el futuro: dependerá de su habilidad para construir modelos que integren razones y causas biológicas.

Palabras clave: racionalidad; razones; racionalismo; psicología evolucionista; mecanismos adaptativos; causas biológicas; Raymond Boudon.

Summary

1. Introduction

Almost a dozen decades after the publication of Durkheim's *Rules of Sociological Method*, most sociologists are still stuck in the foundational error of the discipline: the idea that social facts are *sui generis* and hence irreducible to lower-level facts. Armed with an argument that legitimates a frequently proud ignorance of the developments of the rest of the sciences, most sociologists keep themselves outside of the project that places the hope for a renovation and a better future for the discipline in the microfoundation of social phenomena. Raymond Boudon has certainly been one of the exceptions in this gloomy picture, and his effort to provide sociology with a theory of behavior for the microfoundation of its explanations shined as few others did.

With this project in mind, Boudon addressed the evaluation of the available theories in search for some theory that could be presented as a *general* theory of social behavior. It should be pointed out that the exercise consisting in evaluating theories comparatively is not only a legitimate exercise, but a desirable one. In a normal science, theories do not peacefully coexist, isolated from one another in their ivory towers. Weaker theories surely would interpret the comparative assessment as an aggressive attack on a supposed desirable diversity, but respect for weak theories is not what made science knowledge progress (quite the opposite).

In the case of Boudon, this comparative evaluation is also made with the purpose of identifying a general theory of behavior. Even though Boudon did not explicitly state what he meant by "general theory", one can infer from his writings that he was referring to a theory that, given its strength, can be used "by default" in explanations. For example, if rational choice theory is considered a general theory, then it would make sense to establish the methodological principle of rationality: if a behavior is to be explained, the reasonable thing to do is to start with rationality-based explanations (1998b: 174). Thus, using a type of explanation "by default" would not be a guarantee of success, since other causal forces could be operating, but it would be the least bad of all methodological strategies.

That being said, what is really questionable is the way Boudon makes this comparative assessment. In this paper, I shall try to defend the following arguments. First, I shall state that there are serious problems in the arguments that Boudon uses to dismiss the explanatory strategy based on what he called "biological causes" of behavior. Secondly, I shall point out that some recent findings of several behavioral sciences constitute a radical questioning of the value of his Theory of Ordinary Rationality (TOR from now on), and generally of the rationalist or utilitarian-cognitivist paradigm, as well as a positive revaluation of the Evolutionary Theory of social behavior or Evolutionary Psychology (EP from now on). And lastly, I shall maintain that, in the light of these findings and the current state of our knowledge, neither the reason-based explanatory strategy nor the biological causes-based explanatory strategy can be proposed as a "by default" strategy in the explanation of social behavior, although EP is better placed to do the job. In this sense, the fundamental challenge of the microfoundation strategy of social phenomena will be its ability to generate useful knowledge and criteria to determine which of these two strategies or combination of both is appropriate in each case.

1.1. Evaluating theories of social behavior

In his discussion on the sociology that really matters (2002), Boudon presented a set of criteria for judging what a scientific theory is and how to judge its strength. It seems clear that he used these criteria in his comparative evaluation of different theories of behavior (2006, 2007, 2009). These criteria are

uncontroversial, but could be more accurately developed. This is why we shall present them together with our proposed extension.

First, Boudon proposed the criterion of logical consistency of the propositions that form a theory (2002: 373): a theory cannot house inconsistent propositions. The criterion is indisputable, but it would also be necessary to point out that logical consistency is not a purely internal problem of the theory, so that we could extend the list of Boudon's criteria adding the following one: the propositions of a theory must not conflict with principles and findings that are already well established in other sciences.

Second, Boudon pointed out the criterion of the acceptability of explanatory propositions (2002: 373). There are several reasons for the acceptability of a proposition, but Boudon especially stressed two: propositions must have empirical support and must not include obscure concepts. At least for realists, it is sensible to consider that empirical support must also be on the basis of the principles and assumptions of the theory, at least if we agree that the purpose of the theory is to explain and not only to predict. Therefore, the empirical support for the principles and assumptions of the theory is a criterion that could also be added to the list of those proposed by Boudon.

Thirdly, during his comparative evaluation of theories of behavior, Boudon also used the criterion of the explanatory scope of the theory (2006, 2007): theories can compete in their ability to integrate dispersed empirical results, so we expect that a general theory of behavior would not leave some relevant phenomena unexplained. However, the explanatory scope of a theory is not only an external problem: the process of logical inference from the postulates of the theory must be shown to be fertile in its ability to generate new testable predictions. A fertile theory in this sense is a theory with greater explanatory scope, so that fertility could also be added to the criteria proposed by Boudon.

Therefore, our extended version of the criteria used by Boudon leads us to the following six criteria: a) internal logical consistency, b) external logical consistency (with other disciplines), c) acceptability of its propositions, d) acceptability of its assumptions, e) explanatory scope, and f) fertility. Throughout the paper, we shall use these six criteria for judging Boudon's comparative evaluation of different theories of behavior and for revaluating it in the light of the new findings of several behavioral sciences.

2. Reasons and causes in Boudon's writings

2.1. The Theory of Ordinary Rationality

Boudon was, along with Jon Elster, one of the authors who most acutely addressed the limitations of Rational Choice Theory (RCT from now on). While it is true that replacing an explanation of behavior based on *cultural forces* (such as socialization) for one based on RCT causes an advance in knowledge (see, e.g., Boudon, 2006), several theoretical and empirical reasons led Boudon to argue that RCT could not aspire to be a general theory of social behavior (1998a,

2006). From a theoretical point of view, it seems clear that a) not all actions are instrumental, b) not all instrumental actions are guided by the criterion of utility maximization, and c) RCT does not have much to say about the beliefs, values and objectives on which the action is based. In this sense, RCT has a problem with the fourth of our evaluation criteria (*acceptability of the assumptions*). From an empirical point of view, RCT is incapable of explaining phenomena such as the paradox of voting or several behaviors that are usually observed in the experiments of behavioral economics. RCT, therefore, also has a problem with the fifth evaluation criterion (*the explanatory scope*).

Following the lead of Simon (1982), Boudon grounded his theory in the notion of *subjective rationality*. In objective rationality, the reasons guiding the subject are objectively valid, and therefore there is no mystery as to why the subject perceives them as good reasons. In *subjective rationality*, reasons are not objectively grounded, yet they are perceived as good (Boudon, 1989). Given the demanding conditions to assert that a reason is objectively grounded, it was not difficult to conclude that in most situations the reasons that move us are subjective, and that this is something that our theory of behavior must take into account. In that way, Boudon connects with the Weberian tradition, according to which beliefs and actions can be objectively unfounded and still be understandable (i.e., explainable as a result of reasons perceived as good). Thus, he took on the challenge of building a theory of behavior based on an extended notion of rationality, not on rationality in the strict sense, but rather on reasonableness. Throughout the years, he gave different names to that theory: Cognitivist Model of Rationality (1996), Rational Model in the Broad Sense (2000), Theory of Ordinary Rationality (2009).

TOR seeks to explain the adherence of individuals to "goals, values and representations" (Boudon, 2009: 58). According to this theory, individuals adhere to a goal, value or representation when they perceive it as the consequence of a set of reasons composed of acceptable and compatible elements, and provided that there is no other preferable alternative set of reasons. Boudon called this the *cognitive equilibrium principle* (2012). Thus, the system of reasons cause the acceptance (and the strength of the acceptance) of the individual to that goal, value or representation.

According to Boudon, TOR keeps one of the main advantages of RCT: its final character. When an action is explained as the result of goals, values or representations that are based on good reasons we are also providing a black-box-free explanation. Thus, TOR is presented as a theory that posits that positive or representational beliefs (of the type *X is true*) and normative beliefs (of the type *X is good, fair, legitimate...*) are rational, that is, they are grounded on good reasons (Boudon, 2009). Actions motivated by these beliefs are then rational too.

Especially in *The art of self-persuasion* (1990), but also in other studies (e.g., 1989), and following the lead of Weber, the more empirical Durkheim and some argumentative line of Simmel, Boudon showed that false beliefs, both individual and collective, are also understandable, that is to say, they are the

result of a coherent set of reasons that the individual perceives as acceptable. Undoubtedly, one of the great contributions of Boudon was to note that when we face the existence of false beliefs as scientists, the dead easy recourse to an irrational-based explanation, such as when we "explain" behavior as a result of socialization, offers less satisfactory explanations than those offered by TOR (see, for example, 1989). In fact, Boudon presents this dead easy recourse as part of a "spontaneous sociology" opposed to a scientific sociology (1990: 18). As we shall see, it is questionable that the same argument can be applied to the biological causes-based explanatory strategies.

2.2. Social causes: a demolishing critique

When Boudon developed a typology of theories of behavior (2006), as well as when he developed a typology of theories of values (2001), he classified evolutionary and cultural explanations in the same category. Although both explanations have seemingly little to do with each other, in fact the classification makes sense, since both explanations maintain that mental states can have causes that are unnoticed by the individual, rather than pointing to reasons as TRO does. In the words of Boudon, both theories understand mental states as caused, not as grounded (2001: 32).

Boudon addressed his criticism of *social causes* in his assessment of cultural forces-based explanations. We define *social causes* as those social structures that supposedly shape the individual mental states (Lizón, 2010). Boudon's critique of the explanatory power of these causes is demolishing (see, for example, 1990, 2006). The idea that individual beliefs are mere reflections of collective beliefs, manifested in the individual through socialization, is a surprisingly popular pseudo-explanation, but its fragility and inconsistency is obvious if analyzed in minimal detail. For Boudon, cultural explanations, such as the explanation of LévyBruhl of magical beliefs as a result of a "primitive mentality", are based on cumbersome psychological hypotheses and on *ad hoc* built concepts leading to tautological explanations (the "primitives" confuse verbal associations with causal relationships because they have a primitive mentality, and that mentality consists of a tendency to confuse verbal associations with causal relationships).

Fascinated by the huge diversity of human cultural forms (and to a large extent overestimating it), twentieth-century social science felt compelled to explain certain practices. With the weakness of his conceptual apparatus, the alternative of appealing to the effects of socialization used to generate the false impression of having solved the puzzle. To the question "why do members of culture *x* do *y*?" one could answer "because they have been socialized to do *y*". This apparently deep proposition says practically nothing. The expression "have been socialized to do *y*" is equivalent to the expression "have learned that in culture *x* people do *y*", so that by replacing this proposition for the original, the initial proposal states that "members of culture *x* do *y* because they have learned that members of the culture *x* do *y*". Obviously, the original

question still stands, and rather now the question is twofold: on the one hand, we might wonder about the origin of this cultural practice, and on the other hand, we might wonder about the reasons for the individual's adherence to it (since the mere transmission from one generation to another does not explain its acceptance by the receiving generation: one would need reasons to maintain a belief learned through socialization – Boudon, 2001: 5-6). The problem of circularity in socialization-based explanations points at a problem of the cultural forces theory with the third of our evaluation criteria (*acceptability of the propositions*). This acceptability is further compromised by the constant presence of ill-defined, ambiguous and obscure concepts such as *habitus*, primitive mentality, etc. (Boudon, 2006).

But the problems do not end there. Boudon also notes that these explanations have problems with the first criterion (*internal logical consistency*). For example, in the case of the prevalence of the rule of unanimity in the collective decisions of rural Vietnamese societies, the theory of cultural forces states that in traditional rural areas the individual is subject to the group, and only a unanimous decision can be regarded as legitimized by the group. However, it is not difficult to see that the rule of unanimity is precisely the rule that gives more power to the individual over the group, as unanimity is synonymous with veto power (2006: 153).

Furthermore, cultural theories are weak in generating empirically testable theoretical predictions (sixth criterion). In that sense, Boudon appeals to the uncertainty about the effects of socialization. On the one hand, we know that socialization is not always successful, but the theory does not provide elements to predict when it will not be (e.g., Boudon points to Weber's analysis of the sudden conversion of the Roman civil servants and military officials to Monotheism – 2006: 181). On the other hand, it is known that socialization can have opposite effects: an alcoholic father can either become a role model for his child, driving him to alcoholism, or become a negative model to avoid, leading him to abstemious behavior. Thus, the theory loses its scientific character, since *a posteriori* it is always able to interpret any observable effect as consistent with the proposed cause (or in other words, the theory does not provide tools for its refutation).

This does not mean that there are not *social causes*: socialization obviously exists and has an influence on our beliefs and behavior. But, since cultural theory is affected by so many problems, Boudon rightly concluded that it could not aspire to become a general theory of social behavior (1998a, 2006).

3. Boudon's disdain for the biological causes

In the context of theories of behavior, we define *biological causes* as those neurophysiological processes with a genetically conditioned structure and function, which are activated and modulated by different (material or social) environmental inputs, operate outside the consciousness of the individual,

and have a direct or indirect systematic influence on behavior. In fact, the distinction between *reasons* and *biological causes* is very problematic. Are not reasons neurophysiological processes? Reasons *are* a biological phenomenon. However, we will reserve the term *reasons* to refer to conscious mental representations consisting of arguments for or against a proposition or a set of propositions. It is obvious that these mental representations are the emergent effect of neurophysiological processes, but this is relatively unimportant in the context of this paper. The concept of *biological causes* is reserved here to refer to nonconscious processes operating either on behavior or on mental representations that govern behavior.

Given that there are no doubts that biological causes do exist, the debate for social scientists has focused on their relative importance and on the role they must play in the theory of social behavior. Boudon, and analytical sociologists overall, have played a fundamental role in the erosion of the false belief that social facts are *sui generis*, stating that the microfoundation explanatory strategy should be the proper sociological explanatory strategy, which certainly involves questioning the boundaries of sociology and psychology. But, with some exceptions (Lizón, 2010), the first generation of analytical sociologists stated their preference for intentional explanations of social action, and therefore felt some vertigo when facing the final consequences of the openness to explanations based on biological causes.

As already mentioned, the Boudonian approach to rationality is based on Simon's distinction between objective and subjective rationality (1982). In its definition of subjective rationality, Simon refers to an action that is appropriate to the achievement of given goals within the limits imposed by exogenous (environmental characteristics) and endogenous (characteristics of the organisms) conditions or constraints. Boudon does not seem to develop all the implications of this crucial point of Simon. In his concept of cognitive rationality, he accepts that there are exogenous constraints on what we consider good reasons: different social contexts can result in different sets of reasons being more easily evoked and accepted (2003: 16, 2009: 63). However, he hesitated at how endogenous constraints should enter the model. Faced with this challenge, Boudon moved between questioning the concepts employed in the theory of biological causes and questioning the role to be reserved to these causes.

First, Boudon noted that the theory of "biological forces" had problems with the third of the evaluation criteria (*acceptability of the propositions*). In particular, he pointed out that concepts such as those of bias or risk aversion are just descriptive and *ad hoc* concepts leading to circular explanations. Thus, for example, appealing to the availability heuristic to explain an overestimation of probabilities would provide a tautological explanation: an individual tends to overestimate the probability of an event when it is easily accessible (with known, experienced or easy to remember examples) because he has a tendency to overestimate the probability of an event when it is easily accessible. Now, even though this critique seems solid, it faces a major objection. Consider,

for example, the behavior of some animals (including humans) consisting of preparing the nest (or equivalent) during pregnancy. Biologists have explained this behavior as a consequence of a nesting instinct. If the nesting instinct is defined as a tendency to prepare the nest during pregnancy, does this mean that biologists are offering a circular explanation? Obviously not, provided that it is justifiable to state that the instinct conceivably exists. But this is precisely what evolutionary theory does: to argue that biases (for example) result from a predisposition resulting from an adaptive process. Boudon himself admitted that "It [the notion of bias] could cease to be a mere word if it could be shown that biological evolution, say, has produced a wiring of the brain explaining the bias" (2006:159). But here Boudon confuses the biological and cognitive levels. It is not strictly necessary to identify the neural basis of a cognitive trait to defend its adaptive nature. In fact, the ways to support the plausibility of the existence of a "natural" predisposition are diverse: the neurophysiological basis is one, but also its ontogenetically early appearance, its presence in other primates, its universality in the human species, its functionality for certain adaptive challenges, etc. To the extent that the empirical findings of cognitive psychology have been restated by EP as evidence of adaptive cognitive-behavioral programs (in the next section we will see several examples of this), Boudon's objection is neutralized and the explanatory potential of EP reinforced.

As we said before, Boudon also believes that the acceptability of the propositions of a theory depends on its empirical evidence problems. In that regard, he noted that, in general, evolutionary explanations suggest a phylogenetic conjecture that it is hard to prove (1996: 130). This is a classic critique of EP, but it is generally based on the ignorance of the real heuristic discovery process that this theory uses (see, for example, Machery, forthcoming; Schmitt and Pilcher, 2004).

Second, Boudon attacked the theory of "biological forces" referring to its alleged problems with the first criterion of evaluation (*internal logical consistency*). Thus, for example, he pointed out an alleged contradiction between natural selection and the existence of cognitive biases that systematically lead us to forecast errors (1996: 130). Boudon commits a fundamental error here: either he considers that adaptive designs of the past necessarily have to be adaptive in the present, or that adaptive designs of a context cannot be activated with harmful effects in other contexts. Both possibilities are wrong. As Gigerenzer could see, cognitive biases identified from the work of Kahneman and Tversky are adaptations that take most of the world's regularities. But the world in which our brain evolved into its current form was the Paleolithic, not that of our societies.

And thirdly, Boudon highlighted some problems with the *explanatory scope* of the theory of "biological forces". For example, he noted that this theory cannot explain why in some experiments the answers are so sensitive to changes in the problem formulation (1996: 130). In fact, a main assumption of EP and cognitive psychology is that adaptations are extremely sensitive to contextual cues and, consequently, different cues can trigger very different behavioral

programs (Tversky and Kahneman, 1981). For EP, behavior is extremely context-dependent.

As mentioned, in spite of all these criticisms, Boudon also addressed the question of the role that should be reserved for biological factors in explaining social behavior. His position here is far from clear, but in general, he noted that those forces must play *some* role. Referring to neuroscience, for example, he stated that "it can effectively contribute to the explanation of phenomena of interest to social science" (2009: 112). Surprisingly, the statement was not accompanied by any effort to integrate these contributions into his TOR, probably because his idea of the contribution of these disciplines was wrong. In his text *La racionalidad en las ciencias sociales* (2009) he presented two examples: that of an individual whose optimism was the result of a calcification in his amygdala and that of the acceptance of unfair proposals in an ultimatum game as a result of the neutralization of the activity of the dorsolateral frontal cortex. What is implicit in the text is that neuroscience could explain exotic or strange behaviors that result from the peculiarities of a special brain or from the sectorial paralysis of its normal activity. Thus, Boudon is omitting the real contribution of neuroscience: revealing how everyday actions and decisions of people with normal brains are related to processes that are beyond our consciousness.

Boudon questioned that *biological causes* could be the basis of a general theory of behavior (1996, 2001, 2006, 2007). We shall discuss if he was right or wrong in the last section. The problem is that, based on the errors and the unfounded criticisms presented above, he also ruled out *biological causes* as a key element in the explanation of social behavior. Below, we shall address the implications of this positioning.

3.1. A black box inside the black box

Boudon repeatedly noted that explanations based on the reference to psychological forces (such as those contained in the concepts of *bias* or *frame*) or biological forces (such as those characteristic of sociobiology) are problematic (1998a: 820; 2003: 3). The main reason was that their inclusion was supposed to necessarily derail the final nature of explanations based on ordinary rationality (2009: 116-117). Incorporating notions as *bias* or *module* a black box appeared where initially there was a final explanation, since these concepts relate to elements that are not self-explanatory and whose origin is unknown. Moreover, in many cases such terms would refer to confusing concepts, and Boudon even stated that they were "mere conjectures" (2006: 151) or "mere words" (1998a: 820).

However, Boudon also recognized that various "forces" that are not reasons can affect our beliefs and actions. He did so, for example, when he recognized that a belief can be explained by unconscious mechanisms such as adaptive preferences, or forces as passions (for example, he noted that jealousy can cause the belief in infidelity despite the absence of good reasons to support this

belief) (1990: 4). In *La racionalidad en las ciencias sociales* (2009), he stated that "reasons grounding a belief [...] can be biased under the action of various mechanisms. But adherence to a belief is always the effect of reasons" (2009: 87). Boudon, therefore, was open to recognizing the existence of systematic (biological or not) biases in the reasons grounding our beliefs and actions. Moreover, he acknowledged that there is not a general criterion on the strength of a set of reasons, and all that can be said is that we accept a set when we cannot imagine a better alternative set (2003: 16-17), but this also raises doubts about the ultimate causes of that strength.

The obvious problem for Boudon's position is that, if there are systematic but not studied biases in the persuasion power of a reason, TOR would be assuming a black box in its explanation. How can we assert the existence of systematic biases (on the strength and direction of a reason) the explanation of whose functioning we choose not to consider in our theory, and state at the same time that the main virtue of that theory is the absence of black boxes in its explanations? The inevitable conclusion is that progress in understanding *biological causes* is showing the existence of a black box inside the black box: Boudonian identification of the set of reasons grounding a representation can open the black box that cultural theories and behaviorism blithely assumed, but this is often insufficient to ensure the final character of the explanation. In short, this is not about explanation of behavior being necessarily based either on reasons or on biological causes, but about these two causal forces operating in some combination that 21st-century behavioral science will have to unravel.[1] As we shall see in the next section, EP is offering an evolutionary explanation of our set of adapted cognitive mechanisms, thus letting us go one step further in the process of microfoundation of social phenomena.

4. The challenges of behavioral sciences

As we shall try to argue in this section, the paths of behavioral sciences in the 21st century are inevitably leading us to question the value of TOR as a general theory of behavior. The illusion that a general theory of behavior could do without the so-called "biological forces", despite the recognition of their systematic influence on mental representations and behavior, seems to have its days numbered: the biochemical can no longer be left out of the analysis of the psychosocial. In considering the role of "biological forces", some empirical findings from very different disciplines and research areas are revolutionizing our conceptions of how we perceive, reason, decide, make moral judgments, enjoy the aesthetic, etc. TOR is not only unable to reconcile these results with

1. The distinction between proximate and ultimate causes is a useful analytical tool, but behavioral science cannot settle for an appeal to a supposed difference between two alternative "levels of explanation": an interpretive framework identifying the articulation of both causal forces is required.

the theory, but contradicts them, and is therefore seeing the acceptability of its assumptions and propositions very threatened. Facing TOR, EP is not only providing an interpretive framework to give coherence to all these findings, but is often serving as a generating matrix thereof.

To illustrate this argument, we shall conduct a comparative evaluation of TOR and EP in three different sections. First, we shall address how our understanding of how we process information (how we perceive, think, etc.) has been revolutionized. Second, we shall present some examples of how our explanation of the decision process (especially in economic issues) has been modified. And finally, we shall focus on how our ideas about how we make moral judgments have been challenged. These three fields (information processing, decisions, moral judgments) will serve to illustrate the battle of the two theories. For reasons of space, we cannot address other examples, but this analysis could be extended much further. For example, to assess the ways in which we assess the sex appeal of potential mates and how we choose mates, how we shape our magical and religious beliefs, how we form social hierarchies, etc.

4.1. How we process information

During the 20th century, and under the influence of behaviorism, it was believed that the human being was a blank slate that, under the right stimulus program, could end up showing (within the obvious biological limits) any belief, preference, skill, or behavior. The human being would only bring with itself a general capacity for learning and abstract reasoning, and the rest would be a result of the inputs coming from the social environment.

Despite their differences, both cultural theories and rationality-based theories are part of the Standard Model of Social Sciences (as Tooby and Cosmides called it – 1992). Faithful to this assumption on the human mind, the adherence to beliefs has been understood in TOR as a result of a general reasoning ability: subjects would arrive at different beliefs because they are grounded on different sets of information, but everybody employs the same general intelligence, the same rules of abstract inference. That is how Boudon explains, for example, the "rationality" of the primitive's magical beliefs (see, for example, 1989: 180; 2009: 69), but also scientific beliefs, normative beliefs, and any other type of mental representation. One and the same system of information processing (a general intelligence) would be grounding our good reasons to lend (or not) money to a friend, to morally condemn someone else's behavior, or to judge a potential partner as desirable.

TOR, and in general all the theories based on rationality, are seeing this assumption challenged as a result of the confluence of several disciplines on the same approach: that the human mind, far from being a blank slate, is equipped with a set of psychological modules containing representations and content-specialized processes activated as a result of specific inputs from the environment and prefiguring automatic and predesigned responses (Cosmides and Tooby, 1994; Pinker, 2003; Tooby and Cosmides, 1992). Our mental

architecture has a domain-specific organization. The alleged indifference to the different stimuli is simply not true: the human mind processes different types of information differently, and that process involves many different cognitive-behavioral and partly instinctive, unintentional and nonconscious processes. This conclusion is reached, in fact, having to fight the ideological appeal and the apparent obviousness of the theory of the black slate, and accumulating evidence especially coming (but not only) from neuroscience and cognitive psychology. From a neurophysiological point of view, the evidence that our mind does not process all inputs with the same system comes largely from the study of the effects of neuronal injuries. Localized lesions affect specific functions without overall harm to the general cognitive ability of the individual. For example, there are individuals who maintain their ability to distinguish any two material objects but are unable to distinguish two human faces, two animals of different species, or two fruits. From a cognitive point of view, it has been shown, for instance, that different animals have an instinctive fear of specific predators despite not having seen them in their entire life, or even despite the species having been isolated from them for thousands of years (see, for example, Barrett, 2005). Unfortunately for culturalists and creationists, the human being is not different: a line of research (see, for example, Rackison and Derringer, 2008) has convincingly shown a predisposition to fear of snakes in humans and primates: snakes immediately catch our attention on visual complex arrays, it is easier to induce fear of snakes than induce fear to other objects, and it is more difficult to reverse that fear than the fear of other objects. And beyond this anecdotal and irrelevant predisposition from a sociological point of view, much more relevant evidence for social analysis is being accumulated: our economic choices, our moral judgments and our preferences in mate selection are shaped by specific modules containing inherited predispositions, but the list goes on almost indefinitely. To the dismay of advocates of socialization as a demiurge, a few minutes after birth babies follow stimuli similar to human faces more frequently than other stimuli, and show a difference between men and women, the former showing more interest in mechanical objects and the latter in human faces (Connellan et al., 2000); at two days of life they show a preference for their native language (Moon et al. 1993); at two months they make real social smiles (even if they are blind); at three months they already "know" some basic laws of physics (Baillargeon, 1987; Spelke, 1990); at 9 months they develop without instruction the so-called *joint attention* (using gaze direction of others to set their own) and they start to conceive others as intentional agents; at 18 months (and independently from encouragement and rewards) they show altruistic behavior (Tomasello, 2009), etc. And adults, despite the important role of socialization, also show partly automatic preferences and behaviors when they have to assess the beauty of a landscape (Orians and Heerwagen, 1992) or the artistic value of a work (Dutton, 2009), when they face the challenges of parenthood (see, for example, Gettler et al., 2011, 2012) or the threat of free-riders (Cosmides and Tooby, 1992, 2005), etc. All these predispositions would be nothing without the environment, but the information that comes from it would be nothing

more than an infinite chaos of bits of information if it were not for the existence of innate structuring structures in the human mind: without theories-formation mechanisms there cannot be learning.

Can a serious theory of behavior be grounded in the *tabula rasa* assumption when the idea of the equipotential mind is strongly discredited outside the culturalist stronghold which dominates the social sciences? Probably not. Boudon did not ignore these advances in our knowledge of the mind. In fact, on several occasions he acknowledged the falsity of the theory that states the human indifference to various stimuli, thereby accepting the idea that socialization works with and on innate predispositions (1997: 8, 2001: 77). But he did not develop this argument to its final consequences. From any point of view, TOR is part of the Standard Model of the Social Sciences.

The accumulation of empirical evidence from so many disciplines in favor of the idea of the adapted mind has led to a reversal of the burden of proof: it corresponds to the theory of the blank slate to explain how a general-purpose mind could evolve and produce the effects observed in the empirical work of those disciplines. And it does not seem to be having any success in this work. The consequences are devastating for TOR. A theory of behavior to be used as the basis for the microfoundation of social phenomena cannot contain statements that are inconsistent with those established in other sciences (as the *external logical consistency* criterion states) and cannot be grounded in questionable assumptions about human nature (according to our fourth criteria, the *acceptability of the assumptions*). The idea of not giving any role to biological processes in cognition conflicts with the uncontested evidence on the modularity of our mental architecture. Thus, its assumptions are deemed unrealistic and its propositions on the formation of beliefs are deemed inconsistent with propositions that are well established in other sciences.

Faced with this challenge, EP seems to be a much more solvent theory. From the evolutionary point of view, cognitive modules are conceived as adaptations: if there are specific systems to process specific types of information and containing pre-coded forms of reaction, it is because of their functionality in our past. These systems allowed us to effectively resolve recurring problems affecting our survival and reproduction during the Paleolithic. If, for example, we have a facial recognition module, it is because facial recognition had an adaptive function (to identify our people, remember past interactions, etc.). Our psychological architecture (the integrated set of instincts and general purpose mechanisms) comes from an evolutionary process. The mind is a product of the brain, and there is no special reason why the functional design of this organ has escaped the molding forces of natural selection. Conceiving modules as "designed" by natural selection to solve adaptive problems, EP avoids the problems that TOR had with the second and fourth evaluation criterion.

Moreover, EP has been very capable of empirically substantiating its propositions. Very contrary to the claims of some common critics, who point to a problem with the third of our criteria (*acceptability of the propositions*) because of the lack of empirical evidence, the proposition that a cognitive module is

an adaptation is usually not a mere *just-so story*: EP seeks confirmation of its hypothesis from a surprising variety of sources (see, for example, Schmitt and Pilcher, 2004).

The powerful theoretical framework of EP exceeds that of TOR also in the arena of other evaluation criteria. From the *explanatory scope* point of view, EP (especially because of its modular conception of the mind) appears capable of integrating empirical evidence that TOR could not accommodate in its approach, as we shall see in the following sections. For example, how could TOR explain the asymmetry – and the universality of the asymmetry – between women and men in their mate-choice preferences? EP has successfully done so (Buss et al., 1990). How could TOR explain that in certain social exchange situations evolutionary logic leads us to a logically incorrect but adaptive response? EP has successfully done so (Cosmides and Tooby, 1992, 2005). How could TOR explain why attractive men cooperate less in social exchange while attractiveness does not affect the probability of women cooperation? EP has successfully done so (Takahashi et al., 2006).

Also from the point of view of *fertility*, EP outperforms TOR. Conceiving cognitive modules as adaptations is proving to be a matrix for the generation of novel hypotheses about previously unknown psychological traits that end up becoming part of the explanation for some behaviors, something that TOR cannot claim to be able to do, since it is limited to explaining behavior *a posteriori*. If a psychological mechanism is conceived as an adaptation, that is to say, if we state that it has been shaped by natural selection to perform a specific function, then we can infer some attributes or components (usually referred to as "design features") that the mechanism is logically expected to have. For example, from *error management theory* it could be inferred that women (compared to men) underestimate the levels of romantic commitment that can be inferred from declarations of love. This design feature has been called *commitment skepticism bias*, and its existence has been confirmed by Buss (2000). As noted by Machery (forthcoming), grounding on assumptions about the adaptive nature of a psychological trait, EP is able to infer hypotheses about the existence of psychological capacities, the nature of the process, its development and some situational cues it uses.

4.2. How we make decisions

Simon's transition from objective rationality to subjective rationality was only the beginning of a long process that eventually led to the crisis of the theories that placed reasons as the only relevant causal force over our decisions. At first, the limited power of rationality was recognized, but advances in disciplines such as cognitive psychology and neuroscience took the argument further. It was not only about the existence of cognitive limits in the application of abstract reasoning to particular problems, but about the ubiquity of neurophysiological based cognitive-behavioral programs that, at least in part, precode ways of perceiving, evaluating and deciding.

In this section we present three examples of how the contributions of the behavioral sciences are putting TOR in check. Boudon faced several examples of these investigations (especially those from cognitive psychology – see, for example, 1990) and showed in a relatively acceptable way that they were reinterpretable from TOR. But this is far from being a proof of anything. If, say, the theory *A* is producing a series of results (a_1, a_2, a_3…a_n) that challenge the theory *B*, the exercise showing that a_1 and a_3 are reinterpretable from *B* faces the general problem of falsification: resistance to contrary evidence rather than favorable evidence strengthens the theory. If a_2 is still serving for the falsification of *B*, the exercise is futile. In addition, new empirical findings in favor of *A*'s interpretation of a_1 and a_3 are sufficient to get things back to place and for *A* to claim its superiority over *B*. As I shall try to argue in this section, this is exactly what is happening with TOR: new empirical findings of the behavioral sciences are providing evidence against TOR while EP seems to be in a better position to make sense of them.

Time inconsistency. One of the challenges to RCT that have arisen from these disciplines points to the socalled *time inconsistency*: imminent payments are more valued than future ones. If we get to pick a unit of a good in a month or two units in a month and one day, we will choose the two units waiting for a month and a day, but if we get to choose a unit today or two units tomorrow, many of us would choose a unit today. How could TOR explain time inconsistency? Is it possible to imagine any set of reasons according to which it is different to expect one day today than expecting one day in a month? In the absence of a declared set of reasons, one possibility would be to reconstruct the rationale underlying the decision, presenting behavior *as if* it were the result of this reasoning, but this strategy would lead us to a mere *just-so story*. Therefore, TOR faces in this case a problem with the fifth evaluation criterion (the *explanatory scope*), since it is unable to account for this phenomenon.

EP, however, offers a more satisfactory answer: our behavior is ecologically rational. What moves us is nothing but impatience, and impatience is an evolved mechanism that allows us to manage uncertainty. If there is a possibility that the payment will not be made, and we do not know the likelihood of that possibility, the passing of time can help us to assess it (Sozou, 1998), so that the adaptive response is to choose two units within a month and one day instead of one within a month, but if we have to choose between a unit today and two units tomorrow, we should ensure the profit since we do not know how likely it is that the payment will not be made. Whereas in our evolutionary past the chance that determines access to future resources was presumably high, developing an eager response to such decisions allowed us to profit from regular statistical patterns in the environment, thereby improving our fitness. Note that the decision is not grounded on reasons, but is caused by impatience, and here impatience is ecologically rational.

In the field of the explanation of time inconsistency, the superiority of EP does not only lie in its hypotheses generator matrix offering the explanation presented above, but also in the fact that a set of successfully tested predic-

tions whose results TOR would be unable to interpret have been inferred from that matrix. A first set of inferences are in the field of genetics. If time inconsistency is an adaptation, it necessarily has some support in our DNA. And indeed, there are some variants of genes that correlate with the tendency to show time-inconsistent preferences (Carpenter et al., 2011). Furthermore, by comparing twins a recent study has shown that "delay discounting" has a hereditary component (Anokhin et al., 2011).

Secondly, it would also be evidence for the consideration of time-inconsistency as an evolved cognitive-emotional program that there were a neurological system specifically involved in the phenomenon. Authors like Manuck et al. (2003) and Peters and Büchel (2011) have identified that system. Hariri et al. (2006), for example, have shown that the preference for instant but minor rather than larger and deferred rewards seems to be associated with the ventral striatum activity.

Third, in the study of the function of this neuronal system, the role of hormones is particularly relevant. Time-inconsistent preferences are activated as a result of environmental inputs, but these preferences should have a biochemical support. Here, Kayser et al. (2012) found evidence that dopamine reduces impulsivity in intertemporal choices, showing that hormones play a role in the structure of our temporal preferences.

Finally, another classic source of evidence of the adaptive nature of a psychological trait is primatology. One argument supporting the evolving nature of a trait (though in itself insufficient, like all others) is that it is not unique to humans but shared with our closest relatives. In this regard, it has been experimentally shown that this bias is not unique to humans, although it acquires specific features in each species. Non-human primates are also affected by it (e.g., rhesus monkeys – Hwang et al., 2009).

Therefore, we have evidence that suggests that genes play a role in this type of preferences, that there is a neuronal system involved in the phenomenon, that hormones play a role in this system, and that the trait is already present in other primates. In light of these results, the thesis that time inconsistency can be explained without reference to "biological forces" is untenable. Thus, with regard to time inconsistency, EP outperforms TOR (1) in *fertility*, as it is able to make new predictions (in fact, Boudon never offered evidence of TOR's fertility); (2) *acceptability of the empirical propositions*, as it is able to successfully test those predictions; (3) in *explanatory scope*, as it is able to integrate all these empirical results into a theoretical framework; and (4) to the extent that these results come from different disciplines, also in *external logic consistency*.

Loss aversion. A second example of empirical results that are better resolved by EP than TOR is loss aversion (Kahneman and Tversky, 1984), that is, the preference for avoiding losses rather than obtaining profits. Boudon considered this concept as merely descriptive, and thus the explanations of behavior as resulting therefrom, as circular. Again, this critique of the value of cognitive and evolutionary psychology due to its problems with the third criterion of evaluation (*acceptability of the propositions*) does not take into account some

of the most recent literature on the subject. Contrary to Boudon's assertions, the concept of loss aversion allows us to formulate very clear predictions about certain so far unknown biases, something that seriously questions that the concept has a merely descriptive character. For example, from the concept of loss aversion the existence of another phenomenon has been inferred: the so-called *endowment effect* (Kahneman et al., 1990). TOR does not have this fertility, and moreover it also has difficulties explaining these behaviors: how could ordinary rationality explain the gap between what we would be willing to pay for a product and what we would be willing to receive to sell it (Knetsch, 1989)?

In general, although we shall not go into detail, loss aversion bias has been explained by EP as an adaptation that seeks to maximize the number of offspring (see, for example, Levy, 2010) and initially seeks to maximize the acquisition of food resources (McDermott et al., 2008). If loss aversion is an adapted mechanism, it should be possible to find evidence of its genetic support, its biochemical basis, its neural organization and its relative continuity with non-human primates. And there is evidence in all these directions. Firstly, the variations of some genes are correlated with loss aversion. For example, there is some evidence that the serotonin transporter gene-linked polymorphic region (5-HTTLPR) polymorphism significantly influences performance in a Loss Aversion Task (He et al., 2010). In fact, beyond loss aversion, it seems that, in general, risk tolerance in financial decisions correlates with certain variants of some genes (Dreber et al., 2009; Kuhnen and Chiao, 2009; Zhong et al., 2009). By studying monozygotic and dizygotic twins, the heritability of economic risk preferences has been estimated to be 0.63 (Zyphur et al., 2009). Second, and consistent with the identified genes, the role of serotonin as a hormone that can lead to a reduction of loss aversion has been noted (Litt et al., 2006; Murphy et al., 2009). In general, risk behaviors in financial decisions are associated with the 2D:4D ratio, the ratio between the length of the second and fourth fingers; a ratio that depends on exposure to prenatal testosterone (Garbarino et al., 2011). Other studies have also indicated that risk behaviors in economic investments seem to have a nonlinear u-shaped relationship with endogenous testosterone levels (Stanton et al., 2011). Third, there is a difference in loss aversion between people with and without damage to certain parts of the brain (specifically in the amygdala, the orbitofrontal cortex and the insula, parts of the emotional brain) (De Martino et al., 2010; Shiv et al., 2005), suggesting that such areas perform some function in the phenomenon. And finally, it has been shown that some non-human primates show exactly the same loss aversion behavior (Brosnan et al., 2007; Chen et al., 2006).

Along with all this evidence of the evolutionary nature of this cognitive mechanism, the consideration of loss aversion as an adapted mechanism also allows us to elaborate new hypotheses on the cognitive field itself. EP, for example, has stated the domain-specific character of this mechanism, therefore predicting its variation in different contexts. As shown by Li et al. (2012), loss aversion is accentuated both in men and women when facing challenges in the

domain of self-protection, while it is erased for men facing challenges in the domain of mate selection, as inferred from EP.

In short, in relation to loss aversion, EP outperforms TOR insofar as the former is able to (1) provide an explanation of the phenomenon, (2) infer original predictions, (3) successfully test them, and (4) integrate all these different disciplines resulting in a single interpretive framework.

Social trust. The field of experimental economics, especially in connection with neuroeconomics, is also offering results that challenge the value of TOR as a general theory of behavior. Interestingly, Boudon referred to some of them as evidence of the limited character of RCT, but he could not note how far his theory was also challenged. To argue this point, we focus on a single example: experiments on trust.

In a trust game (Berg et al., 1995), an agent A (investor) receives an amount Y of money from the experimenter and has to send an amount X of money ($0 \leq X \leq Y$) to agent B (trustee). The investor keeps the amount that he does not send to the trustee. The experimenter multiplies X by a factor (for example, he triples it) so that the trustee has $3X$. The trustee must then freely decide how much (Z) he wants to return to the investor ($0 \leq Z \leq 3X$). So, the investor must decide whether to look for his own interest setting $X=0$, or to trust the trustee setting $X>0$.

As in many other experimental designs, RCT prediction is usually not fulfilled, as investors often transfer a positive amount to trustees. Can TOR explain why investors usually transfer a positive sum? Let's say a subject has decided to transfer 50% of his money to his opponent. Given the conditions of the experimental design (anonymity of the parties, the absence of reputation effects and shadow of the future, etc.), the subject probably has no option but to ground his decision on (1) his belief in the general level of people's trustworthiness, so he can treat his opponent under that criterion, and (2) his belief in the level of frustration he expects to experience if the trustee turns out to be untrustworthy. Both beliefs would be grounded on the past experience of the investor. The decision of how much to transfer would therefore be the result of a combination between trustworthiness expectations and betrayal aversion. For the subject, his decision to transfer 50% would be well grounded on beliefs for which he has good reasons (given the conditions of the experiment, there seems to be no other set of reasons that would justify the decision).

So far, TOR seems to be able to explain something that RCT cannot. However, experiments designed to test the role of oxytocin in these decisions can jeopardize TOR's explanation. In those designs all participants inhaled a product whose nature was unknown to them: half of them (the control) inhaled an innocuous substance and the other half (the treatment) inhaled a dose of oxytocin. The results indicate that there is a significant difference between the transfers of the two groups, being higher in the treatment (Baumgartner et al., 2008; Fehr et al., 2005; Kéri and Kiss, 2011; Kosfeld et al., 2005; Van Ijzendoorn and Bakermans-Kranenburg, 2012).

Now let's say that the same subject[2] who decided to transfer 50% of his money in the classic game (equivalent to the control condition in the oxytocin experiment) decides to transfer 70% to his opponent when subjected to the inhalation of the hormone. For the subject, his decision would, in both cases, be grounded on his belief in the general level of people's trustworthiness and his expectation about the level of frustration he would experience in the event of being betrayed. However, under the influence of oxytocin, the same evaluation leads to a different decision. Indeed, experiments show that the belief in other's trustworthiness is not altered between control and treatment, so that the most plausible hypothesis is that oxytocin affects betrayal aversion. When individuals are told to interact with a randomly acting machine, they do not modify their behavior despite oxytocin, which also reinforces the hypothesis of betrayal aversion (Elster, 2007).

Obviously, it is important to point out that what these experiments are showing is not that people become more trusting under the influence of oxytocin, but that *oxytocin levels affect trust levels*. It is not about 70% being a result of inhaling the hormone and 50% being the result of a decision process "free" of biochemical influences. This hormone is naturally produced in all of us,[3] and therefore it is logical to assume that what the treatment is doing is to increase its presence. The logical conclusion of the experiment is that our betrayal aversion in situations that require interpersonal trust is always influenced by oxytocin. As with the 50% transfer, TOR would explain the 70% transfer as the effect of good reasons: the individual believes that he will experience a low level of frustration in the event of being betrayed, which justifies a generous transfer. However, it seems clear that the explanation is inadequate if it simply refers to the belief system that the subject mentioned, since the good reasons grounding those beliefs are actually always biased by biochemical processes acting beyond the subject's awareness.

These experimental results turn the TOR explanation into a black-box explanation. If sets of reasons are not judged solely on their internal properties (consistency, acceptability, etc.) as stated by TOR, but are systematically affected by "forces" that we had not contemplated, the explanation of behavior as a result of a set of reason ceases to be a final explanation: a more fine-grain theory is needed.

Neuroscience and EP provide us the tools needed to open the black box inside the black box. The proposed mechanism is the following: oxytocin is a hormone that inhibits the amygdala, which is a center that is responsible for

2. Obviously, a subject cannot be exposed to the control *and* the treatment. Therefore, experiments analyze the difference in the average responses in the dependent variable between control and treatment groups. However, for the sake of clarity, we present the analysis of these results as if this problem of causal inference were not the case.
3. In reality, the degree of the oxytocin effect depends on its receptors, and the variability on those receptors depends on our genetic information. In the experimental deign, however, the random assignment of subjects to the control and the treatment group ensures an initial equivalence that allow us to test the average impact of a specific dose of the hormone.

emotional reactions, including fear, so that the hormone inhibits social fear, that is, aversion to being betrayed or exploited: it simply makes us more indifferent to the possibility that others do not honor our trusting behavior towards them. And what do the oxytocin levels which we are exposed to depend on? As could not be otherwise, they depend on the environment and genes. On one hand, experiences have an impact on the oxytocin level. Oxytocin levels lead us to a more trusting behavior, but at the same time, being a trustee also increases the levels of this hormone (Zak et al., 2005), leading us to behave as trustworthy. On the other hand, genes also play a role. An already identified gene encoding the protein OXTR, which is an oxytocin receptor, plays a crucial role. Depending on the allele of this gene, our oxytocin levels are higher or lower. Research correlating the three possible alleles of the gene (GG, AG, AA) with different social behaviors are a reality that only the most dogmatic sociologists can ignore (see, for example, Rodrigues et al., 2009; SaphireBernstein et al., 2011; Tabak et al., 2013; Tost et al., 2010; Walum et al., 2012). One of them has already provided evidence that individuals with the GG allele show a more trusting behavior than the rest in a trust game (Krueger et al., 2012).

While TOR has nothing to say about these biochemical processes that bias our perceptions and beliefs, EP offers an interpretative framework capable of integrating all these empirical results. From this theory, trust is interpreted as an adaptation whose function would be to enable cooperation and reciprocity where it is not possible to check the honesty of the other (Dunbar et al., 2007: 122), something that happens very often. Establishing relations of cooperation and reciprocity has obvious advantages for survival and reproduction, so the mechanism that makes this possible can be considered an adaptation. This evolutionary hypothesis allows us to interpret the role of genes and hormones in trust as biochemical processes resulting from natural selection in the environment of our ancestors. In fact, although some authors argue that it is not necessary to identify the biochemical processes that underpin what is proposed as a cognitive adaptation, the fact is that mind-brain unity suggests otherwise, so that the inability to detect a "biochemistry of social trust" would have been a setback for EP (and its presence is non-definitive but important evidence in its favor).

Moreover, the hypothesis is enhanced to the extent that several inferences derived from it have been successfully tested. For example, if trust has a role to play when it is not possible to check the honesty of the other, it is logical to expect that evolution has endowed us with a special sensitivity to scrutinize honesty cues. Since relatives are often trustworthy, phenotypic resemblance may act as one of these cues. DeBruine (2005) designed a trust game in which the investor was shown a picture of the trustee. In the control group a picture of a stranger was shown, while investors in the treatment were shown a photo that mixed the face of a stranger and the investor himself. Members of the treatment group were more likely to trust the trustee, confirming the inference obtained from the evolutionary theory of kin selection. When the phenotypic resemblance is not relevant or is simply absent, we seek other facial cues.

Using different brain imaging techniques, research such as that by Engell et al. (2007), Todorov et al. (2008), and Winston et al. (2002) show that there are specific areas of the brain that are activated to assess trustworthiness, contributing to the idea that trust is a hard-wired mechanism.

4.3. How we make moral judgments

A final example of the challenges that question the validity of TOR as a general theory of behavior comes from the science of the moral, and especially, from the confluence of research on moral philosophy and neuroscience.

Boudon postulated TOR as a theory that could also account for normative beliefs, and therefore moral convictions. For him, both positive and normative representations are always the result of a set of good reasons (2009). In *The moral sense* (1997) he explicitly addressed the question of the existence of innate and universal moral intuitions. In this text and in others, Boudon explicitly accepted the existence of an innate moral sense, but argued that its explanatory power of normative feelings was limited (1997: 9, 2001: 77). Boudon's objections were basically two. First, the theory of an innate moral sense would have difficulties explaining cultural variation. Second, although Boudon recognized that our assessments may be influenced by our human nature, the former generally cannot be deducted from the latter (1997: 9). Therefore, the criticism focuses on the fifth of the evaluative criteria (the *explanatory scope* of the theory). In the lines that follow I shall present some approaches to the science of morality that not only provide solid counterarguments to these objections, but also pose serious problems for TOR (problems that EP have no difficulty in solving).

That evolution has a role in the moral is something that Darwin himself had warned of (1871). However, the research that has shaped the evolutionary perspective of our moral sense is relatively recent. In an influential paper, Steven Pinker summarized this perspective (2008). For Pinker, our biological equipment incorporates a "moral switcher" that, when activated, leads to a special kind of reasoning (if you can call it that); a reasoning other than the one employed to determine if we like something, we are interested in it, etc. The rules that guide this mind-set are comparable to those of Chomsky's universal grammar: they are universal and they structure our moral intuitions in a way that goes unnoticed.

Brown's list of human universals (1991) includes a considerable number of aspects that can be considered typical of the moral: prohibitions such as incest, rape or violence, feelings such as shame, promotion of generous behavior and evil punishing, the distinction between good and evil, etc. Following this line, Haidt and colleagues noted that all cultures considered immoral things like hurting others, inequity, the lack of loyalty to the community, the lack of respect for authority and impurity (Haidt and Graham, 2007a; Haidt and Joseph, 2004). For them, these five principles are considered the ultimate psychological basis of all moral rules. Their universality is already evidence in

favor of its innate character, but so is that 1) all (except for purity) have some continuity in the behavior of other primates (de Waal, 1996), 2) an evolutionary history has been proposed for all of them (for a summary, see Haidt and Kesebir, 2010), and 3) some moral behavior and a distinction between moral rules and social conventions appear ontogenetically early (Tomasello, 2009; Turiel, 1983).

Boudon did not ignore this evidence, he simply pointed out (in his first objection) that the moral instinct could not account for the cultural diversity of moral conceptions. The error in this argument is clear: the moral instinct is universal, but apart from some behaviors that inevitably fall under the domain of morality (for example, rape or murder), others may moralize or amoralize depending on local, cultural processes (for example, in our societies tobacco consumption in public areas is no longer evaluated on pragmatic or instrumental criteria but it became moralized). Furthermore, the relative importance given to each of these principles varies between cultures and even between subcultures of a culture (Haidt and Graham, 2007b; Pinker, 2008). Boudon confused here the existence of a universal biological equipment to the defense of uniformity or universality of behavior; an argument that is clearly a *non sequitur*: at least part of our cultural conceptions are "evoked", that is, resulting from different inputs operating over a universal mental architecture.

In fact, EP has no problem in explaining cultural diversity as a result of universal predispositions (for a distinction between evoked and transmitted culture, see Tooby and Cosmides, 1992). For example, Boudon (1998) recalls the case of Madame de Sévigné, who in the seventeenth century wrote his daughter telling her how much she enjoyed attending a public execution. Today no one would admit to enjoying a public execution, says Boudon, and a "naturalistic" theory could not explain this cultural shift in what we consider moral. However, it is clear that in the case of capital punishment different moral intuitions conflict with each other: on the one hand, not hurting, and on the other, punishing the evil and being loyal to your people. The relative importance given to each of these principles varies between cultures, so understanding variation in moral judgments is not impossible from a "naturalistic" theory. What the "naturalistic theory" sustains (among many other things, as we shall discuss below) is that a) neither then nor now a normal person might consider moral the execution of an innocent, precious and prestigious member of the community (cultural variation has limits), and b) the effort to study the ways in which local variations in inputs that operate on mental architecture produce universal cultural diversity is or can be part of this theory.

The second objection that Boudon pointed out was that although our assessments may be influenced by our human nature, they cannot be deduced from it. The question, again, is whether these assessments can be understood without taking into account those influences. Our position is that they cannot. As we argued above, the recognition of systematic influences on human representations that we choose not to analyze constitutes an explicit waiver to developing a final theory, that is, an acceptance of a black box. Being universal

moral instincts the psychological foundation of all belief or moral judgment, no theory of moral beliefs can seriously do without their consideration. Boudon tried to do so, but to unravel the reasons that support normative judgments he was doomed to employ the fiction of the impartial spectator, who is able to put aside his interests and emotions and base his judgments only in the *common sense* (for example, 2009: 88). Boudon did not seem to wonder about the origins or ultimate foundation of that common sense, something that would have undoubtedly led him to research on the evolutionary basis of human morality. According to these investigations, Boudon's impartial spectator is equipped with innate predispositions to certain moral judgments. These predispositions also determine the type of moral reasoning that we do, and they do it in such a way that the principles of TOR are seriously threatened, as we discuss below.

Automaticity and rationalization. It is particularly important that the moral instincts often lead us to automatic, non-reflective moral judgments. These judgments are in many cases automatic and emotionally charged, and not the result of a conscious and deliberate evaluation. Haidt (2001), for example, conducted an experiment in which subjects were presented the story of a brother and sister who decided to have sex (enjoying it without remorse, making sure that there would be no procreation and keeping it secret). It was a unanimous opinion that the behavior was morally wrong, but people had serious difficulties to argue it: either irrelevant reasons were given (such as the community would feel offended, something impossible as the story clarifies that the sexual encounter was kept secret), or the inability to express the rejection was expressed. This led Haidt to argue that rather than moral reasoning, people make a moral rationalization: unfortunately for TOR, the judgment precedes the reasons.

This interpretation leads to a particularly problematic issue for TOR as a whole, and not just for its explanation of normative beliefs: the pervasive nature of rationalization. Gazzaniga (2011) has provided a solid set of experimental evidence for the existence of a process of the left hemisphere of our brain that he calls "the interpreter", which is responsible for developing coherent *post hoc* explanations of actions and emotions. Neuroscience has indicated that this interpreter "plays" with the perception of time and our own intentions. For example, we can perceive the sequence blow-pain-escape and then explain that the pain we felt led us away, but the truth is that the actual sequence was blow-escape-pain. The interpreter "cannot stand" the idea of our action as caused by something other than desires and beliefs, and our beliefs as being caused by something other than reasons. In the field of moral judgments, the actual and the perceived sequence may not match, the moral emotion preceding the reasons for it, which actually are formulated *a posteriori*.

Unfortunately for TOR, neuroscience is providing strong evidence for this non-reflective, automatic character of moral judgments. Faced with different types of moral dilemmas, individuals experience a conflict between brain areas responsible for emotion (which would be triggered as a result of our moral

intuitions) and areas responsible for logical reasoning. In cases in which the involvement of the subject is colder or distant (like pushing a button involving the death of a person and the salvation of five), the latter take control. In cases where the involvement is more direct (like killing someone with your bare hands to save the lives of five others), the former take control (Greene et al., 2001). This has been confirmed by Koenigs et al. (2007), who show the role of the emotional brain in moral judgments by studying patients with localized brain damage in those areas. Emotional drives, therefore, have a crucial role in much of our moral judgments (Nichols, 2004). In general, these judgments result from an interaction between emotion and cognition (Jeurissen et al., 2014), but some triggers lead to the dominance of one or the other process (see, for example, Hristova et al., 2014).

What can TOR (a theory that aspires to be a *general* theory of behavior) say about these universal moral judgments, their automatic and not reflective character, their dependence on the emotional brain, the blocking of the rational brain and the subsequent rationalizations, the continuity of some moral traits with other primates, and their ontogenetically early appearance? I am afraid it cannot say anything.

An example: altruistic punishment. Some moral judgments mobilized in certain economic decisions also seem to be automatic, visceral. For example, those taking place in the ultimatum game. This game is an experimental design in which two individuals interact anonymously. An individual A (the proposer) has to make a proposition on how to share a certain amount (say € 100) with the individual B (the responder). If B accepts the proposal, the division becomes effective, but if B rejects, both will leave empty handed. The RCT prediction is clear: the proposer will offer to keep € 99 and transfer € 1, and the responder will accept the offer because it would be irrational to reject a positive amount. However, the experimental results show that most of the proposals are between 40% and 50% of the amount to be distributed, and propositions below 20% have a 0.4-0.6 probability of being rejected (Fehr and Schmidt, 2006: 622).

In several texts, Boudon stated that the generally fair proposals are evidence against RCT (see, for example, 1998b:180; 2006:156; 2009:49 and 90), and they certainly are. But Boudon neglects that the behavior of the respondent is equally relevant. Respondents facing unequal proposals usually reject them. Thus, the responder assumes a cost (he waves a positive amount) in what is obviously a punishment to the proposer for his inequity. This behavior has been called "*altruistic punishment*". Although to our knowledge Boudon did not address the interpretation of this behavior, it seems clear that it would be interpreted from TOR as a behavior based on normative reasons such as "X is unfair". Although the proposer is anonymous and the interaction is one-shot, Boudon could argue that the respondent punishes someone who violates a moral principle as equity because the subject observes a moral principle consistent with punishing those who violate a moral principle such as equity, but the explanation here would become circular.

In any case, the stab to TOR comes from the genetic and neurobiological studies showing that there is something more than normative reasons behind altruistic punishment. Boudon knew the experimental results suggesting that the dorsolateral frontal cortex plays a role in altruistic punishment (2009: 112), so if that area is neutralized in a subject receiving a very unequal proposal, the subject still judges the proposal as unfair but he accepts it (i.e., he does not punish the proposer) (Koenigs and Tranel, 2007; van't Wout, 2005). Boudon's conclusion is that RCT is only applicable in those cases when a part of the brain is neutralized. Surprisingly he did not notice that this result also has implications for TOR, since the normal functioning of the dorsolateral frontal cortex (an area of the prefrontal cortex) does not move the subject away from a decision based on instrumental reasons, but away from a decision based on reasons (in general). In that line, and using brain scanning techniques, Sanfey (2004) has shown that rejections are based on visceral disgust: unequal proposals make us feel bad. Against this emotional rejection, punishment feels like a good compensation to us: also by brain scanning, DeQuervain et al. (2004) have shown that people get pleasure from the punishment of norm violators. In other words, we punish because of the negative feelings we experience when we are victims of injustice and because by punishing we experience the pleasure necessary to compensate for those feelings. In fact, subjects reject unequal proposals even when it leads to greater inequality, so that the goal of restoring equity could not explain the punishment (Moll and OliveiraSouza, 2007). As occurs in other situations of moral decision, guts act before reasons. The "moral reasoning" appears *a posteriori* to justify a behavior driven by forces that are not reasons.

Studies on the role of hormones also suggest that altruistic punishment is not driven exclusively by reasons. Research has shown that the probability of rejecting an unfair offer is greater in those with low levels of platelet serotonin (Emanuele et al., 2008) and among men with high testosterone levels (Burnham, 2007). Regarding testosterone, it appears to involve a reduction in the activity of the orbitofrontal cortex (another area of the prefrontal cortex), a brain region responsible for self-regulation and impulse control (Mehta and Beer, 2010), which comes to confirm the visceral, reactive character of this behavior. To the extent that genes play a role in encoding receptors of hormones that are relevant to behavior, genetic studies are also contributing significant evidence. For example, the dopamine D4 receptor (DRD4) gene appears to have a role in the rejection of offers in the ultimatum game (Zhong et al., 2010). In the field of genetics, but by means of twin studies, the heritability of these responses has been estimated at 42% (Wallace et al., 2007).

And again, what can TOR say about the automaticity of these behaviors, their relation to the functioning of certain brain areas, their connection with other hormones and genes responsible for those hormones' receptors, and their relative heritability? Can a theory that aspires to offer final explanations and to be a general theory of behavior ignore all these influences? As a result of these

findings, TOR is negatively affected in several of the criteria for evaluating theories: (1) in the *acceptability of the propositions*, since it establishes reasons as the only causes when empirical evidence points in another direction; (2) in *external logical consistency* and *acceptability of the assumptions*, since it assumes that a capacity for general reasoning applies equally in the formation of positive and normative beliefs when the evidence points to distinguishable cognitive processes; and (3) in *explanatory scope*, since it has nothing to say about the presented findings.

Faced with these problems of TOR, EP is proving to be a fertile matrix from where these evidences arise or to which they can be integrated. From an evolutionary point of view, the role of altruistic punishment appears to be twofold: first, it aims to increase the levels of cooperation (thus making available the adaptive advantages thereof), as the awareness of the existence of this type of behavior can deter defection (Fehr and Gatcher, 2002; Yamagishi, 1986); and second, it reduces the initial adaptive advantage of free-riders (Price, Cosmides and Tooby, 2002). Once this feature has been selected for its adaptive functions, it would remain in us as an instinct, that is, a precoded action tendency (punishing the free-rider) triggered by a disgust/seeking pleasure emotion that has been programmed to be triggered by certain cues (e.g., intentions assessed as hostile or unfair), then taking control of our behavioral reaction. By posing altruistic punishment as a cognitive instinct, the findings of genetic and neurobiological studies are easily integrated into a single framework (when they are not directly inferred from it). Thus, EP is safe from the problems that TOR suffered with several of the theoretical evaluation criteria.

5. Which *general* theory of behavior?

As mentioned above, a general theory of behavior is one that, because of its strength, is entitled to be employed as a "by default" theory in the explanation. The empirical findings of the behavioral sciences that we have discussed so far imply a negative reassessment of the strength of TOR, and especially a rejection of the necessarily final nature of its explanations. According to the idea of "stopping rules" in the microfoundation of social phenomena, sociologists could stop at the level of reasons without accounting for the neurophysiological processes that support them. In my opinion, this is right for many cases of sociological explanation. In these cases, good reasons appear to cause the behavior of individuals. However, the empirical evidence presented in this paper supports the view that in many other occasions, either reasons are systematically biased by biological causes, or these causes cause behavior, thus reasons are mere rationalizations. In either case, a reason-based explanation would be insufficient, and in some of them, wrong. TOR cannot claim the right to be used as a "by default" theory in the explanation of social behavior.

In the comparative evaluation we have made between TOR and EP, the latter is shown to be clearly superior. Can EP appeal to that strength

to stand as a *general* theory of behavior? In short: today probably it cannot, but it could do so in the future. The key will be its ability to accommodate the reason-based explanation in its framework. If EP is able to provide an interpretive framework that clarifies the conditions required for triggering a deliberative route and those required for triggering a more automatic, heuristic route, then choosing this framework "by default" in explaining social behavior would be the least bad alternative. The behavioral sciences of this century will have to work on building models that integrate reasons and biological causes. The evolutionary framework is a serious candidate to do the job.

Bibliographic references

Anokhin, A. P.; Anokhin, S.; Grant, J. D. and Heath, A. C. (2011). "Heritability of delay discounting in adolescence: a longitudinal twin study". *Behavior genetics*, 41 (2), 175-183. <http://dx.doi.org/10.1007/s10519-010-9384-7>

Baillargeon, R. (1987). "Object permanence in 3½-and 4½-month-old infants". *Developmental psychology*, 23 (5), 655. <http://dx.doi.org/10.1037/0012-1649.23.5.655>

Barrett, H. C. (2005). "Adaptations to Predators and Prey". In: Buss, D. M. (ed.). *The handbook of evolutionary psychology*. John Wiley & Sons.

Baumgartner, T.; Heinrichs, M.; Vonlanthen, A.; Fischbacher, U. and Fehr, E. (2008). "Oxytocin shapes the neural circuitry of trust and trust adaptation in humans". *Neuron*, 58 (4), 639-650. <http://dx.doi.org/10.1016/j.neuron.2008.04.009>

Berg, J.; Dickhaut, J.; McCabe, K. (1995). "Trust, Reciprocity, and Social History". *Games and Economic Behavior*, 10, 122-142. <http://dx.doi.org/10.1006/game.1995.1027>

Boudon, R. (1989). "Subjective rationality and the explanation of behavior". *Rationality and Society*, 1 (2), 173-196. <http://dx.doi.org/10.1177/1043463189001002002>

— (1990). *The art of self-persuasion. The social explanation of false beliefs*. Malden: Polity Press.

— (1996). "The 'Cognitivist Model'. A generalized 'Rational Choice Model'". *Rationality and Society*, 8 (2), 123-150. <http://dx.doi.org/10.1177/104346396008002001>

— (1997). "The moral sense». *International Sociology*, 12 (1), 5-24. <http://dx.doi.org/10.1177/026858097012001001>

— (1998a). "Limitations of Rational Choice Theory". *American Journal of Sociology*, 104 (3), 817-828.

— (1998b). "Social mechanisms without black boxes". In: Hedström, P. and Swedberg, R. (eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge University Press.

— (2000). "Reasons, cognition and society". *Mind & Society*, 1, 41-56. <http://dx.doi.org/10.1007/BF02512228>

— (2001). *The origin of values. Sociology and philosophy of beliefs*, New Jersey: Transactions Publishers.

— (2002). "Sociology that Really Matters". *European Sociological Review*, 18 (3), 371-378.
<http://dx.doi.org/10.1093/esr/18.3.371>
— (2003). "Beyond Rational Choice Theory". *Annual Review of Sociology*, 29, 1-21.
<http://dx.doi.org/10.1146/annurev.soc.29.010202.100213>
— (2006). "*Homo sociologicus*: neither a rational not an irrational idiot". *Papers. Revista de Sociologia*, 80, 149-169.
— (2007) "¿Qué teoría del comportamiento para las ciencias sociales?". *Revista Española de Sociología*, 8, 5-21.
— (2009). *La racionalidad en las ciencias sociales.* Madrid: Nueva Visión.
— (2012). "'Analytical Sociology' and the explanation of beliefs". *Revue Européenne des Sciences Sociales*, 50-2, 7-34.
Brosnan, S. F.; Jones, O. D.; Lambeth, S. P.; Mareno, M. C.; Richardson, A. S. and Schapiro, S. J. (2007). "Endowment effects in chimpanzees". *Current Biology*, 17 (19), 1704-1707.
<http://dx.doi.org/10.1016/j.cub.2007.08.059>
Brown, D. E. (1991). *Human universals*. Philadelphia: Temple University Press.
Burnham, T. C. (2007). "High-testosterone men reject low ultimatum game offers". *Proceedings of the Royal Society B: Biological Sciences*, 274 (1623), 2327-2330.
<http://dx.doi.org/10.1098/rspb.2007.0546>
Buss, D. M. (2000). *The dangerous passion: Why jealousy is as necessary as love and sex*. Free Press.
Buss, D. M. *et al.* [and 50 additional authors]. (1990). "International preferences in selecting mates: A study of 37 societies". *Journal of Cross Cultural Psychology*, 21, 5-47.
<http://dx.doi.org/10.1177/0022022190211001>
Carpenter, J. P.; García, J. R. and Lum, J. K. (2011). "Dopamine receptor genes predict risk preferences, time preferences, and related economic choices". *Journal of Risk and Uncertainty*, 42 (3), 233-261.
<http://dx.doi.org/10.1007/s11166-011-9115-3>
Chen, M. K., Lakshminarayanan, V. and Santos, L. R. (2006). "How basic are behavioral biases? Evidence from capuchin monkey trading behavior". *Journal of Political Economy*, 114 (3), 517-537.
Connellan, J.; Baron-Cohen, S.; Wheelwright, S.; Batki, A. and Ahluwalia, J. (2000). "Sex differences in human neonatal social perception". *Infant Behavior and Development*, 23 (1), 113-118.
<http://dx.doi.org/10.1016/S0163-6383(00)00032-1>
Cosmides, L. and Tooby, J. (1992). "Cognitive Adaptations for Social Exchange". In: Barkow, J; Cosmides, L.; Tooby, J. (eds.), *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press, 163-228.
— (1994). "Origins of domain-specificity: The evolution of functional organization". In: Hirschfeld, L. and Gelman, S. (eds.), *Mapping the Mind: Domain-specificity in cognition and culture.* New York: Cambridge University Press.
— (2005). "Neurocognitive adaptations designed for social exchange". In: Buss, D. (ed.) *The handbook of evolutionary psychology*, Wiley, 584-627.
Darwin, C. (1871). *The descent of man*. D. Appleton and Company.
DeBruine, L. M. (2005). "Trustworthy but not lust-worthy: Context-specific effects of facial resemblance". *Proceedings of the Royal Society B: Biological Sciences*, 272 (1566), 919-922.
<http://dx.doi.org/10.1098/rspb.2004.3003>

De Quervain, D. J. F.; Fischbacher, U.; Treyer, V.; Schellhammer, M.; Schnyder, U.; Buck, A. and Fehr, E. (2004). "The neural basis of altruistic punishment". *Science* 27 August 2004: 305 (5688), 1254-1258. <http://dx.doi.org/10.1126/science.1100735>

De Martino, B.; Camerer, C. F. and Adolphs, R. (2010). "Amygdala damage eliminates monetary loss aversion". *Proceedings of the National Academy of Sciences*, 107 (8), 3788-3792. <http://dx.doi.org/10.1073/pnas.0910230107>

De Waal, F. B. M. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge: Harvard University Press.

Dreber, A.; Apicella, C. L.; Eisenberg, D. T.; García, J. R.; Zamore, R. S.; Lum, J. K. and Campbell, B. (2009). "The 7R polymorphism in the dopamine receptor D_4 gene (*DRD4*) is associated with financial risk taking in men". *Evolution and Human Behavior*, 30(2), 85-92. <http://dx.doi.org/10.1016/j.evolhumbehav.2008.11.001>

Dunbar, R; Barrett, L. and Lycett, J. (2007). *Evolutionary psychology. A beginner's guide*. Oxford: Oneworld Publications.

Dutton, D. (2009). *The art instinct: beauty, pleasure, & human evolution*. Oxford: Oxford University Press.

Elster, J. (2007). *Explaining social behavior: more nuts and bolts for the social science*. New York: Cambridge University Press.

Emanuele, E.; Brondino, N.; Bertona, M.; Re, S. and Geroldi, D. (2008). "Relationship between platelet serotonin content and rejections of unfair offers in the ultimatum game". *Neuroscience Letters*, 437 (2), 158-161. <http://dx.doi.org/10.1016/j.neulet.2008.04.006>

Engell, A. D.; Haxby, J. V. and Todorov, A. (2007). "Implicit trustworthiness decisions: automatic coding of face properties in the human amygdale". *Journal of Cognitive Neuroscience*, 19 (9), 1508-1519. <http://dx.doi.org/10.1162/jocn.2007.19.9.1508>

Fehr, E.; Fischbacher, U. and Kosfeld, M. (2005). "Neuroeconomic foundations of trust and social preferences: initial evidence". *American Economic Review*, 346-351.

Fehr, E. and Gächter, S. (2002). "Altruistic punishment in humans". *Nature*, 415 (6868), 137-140. <http://dx.doi.org/10.1038/415137a>

Fehr, E. and Schmidt, K. M. (2006) "The economics of fairness, reciprocity and altruism: experimental evidence and new theories". In: Kolm, S. C. and Ythier, J. M. (eds.) *Handbook of the Economics of Giving, Altruism and Reciprocity.* Vol. 1. Amsterdam: Elsevier.

Garbarino, E.; Slonim, R. and Sydnor, J. (2011). "Digit ratios (2D: 4D) as predictors of risky decision making for both sexes". *Journal of Risk and Uncertainty*, 42 (1), 1-26. <http://dx.doi.org/10.1007/s11166-010-9109-6>

Gazzaniga, M. S. (2011). ¿Quién manda aquí? El libre albedrío y la ciencia del cerebro. Barcelona: Paidós.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

Gettler, L. T.; McDade, T. W.; Feranil, A. B. and Kuzawa, C. W. (2011). "Longitudinal evidence that fatherhood decreases testosterone in human males". *Proceedings of the National Academy of Sciences*, 108 (39), 16194-16199. <http://dx.doi.org/10.1073/pnas.1105403108>

Gettler, L. T.; McKenna, J. J.; McDade, T. W.; Agustin, S. S. and Kuzawa, C. W. (2012). "Does cosleeping contribute to lower testosterone levels in fathers? Evidence from the Philippines". *PloS one*, 7 (9), e41559. <http://dx.doi.org/10.1371/journal.pone.0041559>

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. and Cohen, J. D. (2001). "An fMRI investigation of emotional engagement in moral judgment". *Science*, 293 (5537), 2105-2108. <http://dx.doi.org/10.1126/science.1062872>

Haidt, J. (2001). "The emotional dog and its rational tail: a social intuitionist approach to moral judgment". *Psychological Review*, 108 (4), 814. <http://dx.doi.org/10.1037/0033-295X.108.4.814>

Haidt, J. and Graham, J. (2007a). "Planet of the Durkheimians, where community, authority, and sacredness are foundations of morality". In: Jost; J., Kay; A. C. and Thorisdottir, H. (eds.), *Social and psychological bases of ideology and system justification,* 371-401. New York: Oxford University Press.

— (2007b) "When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize". *Social Justice Research*, 20 (1), 98-116. <http://dx.doi.org/10.1007/s11211-007-0034-z>

Haidt, J. and Kesebir, S. (2010). "Morality". In: Fiske, S. T.; Gilbert, D. T. and Lindzey, G. (eds.). (2010). *Handbook of social psychology*, Wiley.

Haidt, J. and Joseph, C. (2004). "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues". *Daedalus*, 55-66. <http://dx.doi.org/10.1162/0011526042365555>

Hariri, A. R.; Brown, S. M.; Williamson, D. E.; Flory, J. D.; De Wit, H. and Manuck, S. B. (2006). "Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity". *The Journal of Neuroscience*, 26 (51), 13213-13217. <http://dx.doi.org/10.1523/JNEUROSCI.3446-06.2006>

Hristova, E.; Kadreva, V. and Grinberg, M. (2014). "Emotions and Moral Judgment: A Multimodal Analysis". In: Bassis, Simone; Esposito, Anna; Morabito and Francesco Carlo (eds.) *Recent Advances of Neural Network Models and Applications,* 413-421. New York: Springer International Publishing.

He, Q.; Xue, G.; Chen, C. *et al.* (2010). "Serotonin transporter gene-linked polymorphic region (5-HTTLPR) influences decision making under ambiguity and risk in a large Chinese sample". *Neuropharmacology*, 59 (6), 518-526. <http://dx.doi.org/10.1016/j.neuropharm.2010.07.008>

Hwang, J.; Kim, S. and Lee, D. (2009). "Temporal discounting and inter-temporal choice in rhesus monkeys". *Frontiers in behavioral neuroscience*, 3, 9, 1-13. <http://dx.doi.org/10.3389/neuro.08.009.2009>

Jeurissen, D.; Sack, A. T.; Roebroeck, A.; Russ, B. E. and Pascual-Leone, A. (2014). "TMS Affects Moral Judgment, Showing the Role of DLPFC and TPJ in Cognitive and Emotional Processing". *Frontiers in Neuroscience*, 8, 18. <http://dx.doi.org/10.3389/fnins.2014.00018>

Kahneman, D. and Tversky, A. (1984). "Choices, values, and frames". *American Psychologist*, 39 (4), 341-350. <http://dx.doi.org/10.1037/0003-066X.39.4.341>

Kahneman, D.; Knetsch, J. L. and Thaler, R. H. (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem". *Journal of Political Economy*, 98 (6), 1325-1348.

Kayser, A. S.; Allen, D. C.; Navarro-Cebrian, A.; Mitchell, J. M. and Fields, H. L. (2012). "Dopamine, corticostriatal connectivity, and intertemporal choice". *The Journal of Neuroscience*, 32 (27), 9402-9409. <http://dx.doi.org/10.1523/JNEUROSCI.1180-12.2012>

Kéri, S. and Kiss, I. (2011). "Oxytocin response in a trust game and habituation of arousal". *Physiology & Behavior*, 102 (2), 221-224. <http://dx.doi.org/10.1016/j.physbeh.2010.11.011>

Knetsch, J. L. (1989). "The endowment effect and evidence of nonreversible indifference curves". *The American Economic Review*, 79 (5), 1277-1284.

Koenigs, M.; Young, L.; Adolphs, R.; Tranel, D.; Cushman, F.; Hauser, M. and Damasio, A. (2007). "Damage to the prefrontal cortex increases utilitarian moral judgements". *Nature*, 446 (7138), 908-911. <http://dx.doi.org/10.1038/nature05631>

Koenigs, M. and Tranel, D. (2007). "Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game". *The Journal of Neuroscience*, 27 (4), 951-956. <http://dx.doi.org/10.1523/JNEUROSCI.4606-06.2007>

Kosfeld, M.; Heinrichs, M.; Zak, P. J.; Fischbacher, U. and Fehr, E., (2005). "Oxytocin increases trust in humans". *Nature*, 435, 673-676. <http://dx.doi.org/10.1038/nature03701>

Krueger, F.; Parasuraman, R.; Iyengar, V.; Thornburg, M.; Weel, J.; Lin, M. and Lipsky, R. H. (2012). "Oxytocin receptor genetic variation promotes human trust behavior". *Frontiers in Human Neuroscience*, 6. <http://dx.doi.org/10.3389/fnhum.2012.00004>

Kuhnen, C. M. and Chiao, J. Y. (2009). "Genetic Determinants of Financial Risk Taking". *PLoS ONE* 4(2) e4362. <http://dx.doi.org/10.1371/journal.pone.0004362>

Levy, M. S. (2010). "Evolution of Risk Aversion: The "Having Descendants Forever" Approach". Working Paper.

Li, Y. J.; Kenrick, D. T.; Griskevicius, V. and Neuberg, S. L. (2012). "Economic decision biases and fundamental motivations: How mating and self-protection alter loss aversion". *Journal of Personality and Social Psychology*, 102 (3), 550. <http://dx.doi.org/10.1037/a0025844>

Litt, A.; Eliasmith, C. and Thagard, P. (2006). "Why losses loom larger than gains: Modeling neural mechanisms of cognitive-affective interaction". *Proceedings of the twenty-eighth annual meeting of the Cognitive Science Society*, 495-500.

Lizón, A. (2010). «Encrucijadas teóricas en la sociología del siglo xx». *Papers. Revista de Sociologia*, 95 (2), 389-420

Machery, E. (forthcoming). "Discovery and confirmation in evolutionary psychology". In: Prinz, Jesse J. (ed.). *The Oxford Handbook of Philosophy of Psychology*. Oxford University Press.

Manuck, S. B.; Flory, J. D.; Muldoon, M. F. and Ferrell, R. E. (2003). "A neurobiology of intertemporal choice". In: Loewenstein, G.; Read, D. and Baumeister, R. F. (eds.) *Time and decision: Economic and psychological perspectives on intertemporal choice*. New York: Russell Sage Foundation, 139-172.

McDermott, R.; Fowler, J. H. and Smirnov, O. (2008). "On the evolutionary origin of prospect theory preferences". *Journal of Politics*, 70 (2), 335-50. <http://dx.doi.org/10.1017/S0022381608080341>

Mehta, P. H. and Beer, J. (2010). "Neural mechanisms of the testosterone–aggression relation: The role of orbitofrontal cortex". *Journal of Cognitive Neuroscience*, 22 (10), 2357-2368.
<http://dx.doi.org/10.1162/jocn.2009.21389>

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press.

Moll, J. and De Oliveira-Souza, R. (2007). "Moral judgments, emotions and the utilitarian brain". *Trends in cognitive sciences*, 11 (8), 319-321.
<http://dx.doi.org/10.1016/j.tics.2007.06.001>

Moon, C.; Cooper, R. P. and Fifer, W. P. (1993). "Two-day-olds prefer their native language". *Infant behavior and development*, 16 (4), 495-500.
<http://dx.doi.org/10.1016/0163-6383(93)80007-U>

Murphy, S. E.; Longhitano, C.; Ayres, R. E.; Cowen, P. J.; Harmer, C. J. and Rogers, R. D. (2009). "The role of serotonin in nonnormative risky choice: the effects of tryptophan supplements on the 'reflection effect' in healthy adult volunteers". *Journal of Cognitive Neuroscience*, 21 (9), 1709-1719.
<http://dx.doi.org/10.1162/jocn.2009.21122>

Orians, G. H. and Heerwagen, J. H. (1992). "Evolved responses to landscapes». In: Barkow, J. H.; Cosmides, L. E.; Tooby, J. E. *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press.

Peters, J. and Büchel, C. (2011). "The neural mechanisms of inter-temporal decision-making: understanding variability". *Trends in cognitive sciences*, 15 (5), 227-239.
<http://dx.doi.org/10.1016/j.tics.2011.03.002>

Pinker, S. (2003). *The blank slate: The modern denial of human nature*. Londres: Penguin Books.

— (2008). "The moral instinct", *The New York Times Magazine*, 13 January 2008.

Price, M. E.; Cosmides, L. and Tooby, J. (2002). "Punitive sentiment as an anti-free rider psychological device". *Evolution and Human Behavior*, 23 (3), 203-231.
<http://dx.doi.org/10.1016/S1090-5138(01)00093-9>

Rakison, D. H. and Derringer, J. (2008). "Do infants possess an evolved spider-detection mechanism?". *Cognition*, 107 (1), 381-393.
<http://dx.doi.org/10.1016/j.cognition.2007.07.022>

Rodrigues S. M.; Saslow, L. R.; García, N.; John O. P. and Keltner, D. (2009). "Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans". *Proc. Natl. Acad. Sci. U.S.A.* 106 (50): 21437-21441.
<http://dx.doi.org/10.1073/pnas.0909579106>

Sanfey, A. G. (2004). "Neural computations of decision utility". *Trends in Cognitive Sciences*, 8 (12), 519-521.
<http://dx.doi.org/10.1016/j.tics.2004.10.006>

Saphire-Bernstein, S.; Way, B. M.; Kim, H. S.; Sherman, D. K. and Taylor, S. E. (2011). "Oxytocin receptor gene (OXTR) is related to psychological resources". *Proc. Natl. Acad. Sci. U.S.A.,* 108 (37), 15118–22.
<http://dx.doi.org/10.1073/pnas.1113137108>

Schmitt, D. P. and Pilcher, J. J. (2004). "Evaluating evidence of psychological adaptation: How do we know one when we see one?". *Psychological Science*, 15 (10), 643-649.
<http://dx.doi.org/10.1111/j.0956-7976.2004.00734.x>

Shiv, B.; Loewenstein, G.; Bechara, A.; Damasio, H. and Damasio, A. R. (2005). "Investment behavior and the negative side of emotion". *Psychological Science*, 16 (6), 435-439.

Simon, H. (1982). *Models of bounded rationality*. Cambridge: MITT Press.

Sozou, P. (1998). "On hyperbolic discounting and uncertain hazard rates". *Proceedings of the Royal Society B: Biological Sciences*, 265 (1409), 2015-2020. <http://dx.doi.org/10.1098/rspb.1998.0534>

Spelke, E. S. (1990). "Principles of object perception". *Cognitive Science*, 14, 29-56. <http://dx.doi.org/10.1207/s15516709cog1401_3>

Stanton, S. J.; O'Dhaniel, A.; McLaurin, R. E.; Kuhn, C. M.; LaBar, K. S.; Platt, M. L. and Huettel, S. A. (2011). "Low-and high-testosterone individuals exhibit decreased aversion to economic risk". *Psychological Science*, 22 (4), 447-453. <http://dx.doi.org/10.1177/0956797611401752>

Tabak, B. A.; McCullough, M. E.; Carver, C. S.; Pedersen, E. J. and Cuccaro, M. L. (2013). "Variation in oxytocin receptor gene (OXTR) polymorphisms is associated with emotional and behavioral reactions to betrayal". *Social Cognitive and Affective Neuroscience*, nst042. <http://dx.doi.org/10.1093/scan/nst042>

Takahashi, C.; Yamagishi, T.; Tanida, S.; Kiyonari, T. and Kanazawa, S. (2006). "Attractiveness and cooperation in social exchange". *Evolutionary Psychology,* 4, 315-329.

Todorov, A.; Baron, S. G. and Oosterhof, N. N. (2008). "Evaluating face trustworthiness: a model based approach". *Social Cognitive and Affective Neuroscience*, 3 (2), 119-127. <http://dx.doi.org/10.1093/scan/nsn009>

Tomasello, M. (2009). *Why we cooperate*. Cambridge: MIT Press.

Tooby, J. and Cosmides, L. (1992). "The psychological foundations of culture". In: Barkow, J.; Cosmides, L. and Tooby, J. (eds.), *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.

Tost, H.; Kolachana, B.; Hakimi, S.; Lemaitre, H.; Verchinski, B. A.; Mattay, V. S.; Weinberger, D. R. and Meyer-Lindenberg, A. (2010). "A common allele in the oxytocin receptor gene (OXTR) impacts prosocial temperament and human hypothalamic-limbic structure and function". *Proc. Natl. Acad. Sci. U.S.A.,* 107 (31): 13936-13941. <http://dx.doi.org/10.1073/pnas.1003296107>

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Tversky, A. and Kahneman, D. (1981). "The framing of decisions and the psychology of choice". *Science*, 211, 4481, 453-458. <http://dx.doi.org/10.1126/science.7455683>

Van Ijzendoorn, M. H. and Bakermans-Kranenburg, M. J. (2012). "A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group". *Psychoneuroendocrinology*, 37 (3), 438-443. <http://dx.doi.org/10.1016/j.psyneuen.2011.07.008>

Van't Wout, M.; Kahn, R. S.; Sanfey, A. G. and Aleman, A. (2005). "Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making". *Neuroreport*, 16 (16), 1849-1852.

Wallace, B.; Cesarini, D.; Lichtenstein, P. and Johannesson, M. (2007). "Heritability of ultimatum game responder behavior". *Proceedings of the National Academy of Sciences of the United States of America,* 104, 15631-15634. <http://dx.doi.org/10.1073/pnas.0706642104>

Walum, H.; Lichtenstein, P.; Neiderhiser, J. M.; Reiss, D.; Ganiban, J. M.; Spotts, E. L. and Westberg, L. (2012). "Variation in the oxytocin receptor gene is associated with pair-bonding and social behavior". *Biological Psychiatry*, 71 (5), 419-426. <http://dx.doi.org/10.1016/j.biopsych.2011.09.002>

Winston, J. S.; Strange, B. A.; O'Doherty, J. and Dolan, R. J. (2002). "Automatic and intentional brain responses during evaluation of trustworthiness of faces". *Nature Neuroscience*, 5 (3), 277-283. <http://dx.doi.org/10.1038/nn816>

Yamagishi, T. (1986). "The provision of a sanctioning system as a public good". *Journal of Personality and Social Psychology*, 51 (1), 110. <http://dx.doi.org/10.1037/0022-3514.51.1.110>

Zak, P. J.; Kurzban, R. and Matzner, W. T. (2005). "Oxytocin is associated with human trustworthiness". *Hormones and Behavior*, 48 (5), 522-527. <http://dx.doi.org/10.1016/j.yhbeh.2005.07.009>

Zhong, S.; Israel, S.; Xue, H.; Ebstein, R. P. and Chew, S. H. (2009). "Monoamine oxidase A gene (MAOA) associated with attitude towards longshot risks". *PLoS One*, 4 (12), e8516. <http://dx.doi.org/10.1371/journal.pone.0008516>

Zhong, S.; Israel, S.; Shalev, I.; Xue, H.; Ebstein, R. P. and Chew, S. H. (2010). "Dopamine D4 receptor gene associated with fairness preference in ultimatum game". PLoS One, 5 (11), e13765. <http://dx.doi.org/10.1371/journal.pone.0013765>

Zyphur, M. J.; Narayanan, J.; Arvey, R. D. and Alexander, G. J. (2009). "The genetics of economic risk preferences". *Journal of Behavioral Decision Making*, 22 (4), 367-377. <http://dx.doi.org/10.1002/bdm.643>