VOLUME 10
—
JULY 10
2009

Journal of Philosophical Studies

Humana.Mente
Il Pensario della Biblioteca Filosofica

Philosophy of Economics

EDITED BY
MAURO ROSSI & LAURA BERITELLI

QUARTERLY JOURNAL
-
ISSN: 1972-1293

JOURNAL OF PHILOSOPHICAL STUDIES

Biblioteca Filosofica © 2007 - Humana.Mente,
Periodico trimestrale di Filosofia, edito dalla Biblioteca Filosofica -
Sezione Fiorentina della Società Filosofica Italiana,
con sede in via del Parione 7, 50123 Firenze
(c/o la Facoltà di Scienze della Formazione dell'Università degli Studi di Firenze) -
Pubblicazione regolarmente iscritta al Registro Stampa Periodica del Tribunale di Firenze
con numero 5585 dal 18/6/2007.

REDAZIONE - Via del Parione 7, Firenze, presso Biblioteca Filosofica - Facoltà di Scienze della Formazione, Università degli Studi di Firenze

Progetto Grafico: Duccio Manetti
Sito web: www.humana-mente.it
La struttura del sito è realizzata da
Emiliano Mazzetti e Simone Gallorini
Per informazioni e collaborazioni:
redazione@humana-mente.it

Humana.Mente, Issue 10 – July 2009

Introduction

Philosophy of Economics – An Introduction

Mauro Rossi
mauro.rossi@umontreal.ca

The economic science has rarely been immune from contrasts and contradictions.

Originally a branch of moral philosophy, economics emerged as an autonomous field of social scientific enquiry little more than 200 years ago. In the wake of the natural sciences' dramatic success, economists soon embraced a methodology that treated economic phenomena simply as another class of natural phenomena. This methodological stance progressively led to a more and more formal and rigorous analysis, characterised by a growing mathematization of economic models. Since then, economics has not ceased to receive the admiration of other, more recent, social sciences and its methodology has often been considered as an example to imitate by all those who have been impressed by the precision of its formal results.

Nevertheless, the history of economics is far from being a collection of triumphs. Indeed, the economic science's sophisticated formal analysis and its reputation for methodological accuracy often contrasted with poor predictions, explanations and interpretations of real economic events. The systemic crises that periodically shake the world economy cast a dark light on the practical efficacy of the economic science. It appeared as if economics was destined to be a "dismal science" not because of the unwelcome content of its long-term predictions, as Malthus – one of the founder of modern economics – believed, but because of its incapacity to get things right.

Further, these failures offered an easy empirical argument to all those eager to express their dissent against the naturalistic methodology embraced by mainstream economics. While orthodox economists defended their approach by appealing to the yet young state of economic research, it seemed to more heterodox economists that the contrast between theoretical and practical results justified a more serious complaint about the usefulness of the naturalistic approach to economic phenomena. More specifically, it seemed to them that, qua science of human and social behaviour, economics needed a distinctive method of analysis, different from the one employed for studying natural phenomena.

These tendencies are still alive nowadays. However, they somehow take different forms and expressions in the light of a new phenomenon, which has characterised the last two decades, namely, the growing exchange between economics and other disciplines (e.g. psychology, anthropology, neuroscience, ethics), which followed years of reciprocal isolation and mutual disinterest. Indeed, this interdisciplinary exchange has drawn renovated attention on some of the contrasting features that have characterised the history of economics.

On the one hand, the influence that economics exercises on other research areas has assumed a more pronounced and direct form. The appeal of economics is no longer confined to its methodology, but it derives also from the flexibility of its theories and models, which makes them more and more frequently chosen tools for explaining phenomena pertaining to different domains of research. A clear example of this tendency is the development, in recent years, of various accounts trying to explain, for instance, the evolution of moral norms (e.g. Binmore 2005, Bicchieri 2006, McKenzie 2007) and the behaviour of the human brain (e.g. Platt and Glimcher 1999) by means of the extensive employment of game- and decision-theoretic resources.

On the other hand, the results obtained in research areas outside economics (e.g. social psychology, cognitive science and experimental philosophy) have provided further support to the doubts raised by the predictive failures of the economic science. Indeed, the unprecedented availability of empirical studies concerning human brain and behaviour has put into question the core assumptions on which mainstream economics is based, namely, that individuals are self-interested agents and that they are instrumentally rational. A significant example is offered by recent works in psychology and neuroscience (e.g. Camerer and Fehr, 2006) that challenge the traditional assumption according to which the main economic agents, i.e. consumers and firms, act on the basis of selfish motives alone.

The combination of these tendencies produced a seemingly contradictory result: economics seems, at the same time, to conquer and to be conquered by other disciplines. In turn, this interesting phenomenon opens the space for philosophical reflection. Thereby, the philosophy of economics has been emerging as a, more than ever, fascinating and important area of philosophical enquiry. This is the reason why *Humana.mente* has decided to dedicate the current volume to this discipline and to the exploration of some of its liveliest debates. Before illustrating them in more detail, it is worth to briefly illustrate the main areas of analysis with which philosophy of economics is concerned.

First and foremost, the philosophy of economics is concerned with questions about the method adopted for the prediction, explanation and interpretation of economic phenomena. More specifically, this research area investigates the nature of economic theories and models and the role that they play within the analysis of economic phenomena. Second, the philosophy of economics is concerned with the core assumptions and postulates on which economic theory is based. In particular, a great deal of analysis focuses on rational choice theory, broadly considered, that is, as a theory of individual, strategic and collective decision-making. Third, the philosophy of economics examines the links existing, and the mutual contributions, between economics and more normative domains on inquiry, such as political economy and ethics. For instance, key debates in this area are those relative to notions like well-being, freedom, equality and distributive justice.

The current volume of *Humana.mente* explores the most recent tendencies characterising philosophy of economics research by means of a series of articles written by Italian professors, post-doctorate fellows and PhD students. All these articles touch on themes belonging to one of the areas of analysis presented above.

Caterina Marchionni explores the first tendency illustrated above. She argues that a great deal of contemporary economic research can be understood in terms of the role played by the ideal of scientific unification, which she characterises as "the application of the same principles, explanatory schemata, models and tools to the study of phenomena in different domains". Marchionni distinguishes four level at which unification can take place: explanations, theories, fields and disciplines. She then shows how this conceptual framework can help us understand the emergence of new trends in economics research, like the New Economic Geography, and their "imperialistic" effects on neighbouring disciplines.

To a large extent, the explanatory power of economics depends on its ability to tell reliable causal stories about economic phenomena. *Federica Russo* explores the crucial methodological question of how econometric models come to establish that some correlations involving economic variables are indeed causal. Rejecting a strict dichotomy between purely inductivist and purely deductivist approaches, according to which causality is either "inferred from statistical properties of data alone" or directly "given" by the economic theory, Russo defends a mid-way view, according to which the attribution of causality results from a more dynamic interplay of inductive and deductive elements. These elements form a rather rich apparatus,

which is constituted by background knowledge, statistical, extra-statistical and causal assumptions and a hypothetico-deductive methodology.

As Russo shows, the test of well-defined causal hypotheses is an important step within economic theorizing. Often, but not always, testing involves the use of experiments. In his paper, Francesco Guala explores the reasons why the Ultimatum Game (UG) has become a paradigmatic experiment, employed not just in economics, but in other social sciences as well. Guala emphasises two features. First, UG-based experiments possess some virtues (replicability, robustness and disciplinary cohesion), which make them "standardised" experimental designs. Second, the UG can be used as a measurement device that helps us highlight the existence of significant real-world causal relations to which further analysis can then be directed. It is thanks to these features that, according to Guala, the UG "emerged by a process of social selection in experimental game theory, as a robust design that 'tps on' something that seems to matter for us".

Matteo Colombo explores the question how neuroscience can contribute to economics. Moving from Guala's reflection on paradigmatic experiments, Colombo uses the UG as the benchmark for his subsequent analysis. One of the main features of the UG is the contrast between the standard game-theoretic predictions and the observed experimental results. This feature casts more than one doubt on one of the core assumptions underlying game theory, namely, the assumption that individuals are moved simply by self-interest. Colombo examines Cristina Bicchieri's sophisticated attempt to explain the agents' actual behaviour in terms of the notion of 'social norms'. He argues that Bicchieri's account presents both theoretical and practical limitations, which compromise its usefulness as a predictive tool. Colombo then argues that one way to obviate these limitations consists in enriching Bicchieri's model with neurobiologically-informed evidence. This analysis leads Colombo to conclude that neuroscience can contribute to economics by improving its predictions.

Like Colombo, *David Yokum* and *Filippo Rossi* are interested in how disciplines like psychology and neuroscience can combine with economics in the effort of explaining certain problematic features of individual behaviour. Yokum and Rossi focus specifically on charitable donations. By examining public good games, they show that the standard analysis, based on the assumption of self-interested behaviour, leads to predictions that are at odds with the agents' observed behaviour. They propose four different models accounting for the observed anomalies, which leave room for the possession of non-selfish considerations in addition to selfish ones as grounds for behaviour. Such considerations may include feelings of personal satisfaction, or 'warm glow', and preferences for fairness, reciprocity and deservedness. In the last part of their paper, they consider how these models fare with respect of neuroscientific evidence.

While the previous authors focus on the role of self-interest in modelling and explaining individual behaviour, *Giacomo Mollo* examines the other core assumption underlying economic theory, namely, the assumption that individuals are practically (and instrumentally) rational. His goal is not to reject this assumption, but rather to elucidate how we can understand practical rationality. Mollo examines a position in the literature, i.e. cognitivism, according to which we can make sense of the requirements of practical rationality in terms of the requirements of theoretical rationality. After defending cognitivism against several objections, Mollo argues that there is nonetheless one case where cognitivism is unable to explain the failure of practical rationality in terms of a corresponding failure of theoretical rationality. This analysis leads Mollo to conclude that one can be, at best, only a "semi-cognitivist" about practical rationality.

The last three papers deals with issues of relevance for more normative domains of analysis. *Mauro Rossi* considers the problem of interpersonal utility comparisons (IUCs). According to a long tradition in economics, comparing the utilities of different individuals poses insurmountable epistemological difficulties. This is bad news for several areas of analysis (e.g. normative economics, social choice theory, ethics), which are in serious trouble without the possibility of making meaningful IUCs. In his paper, Rossi assesses one promising strategy for solving the problem of IUCs, which focuses on how ordinary people attribute and compare other individuals' mental states. He argues that all the arguments attempting to vindicate this strategy fail. Rossi concludes that the current state of research legitimates, although does not entail, a moderate form of scepticism towards IUCs.

One area of normative debate to which both philosophy and economic analysis have recently brought a significant contribution is the one concerning individual well-being. *Margherita Bottero*'s paper offers an example of how it is possible to fruitfully combine both disciplines in relation to this topic. Bottero's goal is to investigate whether freedom of choice matters for well-being, when the latter is conceived as "life satisfaction" or "happiness". With the help of a meticulous econometric analysis, she shows that freedom of choice is positively correlated with individual's life satisfaction. Bottero notes that the robustness of the results obtained suggests the existence of a causal relation between the two variables. At the same time, she is careful in invoking the need for further research in order to confirm the causality interpretation and clarify some of its features, like the direction of causality.

Volkan Çıdam offers an interesting example of more heterodox research in philosophy of economics. Çıdam focuses on the question of how we can understand Marx's normative critique of capitalism. The question arises because the traditional readings run into a sort of dilemma. Either they adopt Engel's portrait of Marx as a strict positive scientist, in which case Marx's normative critique appears either unintelligible or ideological, or they ground his normative critique in commonsense moral intuitions, which Marx explicitly rejected in all his writings. Çıdam offers an original way out of this dilemma. Revisiting and modifying Luckas's concept of 'reification', he argues that we can better understand Marx's arguments in terms of "the normative ideal of recognition", which sets the standard of his normative critique of capitalism and which thereby motivates the rejection of "all social relations that succumb to a deficient mode of intersubjectivity".

Editoriale

Laura Beritelli
laura.beritelli@humana-mente.it

Mercato mondiale

Quel che adesso ci occorre è una fresa,
un cilindro cimatore, una scheda d'interfaccia,
portati in volo dal Giappone. Ecco qui
tutt'a un tratto souvenir di Timbuctu,
icone, neonati rapiti. Ovunque
ci sono le stesse lamette da barba,
congressisti e api killer.
Circolano autobombe, donne
atterrano dai cataloghi in vendita,
conti bancari si spostano via satellite.
Giungono in volo virus inusitati.
Solo ogni tanto al margine della strada
c'è un mendicante, immobile.

Hans Magnus Enzensberger, *Più leggeri dell'aria. Poesie morali*, Einaudi, Torino, 2001.

Humana.Mente ha sempre avuto interesse alla possibilità di far dialogare la tradizione di pensiero analitica, al cui rigore logico e formale s'ispira, con quella continentale, al cui sguardo critico non può - e non vuole – sottrarsi. Nell'intenzione di presentare i più recenti lavori di filosofia dell'economia in Italia, si è infine realizzata la fortunata occorrenza attraverso cui mostrare le comuni attitudini, le "rassomiglianze di famiglia" tra le due diverse prassi di pensiero come anche, di contro, le difformità che esistono tra di esse. C'è, in effetti, una "critical mass" concettuale su cui l'approccio analitico e quello continentale si scontrano o, meglio, non si incontrano; ma, in definitiva, sembra dipendere dai diversi ordini di valori che orientano le rispettive prospettive di lavoro ed intervento, piuttosto che per una qualche effettiva incompatibilità.

Come ampiamente spiegato nell'editoriale a cura di Mauro Rossi, proponiamo quindi una serie di articoli scientifici, o *papers*, a firma di giovani filosofi dell'economia, che stanno definendo e rendendo grande con il loro lavoro questa innovativa branca del sapere e, con essi, una vasta panoramica delle ricerche *in itinere* nelle più prestigiose università, in particolare quelle del mondo accademico anglosassone; ospitiamo, con grande interesse, un articolo di Francesco Guala, affermato pubblicista e filosofo dell'economia, articolo che l'autore ha tradotto per noi e che, in lingua inglese, è comparso su *Philosophy of Science* (75, 2008, pp. 658 - 669).

A questi lavori associamo una interessante collaborazione con l'Istituto di Studi Sociali di Francoforte. Ringraziamo per aver affrontato assieme a noi questa avventura i giovani filosofi eredi di quella prestigiosissima tradizione, elaborata nel cuore dell'Europa continentale e che tanto ha influenzato la cultura mondiale. Ci ha interessato presentare questi lavori proprio

perché riconosciamo l'importanza che quelle idee rivestono ancora oggi, laddove alcune agiscono ancora e più potentemente "sottotraccia"

Dalla Scuola di Francoforte non ci arrivano dunque lavori di filosofia della scienza economica *tout court*, cui il numero è dedicato; ma un simile accostamento risulta per noi altrettanto fondamentale, perché mostra quanto sia ineludibile l'approccio complessivo, etico e sociale all'economia, indipendentemente che essa sia intesa come scienza sociale o scienza naturale - dibattito questo che si vanifica nel momento in cui è l'economia a dettare il senso del mondo; dacché, come scrive Giuseppe Limoni nella sua introduzione a A. Micocci *The metaphisic of Capitalism* (Lexington Books, Lanham, Maryland, 2008) «l'economia ha oggi acquisito una centralità regia, paragonabile a quella posizione architettonica che un tempo sembrava spettare alla teologia».

Se quindi è vero, con lo stesso autore, che «la scienza economica è oggi [...] scienza della struttura capitalistica», che è la formazione storicamente determinata propria dell'età moderna, allora è di questo che la tradizione continentale si occupa; e se è vero, come prosegue, che «essa non può sottrarsi a tutti i quesiti epistemologici» che ne definiscano e comprovino la scientificità, allora è di questo che si occupa la tradizione analitica. Concludo, citando ancora Limone, che «la scienza economica in quanto descrizione e spiegazione del mondo, non solo descrive e spiega il mondo, perché invece, in quanto creduta scienza, ed in quanto applicata come tecnica, *si fa mondo* diventando di fatto un prodotto *oggettivo* ed *indipendente* del suo modo di pensare». È così che ancora una volta la lucidità della filosofia continentale si manifesta, invitando al contempo all'analisi endogena delle proposizioni della scienza economica cui attendono, per l'appunto, i lavori dei filosofi della scienza e, nel particolare, i filosofi dell'economia.

Suggeriamo quindi come sia possibile assolvere insieme al compito complessivo e più proprio della filosofia, ovvero quello di custodire e vigilare sullo sviluppo dei saperi, unendo - e non tenendo ben separate - le risorse di entrambi gli orientamenti di indagine, sotto l'egida di un agire del pensiero che è già etico e che, essendo intrinseco alla filosofia tutta, appartiene all'operato di entrambi.

È in questo senso che nella monografia *Philosophy of Economics* si parla di etica, di ecologia, di diritto e di tutti gli altri aspetti cui può essere orientato il discorso economico. Se ne parla sia nella colonna dedicata agli economisti, quella delle riletture, o *commentaries*, che propone una selezione di lavori capitali, quelli che hanno fatto la storia dell'epistemologia dell'economia*;* sia nella colonna dedicata ai filosofi dell'economia, quella delle recensioni, o *book reviews*, da cui speriamo sia possibile cominciare a farsi un'idea del perché dell'affermazione di questa nuova disciplina.

ANALOGIE TEMATICHE

Analitici e continentali, nonostante la diversità delle declinazioni, passano entrambi al vaglio di problemi fondamentali in filosofia come quello della libertà dell'uomo, sempre sotteso alle discussioni ma che, esplicitamente, ricorre tanto nell'articolo francofortese del giovane Volkan Cidam, in cui si sottolinea quanto la libertà in Marx fosse una libertà radicale – la riconduce quindi all'idealismo tedesco - e perché sia necessario recuperarla in quell'accezione misconosciuta anche e soprattutto dall'ortodossia marxista; quanto nell'articolo di Margherita Bottero, della Stockholm School of Economics, che si propone di analizzare per la prima volta quanto la libertà condizioni il benessere - e quindi abbia peso nella omonima teoria economica.

Dalla giustapposizione dei due lavori si noterà per esempio quanto l'econometria della Bottero debba scendere a compromessi con quello che lei argutamente rileva essere uno dei due fondamentali problemi cui si è trovata davanti: che la libertà, non è ben chiaro come misurarla. Scrive invece Cidam:

> By 'social mediation' I refer to a form of social relation that plays a constitutive role for subject positions. A constituted subject position denotes a particular attitude towards other subjects and the outside world. This attitude in return depends upon a particular conceptualization of freedom.

La rilettura de *Il Capitale* di Karl Marx, a cura di Martin Sattler, professore emerito (Mannheim) e direttore della Scuola di Heidelberg, ha invece dalla sua parte una eccezionale sagacità: mette infatti il Marx economista fuori mercato e ne sottolinea il linguaggio vago e metaforico, attinto dalle Sacre Scritture. Al contempo, Sattler evidenzia anche l'attualità di quella critica al capitalismo e la sua profeticità: una profeticità sbalorditiva dovuta, probabilmente, al tragico fatto che il capitalismo di oggi non ha attuato alcuna maturazione reale e mostra ancora le stesse contraddizioni operative ed epistemologiche di quello preso in esame nell'Ottocento da Marx. Scrive infatti Sattler: «[...] but the *infancy* of the capitalist system lasted until our days and the elements Marx saw in the *child period* are still with us and they are growing into a super-gigantic scale in our days». Il capitalismo, se osservato nella sua complessità storico-filosofica, risulterebbe quindi cresciuto in modo elefantiaco, ma senza una reale maturazione; e non solo nel senso inteso allora da Marx, ovvero lasciando spazio al comunismo ma, detto con un altro linguaggio partorito dall'idealismo tedesco, quello di Schelling, approfondendosi in vorticosa spirale, a partire dai suoi stessi principii approssimativi, quegli stessi che oggi sono stati apertamente abbandonati tanto dagli economisti quanto dai filosofi della scienza.

L'approccio più proprio della filosofia continentale ci offre dunque ancora quella sua presenza critica irrinunciabile e lo fa sia insistendo di volta in volta sui sottesi più pericolosi di una dottrina economica e sociale che, considerata nell'attuale ampiezza delle sue articolazioni, determina profondamente le nostre vite (*The New Spirit of Capitalism*, recensione di George Hull); sia sottolineando la persistenza, nella comune considerazione del capitalismo e dei suoi strumenti, tanto di una concezione ancora metafisica (talvolta addirittura di un atteggiamento teologico verso il capitalismo, di una fede fatta di credenze e dogmi) quanto di una concezione darwiniana, di organismo vivente (pensiamo solo all'idea di autoregolamentazione del mercato).

L'impianto teologico e l'organicismo sono entrambi retaggi di quelle moderne filosofie della storia che, se da un lato hanno contribuito ad elevare la scienza economica a prassi di pensiero, dall'altro le hanno consegnato quelle stesse criticità che, si è capito poi, hanno impedito loro di essere strumenti di analisi durevoli.

Concludendo, alcune credenze sottese si mantengono in seno alla nostra cultura inascoltate e vanno a costituire il potenziale di nuove derive di senso, che mal si conformano all'ambizione della scienza economica di essere la narrazione privilegiata del reale.

LA COMPRENSIONE DELL'UOMO

Diversi approcci quindi ma non tanto, a nostro parere, nel metodo, quanto nell'individuazione e circoscrizione del proprio ambito di lavoro. La filosofia si attua in entrambi i casi osservando i diversi aspetti fondamentali degli stessi fenomeni, fenomeni che sembrano riguardarci sempre più da vicino (*The Culture of the New Capitalism,* recensito da Marco Solinas). Ci accorgiamo

infatti di quanto i temi dibattuti nella filosofia dell'economia, per quanto ancora una volta molto tecnici, siano anche particolarmente vicini a noi tutti.

È qui, nell'ambito della critica all'economia, che si gioca oggi la comprensione del mondo e dell'uomo. È il luogo in cui i filosofi danno battaglia per fornire, da un lato, un modello scientifico unificante ed un orizzonte comune alle discipline più disparate che collaborano allo sviluppo dell'economia, come possiamo leggere nell'articolo di Caterina Marchionni; dall'altro, per mantenere vigili le coscienze degli scienziati stessi davanti all'avanzare di nuove e vecchie istanze della comunità tutta: istanze esistenziali, sociali ed ambientali.

LETTERATURA SCIENTIFICA E INTERDISCIPLINARIETÀ. IL PRIMATO GNOSEOLOGICO DELL'ECONOMIA

Come filosofia della scienza, la filosofia dell'economia ha una letteratura mista, scritta dagli stessi economisti da un lato e dai filosofi dall'altro. Ancora una volta risulterà chiaro il suo carattere interdisciplinare ed innovativo, direte voi, ma attenzione: essa eredita con una siffatta eterogeneità e ricchezza anche un'impostazione finalistica che le impedisce, se non integrata con una più profonda riflessione, di svilupparsi liberamente. Nonostante infatti l'economia ci convinca oggi di essere la candidata favorita al primato gnoseologico, tuttavia questa interdisciplinarietà che la coadiuva si sta ancora elaborando.

L'economia ha una sua connotazione interpretativa pervasiva ed attinge a piene mani dai modi e dai metodi delle scienza naturali e umane; i temi sono dunque tanto attuali quanto atavicamente legati alla filosofia quali, da un lato, la libertà, la relazione e l'intersoggettività, il desiderio, l'identità, il conflitto, la comunità - che possono farle da premessa; dall'altro, il valore, la libertà di scelta, la preferenza, la definizione di razionalità, l'analisi delle credenze e delle convinzioni e dei pregiudizi delle persone, su cui la filosofia della mente, per usare un altro termine generico ed obsoleto, ha ancora molto da lavorare.

STATUTO ONTOLOGICO INCERTO DEI SUOI OGGETTI

Gli oggetti della scienza economica hanno ancora uno statuto ontologico incerto, sopratutto quando, usati come strumenti tecnici, la discussione sulla loro efficacia sembra rifluire nell'atavico problema dell'affidabilità dell'introspezione. Ciò accade poiché è essa stessa *in primis* una teoresi *borderline* tra scienza e tecnica, da un lato osservazione invasiva dei comportamenti umani e dall'altro affinamento dello strumento. Essa è, nel suo modo più generale, legata alla finanza, alla politica ed alla strategia. Ed è per questo anche che essa può ambire oggi, attraverso il lavoro dei filosofi, a quell'indiscusso primato gnoseologico di cui sopra - cosa questa che comporta una certa assunzione di responsabilità: si tratta infatti di coincidere con un orizzonte semantico esteso all'intera sfera d'influenza della cultura occidentale, di un linguaggio che tutti devono imparano a parlare (e qui si potrebbe aprire un ulteriore dibattito, e non meno importante, sul fattore educazionale).

In conclusione, attraverso la filosofia, l'economica assume un nuovo ruolo nel panorama della ricerca scientifica denotandosi come un'ottima cornice gnoseologica ed una strumentazione euristica incisiva; tuttavia, nell'intento di assurgere davvero al ruolo regio che i filosofi già le riconoscono, ci accorgiamo come essa manchi ancora sul piano epistemologico, cosa da cui è evidentemente inficiata. L'economia dunque non basta a se stessa e non basta per impiantare una sistematica; tuttavia è oggi capace, vista la sinergia che crea tra le tante discipline, di proporre un modello scientifico unificante, atto non solo ad interpretare il reale ma anche a veicolare nuove possibilità, direi uniche, di comprensione tanto della realtà naturale quanto di quella dell'uomo.

Philosophically Incorrect

Crisi economica: una modesta riflessione

Umberto Maionchi

Non so a voi, ma a me i meccanismi di funzionamento della Borsa hanno sempre provocato una certa inquietudine. La semplice idea che un evento, magari anche insignificante e anonimo, accaduto chissà dove, possa influire pesantemente sul valore della moneta che ho in tasca, mi è sempre apparsa francamente assurda. Com'è possibile, mi sono chiesto spesso, che piccoli fatti apparentemente senza importanza o che sembrano non avere alcun legame con i processi economici si rivelino invece così importanti per l'intera economia solo per il fatto di aver influenzato la Borsa? Oggi poi che le transazioni finanziarie viaggiano velocissime attraverso oceani e continenti, l'economia mondiale assomiglia sempre di più ad una pazza lotteria in cui le estrazioni si susseguono ad un ritmo forsennato e senza alcuna logica.

Ma allora, mi sono chiesto, se le cose stanno proprio così, quale grado di affidabilità possiamo attribuire alla cosiddetta scienza economica ed ai suoi principi, ai suoi modelli, alle sue teorie? Quel che è accaduto pochi mesi fa (e che sta tuttora continuando ad accadere) in tutto il mondo ha mostrato chiaramente, anche ai più scettici, la fragilità del sistema economico globale.

La crisi finanziaria che ha dato l'avvio al processo è solo l'aspetto più eclatante e superficiale della crisi complessiva di un modello economico ormai adottato dal mondo intero. E a me, in particolare, appare sempre più evidente che è proprio la nozione classica di economia di mercato a dimostrarsi, ogni giorno di più, una pericolosa utopia. Certo, anche il modello comunista dei piani quinquennali, di sovietica memoria, si rivelò quasi subito una ingenua utopia, ma oggi possiamo tranquillamente affermare che quella era molto meno pericolosa: i danni che avrebbe potuto produrre sarebbero stati sicuramente più lievi. È pur vero, infatti, che un sistema economico caratterizzato dalla bassa produttività determina un mercato meno vivace e una competitività fiacca. Ma è altrettanto innegabile che la catastrofe che stiamo sfiorando è in gran parte dovuta ad un drammatico eccesso di produttività. Per non parlare dello scandaloso spettacolo della distribuzione della ricchezza mondiale e delle differenze, ormai drammatiche ed esplosive, nello sfruttamento delle risorse, elementi che concorrono entrambi alla instabilità generale.

D'altronde è altrettanto innegabile che tutti gli strumenti finora messi in campo per arginare questa deriva si sono dimostrati inefficaci o, addirittura, in qualche caso, capaci di aggravarla. Come esempio tipico mi vengono in mente gli interventi del Fondo Monetario Internazionale (FMI) che, come in molti ormai concordano, non solo non hanno ridotto le distanze tra le economie dei Paesi ricchi e poveri ma, in molti casi, le hanno approfondite, portando intere nazioni sull'orlo del collasso (ricordo, di sfuggita, tutte le campagne umanitarie per l'azzeramento del debito dei Paesi in Via di Sviluppo).

Credo perciò che ormai sia chiaro a tutti (o a *quasi* tutti!) che le CRISI ECONOMICHE non si possano trattare né al pari di eventi naturali come un terremoto o un ciclone (sebbene, anche nel caso del ciclone, si possa invocare la mano dell'uomo, vedi l' "effetto serra"), né come eventi divini ed eccezionali (anche se ci viene chiesto spesso una sorta di atto di fede).

In realtà, infatti, ciò che mi ha colpito nelle analisi e nei commenti di molti economisti (o presunti tali) è la certezza pressoché assoluta che gli eventi economici cui stiamo assistendo non possano che manifestare un'unica catena causale e, dunque, che esista un solo modo di interpretarli. In altri termini, anche se localmente possono essersi presentate delle singolarità, il fenomeno globale viene descritto come un passaggio contingente, magari difficile ma non strutturale: il modello generale non viene posto in discussione. Ed è proprio questa sicumera l'aspetto preoccupante.

Ad un osservatore appena un po' distaccato, le clamorose smentite alle previsioni sbandierate anche solo qualche mese fa non possono che apparire come un importante indizio di fragilità teorica. D'altra parte, che l'economia non potesse fregiarsi a pieno titolo dell'appellativo di "scienza esatta" lo sanno e lo sapevano un po' tutti, eccetto una minima percentuale di fanatici in cerca di una copertura di legittimità!

Se poi cerchiamo di argomentare più liberamente lasciando in secondo piano questioni che potremmo definire ideologiche, vorrei osservare che se dovessimo prendere sul serio la scientificità delle teorie economiche sulla piazza, avremmo anche l'obbligo di sottoporle, com'è usuale, a test di tipo sperimentale secondo criteri e protocolli moderni, ampiamente accettati in ambito scientifico. In realtà, in questi ultimi tempi stiamo assistendo alla nascita di nuove discipline che cercano di fare proprio questo: dotandosi di strumenti più raffinati presi in prestito dalle neuroscienze, in particolare, cercano di analizzare eventi, modelli e teorie economiche in una prospettiva di tipo scientifico allo scopo di formulare previsioni sempre più attendibili e precise.

Ma se ora torniamo a considerazioni di tipo più generale, ci accorgiamo subito che forse la difficoltà più grande nell'affrontare le tematiche economiche sta nel fatto che i principi del liberalismo classico vengono tuttora assunti come verità di fede, come postulati unici ed indiscutibili: una sorta di atteggiamento fideistico-religioso che contrasta platealmente con l'evidenza empirica diretta.

E allora, se le cose stanno così (ed abbiamo buone ragioni per crederlo!) mi permetto di suggerire un modesto rimedio, frutto di familiarità con discipline un po' esoteriche e di esperienze che, storicamente, hanno dato brillanti risultati.

Poiché abbiamo bisogno di un profondo ripensamento dei concetti e dei metodi fondamentali delle discipline economiche, perché non seguire l'esempio della cosiddetta rivoluzione geometrica del XIX secolo e cercare di costruire una sorta di "economia non-euclidea", sulla negazione di qualche assunto di base, di qualche postulato critico (come, ad es. quello che assicura la totale autoregolazione del mercato libero)? Oppure pensare ad una profonda rielaborazione dei principi economici sulla scia della relatività einsteiniana e provare ad elaborare una teoria che

potremmo chiamare: “macro-economia relativistica”? Oppure ancora perché non pensare una “micro-economia quantistica”, che tenga conto dell’indeterminismo dei micro-eventi legati alle scelte soggettive, o delle interazioni non-locali delle Borse, dei ruoli correlati osservatore-evento?

Trascinato dall’entusiasmo mi permetto anche di aggiungere che, per convincere gli scettici più riottosi, sarebbe auspicabile far precedere tutti questi tentativi dalla dimostrazione di uno speciale “teorema di incompletezza” per l’economia di mercato, che riuscisse a mettere in evidenza la sua intrinseca impossibilità di auto-fondazione e auto-giustificazione. Un tale risultato mostrerebbe anche (proprio come nel caso dell’aritmetica formale) che è impossibile una assiomatizzazione finita delle teorie economiche e che, quindi, per definirle e legittimarle, è necessario ricorrere a strumenti e concetti esterni ad esse e, in genere, più potenti (la nozione di “razionalità limitata” o quella di “bene comune”, tanto per fare due esempi molto semplici).

Chiudo notando che la nozione di “economia di mercato” assomiglia sempre di più a quella di “libera caccia”: ambedue si possono praticare finché sussistono condizioni “naturali” molto particolari e restrittive e tali da costituire, curiosamente, un'astrazione vicina all’utopia. Una cosa è comunque certa: il liberismo economico classico, nonostante gli aggiornamenti e le modifiche, resta oggi una delle teorie più vecchie e che hanno tratto meno profitto dal tumultuoso progresso delle scienze. Volendo a sua volta definirsi teoria “scientifica”, questo fatto non rappresenta un piccolo difetto.

Conference

Neuroeconomics: Hype or Hope?

Rotterdam, 20-22 November 2008

Federica Russo

The conference 'Neuroeconomics: Hype or Hope?' was held in Rotterdam on 20—22 November 2008. It was been hosted by the Erasmus Institute for Philosophy and Economics (EIPE), coordinated by an excellent scientific committee (Jack Vromen, Caterina Marchionni, Julian Reiss, Frans de Winden) and an efficient organisation assistant (Joshua Graehl).

Given the fairly recent advent and fast development of neuroeconomics, EIPE sought to set up a platform for discussion on the very foundations, aims, and empirical findings of neuroeconomics.

The organisers invited leading scholars in the field and solicited contributed papers around questions such as: 'What standards of scientific respectability and progress are implied (or invoked) in the claim that neuroeconomics will finally move economics into its proper standing of a modern science?'; 'What consequences does neuroeconomics have for the subject matter, scope and method of economics?'; 'How do the different disciplines of economics and of cognitive neuroscience relate to each other in neuroeconomics? Does the relationship between economics on the one hand and cognitive (neuro)science on the other need to be redefined?'; 'Do we first need to know how different levels of analysis (e.g. of observable choice behaviour, of its underlying computational algorithms and of the neural "hardware" in which they are implemented) relate to each other before we can tell how neuroeconomic evidence and findings bear on economics? If so, what levels are at stake and how are they related?'; 'What light can insights from contemporary philosophy of mind shed on the topics raised here?'; How is neural activity in people related to the various institutions in which they function? How can an improved understanding of neural processes inform institutional analysis?; What is the role and place of evolutionary theory in neuroeconomics?'.

Answers to these and other questions varied greatly, and this comes as no wonder since the very definition of neuroeconomics is currently under debate. The list of keynote speakers was remarkable: Ariel Rubinstein (Tel Aviv University, New York University), Paul J. Zak (Claremont Graduate University), Don Ross (University of Alabama Birmingham, University of Cape Town), John Davis (University of Amsterdam, Marquette University), Uskali Mäki (University of Helsinki), Jack Vromen (Erasmus University Rotterdam), Francesco Guala (University of Exeter, San Raffaele University).

The invited speakers spanned very different themes. Francesco Guala presented the first results the research project 'The Neuroscience of Conventions and Norms' at Exeter; Ariel Rubinstein expressed his deep scepticism about the possible success of neuroeconomics; Jack Vromen was concerned whether looking into the brain/mind would really open the black box of decision making; John Davis discussed what neuroeconomics can contribute to understanding the concepts of individual and of identity; Don Ross considered, inter alia, whether and how game theory can enlighten neuroscience; Uskali Mäki offered a milder view where neuroeconomics is at the same time a hope and a hype; finally, Paul Zak presented experiments used to identify causal relationships that can be used to improve current models.

Ten invited contributions touched on a variety of other topics, yet the dominating theme certainly was the foundational issue: what is exactly 'neuroeconomics' and is it really a

methodological and conceptual revolution? For the full list of contributed talks see the programme online here:

(http://www.eur.nl/fileadmin/ASSETS/fw/Neuroeconomics_Conference_Program.pdf).

All talks were remarkably good. A relaxed and informal atmosphere certainly contributed to having outstanding discussions. It has been a merit of the invited speakers to deliver presentation highly specialised and nonetheless accessible to the non-specialised audience. This, in my view, allowed to set up a common ground for a critical appraisal of neuroeconomics. The format of the conference has been a very well chosen one as well. Contributed papers had a discussant to kick off the discussion. This very much facilitated exchanges between the participants.

It seems that, as often happens in philosophy, no definite answer has been reached to the question of whether neuroeconomics be a hype or a hope. It has been suggested by various people that the fact that it may be well both at the same time is reason enough to pursue the neuroeconomic project accompanied by foundational and methodological discussions.

Conference

Decision, Games and Logic 2009

HEC Lausanne, 15-17 June 2009

Umberto Grandi

The third edition of the workshop on Decision, Games and Logic was held in Switzerland, at HEC in Lausanne, from the 15th to the 17th of June. It was coordinated by an excellent organizing committee composed of Christian Bach and Alessandro Facchini from the local institute, Conrad Heilmann from LSE (London School of Economics) and Olivier Roy from the University of Groningen. The format of the workshop reflected the scope of the past two editions: three high profile tutorials and very participated comment sessions aimed at fostering interaction between graduate students and young researchers in the areas of decision theory, game theory and logic. New features of this edition has been a panel on "What is Rationality?" and the "P. van Emde Boas Swap Session", during which various researchers presented each other's work in a joint presentation.

The first tutorial was given by Jaques Duparc, Professor of Formal Logic at HEC Lausanne, and aimed at introducing the basis of μ-calculus, an extension of modal logic that has a great potential in modeling fixed-point related notions in epistemic logic and in game theory. On the second day of the workshop Pierpaolo Battigalli from Bocconi University gave a detailed lesson on type structures. He presented in detail the mathematical assumptions underlying this model, widely used in game theory to represent interactive beliefs, and outlined the formal characterisation of backward and forward induction using these structures. The argument of the third and last tutorial, given by Luc Bovens from LSE, was an introduction to Bayesian epistemology. The main scope of the lesson was to show how much probabilistic models have to offer to formal philosophy, a scope successfully accomplished trough the use of many insightful examples.

The "PVeB Swap Session" was the main activity during the second part of the days. The idea of this session, originated during the last edition of DGL in Amsterdam, was to have two researchers presenting the main ideas of each other's work, with comments and questions for a final reply of the author. Many talks gave rise to long discussions with the audience. Examples include an exchange between Groningen and Paris (where the next DGL will be held) where Olivier Roy presented a work on dynamic epistemic logic for vague information by Paul Egre, who replied presenting Olivier's work on agreements theorems using weaker assumptions of common belief.

The amount of contributed talks was very much reduced with respect to the previous edition. Between them it is worth noting a presentation by Julien Dutant from the University of Geneva, who proposed a more general model for knowledge as belief based on a safe method, and a thorough formal study in the area of peer disagreement by Giacomo Sillari from the University of Pennsylvania. During the first day of the workshop a very successful poster session provided various young researcher with useful comments.

For a list of all talks and posters consult the programme on the website of the workshop (http:// www.meansandends.com/workshop09/?body=programme).

The last event of the second day of the conference was a very successful panel on rationality chaired by M. Tomassini. Three researchers, representing the full spectrum of opinions about rationality, developed an interesting discussion with good interaction from the audience. Richard Bradley from the LSE presented his middle point view of rationality as

coherence, and has fought hard ("*biting the bullet*" as he have repeatedly said) against his two opponents: Ulrich Hoffrage from HEC Lausanne, whose concept of rationality as adaptation allows inconsistency as an adaptive strategy, and Pascal Engel from the University of Geneve, more entrenched on more classical rationalist positions, stressing the importance of truth side by side with coherence. The day closed with an excellent dinner enjoying a superb view on the lake of Lausanne.

The workshop has been highly successful in fostering interaction between researchers of different communities and institutions. The tutorials were of high quality and very technical, whereas the main approach of the conference remains philosophical. Some areas, however, were not extensively covered,

i.e. foundational and epistemic aspects of decision and game theory, where logic plays a more explicit role, and topics that include interactions from computational logic. The "PvEB Swap Session" was the main focus of the workshop, and resulted in insightful comments for the speakers and interesting discussions for the audience. One defect is that it was sometimes hard to listen to a work not directly from the source. Moreover, this session also limited sharply external contributions to four talks and a poster session. Nevertheless, a great compliment is due to the organizing committee, who succeeded in preparing a high profile programme and an enjoyable social event. The next edition of the workshop has already been announced and will be held in Paris in 2010.

Annuncio
Twist Ontology

Alessandro Bemporad

Roberta Lanfredini

PRESENTAZIONE

Questo lavoro s'inserisce nell'ambito di un progetto che ha recentemente preso avvio presso il Dipartimento di Filosofia dell'Università di Firenze. Si tratta dell'istituzione di ENTIALAB[1], un Laboratorio congiunto fra Università e Impresa (Symmetric[2], software company con sede a Firenze). Il Laboratorio si occupa di Ontologia Applicata, cioè della messa a punto di possibili modelli conoscitivi in grado di rappresentare oggetti o, per meglio dire, particolari domini di oggetti.

Il Laboratorio di ontologia applicata nasce dalla condivisione di due campi di studio e di applicazione tra loro apparentemente eterogenei: da un lato l'ontologia filosofica e la teoria della conoscenza, dall'altro l'analisi, la progettazione e l'implementazione di sistemi software. I membri del Laboratorio provengono quindi da settori professionali molto diversi: innanzitutto i settori maggiormente connessi alla filosofia teoretica e all'ingegneria del software, immediatamente dopo settori di altro tipo, quali, ad esempio, la medicina e la genetica.

ENTIALAB ha fra i propri obiettivi fondamentali quello teorico-filosofico di individuare e definire le categorie fondamentali della realtà, le strutture profonde che si nascondono tra le pieghe del mondo che ci circonda. Strutture che, si suppone, rimangono costanti al variare dell'esperienza che di quel mondo noi tutti facciamo. Esso ha tuttavia, come secondo obiettivo fondamentale, quello di sottoporre le ipotesi così definite al vaglio della loro concreta applicabilità, oltrepassando e, per così dire, "mettendo alla prova" la loro natura puramente speculativa. Il campo di applicazione al quale il Laboratorio fa riferimento è l'implementazione di un insieme di sistemi software di nuova concezione, denominati *knowledge organization systems*, i quali consentono di acquisire, mantenere e elaborare in modo automatico modelli informativi estremamente complessi, attraverso uno schema di rappresentazione formale della conoscenza basato su ontologie.

La base informatizzata di conoscenza di questi sistemi software è alimentata a partire dalle informazioni elementari che possono essere acquisite dal mondo esterno nei modi più svariati, quali ad esempio flussi di dati, segnali provenienti da sensori o dispositivi elettronici, dati linguistici e semantici generati da *full-text retrieval*, insomma informazioni provenienti da qualunque fonte sensoriale digitale.

A questo proposito occorre segnalare come la stragrande maggioranza dell'attuale terminologia informatizzata si basi su nozioni e regole di classificazione vaghe e imprecise. Tale disomogeneità e opacità pone due problemi più specifici. Il primo problema è che una non ben definita determinazione della realtà rappresentata fa "uscire dalle maglie" una grande quantità di dati rilevanti. Il secondo problema è che una traduzione inefficiente fra sistemi alternativi di rappresentazione finisce inevitabilmente per rendere molti dati non traducibili e non comunicabili. A questi due problemi corrispondono quelle che, allo stato attuale, possono essere considerate le due maggiori esigenze dei processi d'informatizzazione: esatta

[1] www.entialab.org.
[2] www.symmetric.it.

determinazione dello schema rappresentativo utilizzato e traduzione degli schemi rappresentativi (spesso taciti e inconsapevoli) già disponibili.

Le ricadute pratiche dell'attività svolta dal Laboratorio sono molteplici. Oltre a contributi di ricerca relativi all'ontologia e alla teoria della conoscenza, il Laboratorio intende apportare elementi di innovazione tecnologica, con particolare riferimento a settori strategici quali la medicina, la clinica e il dominio socio-sanitario. Coadiuvando, sia pure indirettamente, il miglioramento generale dei processi di cura. Il laboratorio di ontologia applicata intende inoltre esplicare le proprie attività, oltre che in ambito accademico, nel campo della formazione e della consulenza, all'interno di una gamma potenzialmente molto ampia di enti pubblici e privati.

È infine una delle finalità dichiarate di ENTIALAB quella di favorire l'introduzione nel mondo del lavoro di giovani dotati di una preparazione mista, umanistica per vocazione ma direttamente esposta all'innovazione tecnologica, in grado di affrontare i mutamenti e l'evoluzione che il mondo sta sempre più rapidamente sperimentando.

Di seguito presentiamo una breve introduzione al modello ontologico denominato Twist messo a punto nell'alveo delle attività del laboratorio.

PRIMA PARTE - INTRODUZIONE

TWIST è un sistema software progettato e implementato per consentire la rappresentazione e l'elaborazione della conoscenza, in conformità a un modello ontologico formale sottostante. Nella parte introduttiva di questa sezione espliciteremo gli elementi concettuali che stanno alla base del sistema, mentre nella seconda parte verrà invece descritto nel dettaglio il modello ontologico sulla base del quale TWIST è stato implementato.

Che cosa è l'Ontologia filosofica

L'ontologia è quella branca della filosofia che si occupa in senso molto generale e astratto di *ciò che vi è*. Il suo scopo è quindi quello di isolare *tipi* e *strutture fondamentali* della *realtà* (oggetti, proprietà, relazioni, eventi, processi), fornendone una chiara caratterizzazione. L'uso che qui facciamo della nozione di ontologia è sostanzialmente descrittivo e intuitivo, nel senso che si sforza di analizzare fedelmente il punto di vista del senso comune, aderendo quanto più è possibile all'"immagine manifesta" del mondo.

La determinatezza del lavoro ontologico non deve essere confusa con la sua univocità. Non esiste, infatti, un solo modo di concettualizzare la realtà.

Nella recente filosofia analitica viene proposta, a questo proposito, una distinzione fra *metafisica* e *ontologia*[3] La distinzione risulta così concepita: l'*ontologia* è la teoria che stabilisce *ciò che vi è*; la *metafisica* invece è la teoria che stabilisce *cosa è ciò che vi è.* Da un certo punto di vista quindi la metafisica, intesa come caratterizzazione descrittiva, è preliminare. Non è affatto chiaro in che misura si possa stabilire se esistano certe entità senza offrirne una caratterizzazione. Questo fatto risulta evidente nelle scienze naturali: non si può, ad esempio, scoprire l'esistenza di una nuova specie senza sapere di che cosa si tratta. Analogamente, non

[3] Si veda A.C.Varzi, Ontologia, Roma-Bari, Editori Laterza, 2005 e E.Runggaldier e C.Kanzian, Grundprobleme der analytischen Ontologie, Paderborn, Verlag, 1998; trad. it. Problemi fondamentali dell'ontologia analitica, Milano, Vita e Pensiero, 2002.

ci si può limitare a dire che esistono oggetti materiali senza darne una qualche descrizione (ad esempio, gli oggetti materiali sono entità concrete, impenetrabili, estese nello spazio, dotati di certe proprietà essenziali, e così via).

Da un altro punto di vista, tuttavia, è l'ontologia ad essere prioritaria. Il carattere prospettico dell'oggetto rimanda infatti in misura essenziale alla convergenza delle varie prospettive *in un unico polo*. Metafisica e ontologia contribuiscono quindi di pari passo alla descrizione e all'identificazione degli oggetti relativi a un determinato *dominio*.

Uno dei problemi fondamentali che l'indagine ontologica si trova ad affrontare è quello di riuscire a fornire una relazione virtuosa fra metafisica e ontologia, individuando le caratterizzazioni che meglio riescono a modellare e concettualizzare un determinato dominio oggettuale. Ciò può avvenire solo mediante un'accurata *caratterizzazione* e un'opportuna *identificazione* degli oggetti appartenenti al dominio che ci troviamo a descrivere. Le descrizioni che non presuppongono una modellazione concettuale, quindi un'analisi ontologica, hanno spesso carattere informale e risultano per questo ambigue, circolari, o addirittura contraddittorie.

La stragrande maggioranza dell'attuale terminologia informatizzata, inoltre, è basata su nozioni e regole di classificazione formulate in modo impreciso, il che finisce per far uscire dalle maglie della descrizione moltissimi dati, e per renderne molti altri non traducibili e non comunicabili.

Ontologia Formale e Ontologia Materiale

Per risolvere questa difficoltà, l'ontologia filosofica propone la distinzione fra *ontologia formale* e *ontologia materiale*[4].

L'ontologia formale è neutrale rispetto a qualsiasi dominio particolare; essa si occupa in senso generalissimo dell'oggetto inteso come un vuoto "qualcosa" e delle sue caratterizzazioni (ad esempio la relazione parte-tutto; la differenza fra astratto e concreto; l'identità). L'ontologia formale, per sua stessa definizione, è svuotata di qualsiasi contenuto determinato. Essa vale quindi per gli oggetti in generale o, il che è lo stesso, per qualsiasi oggetto di qualsiasi dominio considerato.

L'ontologia materiale si occupa, al contrario, di oggetti che appartengono a domini determinati. Essa richiede quindi una caratterizzazione e una modellazione che faccia riferimento alla struttura contenutistica (o qualitativa) dell'ambito considerato (ad esempio oggetti, proprietà e relazioni relative a un dominio ontologico di tipo biomedico). È l'ontologia materiale che permette di formulare, eventualmente, non solo descrizioni ma anche *previsioni* scientifiche. L'analisi può infatti individuare l'emergere di tendenze, di "linee ontologiche", da interpretare, ad esempio, attraverso le metodologie e gli strumenti della medicina predittiva.

La struttura materiale può essere inoltre calibrata su diverse *modularità* o *granularità*: si parla di caratterizzazioni a *granularità fine* (ad es. il numero e l'ordine dei geni in un

[4] Si veda E. Husserl, Ideen zu einer reinen Phänomenologie und phänomenologische Philosophie Erstes Buch: Allgemeine Einführung in die reine Phänomenologie, Husserliana III/1, Nijhoff, Den Haag 1976; trad. it. Idee per una fenomenologia pura e per una filosofia fenomenologica, vol. I, Einaudi, Torino 2002 e E. Husserl, Terza ricerca logica, in Logische Untersuchungen, Zweiter Teil/1: Untersuchungen zur Phänomenologie und Theorie der Erkenntnis. Zweiter Teil/2: Elemente einer Phänomenologische Aufklärung der Erkenntnis, Husserliana XIX/1-2, Nijhoff, Den Haag 1984; trad. it. Ricerche logiche, Il Saggiatore, Milano 1968.

cromosoma) oppure di caratterizzazioni a *grana grossa* (es. la descrizione di una malattia e dei suoi sintomi, del comportamento e delle sensazioni associate).

Ognuno di questi punti – ontologia formale, ontologia materiale o di dominio, granularità della descrizione – devono essere scelti e fissati nel modo più chiaro e netto possibile. Ognuno di essi concorre infatti a delineare la cornice metodologica all'interno della quale si colloca il compito dell'ontologia. Si tratta di un compito che possiamo definire *anti-riduzionista* e *qualitativo*. *Anti-riduzionista*, perché l'analisi descrittiva dell'ontologia è interessata a tutti i livelli di granularità (scientifico e comune; microfisico, mesocosmico e cosmologico). *Qualitativo*, perché le tassonomie ontologiche non si rivolgono solo a ciò che è quantificabile e misurabile, bensì anche (e forse soprattutto) a ciò che è accessibile solo mediante descrizioni di tipo qualitativo.

Universali e Particolari

Un'ulteriore distinzione, che ci sarà utile per comprendere il modello ontologico che prenderemo fra poco in considerazione è la distinzione fra *universali* e *particolari*.

Gli *universali* corrispondono a quanto vi è di generale nella realtà. Essi sono ripetibili nel senso che possono essere esemplificati da più di un oggetto e in più momenti temporali. In virtù di questa loro caratteristica, gli universali non sono localizzati in modo determinato nel tempo e nello spazio. Essi possono esistere in tutti i tempi e in tutti i luoghi e, in quanto tali, si dicono irreali.

I *particolari*, al contrario, non sono ripetibili in quanto possono esistere solo in un luogo e in un tempo determinati. I particolari esemplificano (o *istanziano*) universali e non possono a loro volta essere esemplificati.

È in virtù della relazione di *esemplificazione* di uno stesso universale che due particolari possono essere considerati simili sotto qualche aspetto (ad esempio due individui essere considerati persone). I particolari sono quindi fugaci, contingenti, irripetibili; gli universali sono invece quanto nella realtà vi è di stabile, permanente, ripetibile. Il particolare è un *fatto*, un accadimento. L'universale è un *modello*, un'idea.

Il concetto di universale è distinto da quello di *classe*. Ogni universale ha una classe corrispondente, ma non vale il viceversa, cioè non ogni classe corrisponde a un universale. La classe è infatti un insieme o una collezione di particolari. Vi sono termini generali la cui funzione è di riferirsi a particolari che esistono in un luogo e in un tempo determinati (ad esempio, la classe dei nati a Firenze dal 1959 al 1963; oppure la classe dei laureati in filosofia nell'ultimo decennio). In questo caso si fa riferimento non a universali o a tipi generali, bensì a gruppi o collezioni di entità particolari. La relazione che sussiste fra il particolare e l'universale è di *esemplificazione*; quella che sussiste fra il particolare e la classe è di *appartenenza*.

Oggettualità, determinazioni ed entità

Infine, le categorie ontologiche risultano organizzate gerarchicamente a partire da un'ulteriore distinzione fondamentale: quella fra *oggettualità* e *determinazioni*. Entrambe le categorie rimandano inoltre a una categoria superiore: quella di *entità.* Le *oggettualità* sono entità costituite da *fasci di determinazioni*. Le *determinazioni* sono, in modo complementare, entità che confluiscono in oggettualità.

Al fine di chiarirne la natura, la distinzione fra oggettualità e determinazioni può essere resa anche come distinzione fra *parti indipendenti* e *parti dipendenti*[5]: le *parti indipendenti* sono quelle entità che godono della possibilità di una *autonomia concettuale* e *rappresentativa*; le *parti dipendenti* sono quelle entità che non godono di tale autonomia, ma che necessitano invece di altre parti dipendenti per poter dar luogo a oggettualità indipendenti.

Un foglio sarà quindi un'entità indipendente, mentre il colore sarà un'entità dipendente. Mentre il primo può sussistere da solo, il secondo necessita di integrazione (nel caso del colore, di essere integrato con l'estensione). Analogamente, un individuo sarà una parte indipendente; mentre il suo essere medico (un ruolo di quell'individuo) sarà una parte dipendente (non esistono medici che non siano anche individui). Si noti: ogni entità è, per definizione, anche una parte. Questo vale non solo le parti dipendenti ma anche per quelle indipendenti. Un individuo sarà, ad esempio, parte di un più ampio sistema sanitario, o di un organico, e così via.

Chiameremo la necessità di integrazione fra parti dipendenti relazione di *fondazione*. Le determinazioni devono fondarsi con altre determinazioni per dare luogo a parti indipendenti (oggetti, intesi come interi o frazioni di interi).

Le determinazioni possono essere universali o particolari. Nel primo caso si parla di *proprietà*, nel secondo di *modi*. In modo speculare, anche le oggettualità possono essere a loro volta universali e particolari. Nel primo caso si parla di *tipi*, nel secondo di *oggetti*.

Se interpretate come universali, cioè come proprietà, le determinazioni possono inoltre godere di livelli diversi di generalità: un primo ordine è quello della *specie* (ad esempio la proprietà dell'essere rosso, o della rossezza); un ordine superiore è quello del *genere* (ad esempio la proprietà dell'esser colorato). Le relazioni fra ordini sono relazioni *d'inclusione* (ad esempio l'esser rosso include l'esser colorato).

La combinazione delle due distinzioni categoriali fondamentali – quella fra universali e particolari e quella fra oggettualità e determinazioni introduce il particolare modello ontologico che prenderemo ora in considerazione: *The four-Category Ontology*.

FCO: The Four-Category Ontology

Possiamo mettere a punto un'ontologia che preveda due categorie fondamentali di particolari e due categorie fondamentali di universali. Il rapporto speculare fra le due coppie origina una sorta di *quadrato ontologico*[6].

Le due fondamentali categorie di particolari sono gli *oggetti* e i *modi*. Le due categorie fondamentali di universali sono i *tipi* e le *proprietà* (e *relazioni*). Le Proprietà (universali) stanno ai Tipi (universali) come i Modi (particolari) stanno agli Oggetti (particolari).

Proprietà e Modi hanno come genere superiore il concetto di *determinazione*. Tipi e Oggetti hanno come genere superiore il concetto di *entità*. Il che è un altro modo per dire che esistono entità particolari ed entità generali e che le determinazioni di quelle entità possono essere a loro volta particolari o generali.

In sintesi, possiamo affermare che in ciascun dominio applicativo, all'interno quindi di una determinata comunità semantica, coloro che vi operano distinguono i fenomeni che si presentano nel corso delle loro attività tra *oggetti* (concettuali o fisici), *proprietà* di questi

[5] E.Husserl, Terza ricerca logica, cit. Si veda anche R.Lanfredini (a cura di), A priori materiale, Milano, Guerini, 2006.

[6] Si veda anche E.J.Lowe, The Four Category Ontology. A Metaphysical Foundation for Natural Science, Oxford, Oxford University Press, 2006.

oggetti, e *relazioni* tra questi oggetti. Viene inoltre riconosciuta una differenza tra gli *oggetti individuali* ed i *tipi* (i *generi*, le *specie*) che li concettualizzano.

La determinazione dei tipi, delle proprietà e delle relazioni che contraddistinguono gli oggetti presenti in un determinato dominio, una volta individuati, descritti e classificati, formano quindi un'ontologia, la quale può successivamente essere implementata nei sistemi software per asserire fatti sul mondo, comprensibili ed utilizzabili sia da esseri umani che da sistemi informatici.

Esaminiamo adesso più da vicino il modello FCO, la *Four-Category Ontology*:

Come è possibile osservare nella precedente figura, il modello FCO ruota come abbiamo detto attorno a quattro principali categorie ontologiche, che andremo adesso ad introdurre.

La prima categoria, in un certo senso quella fondamentale, è costituita dagli *oggetti individuali,* o *particolari* (OBJECT). Per oggetto intendiamo una qualunque entità *individuale* (concreta o astratta, dislocata nello spazio-tempo o anche solo nel tempo), dotata di una propria identità riconoscibile, la quale sia dotata di una serie di proprietà che la caratterizzano e di una serie di relazioni che la connettono con altri oggetti. Esempi di oggetti quindi possono essere singole *entità* come una certa persona, un determinato luogo, uno specifico manufatto o un particolare oggetto naturale (come una pietra, o un fiume), così come anche singoli *eventi* quali ad esempio una certa visita medica, un determinato matrimonio, lo specifico pagamento di una multa, un certo prestito di un libro in una biblioteca, etc. In questa prospettiva, gli eventi sono quindi considerati come oggetti a tutti gli effetti, dotati di proprietà e di relazioni.

La seconda categoria è quella dei *tipi di oggetto* (KIND). Un tipo è la rappresentazione concettuale di un insieme di oggetti individuali connessi tra loro da una natura comune, o da un insieme di proprietà possedute da tutti i membri di quel tipo. Rispetto agli oggetti individuali, particolari, i tipi devono essere considerati come concetti *universali*, in grado poi di essere esemplificati attraverso singoli oggetti particolari. Persona, luogo, Organizzazione, Manufatto, Visita Medica, sono tutti esempi validi di tipi di oggetto.

La terza categoria è quella delle proprietà (PROPERTY) e delle relazioni (RELATION). Una proprietà è una specifica caratteristica, o qualità, posseduta da tutti gli oggetti di un certo tipo. In altre parole, una proprietà può essere considerata anche come un *modo di essere* di un

certo tipo di oggetti. Ad esempio, il colore può essere considerato una proprietà dei tavoli, così come l'età, o il nome proprio, possono essere considerate proprietà valide per gli esseri umani.

Le relazioni sono invece da considerarsi come un'estensione del concetto di proprietà, poiché non sono caratteristiche intrinseche di uno specifico individuo (relative, cioè, a un singolo tipo di oggetti), quanto piuttosto *connessioni*, o *associazioni caratterizzanti* tra due o più tipi diversi di oggetto. La relazione d'impiego, ad esempio, caratterizza un'associazione tra un qualunque essere umano ed una qualunque azienda. Altri tipi di relazione sono quella *generativa*, la quale dispone una serie di concetti in ordine ascendente di generalizzazione, oppure quella *partitiva*, che dispone gli oggetti in gerarchie di aggregazione e di composizione tutto-parti.

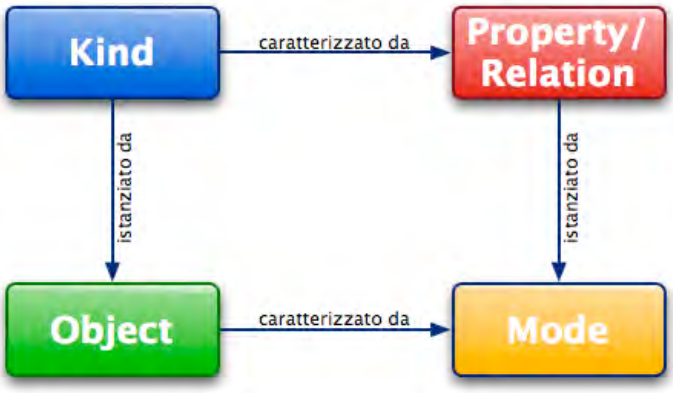
Nel descrivere le proprietà e le relazioni, abbiamo fatto riferimento alla loro definizione in quanto concetti universali, ovvero come caratteristiche o associazioni tra *tipi* di oggetto. Così come per gli oggetti nei confronti dei tipi, anche le proprietà e le relazioni possono essere esemplificate individualmente, nel momento in cui vengono assegnate ai (o meglio possedute dai) singoli oggetti individuali. In questo caso, e questa è la quarta categoria nella nostra ontologia, parliamo più propriamente di modi (MODE), o di *valori* delle singole proprietà.

Ad esempio, mentre il "nome proprio" è una proprietà (PROPERTY) di un generico tipo (KIND) "Essere Umano", "Mario" è uno *modo (*MODE*)*, ovvero un *valore* specifico per la proprietà "nome proprio" quando questa viene esemplificata da uno specifico oggetto individuale (un *OBJECT*, in questo caso una singola persona).

Per riassumere quanto detto, e per concludere questa breve introduzione metodologica, possiamo evidenziare le seguenti immagini del modello FCO:

Universali – Particolari

Abbiamo visto come, nella FCO, i Tipi (KIND), le Proprietà (PROPERTY) e le Relazioni (*RELATION*) siano entità universali, mentre gli Oggetti (OBJECT) ed i Modi (MODE) degli oggetti sono entità particolari.



Nella figura precedente, possiamo osservare le relazioni che sussistono tra le quattro categorie che abbiamo descritto:

0) Gli Oggetti sono istanze dei Tipi (quindi i Tipi sono istanziati dagli Oggetti);
1) I Modi sono istanze delle Proprietà/Relazioni (quindi le Proprietà/Relazioni sono istanziate dai Modi);
2) I Tipi sono caratterizzati attraverso le Proprietà e le Relazioni (le quali a loro volta caratterizzano i Tipi);
3) Gli Oggetti sono caratterizzati dai Modi specifici (i quali a loro volta caratterizzano i singoli Oggetti).

SECONDA PARTE - TWIST ONTOLOGY MODEL

Una volta definito il quadrato ontologico nelle sue linee generali, possiamo adesso esaminarlo un po' più da vicino, allo scopo di introdurre un insieme di estensioni e di precisazioni che ci condurranno a completare lo schema ontologico che stiamo illustrando. Il modello ontologico, progettato dal Laboratorio di Ontologia Applicata, che andiamo adesso a presentare costituisce la base del TWIST ONTOLOGY MODEL, a partire dal quale Symmetric ha implementato , in un prodotto software, il proprio *Ontology Engine*.

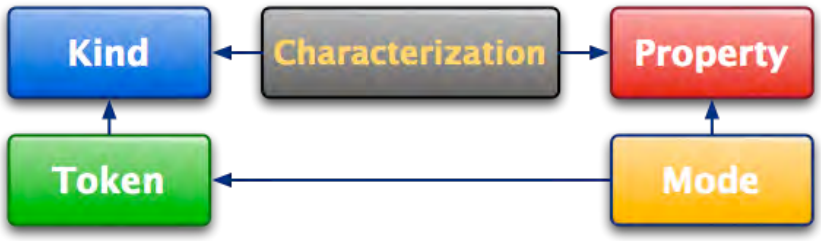
Il concetto di Caratterizzazione

Osserviamo quindi per prima cosa la *relazione* che abbiamo definito sussistere tra le due categorie di *universali* introdotte in precedenza. Come abbiamo visto, gli elementi KIND (i tipi di oggetto) sono *caratterizzati* da una serie PROPERTY, ovvero dalle proprie qualità.

Definiamo quindi il concetto stesso di *CHARACTERIZATION* come un elemento base della nostra ontologia, che va ad aggiungersi alle quattro categorie del quadrato ontologico per

modellare proprio quest'idea di caratterizzazione. Possiamo definire la caratterizzazione osservando come *una proprietà* possa caratterizzare *tipi di oggetti diversi*. Ad esempio, la proprietà Colore potrebbe caratterizzare il tipo *Persona*, se inteso come colore della pelle, così come il tipo *Automobile*, se intendiamo in questo caso il colore della carrozzeria.

Possiamo rappresentare il concetto di caratterizzazione come una *relazione* che connette tra loro *una* PROPERTY ed *un* KIND, nella quale la stessa PROPERTY può essere associata a più elementi KIND diversi, ed uno stesso KIND può essere caratterizzato da più PROPERTY differenti:



L'introduzione della caratterizzazione consente al nostro quadrato ontologico di rappresentare correttamente il legame che unisce gli elementi PROPERTY ed KIND.

Facciamo però adesso un passo avanti, ed esaminiamo il concetto stesso di PROPERTY. Introduciamo nella seguente tabella un TOKEN che esemplifica il KIND *Persona*, l'ipotetico avvocato Giovanni Rossetti, e indichiamo alcune delle possibili PROPERTY che caratterizzano proprio il KIND *Persona*, assieme ai MODE con i quali queste PROPERTY vengono istanziate nel caso del TOKEN in esame:

PROPERTY	MODE
Nome	Giovanni
Cognome	Rossetti
Data di Nascita	25/03/1958
Professione	Avvocato
Titolo di Studio	Laurea in Giurisprudenza
Gruppo Sanguigno	B Rh+
Altezza	178 cm
Assistito da	Dr. Giuseppe Verdi
Stato Civile	Coniugato
Coniuge	Francesca Galli
Comune di Residenza	Firenze

PROPERTY	MODE
Comune di Domicilio	Empoli

E' chiaro che stiamo semplificando (lo faremo continuamente nel seguito del nostro discorso), ed è chiaro anche che queste non sono certo le uniche PROPERTY che potremmo utilizzare per caratterizzare una *Persona*, ma certamente queste sono *caratteristiche*, *qualità* che possiamo associare in generale all'idea di persona (perlomeno quando questa persona è un cittadino dello Stato Italiano).

Le Proprietà Semplici

Scorriamo adesso una per una queste proprietà. Ci accorgeremo che esse non sono proprio tutte dello stesso tipo: esse possono essere in realtà scomposte a loro volta in sotto-categorie della categoria base PROPERTY. Concentriamoci, tanto per iniziare, sul seguente gruppo di proprietà:

PROPERTY	MODE
Nome	Giovanni
Cognome	Rossetti
Data di Nascita	25/03/1958

Ora, senza dubbio potremmo attribuire ai MODE di queste tre proprietà *un valore qualunque*, purché si rispettino alcune regole di base che adesso andremo a definire. Ad esempio, la data di nascita di una persona può essere teoricamente espressa da un qualunque valore, purché rappresenti una data correttamente espressa. Il nome e il cognome di una persona devono essere sequenze di caratteri privi di cifre e di segni di punteggiatura (anche se è chiaro che non tutte le combinazioni di caratteri sarebbero riconosciute come nomi propri validi nella lingua italiana). Attribuiamo a questo tipo di PROPERTY, i cui MODE possono variare liberamente purché all'interno di un determinato tipo di dati (stringhe di caratteri, date, numeri interi, etc.) il nome di **SIMPLE PROPERTY** (proprietà semplici):

Questa è la prima delle sotto-categorie delle PROPERTY che introduciamo nella nostra ontologia: quella delle proprietà semplici, ovvero di quelle qualità degli oggetti che sono espresse da un *valore*, e che possono essere rappresentate nel modo più diretto nel nostro quadrato ontologico, come nel seguente esempio:

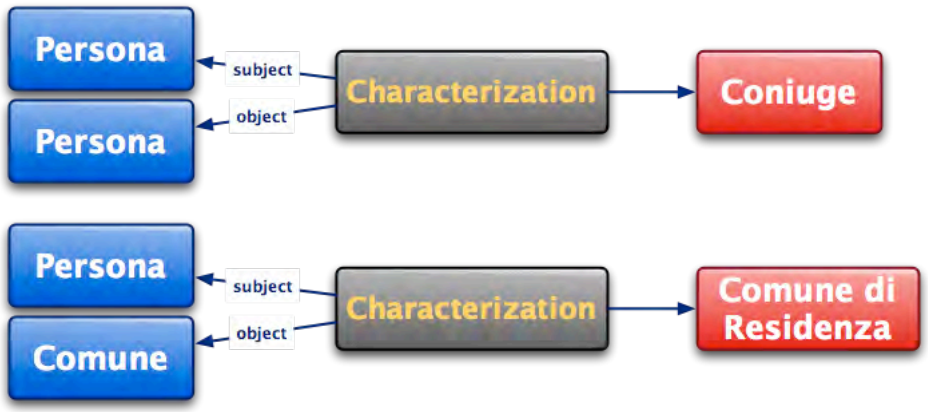
Le Relazioni

Proviamo adesso invece ad osservare il seguente gruppo di proprietà:

PROPERTY	MODE
Assistito da	Dr. Giuseppe Verdi
Coniuge	Francesca Galli
Comune di Residenza	Firenze
Comune di Domicilio	Empoli

È evidente già a una prima analisi superficiale che nel caso di queste proprietà non possiamo parlare di *proprietà semplici*, come nel caso del nome e del cognome. In effetti, i MODE di queste proprietà non sono rappresentati da *semplici valori*, ma piuttosto da *connessioni* verso altri TOKEN, i quali a loro volta quindi sono esemplificazioni di un proprio KIND.

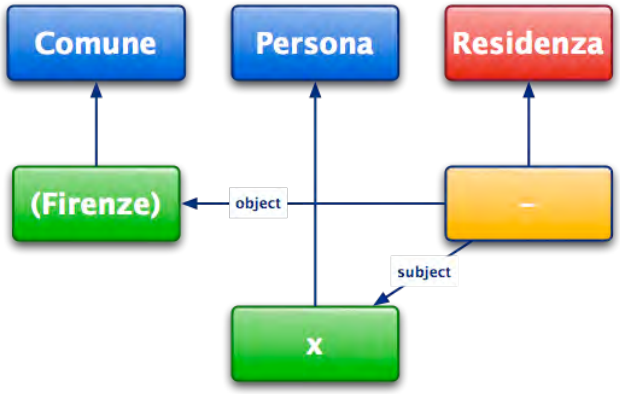
Ad esempio, il Coniuge di una Persona è sempre un'altra Persona; il Comune di Residenza di una Persona è invece sempre un Comune (cioè un oggetto di un altro KIND):



Come possiamo vedere dai due diagrammi precedenti, questa nuova sotto-categoria delle PROPERTY rappresenta le relazioni (chiamiamola quindi RELATION PROPERTY) e stavolta deve

essere espressa come *connessione* tra due elementi KIND, dei quali uno è il *subject* (il soggetto della relazione), e l'altro l'*object* (l'oggetto della relazione).

Le RELATION PROPERTY vengono istanziate da MODE che connettono sempre due TOKEN (anche in questo caso un *subject* ed un *object*) come è possibile osservare nello schema precedente, nel quale si rappresenta la relazione di *Residenza* tra una Persona ed un Comune.



Le Classificazioni

Dopo aver definito le SIMPLE PROPERTY e le RELATION PROPERTY, passiamo adesso a esaminare invece il seguente gruppo di proprietà:

PROPERTY	MODE
Professione	Avvocato
Titolo di Studio	Laurea in Giurisprudenza
Gruppo Sanguigno	B Rh+
Stato Civile	Coniugato

Certamente queste non sono proprietà semplici, in quanto i valori ammessi non sono liberamente esprimibili. Possiamo pensare anche stavolta di trovarci in presenza di *relazioni* fra tipi di oggetto? Concentriamoci in particolare su una delle precedenti proprietà: lo Stato Civile. Che cos'è lo "Stato Civile"? È certamente una proprietà delle Persone ma, più propriamente, lo Stato Civile è la classificazione di un determinato fenomeno, ovvero quello che rappresenta una particolare forma di relazione tra due persone.

In questo caso, quindi, i valori ammessi per questo genere di proprietà non sono *liberi*, come nelle proprietà semplici, ma piuttosto vengono scelti da un *elenco* prefissato, la *classificazione* appunto, che nel caso dello Stato Civile potrebbe prevedere, oltre che "Coniugato", anche gli elementi "Nubile", "Celibe", "Divorziato", e così via.

Dato che sono numerosissime le forme di classificazione previste per descrivere i fenomeni, introduciamo ancora nel nostro schema ontologico una particolare sotto-categoria di

PROPERTY, la CLASSIFICATION PROPERTY, proprio per individuare e descrivere tutte le proprietà del tipo appena visto.

Sorge però adesso un problema, che possiamo evidenziare nella figura seguente:

Se il KIND *Persona* è caratterizzato dalla PROPERTY "Stato Civile", e lo "Stato Civile" è una *classificazione* (nel senso di un elenco predefinito di rappresentazioni codificate di un certo fenomeno), come (e dove) dovremo rappresentare i singoli elementi previsti e ammessi dalla classificazione? La soluzione che abbiamo adottato nella nostra ontologia è stata quella di considerare una differenza tra una proprietà che rappresenta una classificazione e lo schema di classificazione vero e proprio, quello che contiene i singoli elementi della classificazione, il quale viene considerato però non come una PROPERTY, ma piuttosto come un KIND, cioè come un particolare tipo di oggetti.

Object Type e Classification Scheme

Abbiamo così introdotto anche due sotto-categorie di KINDs, che possiamo descrivere attraverso i seguenti due esempi:

Nel caso dei KINDs che rappresentano concetti i cui TOKEN sono *entità individuali enumerabili* (in modo teoricamente illimitato), ciascuna con la propria *identità individuale*, conveniamo di parlare di OBJECT TYPE, ovvero di tipi di oggetto nel senso più pieno. E' questo il caso del concetto di Persona che stiamo esaminando, così come anche il caso di un numero elevatissimo di altri concetti, *continuanti* ed *occorrenti*, quali i luoghi, tutti i tipi di evento e di processo, le organizzazioni, gli oggetti materiali, etc.

Quando invece vogliamo rappresentare gli *schemi di classificazione* come KIND, utilizziamo una particolare sotto-categoria di KIND, il CLASSIFICATION SCHEME. In questo caso, i TOKEN di un certo KIND non sono individui qualunque, come nel caso delle Persone, ma sono proprio tutti e soli gli elementi dello schema di classificazione rappresentato da quel KIND (nell'esempio precedente, tutti i TOKEN che sono elementi della classificazione KIND Stato Civile).

È interessante aggiungere due brevi considerazioni sugli schemi di classificazione, prima di passare oltre:

1) Ogni TOKEN di uno schema di classificazione possiede tipicamente due SIMPLE PROPERTY che lo descrivono: un codice identificativo e una descrizione, così come accade sempre nel caso degli schemi di classificazione;
2) Molti schemi di classificazione sono caratterizzati da una *tassonomia*, quindi con una forma di rappresentazione della conoscenza ad *albero rovesciato*, o anche a *grafo orientato*. In questi casi, che estendono e complicano il semplice schema classificatorio a *dizionario*, è possibile utilizzare delle RELATION PROPERTY per connettere tra loro in vari modi i singoli elementi degli schemi di classificazione (tipicamente attraverso relazioni ISA, cioè relazioni di generalizzazione tra concetti, che formano, appunto, gerarchie di generalizzazione).

Senza approfondire ulteriormente questi argomenti, possiamo comunque evidenziare come l'introduzione di questi nuovi concetti (gli OBJECT TYPE ed i CLASSIFICATION SCHEME, assieme alle proprietà e alle relazioni necessarie alla loro corretta rappresentazione) siano stati aggiunti all'interno del quadrato ontologico, *senza ulteriori categorie di base*.

Torniamo adesso ad esaminare le nostre PROPERTY, in quanto manca ancora un tassello da sistemare per completare l'immagine dell'ontologia che stiamo presentando.

Measurement Property e Measurement Unit Type

PROPERTY	MODE
Altezza	178 cm

In effetti era rimasta ancora una proprietà da analizzare, tra quelle che avevamo indicato nel nostro esempio di caratterizzazione del KIND *Persona*: l'Altezza.

L'Altezza, possiamo affermare con sicurezza, è una qualità degli oggetti che viene espressa come la *misura* di un *fenomeno*, ovvero come una dimensione *quantitativa* che si differenzia dalle altre proprietà viste finora, che erano invece *qualitative*. Ora, la misura di un determinato fenomeno è una proprietà *composta* sempre da *due* elementi:

1) Un valore, tipicamente numerico, che esprime la *quantificazione* della misurazione;
2) Un'unità di misura, ovvero uno specifico elemento di uno schema di classificazione progettato proprio per standardizzare la *misurazione* di quantità fisiche.

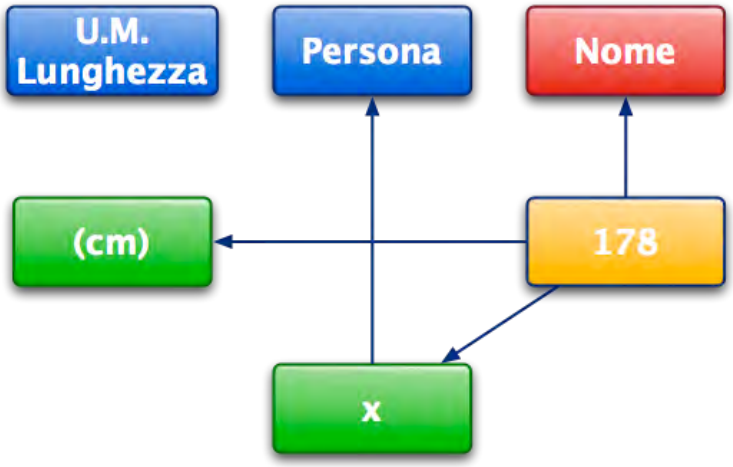
Se guardiamo alla proprietà che abbiamo evidenziato nel nostro esempio, possiamo appunto vedere attive entrambe le caratteristiche della misura: il *valore* (178) e l'*unità di misura* (cm).

Nel definire il nostro modello ontologico, abbiamo ritenuto essenziale poter esprimere il concetto di *misurazione* come concetto originario, introducendo un ulteriore tipo di proprietà: la MEASUREMENT PROPERTY. Attraverso questo nuovo concetto è possibile quindi identificare e definire qualunque misurazione, (come l'altezza, nel nostro esempio) considerandola come proprietà caratterizzante di uno o più KIND.

Per completare la definizione delle proprietà che rappresentano delle misure, è però necessario aggiungere anche l'indicazione di una *unità di misura*. A questo scopo, dato che le unità di misura sono elementi di particolari schemi di classificazione, abbiamo introdotto un ulteriore tipo di KIND, il MEASUREMENT UNIT TYPE (come ad esempio lunghezza, massa,

tempo, forza, energia, pressione, etc.), i cui TOKEN sono proprio gli elementi dello schema di classificazione (cm, kg, mg, sec, V, dyn, e così via):

La valorizzazione di un MODE relativo a una MEASUREMENT PROPERTY prevede quindi sia l'indicazione di un valore, che l'indicazione del TOKEN che rappresenta l'unità di misura associata a quel valore, oltre naturalmente all'indicazione del TOKEN che rappresenta il soggetto della misurazione:



CONCLUSIONE

Abbiamo descritto a questo punto il quadrato ontologico, assieme alle sotto-categorie che abbiamo ritenuto necessario introdurre per rappresentare un modello compiuto di ontologia applicata, ovvero un modello di descrizione della realtà che possa essere compiutamente implementato in un vero sistema software di rappresentazione e manipolazione automatica della conoscenza.
Come spunto per una futura evoluzione del nostro lavoro, introduciamo per concludere il concetto di ENTITY, che può essere considerato come un genere superiore rispetto alle categorie KIND e PROPERTY. Se correttamente interpretato e gestito, il concetto di ENTITY, che modella l'idea di entità in generale, fornirà gli strumenti per affinare ulteriormente il quadrato ontologico, consentendo ad esempio di rappresentare gerarchie di generalizzazione anche tra PROPERTY, oltre che fra KIND.

Obituaries

La redazione si unisce al cordoglio del mondo accademico per la scomparsa di **Aldo Giorgio Gargani**, filosofo e critico illuminante. Nato a Genova, laureatosi in filosofia presso la Scuola Normale Superiore di Pisa e specializzatosi presso il Queen's College di Oxford, Gargani è stato professore ordinario di Estetica presso l'Università di Pisa, *Gast Professor* presso l'Università di Vienna, membro del *Wissenschaftskolleg* di Berlino, della austriaca *Österreichische Ludwig Wittgenstein Gesellschaft,* del *College International de Philosophie.* Non potendo in questa sede nemmeno accennare al suo percorso intellettuale, data la vastità delle sue ricerche ed i numerosissimi interessi, desideriamo tuttavia ricordarne quell'intensa attività d'insegnamento e seminariale in Europa e negli Stati Uniti, che lo hanno reso entrambe indiscusso protagonista della cultura filosofica dell'ultimo quarto del secolo trascorso.

La redazione esprime sentito cordoglio alla notizia della scomparsa di **Ralph Dahrendorf** (1929-2009), filosofo, sociologo e politico liberale. Nato ad Amburgo, Darhendorf conobbe personalmente la prigione e l'esperienza del campo di concentramento per la sua opposizione al regime nazista. Dopo aver conseguito un dottorato in filosofia presso l'Università di Amburgo nel 1952, Dahrendorf ottenne un secondo dottorato, questa volta in sociologia, presso la London School of Economics, nel 1956. Qui venne in contatto con le idee di "Società Aperta" proposte da Karl Popper, di cui fu studente, e che, a detta dello stesso Dahrendorf, davano una risposta alle grandi domande della moderna società industriale poste da Marx. Dopo aver insegnato a Saarbrücken, Tubinga e Costanza nella Germania dell'Ovest e Stanford in California, Dahrendorf decise di entrare in politica. Fu eletto al Parlamento Europeo nel 1970. Nel 1974, fu nominato Direttore della London School of Economics, carica che ricoprì fino al 1984. Dal 1987 al 1997, fu infine Warden del St. Anthony College presso l'Università di Oxford. Da sempre interessato al problema delle divisioni di classe nella società moderna, Dahrendorf fu strenuo difensore del pluralismo liberale, un sistema sociale nel quale vengono riconosciuti e tutelati dalle istituzioni interessi e aspirazioni conflittuali.

La Redazione

Esperimenti paradigmatici: il gioco dell'Ultimatum[+]

Francesco Guala[*]
francesco.guala@unimi.it

ABSTRACT

The Ultimatum Game is one of the most successful experimental designs in the history of the social sciences. In this paper I try to explain this success what makes it a ³paradigmatic experiment² stressing in particular its versatility. Despite the intentions of its inventors, the Ultimatum Game was never a good design to test economic theory, and is now mostly used as a heuristic tool for the observation of non-standard preferences or as a ³social thermometer² for the observation of culture-specific norms.

1. INTRODUZIONE

Immagina di avere appena ricevuto 10 euro. Questa somma dovrà essere divisa con un partner anonimo e invisibile, e dovrai accordarti con lui (o lei) su come dividerla. Lo spazio per la discussione è quasi inesistente: potrai soltanto offrire una divisione della torta, e il tuo partner potrà solo accettare o rifiutare. Se la rifiuta, perderete entrambi l'opportunità di dividervi i 10 euro; se l'accetta, vi porterete a casa la vostra porzione, come indicato dalla divisione che hai proposto.

Questa è sostanzialmente la situazione strategica nota come "gioco dell'Ultimatum" (GU). Probabilmente avrai già incontrato il GU. Se, come Robert Axelrod ha notato, il Dilemma del Prigioniero è diventato l'*E. coli* delle scienze sociali, allora il GU ne è senz'altro la *Drosophila melanogaster*. La battuta di Axelrod va presa sul serio. Suggerisce che la teoria dei giochi sperimentale ha raggiunto un livello di maturità paragonabile a quello di scienze sperimentali affermate quali la biologia molecolare. Suggerisce anche che questa maturità si manifesta in particolare nell'emergere e nel consolidamento di prototipi sperimentali o "esperimenti paradigmatici" analoghi agli "organismi modello" della biologia sperimentale. Gli esperimenti paradigmatici hanno varie caratteristiche epistemiche distintive. In primo luogo, anni di sperimentazione fanno sì che li conosciamo molto bene, sicuramente meglio di qualsiasi altro sistema nel dominio di applicazione della scienza. In secondo luogo, essi vengono usati per formulare una varietà di inferenze ad altri sistemi che sono meno accessibili allo studio sperimentale. E infine, sono solitamente strumenti versatili che possono essere utilizzati per diversi scopi in diversi momenti e in diversi contesti di investigazione scientifica. In questo articolo mi concentrerò principalmente su questa caratteristica, ma anche le altre avranno un ruolo nella storia del GU.

[+] Una versione in inglese di questo articolo è apparsa sulla rivista *Philosophy of Science* 75 (2008) pp. 658-669. Marcel Boumans, Margaret Schabas, Bob Sugden, e diversi partecipanti alla conferenza della Philosophy of Science Association a Vancouver nel Novembre 2006 hanno fornito diversi commenti e suggerimenti. Tutti gli errori rimanenti sono a mio carico.

[*] Dipartimento di Economia Università degli Studi di Milano

2. CONTROLLARE LE TEORIE

Il GU fu concepito e utilizzato per la prima volta da un gruppo di economisti sperimentali tedeschi guidati da Werner Güth nei primi anni '80 del Novecento (Güth, Schmittberger and Schwarze 1982). Güth e colleghi intendevano studiare la contrattazione sequenziale, e scelsero il GU principalmente perché è il problema di contrattazione sequenziale più semplice che si possa immaginare.[1] La semplicità era volta a minimizzare i costi cognitivi di calcolo. È risaputo che gli individui faticano ad analizzare i giochi dinamici complessi, in particolare quando devono operare la cosiddetta "induzione all'indietro" (*backward induction*). Nel GU, non si può non capire che il gioco finisce dopo la mossa del secondo giocatore, per cui il rumore nei dati dovuto alla mancata comprensione della decisione sperimentale dovrebbe essere ridotto al minimo.

Il team di Güth intendeva anche controllare alcuni risultati di esperimenti precedenti. Fouraker e Siegel, due psicologi, avevano realizzato negli anni '50 una serie di studi della contrattazione ormai divenuti classici. Güth e i suoi colleghi trovavano sorprendente che le offerte inique venissero rifiutate raramente negli esperimenti di Fouraker e Siegel. Ipotizzarono che questo potesse essere dovuto ad alcuni dettagli del loro disegno, e decisero di controllarli modificando leggermente la situazione sperimentale.

In effetti, le offerte inique sono rifiutate nel GU. Nel loro primo studio, Güth, Schmittberger e Schwartz scoprirono che in media il primo giocatore offre il 35% della torta. Osservarono una risposta modale al 50% e pochi rifiuti da parte del secondo giocatore (circa il 10%). Quando il gioco venne ripetuto una settimana più tardi (condizione con "giocatori esperti") l'offerta media scese al 31%. L'offerta modale al 50% scomparve, con la maggior parte delle offerte ora fra il 20 e il 30%. C'erano anche più rifiuti con la ripetizione (circa il 30%). Da allora, questi risultati sono stati replicati numerosissime volte, e costituiscono la cosiddetta "anomalia del GU". Ma anomalia rispetto a che cosa?

3. CONTROLLARE LA TEORIA E MANIPOLARE LE PREFERENZE

Secondo la teoria dei giochi classica, il primo giocatore dovrebbe offrire quasi nulla e il secondo dovrebbe accettare. L'idea è che il secondo giocatore dovrebbe affrontare una decisione apparentemente banale: non portarsi a casa niente oppure ciò che è stato offerto dal primo giocatore. Supponiamo che il minimo che si può offrire sia 1 euro. Un euro è meglio di niente, quindi il secondo giocatore dovrebbe accettare. Assumendo la conoscenza comune del gioco e della razionalità, la divisione 1/9 è l'unico equilibrio di Nash del GU.

Questa "previsione standard" per la verità è derivata per mezzo di un meccanismo abbastanza complesso: da una teoria del gioco strategico (il "nucleo" della teoria dei giochi) più un insieme di assunzioni riguardo alle preferenze e alle credenze dei giocatori. Secondo l'assunzione dell'egoismo, i giocatori preferiscono avere più denaro che meno, e non badano ai guadagni altrui. Il GU dunque è un'anomalia rispetto a una previsione derivata da un modello composito (il modello della "razionalità egoista") e solleva il noto problema di Duhem-Quine. La teoria del gioco strategico (l'assunzione di razionalità) è responsabile dell'anomalia, l'assunzione dell'egoismo è inadeguata a catturare il comportamento nel GU, le credenze

[1] La contrattazione è solitamente rappresentata nella teoria dei giochi come un problema di suddivisione, dove il surplus generato dallo scambio fra due beni deve essere allocato fra le parti. I 10 euro nel GU rappresentano in qualche modo questo surplus.

individuali divergono dalle assunzioni classiche, oppure qualche altro elemento del modello non è in grado di spiegare ciò che succede in un GU sperimentale?

La teoria della razionalità che costituisce il nucleo della teoria dei giochi è una teoria del tipo "se… allora…": dice che *se* le loro preferenze e credenze sono così-e-così, *allora* gli individui si comporteranno così-e-così. Se le preferenze non sono egoistiche come postulato dal modello classico, quindi la teoria della razionalità non può essere controllata in un gioco *singolo* come il GU. La teoria dei giochi sperimentale "kosher" dovrebbe dunque cominciare con una misurazione delle preferenze in circostanze non-strategiche, e proseguire controllando se la teoria del gioco strategico formula previsioni corrette utilizzando le misurazioni preliminari come dati iniziali nella procedura di controllo.

Tuttavia la maggior parte della teoria dei giochi sperimentale non è "kosher" da questo punto di vista. La maggior parte degli esperimenti comincia *postulando* il contenuto delle preferenze individuali, invece di individuarle empiricamente. Il GU, come abbiamo visto, non fa eccezione. Ma se il GU non controlla davvero la teoria del gioco strategico (il "nucleo" della teoria dei giochi) perché è diventato così famoso e ampiamente replicato?

Come il Dilemma del Prigioniero, il GU è così semplice che nessuno dubita davvero che i soggetti sperimentali siano razionali, nel senso minimale che le loro azioni seguono dalle loro preferenze e credenze. Piuttosto, il GU è interessante perché rende osservabili le *preferenze non-standard*. Intuitivamente sembra plausibile che il secondo giocatore possa preferire di rinunciare a un po' di soldi per punire le offerte inique, e che il primo preveda che un'offerta iniqua possa urtare i sentimenti del secondo giocatore. I secondi giocatori hanno preferenze "rivolte agli altri", influenzate forse da una preoccupazione per l'equità. Infatti, se consideriamo gli sviluppi teorici indotti da anomalie come il GU, notiamo che l'impegno maggiore è confluito nel tentativo di modellizzare funzioni di utilità non-standard, nelle quali le preferenze individuali sono etero-rivolte piuttosto che strettamente egoistiche come nel modello classico.[2]

Il GU ha il merito di rendere particolarmente evidenti le motivazioni rivolte agli altri che influenzano il comportamento. Se queste motivazioni possano essere incorporate nel quadro della teoria dei giochi oppure no è una domanda diversa, che non affronterò in questo articolo.[3] Ma rispondere a questa domanda non era essenziale per il successo del GU, che fu ottenuto comunque in modo indipendente.

4. ISTITUZIONI

Le prime varianti sul tema del GU erano mirate a controllare la robustezza dell'anomalia, e spesso erano esplicitamente disegnate per farla scomparire. Col senno di poi, possiamo dire che questi tentativi non hanno avuto successo. Lo status del GU non è stato scalfito, come documentano tutte le rassegne pubblicate fin dai tardi anni '80. Nei primi anni '90 tuttavia la storia del GU prende una nuova piega, con la pubblicazione di un importante articolo da parte di un gruppo internazionale di teorici dei giochi guidato da Alvin Roth.

[2] I contributi più noti sono Rabin (1993) Fehr e Schmidt (2000) Bolton e Ockenfels (2000). L'evidenza sperimentale indica in modo chiaro che i modelli puramente consequenzialisti come quello di Fehr-Schmidt e quello di Bolton-Ockenfels sono inadeguati, mentre modelli più complessi come quello di Rabin funzionano meglio, pagando però un costo di indeterminazione predittiva in un vasto insieme di giochi.

[3] Vedi Guala (2006).

Roth, Prasnikar, Okuna-Fujiwara e Zamir (1991) intendevano controllare due ipotesi: (1) la concorrenza è importante?, e (2) ci sono differenze rilevanti nella contrattazione in culture diverse? Per controllare la prima ipotesi essi disegnarono, accanto al GU classico, un mercato nel quale numerosi compratori competono per acquisire un bene posseduto da un singolo venditore. Sia il GU che il gioco di mercato hanno equilibri estremamente asimmetrici (nei quali un giocatore si aggiudica quasi tutta la torta). Però l'equilibrio egoista si ottiene soltanto nelle condizioni di mercato. La concorrenza ha l'effetto di "spazzare via" le considerazioni di equità.

L'esperimento di Roth è uno degli articoli più citati nell'economia sperimentale. Incorpora in un solo articolo quelli che sono generalmente considerati i tre più importanti risultati dell'economia sperimentale: la scoperta che (1) il modello dell'*homo oeconomicus* egoista non riesce a spiegare un gran numero di osservazioni; che (2) il modello della razionalità egoista può nondimeno rendere conto del comportamento in un ampio raggio di situazioni sperimentali; e infine, combinando questi due risultati, che (3) "le istituzioni sono importanti".

Vale la pena distinguere fra due tipi di istituzioni: *norme* informali come quelle di giustizia, uguaglianza, cooperazione, e reciprocità che governano il nostro comportamento in molte circostanze (come il GU); e *regole* esplicite di scambio, aggregazione, trasmissione dell'informazione, ecc. che sono tipiche di istituzioni di mercato relativamente formalizzate. L'importanza delle regole formali di scambio emerge nell'articolo di Roth dal confronto fra la contrattazione nel GU e la concorrenza di mercato. L'importanza delle norme culturali invece è messa in risalto dalle repliche dell'esperimento condotte in quattro diversi paesi (USA, Giappone, Slovenia, e Israele). Roth e colleghi trovarono delle differenze di comportamento significative: i giapponesi e gli israeliani tendono a offrire di meno (l'offerta modale è al 40%) degli americani e degli sloveni, ma le offerte inique a Tokio e Gerusalemme sono rigettate con la stessa probabilità delle offerte 50/50 di Pittsburgh e Lubiana. Nel testare un'ipotesi teorica, per la prima volta il GU veniva anche usato come uno *strumento di misurazione*.

5. MISURAZIONE

Nel 2000 l'antropologo Joe Henrich ha riportato una serie di osservazioni "sorprendenti" ottenute usando il GU fra i Machiguenga, un gruppo di indigeni peruviani. I Machiguenga offrivano divisioni della torta *più inique*, e rifiutavano le offerte inique *meno* spesso dei soggetti occidentali. In altri termini: si comportavano più spesso (anche se non sempre) come degli "uomini economici". I risultati dei Machiguenga potevano apparire anomali soltanto su uno sfondo di conoscenza sperimentale data per acquisita, cioè dopo che i risultati del GU nelle società occidentali erano diventati regolarità "standardizzate" e generalmente accettate. Nei primi anni '80 il comportamento dei Machiguenga sarebbe stato considerato probabilmente poco sorprendente e molto *meno* anomalo (in quanto più vicino alla previsione teorica) di quanto non appaia ora.

Henrich concluse che popoli diversi hanno diverse aspettative culturali e norme di giustizia, e che dobbiamo studiare come queste norme vengono create e sostenute in ciascun contesto sociale. Con questa idea, un gruppo di economisti e antropologi si è proposto di confrontare il comportamento in alcuni esperimenti classici (specialmente il GU e il gioco dei beni pubblici) fra quindici "micro-società" in Sud America, Africa, e Asia. Questo progetto, finanziato dalla MacArthur Foundation, è il tentativo più ambizioso ed esaltante di usare l'economia sperimentale a scopi di misurazione che sia mai stato tentato fino a oggi (vedi Henrich et al. 2004).

Il progetto MacArthur ha diverse caratteristiche metodologiche interessanti. La sperimentazione si combina con un profondo studio etnografico del contesto sociale (strutture politiche ed economiche, credenze religiose, riti, pratiche di divisione del cibo, ecc.). Queste informazioni illuminano alcuni risultati "sorprendenti" (deviazioni dalle previsioni standard) in un modo che è inusuale per le indagini sperimentali classiche. Ma soprattutto, esse sottolineano l'importanza della standardizzazione per il consolidamento dei disegni sperimentali paradigmatici, che possono essere utilizzati come strumenti portatili per misurare fenomeni sperimentali in situazioni diverse.

6. STANDARDIZZAZIONE

La standardizzazione è il processo di consolidamento di un disegno sperimentale paradigmatico. La standardizzazione è determinata da diversi fattori, che possono avere un peso diverso nei vari stadi di un programma di ricerca.

(1) *Imitazione*: come ha enfatizzato Thomas Kuhn, le tecniche di una disciplina scientifica vengono spesso apprese replicando gli esemplari della generazione di scienziati precedente.

(2) *Test di robustezza*: la logica stessa della variazione controllata richiede che i nuovi disegni siano confrontati con una base. Quest'ultima può essere una previsione teorica, ma spesso è semplicemente un risultato sperimentale "classico" che di conseguenza continua a essere replicato indefinitamente.

(3) *Coesione disciplinare*: molte discipline a un certo punto accolgono e impongono delle "buone pratiche" di ricerca sperimentale. Pratiche come l'uso degli incentivi monetari, l'anonimità, la ripetizione, e il divieto di ingannare i soggetti fanno parte di questa categoria nel caso dell'economia sperimentale.[4]

Attraverso questi processi, la standardizzazione porta alla stabilizzazione dei fenomeni, e alla relativa stabilizzazione di un disegno sperimentale. [5]Una volta che un disegno è stato standardizzato, alcuni dei suoi dettagli perdono l'originaria giustificazione metodologica e sono conservati principalmente per ragioni pragmatiche: diventano dei mezzi per ottenere la confrontabilità fra diverse popolazioni di soggetti.

7. VALIDITÀ ESTERNA

Si dice spesso che affrontiamo un GU ogni volta che entriamo in un grande magazzino. Ci viene offerto un prezzo, senza l'opportunità di contrattare. Eppure in un'economia di mercato non percepiamo quasi mai i prezzi come delle offerte del tipo "prendere o lasciare". Abbiamo solitamente altre opzioni, come per esempio entrare in un negozio che vende beni simili e controllare se ci offre dei prezzi migliori. Quando accettiamo un prezzo senza arrabbiarci, è perché crediamo (o assumiamo) che il prezzo risulti da un processo equo di concorrenza di mercato, e che sia più o meno il meglio che ci possa essere offerto data la produzione e altri

[4] Alcune di queste pratiche sono formalizzate nei cosiddetti "precetti dell'economia sperimentale" (Smith 1982) mentre altri sono giustificati in modo informale. Guala (2005, Cap. 11) include un'illustrazione e discussione più dettagliata.

[5] Sugden (2005) cattura entrambi questi aspetti in modo elegante per mezzo del concetto di "esemplare sperimentale" (*experimental exhibit*): un fenomeno accoppiato a un disegno standard, come il modellino di una nave in una bottiglia. Boumans (2005) sottolinea correttamente l'altro aspetto della medaglia: talvolta la robustezza di un fenomeno porta alla standardizzazione del disegno – il disegno che meglio cattura il fenomeno nella sua manifestazione "pura".

costi. (Ovviamente questo non è sempre vero, e, infatti, le associazioni dei consumatori, gli osservatori dei prezzi, e i boicottaggi dei consumatori hanno un ruolo importante nelle economie di mercato.)

Considerazioni di equità emergono in modo diverso in diversi contesti, e il GU può al massimo essere visto come un esemplare rappresentativo di un ampio raggio di situazioni nelle quali le norme di equità giocano un ruolo più o meno diretto nel comportamento economico e sociale. Questo è particolarmente rilevante per il progetto della MacArthur Foundation. Alcune delle piccole società avevano scarsa esperienza delle transazioni di mercato, alle volte erano perfino poco abituate all'uso del denaro. Quando i disegni sperimentali facevano scattare delle norme locali era più per caso che per intenzione dello sperimentatore.[6]

La validità esterna è il problema di generalizzare i risultati sperimentali dalle condizioni di laboratorio ad altre situazioni che ci interessano. È solitamente considerato il "tallone di Achille" della sperimentazione nelle scienze sociali, anche se la sua rilevanza per gli esperimenti nelle scienze naturali è una questione per lo più inesplorata – e forse sottovalutata. Ho sostenuto altrove (Guala 2005, Cap. 7-11) che la questione della validità esterna deve essere risolta caso per caso. Non ha senso dubitare della validità esterna di un esperimento *in generale*. Bisogna prima chiedersi quale sia l'obiettivo dell'esperimento – quale sistema non-sperimentale si intendeva studiare in origine. Spesso gli esperimenti sono fatti senza un obiettivo esterno in mente, e dunque la validità esterna non è il loro principale scopo o preoccupazione.

Intuitivamente la standardizzazione sembra essere nemica della validità esterna. Se il mondo economico e sociale è variegato e complesso, abbiamo bisogno di flessibilità invece che di rigidità. Dobbiamo ritagliare i nostri disegni sperimentali intorno alle specifiche domande poste dal problema che stiamo studiando, piuttosto che il contrario. Ma allora perché la consolidazione di un formato rigido per il GU è stata utile e ne spiega il successo? Perché, come vedremo, i criteri metodologici che si applicano alla misurazione sono diversi da quelli che si applicano al controllo delle teorie.

8. OSSERVARE E INTERVENIRE

I filosofi "neo-sperimentalisti" hanno identificato l'intervento materiale come la *differentia specifica* fra la scienza sperimentale e non-sperimentale. Secondo una versione di questo approccio, sviluppata da Jim Woodward (2003) l'intervento è funzionale alla scoperta di *relazioni causali robuste*. Un intervento, secondo Woodward, è una manipolazione "chirurgica" di una variabile che lascia il resto del sistema sperimentale intatto. Consideriamo la Figura 1. Supponiamo di volere sapere se X causa Y, Y causa X, oppure se le due variabili sono correlate in modo spurio. In un intervento "chirurgico" (I) X è manipolata in modo tale da tagliare tutti i collegamenti con i suoi antecedenti causali, ma nessuno dei suoi effetti putativi è direttamente influenzato dalla manipolazione. Qualsiasi mutamento di Y dovuto a I deve passare attraverso X, in altre parole. Se l'intervento ha successo, e X causa Y, allora la variazione in X si rifletterà in una variazione in Y. Se al contrario Y causa X (il caso di mezzo nella Figura 1) oppure le due variabili sono associate soltanto in modo spurio (caso a destra) la correlazione fra X e Y sarà interrotta.

[6] Per un esempio evidente, cfr. Ensminger (2004) sull'istituzione dell' "*harambee*" fra il popolo Orma del Kenia.

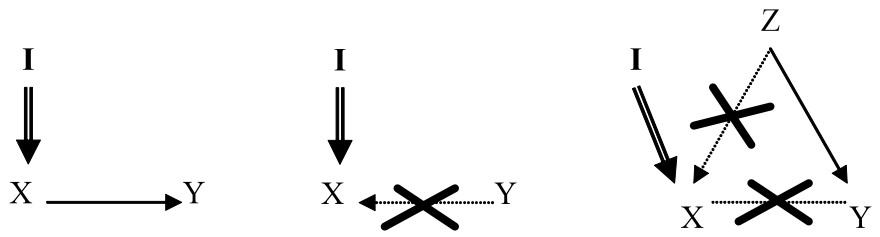


Figura 1

Usando il GU come uno strumento di misurazione, non confrontiamo due (o più) stati di una variabile che possiamo fissare a piacimento. Sicuramente stiamo confrontando *qualcosa* – il numero di rifiuti di offerte inique nelle società A e B, per esempio – ma si assume che il valore della variabile sia determinato da qualche fattore nascosto che non possiamo controllare direttamente. In effetti è estremamente importante che le procedure sperimentali non interferiscano con il funzionamento di questi fattori nascosti, perché altrimenti la misurazione sarebbe considerata "artificiale" e inutile.

Quando il GU è usato come strumento di misurazione, ovviamente, in un certo senso stiamo "controllando" qualcosa. Stiamo controllando l'ipotesi che due insiemi di osservazioni (raccolte negli Stati Uniti e fra i Machiguenga, per esempio) provengano dalla stessa popolazione o processo di generazione dei dati. È proprio per controllare questa proposizione che dobbiamo mantenere il disegno sperimentale il più stabile (o "standardizzato") possibile. Altrimenti, non sarà possibile attribuire differenze nei comportamenti osservati a delle variazioni in qualche cosa di *diverso* (qualunque cosa esso sia) dal disegno sperimentale stesso.

Un'analogia con la termometria può essere utile: immaginiamo di volere dimostrare che l'acqua bolle a temperature diverse al livello del mare e su un altipiano. Abbiamo bisogno di un termometro standardizzato e portatile, che sia robusto a variazioni nell'ambiente esterno quando è trasportato da un sito di misurazione a un altro. Poiché non vogliamo interferire con la variabile che stiamo misurando (la temperatura) saranno permesse soltanto manipolazioni che sono funzionali all'ottenimento di una misura oggettiva e standardizzata che possa essere usata per la previsione e il confronto. Storicamente, la temperatura è stata misurata per molto tempo in modo indipendente da una comprensione anche lontanamente corretta della natura del calore (Chang 2004). Il primo passo nello sviluppo della termometria è la costruzione di un *termoscopio*. Un termoscopio misura la temperatura su una scala ordinale, e può essere usato per determinare uno o più *punti fissi* – per esempio il punto di congelamento e di ebollizione dell'acqua. Una volta fissati questi punti, possiamo costruire un *termometro* che misuri la temperatura su una scala cardinale. Storicamente, un "buon" termoscopio era valutato dalla sua conformità con le sensazioni ordinarie del caldo e del freddo: si deve espandere quando sentiamo caldo e contrarre quando sentiamo freddo. Questo garantisce che stiamo misurando qualche cosa di importante (per noi) ma ovviamente non può garantire che abbiamo incontrato una relazione fisica robusta e significativa.

Lo stesso vale per gli esperimenti economici quando sono usati come strumenti di misurazione. Le cause delle variazioni comportamentali nel GU sono ancora controverse, ed è del tutto possibile che diverse cause siano operanti in diversi contesti sociali. Nonostante questo, il GU può essere usato a scopi di misurazione, se si è ottenuta una standardizzazione appropriata e l'esperimento è stato fatto con cura. In questo caso possiamo concludere che qualsiasi differenza venga trovata, essa deve essere dovuta a differenze nelle popolazioni o nei

processi di generazione dei dati. L'inferenza, per usare l'espressione introdotta da Deborah Mayo (1996) è giustificata perché è stata *controllata severamente*.

Un controllo severo produce dati di un certo tipo D1 se l'ipotesi sotto controllo è vera, e un altro tipo di dati D2 se l'ipotesi è falsa. In termini statistici più familiari, esso minimizza la probabilità di commettere errori di tipo I e II. Il controllo severo, secondo Mayo, è il tratto distintivo della sperimentazione, perché i buoni esperimenti tipicamente supportano inferenze che sono state controllate severamente dall'esperimento stesso.[7] La caratterizzazione di Woodward degli esperimenti volti alla scoperta causale è per lo più coerente con questa concezione. Torniamo alla Figura 1. Se l'intervento ha successo, e X causa Y (caso a sinistra) allora le variazioni in X saranno riflesse in variazioni in Y. Se invece Y causa X o le due variabili sono associate in modo spurio (al centro e a destra) la correlazione fra X e Y scomparirà. In un esperimento genuino e ben fatto, in altri termini, si osserverà un certo tipo di dati se è vero che "X causa Y", e un altro tipo di dati se non lo è.[8] Qualsiasi conclusione causale arriveremo a formulare, essa sarà severamente investigata o controllata secondo il criterio di Mayo.

9. DATI SPERIMENTALI E RACCOLTI SUL CAMPO

Un disegno standard permette alle differenze fra diverse situazioni di emergere. Importa poco quali siano queste differenze. Potrebbero essere dovute a fattori locali, non controllati (abitudini, norme, disposizioni psicologiche); esse sarebbero nondimeno interessanti, perché queste differenze sono proprio quello che stiamo cercando. Infatti una delle lezioni più importanti degli studi interculturali riguarda la difficoltà di generalizzare su tutte le varie popolazioni

Gli studi della MacArthur Foundation mostrano che un comportamento più o meno cooperativo è correlato con diversi fattori in società diverse (ricchezza, genere, età, strutture politiche, esposizione ai mercati, ecc.). Ancora più importante è la dimostrazione che tali relazioni non sarebbero *mai* state scoperte senza la conoscenza etnografica profonda fornita dagli antropologi sul campo. Nel suo studio degli Hadza, un gruppo di cacciatori-raccoglitori in Tanzania, Marlowe (2004) per esempio riporta un grado sorprendentemente basso di condivisione e cooperazione, in una società altrimenti profondamente egalitaria. La spiegazione proposta si appella a una sottile distinzione fra ciò che si vede (per esempio la cacciagione di grossa taglia) e ciò che può essere facilmente nascosto (come il denaro). Le norme degli Hadza impongono la condivisione di ciò che si vede, ma non proibiscono effettivamente il possesso e il consumo individuale di beni più piccoli. L'avarizia è tollerata quando non può essere sanzionata in modo efficace, mentre il controllo sociale si applica alla caccia grossa.

Patton (2004) riporta un livello di comportamento cooperativo più alto fra gli Achuar che fra i Quichua, due gruppi che vivono molto vicini nella foresta amazzonica ecuadoriana. Queste relazioni si osservano sia nei dati sperimentali sia nelle pratiche di condivisione della carne osservate nelle due comunità. Tuttavia, le differenze all'interno dei gruppi mostrano che la maggiore generosità degli Achuar è quasi interamente dovuta a una piccola elite di capi che la usano per scopi politici, in particolare per creare colazioni forti e stabili fra famiglie e villaggi. I Quichua, che hanno una struttura politica meno stabile e fra i quali le coalizioni tendono a essere più effimere, semplicemente non hanno prodotto una simile classe di "condivisori politici". La conoscenza della struttura di una società e delle sue norme è cruciale, perché le

[7] O, più precisamente, inferenze a ipotesi che sono state controllate severamente.
[8] Vedi anche Woodward (2000).

correlazioni "macro" sono spesso ingannevoli o inutili. Ogni caso è diverso; ogni spiegazione è "locale".[9]

10. TERMOMETRI SOCIALI

Un termometro è l'inizio di una diagnosi, non la fine della storia. Misura un sintomo (un sintomo importante e ubiquo) non una causa. Questo non lo rende meno utile – al contrario, staremmo molto peggio senza i termometri. Come un buon termometro, il GU è diventato uno strumento portatile che possiamo trasportare di cultura in cultura per misurare un insieme di fenomeni che riteniamo siano associati alle norme sociali. Quali siano questi fenomeni è ancora controverso. A differenza della temperatura, dove il termometro può essere confrontato con le sensazioni pure del calore e del freddo, non abbiamo delle basi simili per il GU. "Sentiamo" che il GU può essere usato per misurare ciò che appare giusto o ingiusto, ma il nostro concetto di giustizia non è necessariamente una buona guida per comprendere quelli che prevalgono nelle società che stiamo studiando. La nostra concezione dell'equità può funzionare come punto di paragone, tuttavia, e abbiamo proprio bisogno di qualcosa del genere per la misurazione. Il GU è emerso attraverso un processo di selezione sociale nella teoria dei giochi sperimentale, come un disegno robusto che cattura qualche cosa di molto importante per noi. Non abbiamo nessuna garanzia che sia lo strumento di misurazione migliore per studiare ciò che è importante per chi vive in società diverse dalla nostra. Ma è un punto di partenza, e la scienza sociale ha molto bisogno di una qualche piattaforma stabile perché possa decollare la ricerca.

BIBLIOGRAFIA

Bolton G. e Ockenfels A.(2000) ERC: A Theory of Equity, Reciprocity and Cooperation, *American Economic Review* 90: 166-93.

Boumans M. (2005) *How Economists Model the World into Numbers*. Londra: Routledge.

Chang, H. (2004) *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.

Ensminger J. (2004) Market Integration and Fairness: Evidence from Ultimatum, Dictator, and Public Goods Experiments in East Africa, in Henrich et al. (2004) 356-81.

Fehr E. e Schmidt, K. (1999) A Theory of Fairness, Competition and Co-operation, *Quarterly Journal of Economics* 114: 817-68.

Gil-White F. (2004) Ultimatum Game with an Ethnicity Manipulation: Results from Khodovin Bulgan Sum, Mongolia, in Henrich et al. (2004) 260-304.

Guala F.(2005) *The Methodology of Experimental Economics*. New York: Cambridge University Press.

Guala F. (2006) Has Game Theory Been Refuted?, *Journal of Philosophy* 103: 239-263.

Gurven M. (2004) Does Market Exposure Affect Economic Behavior? The Ultimatum Game and the Public Goods Game among the Tsimane' of Bolivia, in Henrich et al. (eds. 2004) 194-231.

[9] Si vedano anche Gurven (2004) Tracer (2004) e Gil-White (2004) per altri esempi del genere.

Güth W., Schmittberger R. e Schwarz, B. (1982) An Experimental Analysis of Ultimatum Bargaining", *Journal of Economic Behavior and Organization* 3: 367-388.

Henrich J. (2000) Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon", *American Economic Review* 90: 973-79.

Henrich J., Boyd R., Bowles S., Camerer C., Fehr, E. e Gintis H. (eds. 2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies.* Oxford: Oxford University Press.

Marlowe F. (2004) Dictators and Ultimatums in an Egalitarian Society of Hunter-Gatherers: The Hadza of Tanzania, in Henrich et al. (2004) 168-193.

Mayo D. (1996) *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Patton J. (2004) Coalitional Effects on Reciprocal Fairness in the Ultimatum Game: A Case from the Ecuadorian Amazon, in Henrich et al. (2004) 96-124.

Rabin M. (1993) Incorporating Fairness into Game Theory and Economics, *American Economic Review* 83: 1281-302.

Roth A., Prasnikar V., Okuno-Fujiwara M. e Zamir, S. (1991) Bargaining and Market Behavior in Jerusalem, Lubljana, Pittsburgh and Tokyo: An Experimental Study, *American Economic Review* 81: 1068-1095.

Smith V. (1982) Microeconomic Systems as an Experimental Science, *American Economic Review* 72: 923-55.

Tracer D. (2004) Market Integration, Reciprocity, and Fairness in Rural Papua New Guinea: Results from a Two-Village Ultimatum Game Experiment, in Henrich et al. (2004) 232-259.

Sugden R. (2005) Experiments as Exhibits and Experiments as Tests, *Journal of Economic Methodology* 12: 291-302.

Woodward J. (2000) Data, Phenomena, and Reliability, *Philosophy of Science* 67: S163-179.

Woodward J. (2003) *Making Things Happen*. New York: Oxford University Press.

Scientific Unification in Economics. The case of the New Economic Geography

Caterina Marchionni[*]
caterina.marchionni@helsinki.fi

ABSTRACT

Scientific unification – the application of the same principles and tools to the study of phenomena from different domains – is a powerful ideal in economics that significantly shapes its internal dynamics and how it relates to neighbouring fields. In this paper I discuss aspects of scientific unification in the case of the New Economic Geography, an approach to spatial issues recently developed within economics. The analysis shows that the drive to unify affects a range of inter-theoretic, inter-field and inter-disciplinary relations, and that it meets with disunities of various sorts, which affect the kind of products unification ultimately delivers.

1. INTRODUCTION

Scientific unification – the application of the same principles, explanatory schemata, models and tools to the study of phenomena in different domains – has been and still is a powerful scientific ideal. Celebrated examples include Newton's theory of universal gravitation, which unified celestial and terrestrial phenomena, Einstein's theory of special relativity, which unified electric and magnetic phenomena, and Darwin's theory of natural selection, which unified phenomena in fields such as comparative anatomy, biogeography, and palaeontology. In economics too scientific unification is a powerful force that significantly affects its own internal dynamics as well as its relationship to neighbouring fields. Consider for example the widespread adoption of economics-style models and explanations in neighbouring social sciences like sociology and political science. The search for unity partly lies behind the emergence of new fields such as neuroeconomics, which integrates economics, psychology and neuroscience to study how the brain makes economic choices; network theory, which studies the properties of networks throughout the social and the natural sciences, and in its economics application it is used to explain phenomena as diverse as scientific collaborations, variations in crime, and differences in social cooperation; or the New Economic Geography, whose unificationist achievements have been celebrated in the awarding of the 2008 Nobel Prize in economics to its founding father, Paul Krugman.

The general intuition behind the ideal of scientific unification is that all sciences are after all concerned with one and the same world. The unification of the entire scientific domain either by means of a grand single theory, or of a small set of fundamental laws is probably unattainable. Nevertheless, dealing with complex interrelations between phenomena seems to require at least some degree of unity and integration of scientific knowledge. Take social phenomena for example. These are closely interconnected through a complex thicket of causal relations. In order to understand them, social scientists might often be forced to transcend disciplinary boundaries. And at least prima facie, it seems plausible that biological theories

[*] Department of Social and Moral Philosophy, University of Helsinki

should also somehow bear on social scientific theories, for the individuals that constitute the social world are after all also biological systems.

In the philosophy of science unification has traditionally been a hot topic of debate (see Bechtel and Hamilton 2008 and Cat 2007 for overviews). Modern discussion began with the classic paper by Oppenheim and Putnam (1958), which envisaged the unity of science in the form of a reduction of all the sciences to physics. More recent accounts have suggested that scientific understanding itself is a product of the unification of disparate phenomena, and accordingly take science to progress by reducing the number of types of facts that we have to accept as ultimate (Friedman 1974; Kitcher 1981). Today many philosophers of science have dismissed these and similar proposals as inadequate both descriptively and normatively.

In spite of the dissatisfaction among philosophers however, scientists continue to pursue unification in some form, and when successful these achievements are widely acclaimed. The philosophical project of finding a one-size-fits-all model has failed not because unification is unimportant, but because it is a local, varied and context-dependent phenomenon. The real challenge for philosophers of science now is to understand how and why scientists engage in various kinds of unifying research strategies; strategies that might prove fruitful in some contexts and not in others, and only for some purposes (e.g., explanation, confirmation, and discovery). Some have already begun to analyze reductionist and unifying strategies from this perspective (e.g., Grantham 2004, Wimsatt 2006, Wylie 1999), but thus far very little attention has been paid to unification in economics. If, as it seems increasingly clear, it assumes different features in different scientific contexts, in economics it is likely to be somehow different from unification in say, biology. We might discover that the unifying strategies deployed within and around economics are misguided. It's more likely, however, that we'll find that they work only sometimes and only for some purposes, and we might be able to understand why they work when they do.

My focus here is on the features of scientific unification in the case of the New Economic Geography. What emerges is that unification drives theoretical development and shapes a variety of inter-theoretic, inter-field and inter-disciplinary relations. At the same time, it co-exists with disunities of various sorts that affect the kind of products that are ultimately delivered. Similar ideas, I believe, could be applied beyond the case under scrutiny. At the end of the paper I briefly speculate on how the pursuit of scientific unification and the sorts of friction it faces play out in the case of neuroeconomics. The objective is to make a case supporting the claim that unification is a powerful driving force in economics and that it takes place often enough to deserve philosophical attention.

2. VARIETIES OF SCIENTIFIC UNIFICATION

Scientific unification encompasses a set of activities that yield the unification or integration of scientific knowledge. I propose to distinguish four levels at which unification can take place, which correspond to items that either enter in the process of unification or are affected by it: phenomena, theories, fields and disciplines. Although differentiating these levels in practice involves some degree of arbitrariness, I believe that each level brings to light important features of scientific unification (though obviously not all four levels will be involved in every case). My case study also shows that, at each level, distinctive frictions influence and constrain the kind and degree of unification established at that level.

First, *explanatory unification* concerns phenomena. It is a matter of explaining several classes of explanandum phenomena with the same, or fewer, explanantia. According to Friedman (1974) and Kitcher (1981), scientific explanation amounts to unification, and hence it

is a matter of deriving many explananda from a small set of premises, or argument patterns. On this view, the explanatory power of a theory is a function of its unifying power and science progresses by reducing the number of types of facts we have to accept as ultimate. In order to appreciate the value of unified explanations however we don't need to endorse Friedman's and Kitcher's proposals in their entirety. It's indeed possible to view unifying power and explanatory power as distinct and still appreciate explanatory unification. For instance, we can hold that unified explanations are to be pursued when and only insofar as they reflect a deep ontological unity among apparently different kinds of phenomena (Mäki 2001): when two phenomena, say optical processes and electromagnetic processes, are identified as being of the same kind. More modestly, unified explanations sometimes make salient commonalities and general patterns, which would be concealed by disunified explanations that focus on the details specific to particular occurrences (e.g. Sober 1999).[1]

Second, *theoretical unification* is a matter of unifying previously separate theories by means of one that possesses all (or most of) their explanatory content. [2] The paradigmatic cases of unification in the history of science have often accomplished unification of both phenomena and theories. For instance, Maxwell's theory unified electromagnetism and optics in showing that electromagnetic waves and light waves were one and the same thing. Similarly, Newton's theory unified Galileo's laws of terrestrial mechanics and Kepler's laws of celestial bodies by showing that the motions of celestial and terrestrial bodies could be explained through the law of universal gravitation. What counts as a class of phenomena is defined by a theory, so that a theory that successfully unifies two classes of phenomena (henceforth the unifying theory) also unifies the respective theories (henceforth the unified theories).

When it is a question of more mundane cases the distinction between the unification of phenomena and of theories nonetheless still proves useful. The reason is that a unifying theory might account for only a subset of the phenomena that the unified ones, taken together, explain. In many cases, a unifying theory includes a set of idealizations and abstractions that differ from that of the unified theories. Each theory, whether unifying or unified, may because of its distinctive idealizations and abstractions be useful for different explanatory purposes. In such cases, explanatory unification may not produce a complete theoretical unification. Theoretical unification turns out to be a matter of degree.

Third, scientific unification can also occur at the level of fields. Let's call this *inter-field integration.* Focusing on the interconnections between fields rather than exclusively on those between theories will produce a more general framework for examining scientific unification. Fields typically include several theories and much more besides:

> A central problem, a domain consisting of items taken to be facts related to that problem, general explanatory facts and goals providing expectations as to how the problem is to be solved, techniques and methods, and sometimes, but not always, laws and theories which are related to the problem and which attempt to realize the explanatory goals (Darden and Maull 1977: 4)

[1] Consider an explanation of an increase in crime rate, which appeals to a higher level of unemployment and abstracts away from the changes of opportunities and motivations of the individuals involved. To see why this is a unified explanation, suppose we have two regions both experiencing a rise in the crime rate as the consequence of an increase in the level of unemployment. The unemployment explanation provides a unified explanation of the two occurrences, whereas the explanation that invokes the changes of opportunities and motivations of the individuals involved would be different in the two cases.

[2] According to Whewell (1847), theories that unify phenomena are also more likely to be true (see also Thagard 2007).

The unification of fields occurs via a number of routes, of which the unification of theories is just one. For instance, it may happen through the development of an 'inter-field theory', in other words a theory that identifies and explains the relations between fields (Darden and Maull 1977), or through the development of a theory that explains the presuppositions of another without replacing it (Kitcher 1984). Fields could also become progressively integrated through the establishment of non-theoretical connections, which Grantham (2004) calls 'practical' unification. This takes place when "one field comes to rely on the methods, heuristics, or data of a neighboring field" (Grantham 2004: 143), such as when the theories or methods developed in one field generate new hypotheses in another, or when data from one field are used to confirm hypotheses developed in a neighbouring field. For the data, methods or heuristics that are distinctive of one field to bear on a neighbouring field there needs to be at least some theoretical understanding of how the two domains are related, but practical unification can occur also with a relatively low degree of theoretical unification. Various practices can therefore contribute to the unification of fields (and thereby of science), although they can hardly, if ever, effect a complete inter-field unification, that is, one in which the new field supplants previously existing ones.

Finally, the pursuit of unification may also affect inter-disciplinary relationships: a theory may enter a domain of phenomena that fall within the purview of a discipline distinct from that in which it was originally developed. This phenomenon is known as *scientific imperialism*. Talk of disciplines rather than fields allows us to emphasize the sociological and institutional aspects that contribute to define disciplinary identities and boundaries, and which are absent from the definition of a field. Although the term 'scientific imperialism' has negative connotations, following an idea advanced by Uskali Mäki (2009), we can distinguish forms of scientific imperialism, some of which can be positively evaluated. If we think of scientists as seeking to increase a theory's degree of unification by way of applying it to new types of phenomena, it is largely a matter of social and historical contingency whether these phenomena are studied within other disciplines. If we endorse the idea that the scope of theories should be determined by the real order of things and not by pre-existing disciplinary boundaries, then it follows that in principle there is nothing wrong with that kind of imperialism that follows from the expansion of a theory's scope. But this is not all there is to scientific imperialism. Scientific imperialism can also affect other aspects of the scientific endeavour such as the standing of one discipline and/or conceptions about the scientific status of one style of research vis-à-vis another. Whereas expanding the scope of a theory in search of real unity among phenomena is potentially progressive, entering another discipline in the attempt to change its theoretical convictions and style of inquiry is a more suspicious form of scientific imperialism. The latter can create or reinforce positions of dominance that have little or nothing to do with recognized scientific merits. Whether breaking disciplinary boundaries turn into suspicious forms of imperialism can only be ascertained on a case-by-case basis.

3. UNIFICATION IN THE NEW ECONOMIC GEOGRAPHY

The New Economic Geography (NEG henceforth) is an approach developed within economics at the beginning of the 1990s that seeks to explain spatial agglomeration (and spatial dispersion) within a unified framework (Krugman 1991, Fujita et al. 1999, Fujita and Thisse 2002, Brakman et al. 2001, Baldwin et al. 2003). The concept of agglomeration refers to seemingly very distinct empirical phenomena: the existence of the core-periphery structure

corresponding to North-South dualism; regional disparities within countries; the existence of cities and systems of cities, which are sometimes specialized in a small number of industries; industry clusters such as Silicon Valley; and finally the presence of commercial districts within cities, such as Soho in London. Although each type of agglomeration could be the result of different types of mechanisms, NEG hypotheses that these apparently distinct phenomena are at least partly the result of similar mechanisms, viz. 'economic mechanisms yielding agglomeration by relying on the trade-off between various forms of increasing returns and different types of mobility costs' (Fujita and Thisse 2002: 1).

The approach rests on two building blocks: the presence of increasing returns at the firm level and transportation costs.[3] Increasing returns at the firm level requires dropping the standard economic assumption of perfect competition and replacing it with that of imperfect competition.[4] On the aggregate level, increasing returns and transportation costs give rise to pecuniary externalities, which are transmitted through the market via effects on prices. In essence, their presence implies that the more firms and workers there are in a locality, the larger the market, the more attractive it becomes as a location for further firms and workers.[5] This generates a cumulative process that may result in the concentration of all economic activity in one location. While pecuniary externalities push towards such concentration, immobile factors, congestion and the like, push towards dispersion. The relative strength of the two sets of forces determines whether spatial agglomeration occurs.

NEG is now a well-established field in economics. In 2008 Paul Krugman was awarded the Nobel Prize in Economics. The Nobel Prize press release mentioned the integration of the fields of trade and geography as one of the remarkable achievements of Krugman's theory:

> Patterns of trade and location have always been key issues in the economic debate. What are the effects of free trade and globalization? What are the driving forces behind worldwide urbanization? Paul Krugman has formulated a new theory to answer these questions. He has thereby *integrated the previously disparate research fields of international trade and economic geography*. (My emphasis) [6]

For NEG theorists too, unification is one of its main accomplishments: it shows that similar kinds of economic mechanisms are at work in bringing about a host of phenomena that have been separately studied, thereby achieving explanatory unification. In doing so, it answers Ohlin's call (1933) for the unification of trade theory and location theory (Brakman et al. 2001). And yet this enthusiasm for the NEG's achievements is not shared by all. Economic geographers, who have been traditionally concerned with spatial issues, received the NEG's

[3] The presence of space implies that the cost of exchanging goods and services across locations increases as (both physical and cultural) distance increases due to physical transport, tariffs, cultural barriers and so on. These costs are generically treated as *transportation costs*.

[4] *Increasing returns* are defined as a decrease in the average costs per unit of output for the individual firm as the level of output increases. In the imperfect competition setting NEG assumes, firms take the price-setting behavior of other firms as given, and do not take into account the effects of changing their own price onto the price index. The products are symmetric, which means that consumers do not prefer one variety to another. However, they prefer variety for its own sake, which means that they always prefer to consume a unit of a new variety than an additional unit of a product they have already consumed. This is one of the reasons why firm agglomeration is beneficial for workers.

[5] *Externalities* are defined as a decrease in average costs as a result of an increase at the level of output of the whole industry. *Pecuniary externalities* are transmitted through the market via price effects for each firm, which, as a consequence, may decide to change its output decisions.

[6] The Nobel Prize press release can be found at http://nobelprize.org/nobel_prizes/economics/laureates/2008/press.html.

appearance with suspicion, if not with outright hostility. Among the accusations economic geographers have levelled against NEG is that it represents yet another instance of imperialism on the part of economics. We can already begin to see that the NEG's pursuit of explanatory unification has affected a range of inter-theoretic, inter-field and inter-disciplinary relations. It is to these issues that the remainder of this section is devoted.

3.1. *EXPLANATORY UNIFICATION*

NEG seeks to explain a host of spatial phenomena previously studied by separate theories in different fields by showing how these are brought about by the same kind of economic mechanisms. The account of explanatory unification advanced by Philip Kitcher (1981, 1989) largely captures the structure of explanatory unification in NEG (Mäki and Marchionni 2009). According to Kitcher, to explain is to unify, and unification (and hence explanation) is a matter of inference and derivation. Explanatory unification proceeds by reducing the number of argument patterns while maximizing the number of explanandum sentences that can be derived.[7] In economics, argument patterns are typically embodied in *model types*, so that unification therein proceeds via the application of a small number of similar model types to an increasing number of economic as well as non-economic phenomena.

In the case of NEG two similar model types have effected the unification: the core-periphery model (henceforth the CP model) and the vertical-linkages model (henceforth the VL model). Both of these adopt the monopolistic competition framework set out in the Dixit and Stiglitz (1977) model with transportation costs, and derive the agglomeration of economic activity between two a priori identical locations. The difference between them lies in the specific foundations for the agglomeration mechanism they postulate; in other words, each type specifies more specific versions of the generic mechanism described in the previous section. In the CP model, the size of the market in each location is determined by the migration decisions of workers: a larger market is a more attractive location to firms and hence to workers via a reduction in the price of the goods. In the VL model, on the other hand, the workers do not move, and the size of the market is made endogenous by the presence of input-output linkages between firms: the more firms there are in a location the larger is the market for upstream firms and the lower are the costs for downstream firms. These model types are then filled in with specific variables in order to derive diverse types of agglomeration phenomena.

Two features of NEG however diverge from Kitcher's account of explanatory unification. First, in Kitcher's, unification is a matter of deriving large classes of explanandum sentences from a parsimonious set of inferential patterns. But explanatory unification can also be seen as a matter of redescribing a large number of apparently independent phenomena as manifestations of a common set of mechanisms, thereby revealing an underlying ontic unity between the phenomena (Mäki 2001). In NEG the two kinds of unification go together. The kind of unity NEG pursues has to do with different phenomena being governed by similar kinds of economic mechanisms. The CP and the VL model types are not merely similar patterns of

[7] Argument patterns include three components: 1) a schematic argument: a sequence of schematic sentences, which are expressions in which most if not all non-logical expressions are replaced with dummy letters; 2) a set of filling instructions that indicate how dummy letters are to be replaced in specific applications; 3) a classification for the schematic argument that provides the inferential characteristics of the argument, dividing the schematic sentences between premises, conclusions, and rules of inference (Kitcher 1981).

derivation, but they embody stylized causal mechanisms, which are fleshed out according to the specifics of each explanandum phenomenon.

The second, related feature is that in contrast to Kitcher, NEG theorists endorse the view, widespread among economists, that a phenomenon is genuinely explained only if it is derived from well-defined microeconomic parameters (Mäki and Marchionni 2009, Marchionni 2009). NEG's contribution vis-à-vis existing theories is held to be twofold. First, it unifies phenomena that were previously treated separately: the *unificationist* contribution. Second, it is the only field within economics that provides a micro-foundation in a general equilibrium framework for the spatial distribution of economic activity: the *explanatory* contribution. If genuine explanation in NEG has to do with the presence of micro-foundations, the derivation of diverse phenomena from a small number of model types is not what the theorists take to do the explanatory work. This is so even though it is the search for micro-foundations (the explanatory contribution) that has revealed that different classes of agglomeration phenomena are governed by the same kind of economic mechanisms (the unificationist contribution).

In the case of NEG, explanation and unification of phenomena have proceeded in parallel, but they trace back to distinct components of the theory. Furthermore, for certain kinds of explananda, they may go in opposite directions: the unification of phenomena demands that explanatory information be abstracted away. As the theorists admit, explanatory unification could not be achieved without neglecting the details that are specific to different kinds of agglomeration phenomena:

> By using highly stylized models, which no doubt neglect a lot of specifics about urban/regional/international phenomena, geographical economics is able to show that the same mechanisms are at work at different levels of spatial aggregation. […]
> In order to lay the foundations for a unified approach, there is a price to be paid in terms of a neglect of institutional and geographical details. (Brakman et al. 2001: 323)

As this quote suggests, unification required that a host of institutional and geographical details specific to the phenomenon be left out. But these specific details are needed to fully account for the spatial phenomena that NEG unifies. It follows that when the theory is used to explain particular aspects of specific phenomena (cities for instance), the NEG unifying mechanism alone cannot do all the explanatory work. It needs to be complemented with information specific to the kind of phenomenon to be explained (such as the presence of land developers or the presence of knowledge spillovers).[8] As we will see in the next section, some of the specific information is provided by other economic theories.

3.2. *THEORETICAL UNIFICATION*

Back in 1933, Bertil Ohlin, a well-known international trade theorist, claimed that the separation between international trade theory and location theory was artificial: "International trade theory cannot be understood except in relation to and as part of the general location theory, to which the lack of mobility of goods and factors has equal relevance" (p. 142). Later on he complained that no serious attempt to "build a general location theory and introduce national borders" had been made (Ohlin 1979: 6). Only twelve years later, Paul Krugman (1991) published the seminal NEG model. NEG promises finally to realize Ohlin's dream of a

[8] Mäki and Marchionni (2009) examine this issue in detail by distinguishing kinds of information and kinds of explananda.

general theory of the location of economic activity, as the scientific background report of the Nobel Prize committee notes:

> Traditionally, trade theory and economic geography evolved as separate subfields of economics. More recently, however, they have … *become more and more united through new theoretical insights*, which emphasize that the same basic forces simultaneously determine specialization across countries for a given international distribution of factors of production (trade theory) and the long-run location of those factors across countries (economic geography) (p. 1, my emphasis). [9]

Furthermore, efforts have recently been made to integrate economic growth in the spatial models of NEG. Thus NEG holds out the promise to unify theories of location, international trade and growth.

According to the standard view of theoretical unification, NEG, the unifying theory, will eventually replace international and location theory, and possibly also growth theory. If NEG had all the explanatory content of the unified theories, then they could be dispensed with. But unification in scientific practice does not always look like this. If the explanatory content of the unifying theory only overlaps with and does not fully cover that of the unified theories, then dispensing with the latter amounts to leaving some explananda unexplained. And at least so far NEG cannot account for a number of stylized facts about location, trade and growth. This is mainly because it focuses exclusively on a kind of economic mechanism (pecuniary externalities) that is believed to operate in bringing about diverse classes of phenomena. In reality there are other mechanisms and forces specific to each class of phenomena that NEG does not encompass. The importance of the alternative mechanisms and forces will vary from case to case, so that whereas for certain stylized facts the NEG mechanism will be more important than the specific ones, in other cases the reverse will be the case. Economists, in effect, generally perceive growth theory, trade theory, location theory and NEG as complementary.[10] Within each field different theories in fact postulate different kinds of economic mechanisms to account for their respective phenomena. Dispensing with one theory basically means dispensing with one kind of mechanism and one possible explanation. Depending on the phenomenon we want to explain, a different mechanism or a different combination of mechanisms acting together will be relevant. In principle further developments within NEG and neighbouring fields could produce a grand theory that tells us when, how and which combinations of mechanisms operate in bringing about certain phenomena. As things stand now, however, there is little reason to believe so. What we now have is a plurality of overlapping, interlocking theories in different subfields, which NEG has helped to render more integrated. Here there is clear tension between the opposing forces for and against unity: on

[9] The report can be found at http://nobelprize.org/nobel_prizes/economics/laureates/2008/sci.html. Note that the report talks of economic geography as one of the fields that NEG unifies. But this is misleading because as I'll explain below the field of economic geography distantiates itself quite considerably from both economics and NEG. It is more appropriate to talk of location theory, an interdisciplinary field, involving regional science, urban and regional economics, concerned with modeling the determinants and consequences of the locational decisions of economic agents (see Marchionni 2009).

[10] Growth theory, trade theory, and location theory include a variety of theories and explanatory models. They are more like fields than theories. In this section for simplicity I use these labels to stand for theories. The point I make about theoretical unification however applies to the theories that within each field seek to account for the phenomena NEG unifies (see Brakman et al. 2001 and Marchionni 2009). Section 3.3 shifts the focus from theories to fields.

the one hand the unifying ambitions of NEG push towards the unification of phenomena and theories, and on the other the presence of theories whose domains only partially overlap with that of NEG maintains a certain degree of disunity.

3.3. *INTER-FIELD INTEGRATION*

Looking at inter-field connections offers additional insights into NEG unification. Trade theory, location theory and growth theory are perhaps more appropriately treated as fields rather than theories because they include several explanatory models that aim to account for a number of stylized facts about, respectively, international trade, location and economic growth. From this broader perspective, NEG can be looked upon as an inter-field theory that examines the relations between the phenomena studied by different fields and thereby contributes to render these fields more integrated.

The development of NEG, in fact, is closely tied to that of contemporary theories of trade and growth in the context of what is known as the 'increasing returns revolution' in economics. NEG is said to be the "fourth wave of the increasing returns revolution in economics" (Fujita et al. 1999: 3). The revolution consists in shifting away from the constant returns to scale-perfect competition paradigm that dominated the discipline of economics until the 1970s and 1980s, and adopt in its stead an increasing returns-imperfect competition framework. Economists have been aware of the crucial importance of increasing returns and imperfect competition for long, and yet they lacked the tools to adequately incorporate them in the formal and rigorous models economists favour. The Dixit-Stiglitz model of monopolistic competition offered precisely that tool. Thanks to its workability and analytical flexibility, the model could be applied to different areas of inquiry. It was originally conceived as a contribution to the literature on product differentiation, but it was later applied to the phenomena of international trade, growth, and geography, all of which are thought to be the result of the presence of increasing returns.

The application of the Dixit-Stiglitz model to phenomena of growth and trade follows a similar path. In both cases, the then dominant neoclassical variant (based on constant returns and perfect competition) was unable to address some stylized facts and the existence of increasing returns was thought to be necessary to explain them. In commenting the situation of trade theory around the 1970s, Paul Krugman (1995: 1244) describes it as "a collection of highly disparate and messy approaches, standing both in contrast and in opposition to the impressive unity and clarity of the constant-returns, perfect competition theory." Thanks to the Dixit-Stiglitz model, theories of growth and trade based on increasing returns ('new trade theory' and 'new growth theory' respectively) became serious complements to the neoclassical ones. Then, a sequence of extensions to Krugman's new trade theory models led to the formulation of the first NEG's model, that is, Krugman 1991 (Brakman et al. 2001).

The role of the Dixit-Stiglitz model of monopolistic competition becomes apparent once we shift the focus from theories to fields and therein to what Grantham calls practical unification. The model functioned as a vehicle towards the integration of several fields in economics: the same analytical tool was employed with the appropriate modifications in various fields, and for precisely the same purpose, namely to deal with increasing returns at the firm level and imperfect competition. Its application to the domain of spatial phenomena finally led to the development of NEG, which in turn significantly contributed to the integration of these fields at the theoretical level. NEG can then be thought as an inter-field theory that is both a product of this process of integration as well as a vehicle of further integration. Both practical and theoretical connections have simultaneously worked towards the integration of several related fields. Even so, the fields continue to proceed largely independently from each other. The drive

towards unification continues to play a prominent role within each field and across them,[11] but as mentioned earlier, so far NEG is unable (and does not aim) to account for a number of stylized facts that theories belonging to these fields seek to explain.

3.4. *ECONOMICS IMPERIALISM*

The development of NEG has also had effects on fields outside economics: economic geography, a subfield of human geography, lays claim to substantial parts of its domain. Economic geography is diverse in both scope and methods. At its core, however, is an emphasis on the complexity of empirical reality, on place rather than space, on concepts such as contingency and specificity and, on the level of method, on the extensive use of case studies and a discursive style of theorizing. Given that the mainstream of the economics discipline has traditionally neglected spatial and geographical issues, it is not surprising that economic geographers perceived NEG as an attempt to invade their own territory.[12]

Mäki and Marchionni (2007) deal with this aspect of the relationship between NEG and economic geography. We propose to consider NEG's foray into the field of economic geography a consequence of its pursuit of unification and suggest that there is nothing inherently problematic in this pursuit. Disciplinary boundaries are not sacred, and it just so happens that the phenomena NEG unifies also fall within the purview of another discipline. On the other hand, we also warn against the danger of unconditionally celebrating unification and the ensuing imperialism. It is indeed possible that the spatial phenomena NEG unifies are not in reality so unified, and/or that the range and significance of the spatial questions NEG can address is very limited (as economic geographers have not failed to point out). All this needs to be established empirically, however. More importantly, as mentioned above, scientific imperialism also affects sociological and institutional aspects of interdisciplinary relations. When what is exported is not only a theory, but a purportedly superior research style and/or the higher standing of a discipline, the mechanisms sustaining these non-epistemic aspects of imperialism may tip the balance in favour of theories for which empirical support is poor at best.

A significant component of the concern expressed by economic geographers indeed has to do with these aspects of imperialism. The danger, as they perceive it, is that NEG might end up enjoying increasing policy influence just because of the higher standing of economics. Similarly, the alleged 'scientificity' of economics could give NEG an extra edge in the academic competition and end up colonizing economic geography at the cost of the latter's theoretical and methodological commitment. As things stand now, it is not at all clear that NEG warrants a superior position vis-à-vis the theories of economic geographers in the domain of spatial phenomena. In sum, the NEG pursuit of unification generates pressures on disciplinary boundaries but in spite of attempts at promoting cross-fertilization, which if successful could promote integration, the two fields largely proceed in mutual disregard for each other's work.

[11] For example trade theory models built to include both neoclassical and new trade theory determinants of international trade, or recent efforts in NEG aimed at the construction of more general models from which specific models can be derived as special cases.

[12] The choice of calling it 'new economic geography' has probably contributed to reinforce the geographers' perception of imperialism. Some NEG theorists proposed to adopt 'geographical economics' instead to underscore the disciplinary origin of the approach, but the original name still remains prevalent.

4. CONCLUSIONS

Philosophical interest in the practice of scientific unification in economics has been scarce. A perspective acknowledging the potential benefits of economists' unifying practices as well as the limitations has largely been missing. This is a gap that needs to be filled because the quest for unity constitutes a powerful driver of theoretical development in economics that significantly affects its internal development as well as its relationship with neighbouring fields and disciplines. We have seen that the NEG's efforts to provide a unified explanation of agglomeration phenomena by showing how they are brought about by the same kind of economic mechanisms has influenced its relationship to other theories, fields and disciplines. From this case, it also emerges quite clearly that unification encounters resistances that ultimately influence the kind and degree of unification that is established at each level. First, unification of phenomena turns out to be a distinct achievement from explanation and for certain explanatory purposes, unification and explanation pull against each other. Second, the stylized mechanism that NEG postulates and on which the unification is based upon does not suffice to do all the explanatory job of existing theories dealing with the phenomena NEG unifies. Third, as an inter-field theory NEG has further advanced the integration of a number of related subfields in economics, but for the most part these continue to proceed autonomously. Finally, insofar as inter-disciplinary relations are concerned, in spite of a few attempts at cross-fertilization, NEG and economic geography continue to proceed in mutual ignorance of each other's work.

I believe that other cases of unification in economics can be usefully examined by paying attention to the levels at which unification occurs and the frictions it faces at each of these levels. Consider for example neuroeconomics. It is often seen as an attempt to reduce economic theories to theories about the brain, and neuroeconomists themselves have sometimes expressed unifying ambitions.[13] The following quote from two major representatives is telling in this respect:

> Economics, psychology, and neuroscience are converging today into a single, unified discipline with the ultimate aim of providing a single, general theory of human behaviour. This is the emerging field of neuroeconomics in which consilience, the accordance of two or more inductions drawn from different groups of phenomena, seems to be operating. Economists and psychologists are providing rich conceptual tools for understanding and modelling behaviour, while neurobiologists provide tools for the study of mechanism. (Glimcher and Rustichini 2004: 447)

Unlike NEG, however, the main vehicles of unification are the new technologies (PET and fMRI for example), which make it possible to generate data on the brain level, and a batch of methods, including experimental templates (e.g., the Ultimatum Game), which have become stable enough to travel across disciplinary boundaries (cf. Guala 2009). These have made available a plethora of new data in a situation in which empirical evidence has been chronically lacking. Neuroeconomics potentially increases the degree of integration among economics, psychology and neuroscience, which in turn could lead to improved knowledge about decision-making. For instance, neuroscientific knowledge may generate new behavioural hypotheses and introduce (mechanistic) constraints on models of individual choice (Clithero et al. 2008).

From the disunity side, many have claimed that economics is and should remain unaffected by neuroscientific findings (e.g., Gul and Pesendorfer 2008). I suspect that this dismissive reaction is, to some extent, due to worries concerning aspects of neuroscience imperialism

[13] The reduction of higher-level theories to theories at the lower level also contributes to the unification of science. This is in fact Oppenheim and Putnam (1958)'s model of scientific unity mentioned above.

that could threaten economists' own disciplinary identity. The idea that economics should be insulated from whatever is going on in neuroscience is perhaps too extreme. Nevertheless, although neuroeconomics is still in its infancy and it is too soon to tell how it is will develop, there are good reasons to doubt that it will ever replace standard ways of doing economics. For one thing, the contribution of neuroscience to economics has to pass through psychology (e.g., Padoa-Schioppa 2008), and a wealth of arguments, which I won't go into here, has made it clear that the global reduction of psychology to neuroscience is unfeasible and possibly detrimental[14] – even though successful local reductions occur. As in the NEG case, here, too the quest for unity plays an important role, but the frictions exerted by disunities of various sorts crucially influence the kind of products that are delivered in practice.

In fact the products that the pursuit of scientific unification delivers are often a far cry from the splendid achievements that standard philosophical models presuppose. Still, I believe, economists' unifying practices, and their effects within and around economics, are worth philosophical attention. If we think of unification as a set of activities good for some purposes and only in some contexts, analyzing the way in which economists actually practice them will help us to identify their strengths and their limitations, and possibly to "calibrate them as tools" (Wimsatt 2006: 446). Some features may well remain stable across contexts and locales even though the specific activities and the frictions they face are different from case to case. The analysis of the NEG unification should be seen as a contribution to this more ambitious project.

Acknowledgments: Part of this paper is based on work done jointly with Uskali Mäki. I thank Emrah Aydinonat and Alessandro Lanteri for useful comments on a previous draft of this paper. All remaining mistakes are obviously mine.

BIBLIOGRAPHY

Baldwin R., Forslid R., Martin P., Ottaviano G. and Robert-Nicoud F. (2003) *Economic Geography and Public Policy*. Princeton University Press.

Bechtel W. and Hamilton A. (2008) Reduction, integration and the unity of science: natural, behavioral, and social sciences and the humanities. In T. Kuipers (ed.) *Philosophy of Science: Focal Issues*. New York: Elsevier, 377-430.

Brakman S., Garretsen H. and van Marrewijk C. (2001) *An Introduction to Geographical Economics: Trade, Location and Growth*. Cambridge: Cambridge University Press.

Cat J. (2007) 'The unity of science.' *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/scientific-unity.

Clithero J.A., Tankersley D. and Huettel S.A. (2008) Foundations of neuroeconomics: from philosophy to practice. PL.

Darden L and Maull N. (1977) Interfield theories. *Philosophy of Science* 43, 44-64

[14] The standard defence against the reducibility of psychology to neuroscience is the 'multiple realisability' argument (Fodor 1974). How exactly mental states and brain states are related and what consequences this relation has for the sciences involved is still a hotly debated issue.

Dixit A.K. and Stiglitz J.E. (1977) Monopolistic competition and optimum product diversity *American Economic Review* 67, 297-308.

Fodor J.A. (1974) Special sciences (or: the disunity of science as a working hypothesis.) *Synthese* 28, 97-115.

Friedman M. (1974) Explanation and scientific understanding. *Journal of Philosophy* 71, 5-19

Fujita M. and Thisse J-F. (2002) *Economics of Agglomeration. Cities, Industrial Location and Regional Growth*. Cambridge: Cambridge University Press.

Fujita M. Krugman P. Venables A. (1999) *The Spatial Economy. Cities, Regions and International Trade.* Cambridge Mass: MIT Press.

Glimcher P.W. and Rustichini A. (2004) Neuroeconomics: the consilience of brain and decision. *Science* 306, 447-452.

Grantham T.A. (2004) Conceptualizing the (dis)unity of science. *Philosophy of Science* 71, 133-155.

Guala F. (2009) Paradigmatic experiments: the ultimatum game from testing to measurement device. *Philosophy of Science* 75, 658-669

Gul F. and Pesendorfer, W. (2008) "The case for mindless economics." In A. Caplin, A. Schotter (Eds.) *The Foundations of Positive and Normative Economics. A Handbook*. Oxford: Oxford University Press, 3-39

Kitcher P. (1981) "Explanatory Unification." *Philosophy of Science* 48: *251-81*

Kitcher P. (1984) "1953 and all that: a tale of two sciences." *Philosophical Review* 93, 335-373

Kitcher P. (1989) "Explanatory unification and causal structure." *Minnesota Studies in the Philosophy of Science* Vol. 13. Minneapolis: University of Minnesota Press, 410-505

Mäki U. (2001) "Explanatory unification: double and doubtful." Philosophy of the Social Sciences 31 (4), 488-506.

Mäki U. (2009) Economics imperialism: concept and constraint. *Philosophy of the Social Sciences.*

Mäki U. and Marchionni, C. (2007) Is geographical economics a form of intellectual imperialism? *Unpublished manuscript*.

Mäki U. and Marchionni, C. (2009) On the structure of explanatory unification: the case of geographical economics. *International Studies in the History and Philosophy of Science.*

Marchionni C. (2009) Geographical economics & its neighbors – forces towards and against unification. In U. Mäki (ed.) *Elsevier Handbook of the Philosophy of Science*, vol. 13: Philosophy of Economics. New York: Elsevier.

Ohlin B. (1933) *Interregional and International Trade*. Cambridge, Mass: Harvard University Press.

Ohlin B. (1979) *Some Insufficiencies in the Theories of International Economic Relations.* Princeton: Princeton University.

Oppenheim P. and Putnam, H. (1958) The unity of science as a working hypothesis. In H. Feigl, M. Scriven, and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 3-36

Padoa-Schioppa, C. (2008) The syllogism of neuro-economics. *Economics and Philosophy* 24, 449-457.

Sober E. (1999) The multiple realizability argument against reductionism. *Philosophy of Science* 66(44) 542-564

Thagard P. (2007) Coherence, truth, and the development of scientific knowledge. *Philosophy of Science* 74, 28-47

Whewell W. (1847) *The Philosophy of the Inductive Sciences Founded on their History*. London: John W. Parker.

Wimsatt W. (2006) Reductionism and its heuristics: making methodological reductionism honest *Synthese* 151, 445-475.

Wylie A. (1999) Rethinking unity as a working hypothesis for philosophy of science: how archaeologists exploit the disunities of science. *Perspectives on Science* 7(3): 293-317.

Causal Arrows in Econometric Models

Federica Russo[*]
f.russo@kent.ac.uk

ABSTRACT

Econometrics applies statistical methods to study economic phenomena. Roughly, by means of equations, econometricians typically account for the response variable in terms of a number of explanatory variables. The question arises under what conditions econometric models can be given a causal interpretation. By drawing the distinction between associational models and causal models, the paper argues that a proper use of background knowledge, three distinct types of assumptions (statistical, extra-statistical, and causal), and the hypothetico-deductive methodology provide sufficient conditions for a causal interpretation of econometric models.

---

1. INTRODUCTION

A *vexata quaestio* in philosophy of economics is the extent to which econometric models tell causal stories. The question is certainly not new and occupies much of philosophically-minded discussions of the methodology of economics and, more generally, of the social sciences.

This paper adds to the literature offering its own position. The position I defend stems from the dissatisfaction with a certain class of arguments—which I shall call 'metaphysical arguments'—and from the recognition that another class of arguments—which I shall call 'methodological arguments'—are on the right track but only go half way through.

The paper is organized as follows. In section 2, I explain the particular epistemological viewpoint I adopt in looking at econometric models. The motivation for an epistemological approach to causality comes from the dissatisfaction with 'metaphysical arguments' that assume causal structures generating the observed data instead of explaining how we come to establish whether a given correlation is causal. In section 3, I review traditional 'methodological arguments' given by philosophers of economics and philosophically-minded econometricians. In those arguments two trends can be identified: those who stress the importance of causal mechanisms and those who put the whole burden of the causal interpretation in the assumptions of the model. I then offer in section 4 my own answer. I shall do that in two steps. First, I present an account of the conditions under which econometric models can be given a causal interpretation by defending the following interrelated ideas: (i) there is an important distinction between associational models and causal models; (ii) the difference lies in their respective features, notably whilst associational models just have statistical assumptions, causal models also have extra-statistical and causal assumptions, employ a hypothetico-deductive methodology, and use background knowledge in an essential way at each stage of the model building and model testing process. Second, I discuss the methodology, results, and critiques of a case study on the relations between health and wealth in elderly Americans.

---

[*] Honorary Research Fellow, University of Kent

2. AN EPISTEMOLOGICAL INVESTIGATION INTO ECONOMETRIC MODELS

Let me make clear from the outset what perspective I shall take in looking at econometric models. In the philosophy of causality three broad areas of investigation may be distinguished. The metaphysics of causality is interested in what causality in fact is, in what kind of entities causes are, or in what a causal claim means. The epistemology of causality, instead, investigates how we come to know about causal relations. There is a fleeting borderline between epistemology and methodology, but a line between the two can be drawn nonetheless. Whilst methodology is concerned with problems of scientific methods and aims at developing successful methods for the discovery and confirmation of causal relationships, epistemology is rather interested in the conceptual issues behind those methods. Metaphysical, epistemological, and methodological issues ought not to be conflated and a joint investigation from these three perspectives will hopefully allow us to get a better grip on causation.

This paper entirely locates within the domain of epistemology and methodology. Namely, the focus is on how we come to know about causal relations, regardless of the position one may take about the metaphysics of causation, i.e., about what causation in fact is. The motivation for an epistemological perspective comes from difficulties in 'metaphysical arguments', according to which, simply put, genuine probabilistic dependencies are causal dependencies, i.e. genuine correlations are per se causal. An example is Hausman (1998) who reiterates the idea that true correlations are causal and that if, eventually, the correlation turns out to be spurious this means that we picked out the 'wrong' correlation (see, e.g., Hausman 1998, p.33, 56). Differently put, correlation does prove causation—accidental correlations, in Hausman's view, simply aren't true correlations. To be sure, what is meant by these kinds of argument is that modulo certain conditions—most typically the so-called Markov condition—correlations are causal. However, it is highly a controversial matter whether the Markov condition ensure causality (see for instance Hausman and Woodward 1999 and 2004 and Cartwright 2002 on this point).

There is also another problem with 'metaphysical arguments'. A stock example is the positive correlation between the increasing number of storks and the increasing of birth rates in Alsace. According to arguments à la Hausman, this correlation is accidental, not genuine, hence not causal. Agreed, it might well be the case that genuine correlations be causal, but why we do not believe that the increasing number of storks is causally related to the increasing number of births? The answer, it seems to me, is that our background knowledge does not contain any theory or piece of information that makes this correlation plausible, let alone plausibly causal, in any possible way.

In a similar vein, Cartwright (1989) believes that capacities are responsible for, or give raise to, stable regularities and therefore are the very ontological basis of any observed statistical correlation. Capacities have the peculiar feature of being stable across different background conditions, but the question is, in the first place, how do we know that something is a capacity. At times, Cartwright (1989, pp.148 ff) seems to introduce a circularity in the account because the main point of econometric model is to test stability of relationships between variables; but those relations will be stable in case they are the effects of capacities, therefore stability cannot be a test to know whether something is a capacity in the first place. Steel (2008, pp.82-85) rises a similar worry concerning extrapolation, because if a causal relation has to have the feature of being stable across different contexts, then capacities cannot be what grounds extrapolation inferences exactly because what we are trying to establish is whether (stable) capacities are at work.

Thus, even if we assume that there are true causes that give rise to the observed correlations, the question still remains as to how we come to establish that some correlations, and not others, are causal. Echoing Granger (2003), the question is whether 'causality' applies to the model or the data generating process. I'd opt for the former option, in particular, my interest then lies in the process of model building and model testing, and, within in this process, I aim to highlight the conditions under which econometric models can be given a causal interpretation. This task I shall undertake in section 4.

3. STANDARD VIEWS

The literature in econometrics and social science is indeed vast, but two broad families of arguments can be identified nonetheless. A first family of arguments is centred around the notion of mechanism and the second around the assumptions of econometric models. To be sure, those two families are not mutually exclusive and there in fact a version of the 'mechanism argument' that eventually collapse into the 'assumption argument. Let me explain further.

Mechanisms are usually central for those who advocate a structural approach in econometrics. Simply put, and leaving aside the technicalities, structural models explain economic phenomena by means of (sets of) equations that describe causal mechanisms.

In the philosophy of economics and social science, a noteworthy partisan of the mechanistic approach is Little (1990), who claims that causal analysis in the social sciences is legitimate insofar as models identify social mechanisms. Little believes that such social mechanisms work through the actions of individuals—a position also known as methodological individualism. Hoover (2001), instead, stresses the causal import of the structural approach in econometrics arguing for a reality of macroeconomic structures that does not boil down to the reality of microeconomic relations. Hoover also tags along with mechanistic approaches because a causal structure is, in his view, a "network of counterfactual relations that maps out the underlying mechanisms through which one thing is used to control or manipulate another" (2001, p.24). The idea behind approaches à la Little (but not necessarily à la Hoover) is that there is a causal relation between A and B if and only if there is a mechanism that links A to B—a position hold, in slightly different ways, also by other 'mechanical philosophers'. Nevertheless, the emphasis on mechanisms is famously criticized by Kinkaid and more recently by Reiss. Kinkaid (1996), for instance, thinks it is false that in order to know whether X causes Y at least a mechanism linking X to Y has to be identified. Reiss (2007b and 2008, ch.6), along the same lines, argues that mechanisms might not be the most useful strategy to achieve other goals, for instance measuring concepts such as 'inflation rate' or 'unemployment rate'.

A different emphasis on mechanisms is given by economic theorists, who describe economic mechanisms that work 'ceteris paribus'. Such position is typical of classical economists such as Adam Smith, David Ricardo, Thomas Malthus, and John Stuart Mill, or the Chicago School of Economics. One way or another, it is the economic theory that 'dictates' the mechanism, instead of conceiving the whole modelling procedure as the attempt to identify what the mechanism is and how it works. It is in this vein that, according to Heckman (2008), structural econometric analysis has the following peculiar feature: it aims to model the generation of the outcome (i.e., the dependent variable) taking into account the agent's decisions to undertake a treatment. Thus, the outcome is explicitly modelled "in terms of its determinant as specified by [economic] theory" (p.18, my emphasis).

Within literature of econometricians and statisticians, instead, another trend can be recognized. The goal here is to model the stochastic mechanism that generates the data—in their jargon, the data generating process. The specification of the data generating process, in

turn, depends on the statistical model, and, consequently, the assumptions of model play a crucial role. Freedman (2004), for instance, distinguishes between statistical and causal assumptions and requires interventions to grant causal inferences. The crucial assumptions, in his account, are the causal ones, which eventually consist in assuming that structural equations unveil the causal mechanism that generate the observed data. This way, however, there isn't much difference between this methodological argument and the metaphysical arguments mentioned in the previous section. In fact, under this account the causal interpretation of structural equation consists in assuming that there is mechanism behind but stay silent on how we come to establish whether there is such mechanism. Holland (1986) goes a step further and draws a distinction between associational models and causal models, where the former simply make descriptive claims about conditional distributions and the latter also aim to quantify the causal effect of a treatment or intervention. Stone (1993), finally, focuses specifically on the causal assumptions, ranking them from the strongest—i.e., covariate sufficiency—to the weakest—i.e., ignorable treatment assignment.

All these arguments certainly get right part of what is at stake but not, I contend, the whole story. Mechanisms certainly play a role here. But why? Is it because we assume the existence of a given causal structure that we believe gives rise to the observed distributions? Or is it because we aim to model a causal mechanism? Assumptions are certainly central too. But what is exactly their import in justifying the causal interpretation? The story I want to tell somehow embraces both those views. Notably, I will defend the idea that we have to model mechanisms paying particular attention to the different types of assumptions made in the model.

Somehow, the view defended here is mid-way between the 'deductivist' and 'inductivist' approaches in econometrics (Moneta 2007). In the former, causes are 'given' by the economic theory; although there is some degree of freedom as to what economic theory to choose, once a choice is made, this imposes the restrictions on the model. Econometrics is thus reduced to measuring (statistical) relations between variables, rather then (dis)confirming causal hypotheses. In the latter, causes are inferred from statistical properties of data alone, by imposing to the model the simplest causal structure that allows identification—a methodology that strongly resembles present-day graphical models. The view defended here locates in between those two positions (i) because economic theory has to play a role in the model building and model testing process as part of background knowledge, but does not have to be the sole element to determine the choice of variables and the interpretation of results; (ii) because statistical analyses, where we let "the data speak as much as possible"—to echo Moneta (2007, p.119)— independently of any a priory economic theory can also play a role in the model building process but do not exhaust the causal analysis itself.

4. 'CAUSAL' ECONOMETRIC MODELS

The arguments offered next hinge upon the distinction and comparison of two classes of models: associational models and causal models. It is commonly agreed that associational models just make descriptive claims about conditional distributions, whereas causal models, in addition, aim at evaluating statistical relevance relations to 'quantify' the causal effect of the explanatory variables on the response variable. However, this cannot be the whole story, since it still stays unclear how, in causal models, correlations suddenly turn into causal relations and probabilistic dependencies into causal dependencies. In the following, I argue that differences between associational models and causal models can be identified at three levels: (i)

background knowledge, (ii) assumptions, and (iii) methodology. The difference between associational and causal models is schematically represented in table 1.

Associational Models	Causal Models
Statistical Assumptions	Background knowledge/Causal context Statistical Assumptions Extra-Statistical Assumptions Causal Assumptions Hypothetico-deductive Methodology

Table 1: Associational Models vs. Causal Models

ASSOCIATIONAL MODELS

The goal of associational models is to describe how a given variable (the dependent variable) varies depending on other variables (the independent ones). Associational models are typically used to make exploratory analyses of data in order to see what correlations between variables hold. Background knowledge does not play any particular role in associational models and variables do not play specific causal roles. Associational models rest on a number of standard 'statistical assumptions'. Leaving technicalities aside, we suppose that the model have some characteristics (usually, linearity and normality) such that it is easy to manipulate, easy to estimate statistically, and the resulting estimates have nice properties. We also assume that variables are measured without error, that the errors are not correlated with the independent variables. When these assumptions are satisfied, the conditional distribution correctly describes how variables co-vary. But at this stage there is no necessary causal information conveyed by the parameters, nor it is generally valid to give the regression coefficients the causal interpretation. Also, dependencies described in associational models are symmetric, which doesn't tell much about causation, in fact.

To interpret the coefficients causally means that the coefficients appearing in the equations measure the effect on the dependent variable caused by a change in the independent variable(s). To go beyond the descriptive claims we need (i) accurate knowledge of the causal context, (ii) further assumptions, and (iii) a methodology to confirm/disconfirm causal hypotheses.

CAUSAL MODELS

Causal models are equipped with a much richer apparatus than associational models simply do not have. This involves: (i) background knowledge, (ii) further assumptions, and (iii) hypothetico-deductive methodology.

BACKGROUND KNOWLEDGE

Background knowledge certainly includes the economic theory but also includes general knowledge of the socio-political context, knowledge of demographic characteristics of the population under investigation, or 'institutional' knowledge (i.e., knowledge of the functioning and procedures of an institution such as the Central Bank). This is what many social scientists usually call 'field knowledge'. In some cases, notably when dealing with disciplines that need to include biological variables, background knowledge may also include knowledge of the physical-biological-physiological mechanism. This is the case, for instance, in epidemiology, where one of the objectives is to understand how health variables affect socio-economic variables, or the other way round. Well established scientific theories also belong to filed knowledge. No doubt evidence is important for background knowledge. In particular, evidence of the same putative causal relations operating in different populations may justify further research, or evidence about different causal relations operating in other populations may justify a different modelling strategy. Thus, the use of different/similar data and/or models also belongs to background knowledge.

ASSUMPTIONS

Let us now turn to the assumptions needed in causal models. Beside standard 'statistical assumptions' also made in associational models, causal models have two extra sets of assumptions that I shall call the 'extra-statistical' assumptions and the 'causal' assumptions.

By 'extra-statistical' assumptions I mean all those assumptions that either are not related to the statistical properties of the distributions or have causal meaning but are not subject to statistical test. Among these assumptions we find, for instance, the direction of time, the direction of the causal relations, the causal mechanism.

By 'causal' assumptions I mean all those assumptions that are subject to statistical test and that contribute to interpret causally the relations between variables in the model. Two causal assumptions are particularly important: 'covariate sufficiency' and 'no-confounding'. Covariate sufficiency states that the independent variables are all the variables needed in order to explain the variance of the dependent variable. The assumption of 'no confounding' then plays a complementary role and means, simply put, that we ruled out other factors liable to 'screen off' the variables we took into account. These assumptions are causal because the causal interpretation crucially depends on them in the following sense.

Causal modelling rests on the so-called 'closure of the system'. This assumption says that the system under analysis is not subject to external influences and thus can be separated, so to speak, from the larger web of interrelations in which it is located. Thanks to this assumption we can, at least in principle, detect the causes acting in the system under investigation. The assumption of the closure is related to covariate sufficiency and no-confounding thus: if we put those three pieces of the puzzle together—(weak) closure, covariate sufficiency, and no-confounding—we end up with a causal model that describes a (quasi) hermetic mechanism where all and only the chosen explanatory variables play a role.

HYPOTHETICO-DEDUCTIVE METHODOLOGY

In order to tell causal stories, econometric models have to employ a hypothetico-deductive methodology. Simply put, hypothetico-deductivism is the view according to which scientists

first formulate hypotheses and then test them by seeing whether or not the consequences derived from the hypotheses obtain. Popper (1959), who developed the H-D methodology, was motivated by the need of creating a scientific method in a non-inductive way. However, in causal analysis, hypothetico-deductivism takes a slightly different facet specifically concerning deduction and, mostly, is not strictly falsificationist in character. Yet, it does borrow from the Popperian account the primary role of the hypothesis-formulation stage. I shall get back to this point later. According to the H-D methodology, model building and model testing essentially involve three stages:

1. formulate the causal hypothesis;
2. build the statistical model;
3. draw consequences to conclude to the empirical validity or invalidity of the causal hypothesis.

The causal hypothesis to put forward for empirical testing does not come from a tabula rasa, but emerges from a causal context, namely from background knowledge. The hypothesis formulation stage may also be informed by results of associational models as they suggest interesting correlations to submit to further scrutiny. The rest of the model building and model testing process is meant to (dis)confirm the causal hypothesis. In practice, this requires building the statistical model, and then drawing consequences from the hypothesis in order to test the hypothesis against empirical data.

It is through the estimation of the statistical model and through hypothesis testing that we will (dis)confirm the causal hypothesis. If the model is correctly estimated and fits the data, and if certain conditions are satisfied (notably, exogeneity and invariance) the hypothesized causal link is accepted, rejected otherwise. The hypothetico-deductive structure of causal modelling is thus apparent: a causal relation is first hypothesized and then put forward for empirical testing. That it to say, the causal hypothesis is not directly inferred from the data, as is the case with inductive strategies, but accepted or rejected depending on the results of tests and on background knowledge.

As anticipated above, hypothetico-deductivism in causal modelling does not involve deduction *stricto sensu*, but involves a weaker inferential step of 'drawing consequences' from the hypothesis. Once the causal hypothesis is formulated out of the observation of meaningful co-variations between the putative cause(s) and the putative effect and out of background knowledge, we do not require data to be implied by the hypothesis but just that data conform to it. Here, 'conform' means that the selected indicators and relations among them adequately represent the conceptual variables appearing in the causal hypothesis and the relations among them. Thus, this way of validating the causal hypothesis is not, strictly speaking, a matter of deduction, but surely is, broadly speaking, a deductive procedure. More precisely, it is a hypothetico-deductive procedure insofar as it goes the opposite direction of inductive methodologies: not from rough data to theory, but from theories to data. For a discussion of the H-D methodology, see also Little (1998, ch.9), Cartwright (2007, ch. 2) and Russo (2008, ch. 3 and 2009).

The two main tests of model parameters concern invariance and exogeneity. Invariance tests check whether the relation between two variables is stable across different environments, for instance under intervention or across different panels of the data set. The invariance condition acts as a constraint on the causal relation. This means that invariance does not constitute causation, but is a feature we require in order to interpret the parameters as causal parameters. The idea behind exogeneity is the following. Exogeneity is a condition of separability of inference. Given a data set containing a number of variables of interest, the goal

of modelling will be to decompose an initial joint probability distribution of all the variables into a sequence of marginal and conditional probability distributions. This (statistically) corresponds to identify and specify the variables playing a specific causal role and the way those causal variables interrelate. Background knowledge is of primary importance in guiding the choice of the marginal-conditional decomposition.

A CASE STUDY: HEALTH AND WEALTH IN ELDERLY AMERICANS

A study published in the Journal of Econometrics (Adams et al 2003) examined the possible causal links between health and socioeconomic status (SES) in elderly retired Americas. The links between health and SES have been the object of numerous studies. A survey of the literature shows that a significant association holds for a variety of health variables and alternative measures of SES. Also, much discussion has been devoted to the mechanisms behind this association, and yet there have been relatively few studies that allow causal paths to be definitively identified. This study adopts a particular methodology and applies statistical methods to test for the absence of direct causal links from SES to health, and from health conditions to wealth.

Adams et al. examine a sample from the population of elderly Americans aged 70 and older, in particular they examine whether innovations in health and wealth in a panel are influenced by their past values and by the past values of other relevant variables. Data come from three surveys performed between 1993 and 1998, providing information about health, wealth, and demographic information of individuals. A first analysis of data shows a significant association between health and SES. On the basis of background knowledge, including previous studies and knowledge of the socio-demographic conditions of elderly Americans, Adams et al. hypothesize that the links between health and wealth conditions can be represented as (roughly) sketched in fig. 1.

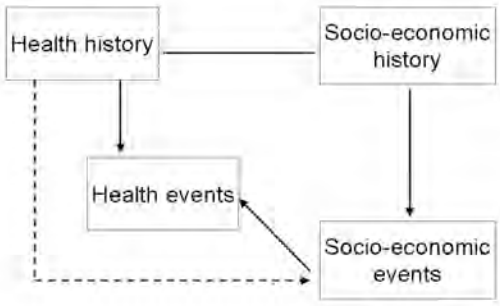


Fig.1 - Possible causal links between health and wealth

In the graph, arrows represent causal relations; when the line is dotted, the causal relation is absent. Thus, Adams et al., based on background knowledge and preliminaries analyses of data, expect that no significant causal link be found from health history of retired people to their present wealth status. On the contrary, they expect that socio-economic history, through their current wealth status, does affect their current health.

Following the approach of Granger (1969), two specific causal hypotheses are put forward for empirical testing: (i) there is no causal link from health to SES, and (ii) there is no causal link

from SES to health. Those two hypotheses are evaluated by means of non-causality test and invariance tests. Very informally, the structure of non-causality tests, in the approach of Granger (1969), is as follows: the present state of a variable (say, current health of individuals in the sample), causally depends on all its history (that is all the health history of individuals in the sample) and on the history of other relevant variables (say, SES of individuals in the sample). Thus, if SES and health turn out to be independent, then we say that SES does not Granger-cause health. Invariance tests, instead, aim to check whether model parameters stay stable across different panels.

In testing the significance of SES on a number of health outcomes, the authors draw the following conclusions: SES is, usually, strongly associated with health; however, the link is weaker when they control for previous health problems; also, the association between SES and mental and chronic illness is stronger than the association between SES and acute and sudden outset health conditions. Adams et al. suggest that those findings reflect the differential coverage of Medicare; in fact, insurances typically do not cover mental illness.

This simple presentation of the case study already illustrates the hypothetico-deductive methodology used. Look at where causality is; causality is not at merely the output, pulled out of the statistical model. Instead, the two causal hypotheses are put forward for empirical testing within the model, and then accepted or rejected depending on the results of test and on whether the results are congruent with background knowledge. To illustrate the features of causal models discussed above, let us discuss each feature separately.

Background knowledge and causal context. In Adams et al. work, it includes previous studies on the relations between health and wealth. In particular, the association between health and wealth holds for a variety of health variables and alternative measures of SES. Therefore, a causal interpretation of those correlations is plausible and worth testing.

Causal hypothesis. In the health and wealth case study the causal hypothesis is split into two: (i) is no causal link from health to SES, and (ii) there is no causal link from SES to health. In order to evaluate the causal impact of SES on health and vice-versa, Adams et al. build a model following the approach of Granger and perform non-causality and invariance tests on the causal hypotheses specified.

Statistical assumptions. The statistical analysis fits the approach of Granger, which is essentially based on regression methods and where standard statistical assumptions hold.

Extra-statistical assumptions. In this case study two extra-statistical assumptions are at stake: causal priority and causal ordering. In fact, in Granger-approach it is explicit that the history of variable causally determines the current value of the variable. Accordingly, statistical tests are about whether heath histories determine current wealth and whether wealth histories determine current health.

Causal assumptions. For the most part, Adams et al. discuss the assumption of covariate sufficiency, that is all the relevant explanatory variables have been included. Another important assumption concerns and structure of the causal relation: by adopting Granger's approach, instantaneous causality is ruled out, at least for sufficiently brief time intervals.

Tests. Within a 'Granger-causality' framework, Adams et al. test model parameters for invariance across different panels. More explicitly, they test (i) whether the conditional distribution 'current health given wealth history' and (ii) whether the conditional distribution 'current wealth given health history' hold stable across different panels analyzed. The invariance conditions (largely) fails in the first case, i.e. variations in health generally do not correspond to variations in wealth, hence, the hypothesis of no causality is accepted. On the other hand, the conditional distribution 'current wealth given health history' results (relatively) stable and therefore the hypothesis of no causality is rejected.

Hypothetico-deductive methodology. Adams et al. are clearly using a hypothetico-deductive methodology in their analysis. They first formulate the causal hypotheses out of background knowledge and then, iteratively, build the model and perform statistical tests to confirm or disconfirm the causal hypotheses. Background knowledge and causal context play a primary role at each stage: from the choice of variable until the evaluation of results. A clear identification of the steps in the process of model building and model testing allows to identify possible leaks or points of disagreement between scientists. Whilst there is a shared consensus on the background of this study (see the commentaries in the same volume of Adams et al.'s paper), a number of commentators suggested that a structural approach would be better suited to study the links between health and wealth (Hausman (2003), Heckman (2003), Hoover (2003)). The particular form of the invariance tests is also debated. For instance, Hoover (2003) agrees that invariance is an important indicator for the presence of a causal relationships, but doubts on the cogency of the evidence used by Adams et al. Analogously, remarks on the chosen definition and tests of non-causality also come from Florens (2003). Adda et al (2003) object to their suggestion that the observed links between SES and health are the effect of the access people have to health infrastructures. The reason is that Adda et al. apply the very same methodology to two different data sets, one from UK and one from Sweden, and get very similar results. However, since medical systems are very different in those countries, coverage of Medicare cannot be what produces these results. As a consequence, policies intervening in that direction would be, in Adda et al.'s view, misplaced.

EPISTEMOLOGICAL MORALS

Let me know state more precisely the epistemological morals to draw from the arguments given above. I undertook this path of investigation to overcome the dissatisfaction with available arguments that either assume a causal structure instead of saying how we come to establish it, or with methodological arguments that stop too early in praising the role of assumptions in causal models. In order to make causal claims we need different types of assumptions. The distinction I drew between statistical, extra-statistical, and causal assumptions is meant to make clear the import of each of them towards establishing causal relations. For instance, extra-statistical assumptions such as causal ordering, usually come from knowledge of the causal contexts, are typically not tested in the model, and yet participate in causal attribution because they serve, for instance, to fix which variables are the causes and which variables are the effects. Nevertheless, the main point of this distinction is to show that it is not the case that the stricter the statistical tests the greater our confidence in the causal interpretation. Causal relations are not pulled out of statistical hats. The hat, so to speak, is itself causal. The conceptual hypothesis is in fact a causal claim the evaluation of which demands for empirical testing. In other words, causality is a matter of confirmation, or borrowing the statistical vocabulary, it is a matter of accepting or rejecting a given causal hypothesis.

Let me linger more on this 'causal hat'. Model building and model testing take place in a causal context. Background knowledge plays a major role in the preliminary analyses of data, in the selection of variables, and in the formulation of causal hypotheses. It is also on the basis of background knowledge that the choice of testing for invariance some conditional distributions, but not others, relies. It is, finally, on the basis of background knowledge that results are evaluated. The importance of background knowledge, however, goes beyond the choice of variables and for the formulation of the causal hypothesis. It is also important in order to rule out highly implausible causal relations before starting the whole machinery of

model building and model testing. In manipulationist accounts à la Woodward (2003), the relation between the increasing number of storks and the increasing number of births is not causal because it would turn out not to be invariant under intervention; namely, no manipulation on the variable 'number of storks' would lead to changes in the variable 'number of births'. Nevertheless, should we go as far as testing for invariance in cases where we can exclude that the correlation is causal on the basis of background knowledge?

In the literature, the crucial role of the causal context and of the conceptual framework has been stressed by many eminent scientists and philosophers—jut to name a few: Fisher (1925), Haavelmo (1944), Kendall and Stuart (1961), Suppes (1970), Humphreys (1989). And yet, this emphasis hasn't been given the proper importance. It has oft been taken as a platitude about causal modelling—causal methods rely on background knowledge in an obvious sense—or as a 'hidden challenge'—what in fact background knowledge taught us that we did not already know? In other words, if model building and model testing are largely based on background knowledge, how do we go beyond it and gain causal knowledge?

What allows us to go beyond background knowledge is the hypothetico-deductive methodology of causal models. This is a dynamic methodology that allows a *va et vient* between established theories and establishing theories. Established scientific theories are (and ought to be) used to formulate the causal hypothesis and to evaluate the plausibility of results on theoretical grounds. But causal models also participate in establishing new theories by generalizing results of single studies. This reflects the idea that science is far from being monolithic, discovering immutable and eternal truths. If the model fits the data, model parameters are (sufficiently) invariant and the relations are congruent with background knowledge, then we can say, to the best our knowledge, that we hit upon a causal relation. But what if one of these conditions fails? A negative result may trigger further research by improving the modelling strategies, or by collecting new data, thus leading to new discoveries that, perhaps, discard background knowledge.

A last corollary of this epistemological regard on econometric models is that, by and large, causal models model mechanisms. As mentioned above, hypothetico-deductive causal models aim to (dis)confirm a causal hypothesis; the causal hypothesis, however, states something more than simply a causal relation—it's about a causal mechanism. Let me develop this idea further.

What do causal models do? Causal models model the properties of a social system. In particular, they model the relations between the properties or characteristics of the system, which are represented by variables. To model the properties of a social system means to give the scheme, or the skeleton, of how these properties relate to each other. However, this causal mechanism is not modelled in terms of spatio-temporal processes and interactions à la Salmon (1984 and 1990) but is statistically modelled. Concepts typical of statistical causality—e.g., statistical relevance, comparison of conditional probabilities, and screening-off—are used in order to identify the types of relationships that hold among the variables of interest. This is how the causal model models the causal mechanism governing the social system. In particular, causal models seek to uncover stable variational relations between the characteristics of the system.

This view of mechanisms is not perfectly in line with more 'physical' views such as those of Glennan (1996), Machamer et al. (2000), Bechtel and Abrahms (2005), or Craver (2007). However, it is not in opposition either. Craver, for instance, echoing the definition given by Machamer et al (2000), conceives of mechanisms as "a set of entities and activities organized such that they exhibit the phenomenon to be explained" (Craver 2007, p.5), and calls it a "skeletal description". Such characterization of mechanisms, I believe, is broad enough as to account for mechanisms in various domains. Should we take the entities to be neurons and the

activities neurotransmitters release, the above skeletal description will well fit neuro-mechanisms. Should we take entities to be socio-demo-economic variables, and activities to be their influence on other socio-demo-economic variables, the above skeletal description will fit equally well social mechanisms. The degree of 'physical' reality one wishes to give to entities and activities may lead to different accounts — notably, to a different ontological commitment to the existence — of mechanisms. In the social contexts we do not necessarily need to endorse the view that elements and relations should always have 'physical' counterparts. This would lead us into debates concerning social ontology and methodological individualism, which, of course, are far beyond the scope of the present discussion.

5. CONCLUDING REMARKS

Econometrics applies statistical methods to study economic relations. By means of structural models, econometricians typically account for the response variable in terms of a number of explanatory variables. Under what conditions can econometric models be given a causal interpretation? This paper defended the view that (i) a proper use of background knowledge, (ii) statistical, extra-statistical, and causal assumptions, and (iii) hypothetico-deductive methodology provides sufficient conditions for the causal interpretation. On the contrary, associational models do not allow any causal interpretation as they are just equipped with statistical assumptions. This goes against the views that causal structures are simply assumed and are that which is responsible for the observed correlations, and against the views that 'bootstrap' causal relations from stringent statistical tests alone. The causal interpretation is instead a matter of confirmation. In hypothetico-deductive terms, the goal is to confirm (or disconfirm) a hypothesis that is explicitly causal. Nancy Cartwright famously expressed this idea with the motto 'No causes in, no causes out'. In spite of appearance, the account is not circular, and if there is a circle at all, it is virtuous and not vicious. To put causality in the model means that the evaluation of causal relations is relative to some conceptual and formal framework; in this framework we find background knowledge, causal assumptions, and hypothetico-deductive methodology. The account offered gives sufficient conditions that allow to interpret econometric models causally. It is worth emphasize that such account deeply depends on the epistemological perspective I adopted. By focusing on how we come to establish causal relations rather than on what causation in fact is, I implicitly endorsed an epistemic view of causality very much in line with the view defended by Williamson (2005). According to this view, results of econometric models are not immutable and eternal causal truths, but justified beliefs, which are fallible. This fact is indeed mirrored in the hypothetico-deductive methodology of causal models which, being dynamic and flexible, allows us to revise our causal beliefs by building and testing new 'causal' econometric models.

Acknowledgements. Very helpful and stimulating comments came from the fellows at the Center for Philosophy of Science (University of Pittsburgh)—Claus Beisbart, Erik Curiel, Laura Felline, Ulrich Krohs, Flavia Padovani, Chris Pincock, John Norton (director). I am indebted to Alessio Moneta for making a number of valuable suggestions, especially concerning the role of economic theory in causal analysis. Financial support from the FSR-FNRS is also gratefully acknowledged.

BIBLIOGRAPHY

Adams P., Hurd M., McFadden D., Merrill A., Ribeiro, T. (2003) Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112, 3–56. With discussion.

Adda J., T. Chandola, and M. Marmot (2003) Socio-economic status and health: causality and pathways. *Journal of Econometrics*, 112, 57-63.

Bechtel W. and A. Abrahamsen (2005), Explanation: a mechanist alternative, *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, 36, 421-441.

Cartwright N. (1989) *Nature's capacities and their measurements*. Oxford: Clarendon Press.

Cartwright N. (2002) Against modularity, the causal Markov condition, and any link between the two: comments on Hausman and Woodward. *British Journal for the Philosophy of Science*, 53(3), 411–453.

Cartwright N. (2007a) Hunting causes and using them: approaches in philosophy and economics. Cambridge: Cambridge University Press.

Craver C. F. (2007), Explaining the brain. Mechanisms and the mosaic unity of neuroscience. Oxford: Oxford University press.

Engle R. F., Hendry D. F., Richard J. -F. (1983) Exogeneity. *Econometrica*, 51(2), 277–304.

Fisher R. A. (1925) *Statistical methods for research workers*. London: Oliver and Boyd. http://psychclassics.yorku.ca/Fisher/Methods/. Accessed 14 March 2008.

Florens J-P (2003) Some techical issues in defining causality. *Journal of Econometrics*, 112, 127-128.

Florens J-P, M. Mouchart and J.-M. Rolin (1980), Réductions dans les Expériences Bayésiennes Séquentielles, paper presented at the Colloque Processus Aléatoires et Problèmes de Prévision, held in Bruxelles 24-25 April 1980, Cahiers du Centre d'Etudes de Recherche Opérationnelle, 23(3-4), 353-362.

Freedman D. A. (2004) On specifying graphical models for causation, and the identification problem. *Evaluation Review*, 26, 267–93.

Glennan S. (1996) Mechanisms and the nature of causation. *Erkenntinis*, 44, 49-71.

Granger C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438.

Granger, C. W. J. (2003) Some aspects of causal relationships. *Journal of Econometrics*, 112, 69-71.

Haavelmo T. (1944) The probability approach in econometrics. *Econometrica*, 12, iii–vi + 1–115.

Hausman D. (1998) *Causal asymmetries*. Cambridge: Cambridge University Press.

Hausman D., Woodward, J. (1999) Independence, invariance, and the causal Markov condition. *British Journal for the Philosophy of Science*, 50, 521–583.

Hausman D., Woodward, J. (2004) Modularity and the causal Markov condition: a restatement. *British Journal for the Philosophy of Science*, 55, 147–161.

Hausman J. A. (2003) Triangular structural model specification and estimation with application to causality. *Journal of Econometrics*, 112, 107-113.

Heckman J. (2003) Conditioning, causality, and policy analysis. *Journal of Econometrics*, 112, 73-78.

Heckman J. (2008) Econometric causality. IZA DP No. 3425. http://www.nber.org/papers/w13934 accessed 11 April 2009.

Holland P. W. (1986) Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–970.

Hoover K. D. (2001) *Causality in macroeconomics*. Cambridge: Cambridge University Press.

Hoover K. D. (2003) Some causal lessons from macroeconomics. *Journal of Econometrics*, 112, 121-125.

Hoover K. D. (forthcoming) Probability and Structure in Econometric Models. In The Proceedings of the 13th International Congress of Logic, Methodology and Philosophy of Science. London: King's College Publications.

Humphreys P. (1989) *The chances of explanation*. Princeton: Princeton University Press.

Kendall M. G., & Stuart, A. (1961) The advanced theory of statistics. Vol. 2, London: Griffin & Co

Kinkaid, H. (1996), *Philosophical Foundations of the Social Sciences*, New York: Cambridge University Press.

Little D. (1990) Varieties of social explanations: an introduction to the philosophy of social science. Boulder: Westview Press.

Little D. (1998) Microfoundations, method and causation. On the philosophy of the social sciences. New Brunswick, NJ: Transaction Publishers.

Machamer P. L. Darden and C. F. Craver (2000), Thinking about Mechanisms, *Philosophy of Science*, 67, 1-25.

Moneta A. (2007) Mediating between causes and probabilities: the use of graphical models in econometrics (pp. 109-130) In F. Russo and J. Williamson (eds), *Causality and Probability in the Sciences*. London: College Publications.

Mouchart M. and Russo, F. (2010) Causal explanation: recursive decompositions and mechanisms. In P. McKay Illari, F. Russo, and J. Williamson (eds) *Causality in the Sciences*. Oxford University Press.

Mouchart M., Russo F.,Wunsch G. (2008) Structural modelling, exogeneity and causality. In H. Engelhardt, H. -P. Kohler, A. Prskwetz (Ed.), *Causal analysis in population studies: concepts,methods, applications* (Chapter 4) Dordrecht: Springer.

Pearl, J. (2000) *Causality. Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Peto R., Darby, S., Deo, H., Silcocks, P., Whitley, E., Doll, R. (2000) Smoking, smoking cessation, and lung cancer in UK since 1950. *British Medical Journal*, 321, 323–329.

Popper, K. (1959) *The Logic of Scientific Discovery*. Hutchinson, London.

Reiss, J. (2007a) Time series, nonsense correlations, and the principle of common cause (pp. 179-196) In F. Russo and J. Williamson (eds), *Causality and Probability in the Sciences*. London: College Publications.

Reiss J. (2007b) Do we need mechanisms in the social sciences? *Philosophy of the Social Sciences*, 37(2), 163–184.

Reiss J. (2008) Error in economics: towards a more evidence-based methodology. London: Routledge.

Russo F. (2008) Causality and Causal Modelling in the Social Sciences. Measuring Variations. New York: Springer.

Russo F. (2009) Explaining Causal Modelling. Or, What a Causal Model Ought to Explain. In Proocedings of the SILFS Conference, Milan 8-10 October 2007.

SpirtesP., Glymour, C., Scheines, R. (1993) Causation, prediction, and search. New York: Springer.

Steel D. (2008) Across the boundaries. Extrapolation in biology and social science. New York: Oxford University Press.

Stone R. (1993) The assumptions on which causal inferences rest. Journal of the American Statistical Association, 55(2), 455–466.

Suppes P. (1970) *A probabilistic theory of causality*. Amsterdam: North Holland.

Williamson J. (2005) Bayesian nets and causality. Philosophical and computational foundations. London: Oxford University Press.

Woodward J. (2003) Making things happen: a theory of causal explanation. Oxford: Oxford University Press.

What can Neuroscience offer to Economics?

Matteo Colombo[*]
M.Colombo-2@sms.ed.ac.uk

ABSTRACT

The specific regions in the brain that are active when some behaviour is observed is a kind of information that may be interesting for neuroscientists, but how could it be fruitful for economic theory? The thesis defended in the essay is that the brain matters to prediction. By using the Ultimatum Game as a benchmark, it is argued that if the goal of a model of human behaviour is to yield good predictions about important classes of choices, then models that incorporate neurobiological variables may have some advantages over alternative models. The essay comprises two parts. Part I first analyses the Ultimatum Game and illustrates some of its experimental results. Then, it evaluates in detail the merits and shortcomings of Cristina Bicchieri's model based on social norms. It centres on the predictive power of the model, and articulates some challenges it faces. Part II begins with a review of some neurobiological findings which suggest a different approach to construct predictive models of human behaviour. Drawing upon these findings, it gives some reasons why the predictions of a neurobiologically-informed model seems to have some advantage over those of Bicchieri's model. A critical discussion of the thesis against possible objections terminates the essay. The conclusion follows that one way in which the study of the neurobiological foundations of decision-making might be fruitful for economic modelling is in enhancing the predictive quality of its models.

INTRODUCTION

The specific regions in the brain that are active when some behaviour is observed is a kind of information that may be very interesting for neuroscientists. But what does it add to economic theory? The claim defended here is that the type of knowledge of brain processes offered by neurosciences matters to prediction (see also Camerer 2007). To the extent I am right, the attempt to integrate evidence, concepts and tools from the fields of economics, psychology and neuroscience within the new domain of neuroeconomics will turn out to be a realization of the methodological ideal described by Milton Friedman in "The Methodology of Positive Economics". There, Friedman advocates a requirement of predictive success for judging a "positive" scientific theory: "The ultimate goal of a positive science is the development of a "theory" or "hypothesis" that yields valid and meaningful (i.e. non truistic) predictions about phenomena not yet observed." (Friedman 1953, p.7)

An important cautionary note right from the beginning: It may sound that I focus on prediction in order to avoid "the real" problems – e.g. issues about explanation, and understanding. I would like to reply to this charge by giving it a different twist. One way to read my claim is that *however* you want to explain, or understand human behaviour, a model that takes into account neurobiological parameters seems to have an advantage over competitors for predictions. If we consider the importance that predictions have in our lives, it may be easier to acknowledge that the choice to limit a research on prediction is still worthwhile. Prediction serves at least two crucial goals: one pragmatic, the other epistemic

[*] Department of Philosophy, University of Edinburgh, UK

(see Rescher 1997). On the one hand, predictions are necessary to interact successfully with our environment. E.g., predicting that if we jump off the top of the Cathedral of Milan, nasty consequences for our health will be extremely likely to ensue, we may prefer to take the stairs to go down. On the other hand, prediction is a test of scientific theories. It is a common view that 'in assessing the confirmation or evidential support of a hypothesis, we must take into account especially (and perhaps exclusively) the predictive success or failure of its *predictions*' (Musgrave 1974, p. 2).

Before spelling out the two criteria that will orient my assessment of the comparative merits and problems of the predictions given by concurrent models, it is worth making clear what it is meant by 'prediction'. Following Forster (2008), '[t]he term prediction is always used to refer to the "diction" of past, present, and future events', the "diction" of a claim that previously we had no reason to believe. Hence, a good prediction is not necessarily to refer to the future. I take a good prediction to be one that is both *secure* and *informative*. A secure prediction is based on reliable, well-evidentiated grounds. Because of this, it is likely that the prediction turns out to be correct. E.g., the prediction that there will be a full moon within the next thirty days is secure. The more adjustable parameters a model has, the more secure its predictions, but the greater the risk of accommodation. *Accommodation* is one of the risks of a secure prediction. A model accommodates the data when it is merely consistent with them. The typical case of accommodation is when a set of data is deduced from the model (hence, model and data are consistent), *and* the same set of data was used in the construction of the model. E.g., a model of the form (H&Y) does not genuinely predict that Y: It accommodates Y. A good prediction is not trivial, is *informative*. Informative predictions are not vague, preferably they are quantitatively accurate. E.g., the prediction that the comet Hale-Bopp had its closest approach (at a distance of 1.315 AU) to Earth on March 22, 1997 is informative.

With these conceptual tools at hand I can now make clear the structure of my argument. The Ultimatum Game (UG) is my benchmark for evaluating the predictive quality of a model. I begin by detecting an "anomaly"[1] in the UG. Standard game-theoretic prediction is at loss in the face of the behaviours of people playing the UG. Thus, if we want to stick to the predictive success requirement, we must revise our model. Several enriched models have been proposed. Some of them, because of their flexibility, seem promising to predict well. However, I argue that this flexibility comes at some cost: Because it is problematic to obtain a reliable measure of the adjustable parameters they contain, these models are particularly subjected to merely *accommodate* the evidence. That's the main reason why we should try to develop a model with adjustable parameters flexible enough to predict the variety of "anomalies" observed in the UG, *and* specific enough to figure out how to reliably measure them so that the model enables us to give predictions that are both informative and secure. I argue that a model that incorporates neurobiological variables seems to satisfy these requirements. However, neurosciences offer no panacea. I discuss some of the limits of the type of neuroscientific results that back my argument, and I provide positive suggestions on how these shortcomings might be tackled.

If my argument is correct, it follows that *if* we want to take Friedman's advice seriously, then we have good reasons to try to incorporate neurobiological parameters in our models of decision-making in economic environments.

[1] Borrowing Richard Thaler's words: an anomaly is 'an empirical result … if implausible assumptions are necessary to explain it within the [rational choice] paradigm' (Thaler 1988, p. 195).

The essay is organized in two parts.

The first part is divided into two sections. Section *I* explains why I believe that the UG is interesting and worth studying. I describe the UG, its game theoretic analysis, and how people play the game. Section *II* introduces the concept of "enriched models". These models try to account for the actual behaviour of the people who play the UG by appealing to such concepts as "fairness", "warm glow", "envy", "social norms". I restrict my analysis to one of these proposals. After having motivated my choice, I analyze in detail how the norm-based approach defended by Bicchieri (2006) works when it's called for accounting for the UG results. I argue that this kind of account risks to merely accommodate evidence and predictions because of the nature of its adjustable parameters.

The *second* part of the essay is shorter. The reason is simple, and it is important to be clear: Currently, there is no neurobiologically-informed model. Drawing on both the detailed analysis of the first part and on some recent neurobiological findings, my goal is to give some reasons why *if* we want good predictions, then we should try to point to quantifiable biological variables which have a large influence on behaviour and are underweighted or ignored both in game theoretic, and enriched models such as Bicchieri's. After having described one study on the neural basis of economic decision-making in the UG, I critically discuss possible objections to the potential predictive significance of neurobiological variables.

I. THE ULTIMATUM BARGAINING

Game theory is a collection of models attempting to understand situations in which decision-makers interact with one another. Game theoretic analyses predict that *rational*, *self-interested* players will make decisions to reach outcomes, known as *Nash equilibria*, from which no player can increase her own payoff unilaterally. Strategic bargaining behaviour is one of the concerns of game theory.

To see the role played by the assumptions of rationality and self-interest in game theory, let us consider the Ultimatum (or "take-it-or-leave-it") Game which is one of the simplest form of bargaining. This two-stage, two-person game is defined as follows. A sum of money m is provided. Player 1 proposes that x units of the money ($x \leq m$) be offered to player 2. Player 1 would retain ($m - x$). Player 2 responds by either accepting or rejecting the offer x. If player 2 accepts, player 1 is paid ($m - x$) and player 2 is paid x; if she rejects, each player receives nothing (0, 0). In either case the game is over.

Two features of the UG are worth emphasizing. First, the UG is a non-cooperative game: Players cannot make binding agreements about what to do. They have to form expectations about other players' action without communicating. Second, the UG is very simple in its game-theoretic analysis: It requires only two assumptions to make a prediction. It is simple in its instructions: Players understand quickly and without effort the rules of the game.

'The UG is one of the most successful experimental designs in the history of the social sciences' (Guala 2008). According to Guala, the success of the UG can be explained if we focus on some of the epistemic features that qualify it as a "paradigmatic experiment". One of these features is its versatility: A paradigmatic experiment enables us to make comparisons and draw a variety of inferences in different contexts of scientific inquiry. This versatility makes the UG suitable for my argument, where I evaluate the comparative merits and problems of the predictions given by two models built in two different contexts.

The game theoretic analysis of the game yields a precise prediction about what players will do. The prediction of a restrictive concept in game theory, *the subgame perfect equilibrium*, is that for any positive amount offered by Player 1 (the proposer), Player 2 (the responder)

knows that she faces a choice between gaining nothing (if she refuses the offer) or something (if she accepts). If the responder maximizes her own payoff, she will accept any positive amount. If the proposer maximizes her own payoff *and* expects the responder to maximize, she will offer the smallest amount possible. Hence, the proposer will offer the minimum possible split to the responder, who will accept.

The assumptions of *rationality* and *self-interest* entail this prediction. According to the self-interest assumption, players prefer more money to less, and don't care about the outcomes or preferences of others. Notice that this is a narrow conception of "self-interest", concerned with money alone. Rationality has to be understood as *practical, instrumental, rationality*. Given the agent's perfect knowledge of the outcomes of the alternatives open to her and to the other player, and given that she can identify the best of them, she will be practically rational in choosing action *A* if *A* is the action she believes will lead to the consequences she prefers.

The experimental literature on the UG indicates a robust behavioural pattern at odds with the game theoretic prediction. Since the first experiment which studied the UG (Güth, Schmittberger, & Schwarze 1982), the UG has been studied in many diverse settings where different parameters of the game were modified. Supposing that the total sum of money is 10, the split offered is typically around (6,4). And low offers, namely offers around 2 or less out of 10, are very likely to be refused (Camerer 2003).[2] The UG is typically *anonymous* and *one-shot*: Players don't know the identity of the other, and play only once. The rationale for these two characteristics is to abstract away from the possibility of incentives for reciprocity and cooperativeness - which would be involved in repeated games with the same partner - and to keep players' behaviour insulated from such influences as the desire to please the experimenter, or the fear of ruining a friendship, or the incentive to build a good reputation – which would be involved if the identity of the player was known.

The behavioural pattern displayed by people playing the UG immediately raises the well-known Duhem problem: Why do people tend *not* to play the subgame perfect equilibrium, and instead tend to coordinate on 50-50 or 60-40 splits? What part of the standard game-theoretic model has to be blamed for the anomaly? And, once we have identified the culprit, what modification has to be done so as to enable the model to predict well?

II. BICCHIERI'S SOCIAL NORM MODEL. AN ANALYSIS

By virtue of the simplicity of the UG, we have reason to maintain that the players are rational in the minimal sense specified above: their actions follow from their preferences and beliefs. What seems to be revised is the assumption of strict self-interest that individual preferences are concerned with money alone. Accordingly, the target of most of the theoretical developments in the game-theoretic literature of the last two decades has been the assumption of strictly self-interested preferences. The most common move has been to build models with non-standard utility functions, according to which individuals have "other-regarding" preferences. This kind of models allows a player's utility function to take into account the outcomes, preferences, and expectations of *other* players. Notice that a common feature of these "enriched" models is the preservation of the logical framework of expected

[2] There is however some variation in the findings. The most remarkable variations are found either across cultures or across subjects with a neurological condition such as autism. (see Roth *et al* 1991, and Henrich *et al*., eds 2004 on cross-cultural variations; Sally & Hill 2006 for a research with autistics).

utility theory: They do not reject the rationality assumption, they point to the maximization of a non-classical utility function whose empirical substance is provided by the new parameters.[3]

In this section, I critically analyze the "enriched" model developed by Cristina Bicchieri in *The Grammar of Society* (2006) (henceforth, GS). I focus on one model to facilitate a more detailed discussion than it would be possible if I were to consider a wide range of current views. I focus on Bicchieri's for two reasons. First, hers is one of the most recent and promising proposals; second, and importantly, she argues that her model fares better than the alternatives when it comes to prediction. In my discussion I pay special attention to whether, and to what extent, this last claim is justified.

Every known society has a multitude of social norms that regulate the behaviours of individuals in a variety of situations. In a given social context, the same kind of social norm might produce different behavioural patterns across individuals. If most people abide by social norms, then norms can account for behavioural patterns observed in a population. And different norms can account for behavioural variations across societies. Consequently, social norms are important for predicting the behaviour of individuals. These observations call for elucidation:

1) What is it meant exactly by a 'social norm'?
2) What are the *mechanisms* that regulate the power social norms have to influence human behaviour?
3) What are the *conditions* under which individuals are likely to follow a social norm?

In GS, Bicchieri attempts to account for these three problems by providing the foundations for a new model of human behaviour based on a precise characterization of a *social norm*.

Definition According to Bicchieri, a rule is a social norm in a population if and only if a sufficient number of people in that population:

(i) know that the rule exists and applies in situations of a certain type S, and
(ii) prefer to conform to it, in situations of type S, on the condition that
  (a) it is believed that a sufficient number of others conform to it in situations of type S, and, either
  (b) it is believed that a sufficient number of others expect one to conform in situation of type S, or
  (b`) it is believed that a sufficient number of others expect one to conform in situation of type S, prefer conformity, and may sanction one if one does not conform (GS, p.11).

Suffice here to clarify three points about Bicchieri's definition. First, a social norm is a set of mutual expectations, and a communality of beliefs is a precondition for its existence. A social norm like "tipping for service in a restaurant" has no reality other than our expectations that others leave a tip in a restaurant, *and* that others expect us to tip in the same type of circumstance. Second, a social norm cannot be simply identified with a recurrent collective behavioural pattern. Taking a shower in the morning is not a social norm. I take a shower in the morning whether or not I expect others to do the same. Moreover, a rule can be a social norm in a population, even if compliance to it is not observed. Imagine the social norm in a population that whoever first makes a proposal that something has to be done is directly responsible for making sure that the proposal is carried out. During a seminar students may

[3] Well known "enriched" models are Rabin (1993), Fehr & Schmidt (1999), Bolton & Ockenfels (2000).

avoid suggesting a certain topic of discussion fearing that that social norm will be followed, and, hence, they would have to prepare the talk. In this situation nobody is violating the norm. Everybody is eluding it. Third, the conformity to a social norm is *conditional*: One has a preference to conform to a norm *N* in a situation of type S, under the conditions that one expects others to conform to *N* in S (*empirical expectations*), and one believes that others think one ought to conform to *N* in S (*normative expectations*). This condition is discussed in detail later. Notice, for the moment, that the conditionality of the preference to conformity to a norm lends itself to empirical testing. Were the expectations that underlie a social norm to be different, we would predict behaviour to change in determinate ways. Clearly, unless we have an account of when, how, and to what degree the expectations that constitute a social norm affect behaviour, there is little hope in drawing precise, informative predictions. We need understand the *mechanism* by which social norms influence our behaviour, and the *conditions* under which individuals are likely to follow a social norm.

Mechanism Bicchieri undertakes the first task by relying on findings from experiments in cognitive psychology (GS, Ch.2). The mechanism works as follows. Subjects interpret and categorize a given context as a function of the situational cues that spark their attention. The process of categorization relies on *spreading activation*. That is, the activation of the representation of a certain concept spreads to representations of concepts related to it. E.g., when we are presented with the stimulus word *tiger*, we retrieve not just the representation of a tiger, but also related representations like *feline*, *predator*, etc. Then, depending on how subjects categorize the context, a script of a certain type activates. *Scripts* are cognitive structures we acquire through personal experience and habit that represent stored knowledge about people, objects, events, and roles relevant to the situation at hand. Scripts prompt beliefs and expectations about social roles and sequences of actions appropriate in that situation. A "restaurant script", e.g., represents roles (waiters and diners) and sequences of appropriate actions (diners enter the restaurant, wait to be seated at a table; waiters take their order; diners eat, ask for the bill, pay, leave a tip, and leave the restaurant). '*Social norms are embedded into scripts*' (GS, p.94): Social norms are among the set of beliefs, expectations, and preferences prompted by a script. To see this mechanism in action, consider an UG framed in terms of the gains from a transaction between a buyer and a seller. Player 1 is said to be the seller; she is endowed with €10. Player 2 is said to be the buyer. A table registers the profit of the seller and of the buyer for each price (€0, €1, €2,…, €10) charged by the seller when the buyer decides to purchase. The profit of the seller is equal to the price she states; the profit of the buyer is €10 minus that price. The profit of each is zero if the buyer refuses to purchase at the price stated by the seller. In such context, the cues provided are likely to guide the interpretation of the situation in terms of a market situation. Presumably, the categorization "market exchange" activates a script that defines determinate mutual expectations underlying a social norm (if it exists) among the players.

When compared with the results of an UG with standard instructions, it is found that the "buyer-seller" manipulation elicits lowered offers, whereas the rejections rates remains unchanged (Hoffman *et al* 1994). Indeed, in western culture the right of sellers to quote a higher price is not usually questioned, nor that of the buyer to decide to purchase or not to purchase. Two points are worth noticing. First, consistently with Bicchieri's account, the example just provided unambiguously shows that context matters. Two UGs with the same logical structures, but embedded in different contexts, with different situational cues, are likely to elicit different behaviours because they are likely to prompt different social norms. Second, as acknowledged by Bicchieri, '[*t*]*he predictive power of a theory of norms therefore depends*

on knowing which situational cues trigger which norm' (GS, p.76). But it also depends on a number of other conditions. In the remaining of this section it is argued that the fulfillment of these conditions represent important challenges to the predictive success of her model.

Conditions for prediction There are two conditions under which an individual is likely to follow a social norm *N* in a context of type *S*. First, *N* is correctly identified. Second, individuals in *S* have the right kind of expectations, and therefore *N* exists in *S*.

Identifying a norm There are at least three reasons why *identifying N* in a certain context might be problematic. The first is that, if we exclusively focus on the behaviour of people e.g. in the UG, there might be alternative social norms, *N*, *N`*,…,*N^n*, that entail the same behavioural pattern. This is the problem of (deductive) *underdetermination*, and it would threat the informativeness of the prediction given by the model. If we want our model to give informative predictions, we need an assessment of the social norm likely to be in place in the UG *independent* of the behaviour observed in the game itself. Independent measurement is one of the remedies for limiting the bite of the underdetermination problem. With social norms, however, independent measurement is made difficult because they might be too vague, and consequently it might not be clear how to assess them independently of behavioural data. This is the second problem. The third is that individuals of a population might be unaware to be in a context where *N* applies. This has to do with the situational cues, crucial in priming *N*.

Underdetermination Consider this slightly different form of the UG. Before the responder hears the offer, she must set an acceptable offer range: she is asked whether she would accept a 100-0 split, and then whether she would accept a 90-10, a 80-20, a 70-30, etc, until a point is reached where she would accept anything higher. If player 1's offer is below her acceptable range, her response would count as a rejection. Player 1's proposal is finally revealed. The players belong to the same population, and there are no contextual cues that affect their expectations. An experiment with Gypsies in Vallecas, Madrid, shows that in this situation although 97% of proposers offered an equal split, as responders the Gypsies were willing to accept completely unfair offers: The acceptance of the *zero* offer was the modal value (Pablo *et al* 2006). In order to have predicted such behavioural pattern on the side of the respondent, we would have needed to know about the social norms existing in that population relevant to an UG-situation. One possible way to make sense of this behavioural pattern is by means of a norm of hospitality. Hospitality seems to be a social norm that primarily affects the behaviour of those who can offer: The proposer offers half and the other accepts whatever is offered. Once this norm has been recognized as obtaining in that situation, Bicchieri's model wouldn't have problems in predicting the behaviour of the Gypsies. However, the same behavioural pattern is also compatible with a different social norm like "help the needy". According to this norm one would expect that if the proposer offers zero, then she is needy, and this would motivate the respondent to accept a zero offer. With some imagination we can devise other social norms compatible with the same behavioural pattern. Because of underdetermination, it would be logically possible to find an infinite number of models that will accommodate the same behavioural pattern without genuinely predict it. One way to tackle the problem is to "measure" the expectations underlying the social norm we assume to be in place *independently* of the observed behaviour. As convincingly argued by Larry Laudan in different places (e.g. Laudan 1990), deductive underdetermination does *not* entail that the *choice* of a model is underdetermined. One of the criteria used to decide which model is better, and consequently to be preferred over rivals, is independent measuring of its assumptions. Independent evidence about a social norm in a population would facilitate us to uncover the

real expectations and motives behind the behaviour in an UG. And knowing motives and expectations would enable us to prefer a model that yields informative predictions.

Vagueness and Independent Measures How we measure a social norm independently of the observed behaviour is problematic since social norms are complex, vague, and might be impossible to define them in a precise way. A norm of hospitality can be regarded as an example of a vague social norm. I use 'vague' meaning that hospitality cannot be precisely defined because it has many diverse aspects. In a typical situation where hospitality exists we might expect that the host invites the guests in her house; the host entertains them with kindness; she offers food and drinks; the guest accepts whatever is offered with goodwill, etc. Hospitality might exist in a context even if any one of these expectations is missing as long as people in that context share some of the other expectations. Yet, there are at least three methods to measure a social norm: first, questioning people; second, inferring norms from behaviour (other than that displayed in the UG); third, in-depth ethnographic research. Each of these methods has biases that may undermine the reliability of the measure.

"Which split proposed in an UG would be fair?", or "What is the norm to follow in situation of type S?" are the kind of questions that may enable us to determine the content of a social norm of fairness in a population. Questioning people, however, may yield an unreliable measure because they may lie, may give the answer they suppose the research wants, may understand the question in different ways.

Apart from the reasons given above for why a social norm cannot be simply identified with a recurrent collective behavioural pattern, there are at least two further challenges in measuring a social norm from behaviour. First, to account for behaviour by citing a social norm may lead to circularity if the norm is first measured from the behaviour in question. Second, there is the problem of simultaneous attribution of belief and desire, namely the problem of 'discriminating the respective roles played by an agent's beliefs and desires in the production of the actions we observe her to perform' (Bradley unpublished). On the basis, e.g., of the observation that a number of people leave a tip at a restaurant, we might attribute to them the belief that there exists a social norm of tipping at a restaurant, and the desire to follow it. Alternatively, we might attribute to them a belief that the service has been really good, and the desire to reward good service regardless of there being a social norm of tipping at a restaurant. The two alternatives entail different behaviours under different circumstances: After an unsatisfactory service, according to the first belief-desire attribution one may still leave a tip following the social norm; instead, according to the second belief-desire attribution one will not. The choice as typically observed does not *alone* allow us to decide between these two alternatives, and the many other possible ones. To be sure, this is not a knock-out challenge since each hypothesis is testable in principle; that is, there are conditions under which we can determine whether one of them is probably false. Nonetheless, the problem remains to specify an underlying theory of beliefs and desires which enables us to systematically infer one's mental states from her observed behaviour.

The third method, that of ethnographic research, has been adopted by a group of economists and anthropologists who set out to study the foundations of human sociality through classic economic experiments like the UG in fifteen "small-scale societies" in South America, Asia, and Africa (Henrich *et al* eds. 2004). For each population experimental research was flanked by independent information concerning social context, political and economic structure, religious beliefs, etc. It turned out that the between-group behavioural differences in the UG were related to indicators of patterns of social and economic interaction (e.g. level of market exchange, importance of anonymity in commercial transactions) that framed the daily life of the population. Yet, there is no guarantee that a significant correlation will be

always found between behavioural patterns in the UG, and the socio-economic variables that researchers decide to study. This may be due either because there is in fact no correlation between UG decisions and variables of socio-economic patterns, or because the researchers don't collect the appropriate information, or because the information they collect from certain individuals is not representative of the whole population but of only some subgroup which has to be identified.

Context and Cues The third condition that would enable us to predict that an individual is likely to follow a norm is that the individual has to be aware to be in a context where the norm applies. Situational cues govern the mapping between context, recognized as being of a certain type courtesy of the cues, and activation of the social norm appropriate to that context. But, what drives people's attention to situational cues? The key notion to tackle this question is *salience*. A salient item stands out relative to neighbouring items, thereby sparking people's attention. A red dot surrounded by green ones is salient. Saliency is perhaps best understood in the field of visual perception where it is defined in function of such cues as colour, intensity, orientation, and motion. In the case of norms, salient cues – Bicchieri suggests (GS, p.112) - 'may involve a direct statement or reminder of the norm, observing others' behaviour, similarity of the present situation to others in which the norm was used, as well as how often or how recently one has used the norm.' Salient cues prime the expectations underlying the norm. Conditional on these expectations, certain behaviour is likely to ensue.

Bicchieri tested her model by manipulating salience. In an UG, people's attention has been cued by information about the normative expectations of others that had played the game before. The prediction obtained that more players would follow a norm of fairness, which in that context would dictate an equal split (Bicchieri & Xiao 2008).

The potential problem for prediction here stems from a tension within the model. On the one hand, Bicchieri's model makes reference to *type*-situations. On the other hand, the appeal to situational cues makes the interpretation\activation of social norms dependent on specific, *token*-situations. The interpretation of a social norm is *local*: Fairness, e.g., has different meanings in different circumstances, depending on the people, objects and environment that define the situation. A *type-situation* is a general type of situation like "football match", "bargaining", "theatre-play". A *token-situation* is a specific situation of a certain type performed in a particular context. Hamlet played at La Scala now, with a certain setting, certain actors, and a certain audience, is a token-situation of the type "Hamlet-theatre-play". We can *judge* that the individual *i* is in a situation of a certain type; e.g. we can judge that *i* is at a theatre play of Hamlet. But *i* makes decisions, has expectations, follows a social norm, always in token-situations. E.g. *i* expects to smile when the Hamlet she is attending is being performed by a company of funny comics wearing fancy dresses; instead, she expects something deep when the play is being performed by a company of continental philosophers. The two situations are of the same type, but differ in situational cues. The difference is likely to prompt different expectations, preferences and actions. Consider again the UG. It has been shown that in a common UG-token-situation, a bargaining type-situation, attractive people are treated differently by others in that they are offered more, and they are expected to give more (Solnick & Schweister 1999). Now, in a token-situation of a certain type there may be innumerable situational cues that can spark one's attention. The attractiveness of the people in a bargaining type-situation can be one. In function of our personal history, we will interpret the context in one way or another, we will find one person attractive or not. The interpretation of the context will affect our expectations and preferences, and hence may prime this or that norm we may follow in that token-situation. Even if we can judge that *i* is in a type-situation *S*, we still need careful inquiry into the token-situation. The potential problem for prediction is here with informativeness. For we would like a model that could be applied to type-situations

besides the token-ones from which it was deduced. We would need a general underlying theory that specifies some systematic functional relationship between situational cues, which are adjustable variables of the model, and experimental results. We would like something analogue to a saliency map, which represents visual saliency of a corresponding scene, made available by neurocognitive research on visual perception (e.g. Koch & Ullman 1985). As far as I know there are no such "maps" in the field of decision-making yet. The challenge is in measuring, or quantifying, the internal state of an individual, such as her personal history, the goals, beliefs and motivations she has at a time.

Having the right kind of expectations One has a *preference* to conform to a norm *N* in a situation of type S, provided *N* exists and has been correctly identified, under the conditions that she expects others to conform to *N* in S (*empirical expectations*), and she believes that others think she ought to conform to *N* in S (*normative expectations*). There are at least two problems here. First, the conditions are *not* sufficient to determine one's following a rule. This may threat the security of the predictions the model gives. Second, there is no hint as to how (quantitatively) determine the *likelihood* of observed behaviours. This is a drawback for predictive informativeness.

The insufficiency of having the right kind of expectations is apparent: We often have other personal motives for not following a norm. I may live in a society with a strong norm of revenge. I expect that most people take vengeance on those who have wronged them, and I believe that others think I ought to take revenge on those who have wronged me. However, I have also a stronger preference for not harassing others. Bicchieri acknowledges that the presence of a norm of revenge, 'and its salience in a particular situation, motivate me to act in a congruent manner, but my behaviour is ultimately explainable [and predictable] only by reference to my preferences and expectations' (GS, p.22). Suppose we ask the responder in an UG what she believes is fair in that UG-situation, and whether she expects others to follow it. This might give us information that this person believes a norm *N* applies in that situation, and this would lead us to predict her action in accordance with *N*. However, such action might still not occur, due to the presence of other motives and incentives which, unbeknownst to us, bear on her decision more than her expectations related to *N*. To remedy this problem, Bicchieri suggests (Bicchieri & Chavez 2008) that a fine-grained account of individuals' sensitivity to specific norms would give us more reliable grounds to predict their behaviour since it would enable us to compare the strength of concurrent motivations in a given situation - e.g. the preference to follow a norm versus other preferences that may overcome that preference. How does the suggestion work exactly? Consider responders and proposers in an UG. Each player can be of different types. In order to measure the expected norm-sensitivity, call it K, of a type of proposers (P) 'we may ask responders about the expected type of offer versus the offer that is fair. Then we would have indirect information about the expected value of KP: If the two values differ, the expected KP is low. If they are the same, the expected KP will be high or low depending on the (high or low) values of the expected type of offer and the offer that is fair. If we ask proposers about the distribution of normative expectations on the part of responders (R), and we then observe their offers, we can get information about the expected KR.' (Bicchieri, personal communication). I see three problems with this suggestion.

First, Bicchieri's reasoning entails that we can *only* have indirect evidence about one's K. The information we obtain following her suggestion is a measure of the type of individuals expected by the *others*. The grounds to ascribe a certain K to a responder in an UG would be the proposers' beliefs about the responders' normative expectations, seeing whether the

proposers behave according to their beliefs. The problem is whether this kind of evidence gives us grounds strong enough to reliably ascribing Ks to respondents. My suggestion is that direct evidence about one's K would do a better job. For a proposer may believe e.g. that a responder has strong normative expectations of fairness, and accordingly she makes a fair proposal; yet, the respondent's K doesn't match with the proposer's expectations since she also accepts unfair offers. The contrast between having direct evidence that the individual *i* is sensitive to a norm and having indirect evidence thereof is between *i*'s having certain expectations and preferences and *her* saying so, and *another*'s believing that *i* has certain expectations. A direct way of knowing what a respondent believes and prefers is thus her own sincere confession. But this would pose again the threat of accommodation since the model would simply describe the situation at hand.

Second, there is a problem with informativeness since Bicchieri's approach seems to have a difficulty in providing a quantitative answer to the question: *How well* the model predicts? We would like to know about the predictive strength of a model by means of probabilities that it would assign to the predicted outcomes. If our predictions depend on a parameter K, we would like to know a systematic way to determine to *what degree* one is sensitive to a norm in a context. In default of such positive proposal the predictions of our model lacks in informativeness.

The third problem is conceptually prior. It is that of determining quantifiable *individual*-level variables statistically correlated with players' decisions. Bicchieri's model takes into account *types* of individuals, defined by their beliefs about the relevant norm in a type of context, their expectations about their opponents, and their norm-sensitivity (Bicchieri, personal communication). The model seems promising with predictions *across* groups of individuals where we can take group-membership as defining a type of individual. However, it is unclear how it would predict the behaviour of specific individuals *within* a group. It is, in fact, possible that individuals of the same type behave differently. For the way Bicchieri construes a type of individual allows that they can have different motives and preferences. Although variation between types of individuals can be accounted with economic, social, or cultural differences, the same does not apply within the same type of individuals.

Such a charge may seem unfair: the social sciences (and to a great extent also psychology) are less concerned with single individuals than with statistical tendencies across individuals. Also in physics it is customary to take into consideration the behaviour of many particles rather than a single one.

My point however can be read in a weaker way. Even if the behaviour of a single individual within a group may not be that interesting, collecting appropriate individual-level information systematically correlated to one's choice may be useful to make predictions more secure and informative *across* groups of individuals after statistical generalization. As usual in science, we may examine a sample of a certain class of individuals and then generalize our findings to the class as a whole. This information may be neurobiological. The second part of this work articulates such proposal.

III. NEUROBIOLOGICALLY-INFORMED MODELS

Twenty-one years after Güth *et al.*'s seminal work on the UG, a group of psychologists and neuroscientists led by Alan Sanfey analyzed subjects with functional magnetic resonance imaging (fMRI) as they played the UG. Sanfey *et al.*(2003) compared the brains of subjects responding to 50-50, 60-40 offers (in the experiment the total pie to split was $10), and 90-10, 80-20 offers. Three brain areas were found to be differentially activated: the Dorsolateral

prefrontal cortex, the anterior cingulate, and the anterior insula. For the purpose of this essay, the crucial finding is the correlation between the activation of the anterior insula with choice-behaviour. Specifically: First, the activation of the anterior insula was significantly correlated with the rate of rejection; Second, the magnitude of activation was "a function of the amount of money offered to participants" (p.1756). Third, the activation was "uniquely sensitive to the context" (Ibid.): there was greater activation for a 80-20 offer from human partners than the same offer from computers. From the first finding follows that whether players reject an offer or not may be *predicted* with a certain accuracy by the level of their insula activity.

Sanfey *et al*'s experiment is quite typical in the world of so-called "neuroeconomics": From a subject who confronts a choice problem D, a pair (d, *x*) is drawn. d is the alternative in D chosen by the subject; *x* is a vector of numbers representing the activities measured in various areas of the subject's brain during the period between the moment she is presented with the choice-set of D and the moment she chooses d. After statistical analysis, a correlation between choice-behaviour d, and the activity of a specific brain area may become apparent. Building upon this kind of correlation, it may be possible to integrate neurobiological variables into behavioural models. The predictive power of such models would be bound to the significance and robustness of the neurobiological finding. The remaining of this essay is a critical assessment of the comparative merits, and limits, of the neurobiological approach to modelling I wish to suggest. Security and informativeness of the predictions that this kind of models would give orientate my discussion as with Bicchieri's.

Security Sanfey *et al*'s experiment involved 19 American subjects that completed 10 rounds playing the game with a human partner, and 10 with a computer partner. Rubinstein (2006) and Harrison (2008) crisply point out that the inferences drawn from experiments of this kind are highly problematic because of their small, culturally homogeneous samples, the methodology of pooling subjects, and the heavy statistical machinery required to make raw neurological data amenable to analysis. How can one feel comfortable about building upon such weak grounds? I would like to make two points about Rubinstein's and Harrison's critiques.

First, their charges about samples and statistical analyses are empirical in nature. As such, there seems to be no principled reason why the methodological standards that enable us to warrant the inferences from a neuroeconomic experiment couldn't be raised. Whereas statistical flaws can be solved with careful data analysis made explicit and open to scrutiny, the problem with samples size can be faced in the long run, when many *replications* will tell us about the *robustness* of regularities like that lying below Sanfey *et al*'s set of data. I use 'replication', as opposed to 'exact repetition' (see Radder 1996; Guala 2005, Ch.2). A replication involves some (slight or radical) modification of the original experimental design. A nice replication of Sanfey *et al*'s experiment might involve subjects from different cultures. At the moment we can only speculate e.g. that in Israel, where lower offers are accepted more often (Roth *et al* 1991), we may expect less insula activity for the same sized offer than in the US, and less rejection. If the regularity firstly observed survives across replications, then we have reason to conclude that it is *robust*: It doesn't depend on details of the situation or on the particular statistical assumptions used to derive the results. Therefore, by emphasizing the evidential value of replications, and by noticing that we have no reason why we couldn't expect replications of neuroeconomic results, and higher statistical standards, the bite of these charges can be limited.

Second, a better way of evaluating Rubinstein's and Harrison's critiques is to ask whether the *evidential grounds* for neurobiological predictive claims may be more reliable than the grounds for psychological ones. Analyzing Bicchieri's model we discussed some of the

evidential problems with measurement and ascription of expectations to people. I argued that: First, because of the nature of social norms, secure predictions can be gained at the risk of accommodation; Second, the conditionality of preference to follow a social norm seems to open an epistemic gap between the actual motives, in a *token*-situation, that ultimately render a prediction correct\incorrect, and the information about *type*-situations and *type* of individual, necessary to ground the predictions of Bicchieri's model.

The grounds for predictions for a model which integrates neurobiological parameters consist in the empirical finding that two quantitative factors (e.g. insula activation and rate of rejections in the UG) are correlated in such a way that the behaviour of the one foreshadows the behaviour of the other with statistical significance. The security of the predictions is bond to the stability of the functional specialization of the target neurobiological structure *across brains and situations*. On the one hand, however, the activation of the target structure may not be sufficient to warrant the prediction since it may be overcome, or blocked, by the activities of other structures. However, the problem is not insoluble: We can identify *to what degree* the insula activity is sufficient to warrant the prediction of rejection in the UG by inquiring its relation\interaction with other brain structures. On the other hand, the activation of the structure may not be necessary. As before, the problem can be solved with further, accurate experimentation. In order to assess to what extent the insula activity is necessary in predicting the rejection rate, it would be interesting, e.g, to study the behaviour in the UG of neurological subjects with lesions in the insula; but, as far as I know, such experiment has not been set up yet.

A more serious threat to the reliability of the evidential grounds for neurobiologically-based predictions is the variability of brains across individuals. The problem, here, is not with token vs. type-situations, which would be bypassed by focussing on brain activations since brain activations are *already* sensitive to the token-situation where a person acts. The problem, instead, has primarily to do with the *plasticity* of the brain: Plasticity, or neuroplasticity, is the capacity of the brain to reorganize neural pathways based on new experiences. Both structure and function of developing brains are shaped, to some extent, both by the environment and by cultural experience. Differences in neural responses to a given stimulus are also likely to exist All brains, then, are different from one another to some degree. To the extent that different cultural norms and practices, and different environments exist across social groups differences in neural responses to a given stimulus are also likely to exist (Chiao & Ambady 2007). This likely variation poses at least one substantive problem to a model neurobiologically-informed. We might expect to find differences in the *type* of neural activity correlated to the same *type* of behaviour across subjects. Were this actually so, then the basis for our neurobiologically-informed model of human behaviour would be intrinsically unstable. However, we have reason to believe that the brain variance across groups is not so dramatic.

First, it is worth noticing that despite notable progress in describing cultural variation at the behavioural (and genetic) level, relatively little is known about how the structure and function of the human brain vary across subjects and cultures. Hence, sweeping claims are not justified by current evidence. Second, although there is no arguing that there is some kind of plasticity in our brains, there are clear limits on plasticity (Gazzaniga, Ivry, & Mangun 2002, ch.15). Different types of neuroplasticity occur during certain critical periods, notably: 1. During normal brain development when the brain begins to process sensory information, when it is still immature; 2. After brain injury, especially when the lesion involves the somatosensory system, to compensate for lost function; 3. Through adulthood underlying learning and memory mechanisms. After these critical periods, the central nervous system is characterized by a relative rigidity. Moreover, when it occurs, neuroplasticity seems to be particularly

significant at micro-levels of organization, e.g. at the synaptic level (Elman *et al*. 1996; Quartz & Sejnowski 1997). Here, plasticity involves subtle changes in the strength of synaptic connections between neurons which underlie the mechanisms of learning and memory. For these reasons, it appears unlikely that we find dramatic differences in the *type* of brain activity correlated to the same type of behaviour across adult individuals who have not suffered brain lesions – the target of brain imaging experiments indeed. Brains are shaped, to some extent, both by the environment and by cultural experiences. All brains, then, are different from one another to some degree. To the extent that different cultural norms and practices, and different environments exist across social groups, Given this relative stability in the brain activity underlying a given type of behaviour, more reliable grounds for a neurobiologically-informed predictive model are to be expected when we will have a general, detailed theory of the brain, and as we come to understand better how the instruments and techniques employed by neuroscientists mediate their observations (see Bechtel forthcoming, for the epistemology of evidence and instruments in neuroscience). At the moment, as with any other immature science, there is lots of work to do in the neurosciences: We know a good deal about brain neuroanatomy, and cytoarchitecture. We know less about the physiology and functioning of the brain than we know about other organs. Little is known about the mechanisms underlying the exercise of our cognitive abilities. Currently, the level of detail and quality of our knowledge about the brain is essentially constrained by the improvement of the technological apparatus used during experimentation. How and to what extent we will be able to integrate in a comprehensive theory of the brain the huge amount of data collected during experiments is still to be determined.

Informativeness Consider the following experiment: People are asked to say aloud, sincerely, their reasoning during an UG. The experimenter records everything, and at the end of the game comes up with a model based on the notes she has written down. Her model is highly successful since it fits exceptionally well the data: It gets the behaviour of those subjects right with a degree of accuracy of 99% -- 1% being the error rate due to her disattention. Should we regard that a good predictive model? Clearly not. For it is a mere record, or description, of what was going on during the UG. The evidence that that model purports to predict was already used in the construction of the model itself. The evidence is written into the model: This is a clear case where the model accommodates the evidence without predicting it. Now, what of the objection that the situation of a brain imaging experiment during an UG is analogous to that case? In the imaging experiment, so the charge would go, we would be just *witnessing* the brain activity correlated to the thought processes leading to decision; moreover, if mapped closely enough there may be a 100% correlation between patterns of brain activity and observed behaviour. Hence, the evidence would be written into the model, and, as before, the model would merely accommodate the evidence.

There are at least three main reasons why the case of a neurobiologically-informed model is unlike the one above. All three reasons draw on the special weight of independent assessment. First, the insula, the brain structure that correlates with rejection in the UG, is known to be involved in a variety of behaviours beside the one observed in the UG. Significant activations of the insula have been observed in response to disgusting gustatory and olfactory stimuli (Small *et al*. 2003; Zald & Pardo 2000), to the sight of disgusted facial expressions of others (Phillips *et al*. 1997). Following stimulation of the anterior insula, it has also been found that subjects report feeling nauseous (Krolak-Salmon *et al.* 2003). If we compare these findings with those of Sanfey *et al.*'s fMRI experiment during the UG, "repulse" is the feature common to all the correlated observed behaviour. This commonality would lend predictive

informativeness to our neurobiologically-informed model since it could be applied to situations *beside* the UG from which it was deduced to predict rejection. To this point it may be objected that the remark by a subject 'I am going to reject the offer' is also applicable to a wide variety of situations besides the UG, e.g. selling a car or renting an apartment. So this doesn't distinguish brain imaging studies from the imagined 'write down everything they say about their reasoning' study. Therefore, the point made above seems irrelevant. However, this objection seems to me to overlook a fundamental difference between the two cases. Suppose the ideal situation where a subject is sincere in her saying 'I am going to reject the offer', there is no physical impediment to the realization of her intention, *and* she has no other intentions that may override that one. Contrast this situation with another one where a subject's insula is significantly activated *and* there is no other activity in her brain that may override that one. Now, considering the first case, it seems to me that it is not an empirical discovery that when the subject states that she is going to reject the offer, she has the intention to do so, and she will do so *regardless* of the context of her utterance. Because her behaviour is *conceptually* bound up with the meaning of the predicate 'to be going to', the 'write down everything she says' study would simply describe the situation at hand, and would not provide us with any significant empirical discovery. Consider instead the second case. There we witness a neural event correlated with a person's repulse-behaviour. Unlike the former case, the correlation is not conceptual, but *inductive*: We first observe the correlation in a certain context; then, given numerous observations of the same correlation in a variety of contexts *beside* the first one we infer that a certain brain region is correlated to a certain type of behaviour. Unlike the first study, then, we have made an empirical discovery, a discovery revisable in light of further empirical evidence.

The second main reason why the two cases differ is the following. Courtesy of the vast array of technologies used in neuroscience we might use the data from one experiment to predict *independently* what is likely to happen in another. We might activate or disrupt the insula by using e.g. rTMS (repetitive magnetic stimulation, which temporarily disrupts brain activity in a target region of interest), predict differences in rates of rejection that would follow, and see whether we get them right. That, unfortunately, has not been done.

Finally, as emphasized by Worrall (2006), the interesting cases in which some evidence follows from a model, all involve a general underlying theory which itself stands in need of confirmation from evidence. Although, as indicated earlier, there is currently no comprehensive theory of the brain, there is no reason why after accumulating evidence such theory, which would specify the functional relationship of the insula with other brain structures and other observed behaviours, cannot be articulated.

Provided the model would not simply accommodate the data, and that it would give quantitative predictions by relying on correlations whose significance can be determined statistically, to what extent would it mark a substantive improvement in predicting behaviour? I would like to answer by borrowing an analogy from Rescher (1997, p.127). Take meteorological forecasting in Britain. The easiest route to predictive success seems to stick with the hypothesis that tomorrow's weather will be the same as today. In recent years, this prediction has had a rate of success of 75%. The British meteorological Office, courtesy of its high-tech resources, has been able to push this rate up to 85%. It doesn't look like a terrific improvement. Nonetheless, it is comparatively significant and further improvable in the level of detail. The same might be said for how a neurobiologically-informed model would fare in the UG: It would give comparatively more precise predictions sensitive to potentially significant details ignored by alternative models.

IV. CONCLUSION

Ariel Rubinstein (2008) asks us to assume the following scenario: 'We are able to map all brains onto a canonical brain. The functions of the different areas of the brain are crystal clear to us. The machines used in experiments are cheap enough that thousands of subjects can be experimented on. And finally the data are unambiguous and double-checked. What would be the potential role of brain studies in economic theory?' The suggestion this essay has defended is that the brain would matter to prediction.

A neurobiologically-informed model would not be in opposition to other models such as Bicchieri's. It would provide predictions more secure and informative by incorporating parameters neglected by current models, which may significantly correlate to human behaviour. When compared to models like Bicchieri's, because of the nature of its parameters, the potential advantage of a neurobiologically-informed model is that its grounds would be more easily warranted by independent lines of research without the risk of accommodation, and its predictions would be quantitatively accurate.

One way to summarize my argument is that by observing brain area *x* activation we may bypass such problems as the "specification problem" in models like Bicchieri's. The "problem of specification" is the problem of identifying the norms that prime a certain behaviour. For example, we may have a situation such as:

NORM A -> BRAIN ACTIVATION *x* -> UG BEHAVIOUR Y
NORM B -> BRAIN ACTIVATION *x* -> UG BEHAVIOUR Y

If different norms activate the same brain areas, and these are correlated with the same types of behaviour, then by observing the brain we would have an instrument for prediction more simple and effective. The importance of making good predictions is thus the reason to explore alternative models informed by neurobiological evidence. The hope is that these models will account for anomalies and make interesting new predictions.

However, the scenario just envisaged is ideal. Neuroscience is still an immature science, and the work to do before trying to integrate neurobiological parameters into a model of economic decision-making is a lot. This essay has tried to argue that this work would be important and worthwhile.[4]

BIBLIOGRAPHY

Bechtel W. (forthcoming) The epistemology of evidence in cognitive neuroscience. In R. Skipper Jr., C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson, and R. Richardson (eds.)*, Philosophy and the Life Sciences: A Reader. Cambridge,* MA: MIT Press.

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

[4] The bulk of this paper was written at the London School of Economics (UK) during my M.Sc. Alex Voorhoeve, Francesco Guala provided generous comments on previous drafts. Cristina Bicchieri helped me very kindly by answering many questions I had about her theory. At an earlier stage I discussed some of the ideas contained here with Matteo Motterlini. A sincere thank you to all of them for their help. The usual disclaimers about the remaining errors apply.

Bicchieri C. & Chavez, A. (2008) Behaving as Expected: Public information and fairness norms, unpublished. URL: < http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1082264>, July 2008.

Bicchieri C. & Xiao, E. (2008) Do the Right Thing: But Only if Others Do So, *Journal of behavioural decision making*, 21: 118.

Bolton G., & Ockenfels, A. (2000) ERC: A Theory of Equity, Reciprocity and Cooperation, *American Economic Review* 90: 16693.

Bradley R. (unpublished) 'Attributing Mental States'.

Camerer C. (2003) *Behavioral Game Theory: Experiments on Strategic Interaction,* Princeton University Press.

Camerer C. (2007) Neuroeconomics: Using Neuroscience to Make Economics Predictions. *Economic Journal*, 117: C2642.

Chiao J.Y. & Ambady N. (2007) Cultural neuroscience: Parsing universality and diversity across levels of analysis. In Kitayama, S. and Cohen, D. (Eds.) *Handbook of Cultural Psychology*, Guilford Press, NY.

Elman J. L., Bates E. A., Johnson, M. H., KarmiloffSmith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking Innateness*. MIT Press.

Fehr E. & Schmidt, K. (1999) A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics* 114: 81768.

Forster M. (2008) Prediction, in S. Psillos & M. Curd (eds.), *Routledge Companion to Philosophy of Science*,London: Routledge.

Gazzaniga M. Ivry R., Mangun G., (2002), *Cognitive Neuroscience*, W.W. Norton & Co. Inc., New York.

Guala F. (2005) *The Methodology of Experimental Economics,* New York: Cambridge University Press.

Guala F. (2008) Paradigmatic Experiments: The Ultimatum Game. *Philosophy of Science*, forthcoming.

Güth W., Schmittberger, R., & Schwarze, B. (1982) An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*(4), 367-388.

Harrison G. (2008) Neuroeconomics - A Critical Reconsideration. *Economics and Philosophy*, 24: 303 – 344.

Henrich J., Boyd R., Bowles, S., Camerer C., Fehr E., & Gintis H. (eds. 2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen SmallScale Societies.* Oxford: Oxford University Press.

Hoffman E., McCabe K., Shachat, K., & Smith, V. (1994) Preferences, property rights and anonymity in bargaining games. *Games and Economic Behavior*, 7:346-380.

Koch C. & Ullman S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry, *Human Neurobiology* 4:219-227.

Krolak-Salmon P. *et al.* (2003) An attention modulated response to disgust in human ventral anterior insula. *Ann. Neurol.* 53, 446–453.

Laudan L. (1990) De-Mystifying Underdetermination, in W. Savage, ed., *Scientific Theories*. Pp. 26797. Minneapolis: University of Minnesota Press.

Pablo B.-G., Ramón, C.-R., & Almudena, D. (2006) Si él lo necesita: Gypsy fairness in Vallecas. *Experimental Economics, V9*(3), 253-264.

Phillips M.L. *et al*. (1997) A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389, 495–498.

Quartz S.R. & Sejnowski, T.J. (1997) The neural basis of cognitive development: A constructivist manifesto. *Behavioural and Brain Sciences* 20: 537–596

Rabin M. (1993) Incorporating Fairness into Game Theory and Economics, *American Economic Review* 83: 1281-302.

Radder H. (1996) *In and about the World: Philosophical Studies of Science and Technology*, Albany: State University of New York Press.

Rescher N. (1997) *Predicting the Future*. Albany: State University of New York Press.

Roth A., Prasnikar V., Okuno-Fujiwara, M. & Zamir, S. (1991) Bargaining and Market Behavior in Jerusalem, Lubljana, Pittsburgh and Tokyo: An Experimental Study, *American Economic Review* 81: 1068-1095.

Rubinstein A. (2006).”Comments on Behavioral Economics”, in *Advances in Economic Theory* (2005 World Congress of the Econometric Society), Edited by R. Blundell, W.K. Newey and T. Persson, Cambridge University Press, 2006, vol II, 246-254

Rubinstein A. (2008) Comments on Neuroeconomics, *Economics and Philosophy*, 24: 485 – 494.

Sally D. & Hill E. (2006) The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness, *Journal of Economic Psychology* 27(1):73-97.

Sanfey A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003) The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755-1758.

Small D.M. *et al*. (2003) Dissociation of neural representation of intensity and affective valuation in human gestation. *Neuron* 39, 701–711.

Solnick S.J. & Schweitzer, M.E. (1999) The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes* 79(3): 199-215

Thaler R. H. (1988) Anomalies: The Ultimatum Game, *Journal of Economic Perspectives*, 2, pp. 195–206.

Worrall J. (2006) History and Theory-Confirmation in J. Worrall and C. Cheyne (eds) *Rationality and Reality: Conversations with Alan Musgrave*. Pp. 31-61 Kluwer Academic Publishers, 2006.

Zald D.H. & Pardo, J.V. (2000) Functional neuroimaging of the olfactory system in humans. *Int. J. Psychophysiol*. 36, 165–181.

A Neuroeconomic Perspective on Charitable Giving

David Yokum[*]
dyokum@email.arizona.edu

Filippo Rossi[^]
filippor@email.arizona.edu

ABSTRACT

Psychologists and economists, particularly those assuming that people are rational egoists, have struggled to understand the causes of voluntary donation for decades. Why would a person decide to sacrifice part of his or her material payoff in order to increase the wellbeing of others? In the first part of this paper, we outline a core set of possible motivations, and then consider how those motivations can be used to construct behavioral models that can also be tested in terms of what we know about brain function. We emphasize the role of other-regarding preferences and argue that there are *moral* judgments, independent of any consideration of payoffs, that partially determine when and to whom such preferences exist. In the second part of the paper, we argue that a neuroeconomic perspective can help understand charitable giving, and then discuss recent neuroimaging studies that demonstrate this potential.

---

1. THE CHALLENGE OF VOLUNTARY DONATION

Homo economicus, that rational creature concerned only with its personal, material payoffs, is not an ideal candidate from whom to elicit a charitable donation. Unless there are offsetting material benefits, it will refuse to contribute anything to the public good. In this paper, we outline evidence demonstrating that people are more charitable than predicted by rational egoism alone, and then explore alternative models that use other-regarding motivations to explain the difference. Attention is given to unpacking the features determining when and to whom such motivations exist and, in particular, how neuroscientific evidence can inform the debate.

Two caveats before we get started. First, for the moment, we are restricting egoism to mean a self-centered focus on achieving *material* gains for oneself, without any regard for the interests of other people. A more inclusive definition of egoism would allow any self-centered focus, material or otherwise, to count. For example, it might be that the decision maker does not expect a material payback, but instead wants to feel the *pleasure* of giving money to the needy. This desire is still selfish in that it seeks to obtain personal benefit – the pleasurable sensation – rather than advance the welfare of the beneficiary. We will return to this possibility in detail later.

Second, egoism and altruism are defined differently by different researchers. For instance, biologists often discuss selfishness and altruism in terms of evolutionary fitness without any reference to intentionality. A different question, the one which we will focus on, is whether the proximate, psychological system motivating the decision to act altruistically is selfish, altruistic, or both. It is a question of what beliefs and desires are driving behavior. To put it

[*] Neural Decision Science Lab Department of Psychology The University of Arizona
[^] Neural Decision Science Lab Department of Psychology The University of Arizona

loosely, do humans have a cognitively represented, ultimate desire of the form "I want others to be well-off" – a desire that exists independently of their concern for their own wellbeing? Or is the only ultimate desire instead "*I* want to be well-off," perhaps coupled with auxiliary beliefs that acting charitably will somehow advance *that* goal? To say that humans act charitably because they are psychologically motivated by altruism is to say that they possess an ultimate desire to advance others' wellbeing, irrespective of their own wellbeing. They have other-regarding preferences. Non-altruistic explanations of charity, in contrast, either deny altruistic impulses altogether or explain them away as instrumental desires: they exist only to advance ultimate desires that are selfish.

Returning to the original discussion, whatever its descriptive limitations, the assumption of rational egoism has stimulated the creation of formal models, such as in game theory,[1] [1]that have precise, testable predictions about how persons will behave toward one another. This provides a useful benchmark from which to compare real behavior and, as we will see, helps orient research aimed at understanding underlying neural mechanisms. *Public good games* are immediately relevant for our purposes. Within a typical public good design, each individual is given a monetary endowment and must decide whether or not to contribute some portion of that endowment to a common pool from which all participants will benefit; the transferred amount is often increased by a stated factor (Leydard, 1995; Gächter et al. 2009). For example, each subject receives ten euro, and any money donated to the common pool is multiplied by two before being evenly redistributed back to the participants. If there are four subjects, and everyone contributes his or her entire endowment, then the final common pool is 4[10 x 2] = 80 euro, and each individual receives back 20 euro – a two-fold increase from the initial endowment. This is obviously a better outcome for all parties than if no one contributes, for in that case each individual remains with only ten euro. Nevertheless, *Homo economicus* will contribute nothing. The reason is that the highest payoff is achieved if one defects while everyone else contributes. In that case the payoff is 10 + (3[10 x 2])/4 = 25 euro – a five euro increase from when personally contributing as well. The other egotistical players, of course, will also realize the benefits of unilateral deviation. The end result is that no one contributes to the common pool (Dawes & Thaler, 1988). Such an undesirable outcome is a Pareto-deficient equilibrium, since everyone would be better off if no one defected, but no one actually does so because of the even higher potential benefits of unilateral deviation.

Pareto-deficiency is a classic obstacle in many one-shot games, and it is also the basis of the free-rider problem. In the context of charitable giving, it means that each person is predicted to refuse donating time or money, relying instead on the contributions of other people – thereby getting a "free-ride" to a public good made possible by *others'* donations. However, real behavior in laboratory and naturalistic settings negates this prediction (Leydard, 1995; Camerer, 2003; Andreoni, 1995). Rather than transferring zero euro in the one-shot public good game, for example, subjects typically donate 40-60% of their endowment (Marwell & Ames, 1981). Comparable rates are also contributed in the first round of a finitely repeated version of the game (Isaac & Walker, 1988; Kim & Walker 1984).[2]

Charitable behavior is not restricted to economic games either. A recent survey in the United States revealed that 89% of households donated an average of $1,620, or 3.1% of household income, to charitable organizations, and almost half of all American adults volunteered their time to participate in such groups (Independent Sector, 2001). Such

[1] See von Neumann and Morgenstern (1944) and Osborne and Rubinstein(1998).

[2] For a more general discussion on the public good game, see Ledyard (1995) and Camerer (2003).

estimates fluctuate across surveys and countries (Andreoni, 2006; 2008), but in all cases charitable giving is well above the predicted level of zero contribution.

Initially it was thought that the egoism assumption could be preserved at the expense of rationality. For one reason or another, people make mistakes and donate to the public good against their own best interest. Such a hypothesis is partially consistent with data from repeated public good games. Although subjects in an iterated public good game make first-round donations that are comparable to those in a one-shot interaction, the donation rate begins to decay with each subsequent round until, in the final round, most subjects are actually behaving as egotistical players (Dawes & Thaler, 1988; Fehr & Schmidt, 1999; Kim & Walker, 1984). It was therefore hypothesized that subjects were initially confused by a rather odd experimental setting, and that behavior during later rounds, once the rules are learned, is more informative of what would actually happen in naturalistic settings.

Andreoni (1995) directly addressed the possibility of confusion. He used three different public good games: (a) a standard version; (b) a version where players were paid according to the relative rank of their earnings (it was, in other words, transformed into a zero-sum game); and (c) a modified standard version with rank information (which was inconsequential with respect to payment, and meant only to control for the effects of information about rank on behavior). Note that in condition (b) there is no incentive to cooperate, due to its zero-sum nature. Donation in this case can therefore be attributed to confusion, for it benefits no one. The results replicated the common finding for the standard condition (a): cooperation decayed with each additional round, although donation rates never dropped below 25%. Subjects in condition (b), however, quickly and drastically altered their behavior, donating less than 10% by the fourth round. Analyses across conditions revealed that confusion was indeed *one* explanatory factor for early round behavior, but that it was unable to fully explain the persistent willingness to donate. Moreover, the observed late round defection was best explained, not by learning, but rather as a reaction to free-riding: players initially try to cooperate, but after their attempts are exploited by others, they turn to defection as a form of retaliation.[3]

The so-called "re-starting effect" (Andreoni, 1988; Cookson, 2000) provides further evidence against the confusion hypothesis. If subjects are learning that donation is against their own interests with each subsequent round of a public good game, then it should not matter if the game is interrupted and then restarted. If players are learning that the best strategy, for a rational egoist, is not to donate then we would predict that cooperation should constantly decline across rounds of the repeated game. If, on the other hand, subjects are sensitive to the decisions of their partners (rather than merely ignorant of the game rules), then a new start may re-open possibilities for cooperation. The experimental data confirms that donation rates, contrary to the confusion hypothesis, do in fact bump back up to high rates when the game is stopped and restarted. The upshot is that confusion *per se* is unable to explain the observed rates of contribution.

[3] See also Goeree et al. (2002) for an investigation of this issue within the framework of stochastic game theory. They used a series of (quasi) one-shot public good games that varied: (a) the number of players in a group; (b) the magnitude of how much a player received in return for contributing one unit of his or her endowment to the public good; and (c) how much each *other* player in the group received from the unit contributed by the player. Their data revealed that both altruism and behavioral noise are explanatory factors of behavior in a public good setting. See also Palfrey et al. (1997).

Substantial evidence therefore indicates that *Homo economicus* is an endangered species, and that for most people the assumption of raw egoism focused on material payoffs simply does not hold. So why, then, do people voluntarily donate their time and money to a public good even when there are not obvious offsetting material benefits? The terrain of candidate explanations is complex, spanning disciplines such as economics, psychology, and evolutionary biology, and there are a variety of terms used by different authors that overlap in some ways but not others. In the next section, we attempt to outline several major conceptual possibilities, but make no claim to being exhaustive. Our aim is to capture a set of essential motivations, even if necessarily oversimplified, in order to begin constructing behavioral models that can also be tested in terms of what we know about brain function.

2. COMPONENTS OF CHARITABLE DONATION

It is worth noting that the demise of *Homo economicus* in no way obliterates the role of selfishness. Most people act charitably under the right circumstances, but not everyone does. And for those with other-regarding preferences, there is no reason to think that some degree of selfishness no longer exists. The motivations outlined here should therefore be understood as expanding rather than replacing selfish models. People act charitably for selfish reasons at times, but non-selfish reasons are also necessary to explain the level of generosity that exists, which is well above that predicted by selfish models, as we saw in the previous section. Bearing this in mind, we now turn to four components of charitable decisions that are different from seeking personal material payoffs and warrant special scrutiny: warm-glow, preferences related to fairness, reciprocity, and deservedness.

2.1 *WARM-GLOW*

James Andreoni (1989; 1990) provides a useful distinction, often cited in the charitable giving literature, between pure and impure altruism. The dichotomy captures the possibility that a person might donate either because an increase in a public good is desirable *per se* or, alternatively, because he or she experiences a sort of selfish, personal satisfaction from the very act of giving. To consider this formally, let w_i represent person *i*'s wealth, and assume that there is one private and one public good. Wealth can be spent either on personal consumption of the private good, denoted x_i, or given as a charitable donation to the public good, denoted g_i. Let G represent the overall public good and be equal to the sum of all donations, that is, person *i*'s donation (g_i) plus everyone else's donation. People can then be assumed to have a utility function $U_i = U_i(x_i,G,g_i)$, which they seek to maximize. What this means is that people derive satisfaction from their private consumption (x_i), the level of public good (G), and their personal contribution to that public good (g_i), and thus their decision of how much to donate, if at all, will reflect a balance between these three considerations.

A person is said to be purely altruistic if he or she cares about the status of the public good but not how it is achieved, except insofar as it affects their level of private consumption. The utility function in this case reduces to $U_i = U(x_i,G)$. A person is a pure egoist, on the other hand, if he or she cares only about private consumption and personal contribution. In this case, $U_i = U(x_i,g_i)$. The g_i term, as part of the utility function, captures a phenomenon referred to as "warm-glow." It reflects a type of satisfaction that is independent of that derived from achieving a given level of G. It is instead a satisfaction derived immediately from the giving act itself – a good feeling from personally helping out, so to speak. Someone who is neither purely

altruistic nor purely egotistical, that is, a person who is sensitive to both *G* and g_i, is considered to be an impure altruist.

What we want to stress here is the "warm-glow" component of the model. The variable for private consumption, x_i, represents the continued presence of selfish motivations in charitable decisions, as we stressed at the onset of section two. And the sections 2.2 and 2.3 below could be interpreted as further elaborating what is represented by *G*, in other words, what it means to have truly other-regarding preferences. Warm-glow is unique, however, in that it does not cleanly fit into either of the traditional categories of selfishness and other-regarding preference. There is no material payoff at stake, and it seems odd to call someone who enjoys helping others selfish. Nonetheless warm-glow is, in an important sense, entirely selfish. The act of giving is merely a means to the end of personal satisfaction, and at bottom the well-being of others is not a concern. To see that this is the case, consider the following decision: should you personally donate one euro to a charitable group or allow the charitable group to receive instead one-thousand euro from someone else. The person motivated only by warm-glow will select the former, despite the substantial monetary loss to the charitable organization.

2.2 SOCIAL PREFERENCES

Theories of fairness assume that subjects have preferences regarding the way in which certain resources are allocated. An allocation that satisfies these preferences is called fair. We assume that an unfair allocation can motivate a decision-maker to act in a compensatory manner, that is, to seek the obtainment of a fair allocation. In this sense, fairness enters as a motivational force in a person's preferences (Camerer, 2003, p. 114). A theory of charitable giving should therefore accommodate the relationship between voluntary donation and the theory of justice of the donators.

In the economical literature, four main types of social preferences are usually discussed (see Charness & Rabin, 2001, for a review): self interested, competitive, social-welfare, and differences-aversion motives. Selfishness was considered above. A competitive motive is not particularly interesting from the point of view of voluntary donation: if a subject's preferences are such that she prefers to be better off than her opponents, it is unlikely for her to donate. Let us then focus on the last two preference types.

Consider the linear model below, adapted from Charness and Rabin (2001), which can be used as a simplified representation of different kinds of distributional preferences:

$$U_i(\pi_i, \pi_j) = \pi_j(\rho \cdot r + \sigma \cdot s) + \pi_i(1 - \rho \cdot r - \sigma \cdot s) \qquad (1)$$

In a two-player interaction, equation (1) entails that the utilities derived from material payoffs for players *i* and *j*, denoted π_i and π_j, respectively, depends on: (a) the magnitude of the two payoffs and (b) the parameters ρ and σ, which provide a way of modeling different kinds of distributional preferences. The strategic interaction is described through *r* and *s*. If *i*'s material payoff is bigger than *j*'s (i.e., $\pi_i > \pi_j$), then $r = 1$; otherwise it is zero. If instead $\pi_i < \pi_j$, then $s = 1$; otherwise it is zero.

Andreoni and Miller (2002) and Charness and Rabin (2001) proposed a quasi-maximin model of social-welfare preferences, where subjects are concerned about their material payoff, the material payoff of the individual with the lowest payoff in the society, and finally the public good (in terms of the model in (1), this translates into the following constraints: $1 \geq$

$\rho \geq \sigma > 0$, $\sigma \leq 1/2$). The upshot of this model is that subjects prefer more for themselves, are more likely to help when they are better off, and prefer Pareto-improvements.[4] Charness and Rabin (2001) found that the quasi-maximin model was capable of organizing the data from a battery of thirty-two dictator and response games, and thus that it may capture important features of people's distributional preferences.

Another important theory has been proposed by Fehr and Schmidt (1999), one that is particularly pertinent to the analysis of public good games.[5] They present a two parameter model in which subjects are concerned about the differences between their payoff and the payoffs of the other players; it is essentially a type of "self-centered inequity aversion." The intuition behind Fehr and Schmidt's proposal is that when subjects have more of a given resource, they are more willing to sacrifice that resource in order to compensate for payoff differences. On the other hand, when subjects are behind, they are willing to hurt their opponents in order to shrink the distance between the material payoffs. This can be written formally as:

$$U_i(\pi) = \pi_i - \alpha_i \max\{\pi_j - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_j, 0\} \qquad (2)$$

The parameter β_i captures the degree to which subjects care about the inequality of the payoff when they are ahead, while α_i represents their concern when they are behind. A reasonable constraint is that α_i should be greater than or equal to β_i.[6] Fehr and Schmidt also assume that $0 \leq \beta_i < 1$, which distinguishes their theory from a competitive preferences model, where subjects prefer to be better off than other players. This model is also capable of explaining a wide variety of experimental data.

The above models are highly simplified, and there is no agreement on which is best for representing social preferences. Charness and Rabin therefore provide some sound advice:

> Too much will be lost if experimentalists jump too quickly to calibrating highly simplified model […] At this stage, models ought to be developed that help to interpret psychologically sound and empirically prevalent patterns of behavior common in a broad array of games. (2001, p. 821)

What is most important for our purposes is the explicit representation of some sort of fairness preference, namely, a concern about how resources are distributed that is sensitive to the payoffs of others. Whether this motivation reflects a desire to enhance the well-being of the worst off, to simply avoid inequity, both, or some other sort of other-regarding preference is, at this stage, to early to determine. However, as experimentalists tackle this issue, we think it necessary to keep in mind that people are unlikely to have social preferences that are uniformly applied across people and situations. Such preferences are not blind, so to speak, and as such models of charitable giving should be sensitive to qualifications about when and to whom charitable motivations apply. We turn now to this possibility.

[4] See also Yaari and Bar-Hillel (1984).

[5] See also Fehr and Fischbacher (2002) and Bolton and Ockenfelds (2000).

[6] It would be surprising, after all, if people cared about inequality when they were ahead *more* than when they were behind, for in the latter case selfish motivations would also be predicted as a motivating factor to achieve equality. Also, as Fehr and Schmidt note, this assumption is in line with the loss aversion literature (e.g., Tversky and Kahneman, 1991), which reflects the common saying that "losses loom larger than gains."

2.3 RECIPROCITY AND DESERVEDNESS

The theories outlined above do not specify anything about the qualities of the recipients of a charitable act, but instead focus exclusively on their material payoffs. Nonetheless, issues such as whether the potential recipients *deserve* your help or satisfy some moral criteria are important candidates for inclusion in the consideration of whether or not to act charitably (see Rabin, 1993; Dufwenberg et al., 2004; Charness & Rabin, 2001). As Dufwenberg et al. (2004) note:

> The assumption that individuals only care about final distributions implies that they must be indifferent concerning *how* distributions come about. This is problematic if in fact individuals regard information about their co-players' specific choices or intentions as important to their decision-making. (2004, p. 260-70)

What determines whether someone deserves help? In the economic literature, deservedness is strictly related to fairness in the form of reciprocity (Rabin, 1983; Dufwenberg et al., 2004). Charness and Rabin (2001), for example, focus their attention on *withdrawal reciprocity*. The idea is that subjects will withdraw their willingness to act charitably toward people who are unwilling to sacrifice for the sake of fairness. They observed that this kind of reciprocation is particularly important in the case of simple dictator and response games.

In the game theoretic literature, Rabin (1993, p. 298) developed the notion of "fairness equilibrium" on the basis of three intuitions: (a) people are willing to sacrifice their own material well-being to reciprocate kindness; (b) they are also willing to sacrifice well-being to punish unkindness; and (c) the first two conditions significantly affect human behavior. Dufwenberg and Kirchsteiger (2004) developed this notion for extensive form games,[7] and their formulation provides an example of the kind of formalism that may be implemented in order to qualify when and to whom other-regarding preferences apply.

Call $A_{i \in N}$ the set of player *i*'s strategies, with *N* the set of players. Denote with *A* the set of strategy profiles (the Cartesian product of the sets A_i). Dufwenberg and Kirchsteiger define two components of the utility function for player *i*: his material payoff π_i and the *reciprocity payoff*. This latter component depends on the players' beliefs at different stages of the game (Battigalli & Dufwenberg, 2009). In particular, there are: (a) the set of possible beliefs of player *i* on *j*'s strategies, $B_{ij} = A_j$ and (b) player *i*'s beliefs about player *j*'s beliefs about player *k*'s strategy, $C_{ijk} = B_{jk} = A_k$. Players update these beliefs throughout the game on the basis of the previous actions of their opponents. For example, $b_{ij}(h)$ represents an updating of B_{ij} based on *h*, a possible history in the game.

From here we can define a kindness index, which can be used to discriminate the players with whom to be kind. Following (in part) Rabin's original formulation (1993), Dufwenberg and Kirchsteiger define a *references* payoff (π^e_j) as the average between the highest and the lowest payoffs that player *i* can give to player *j* by playing one of his efficient strategies (Dufwenberg et al., 2004, p. 276). The distance and sign of the material payoff of *j* from the reference payoff defines the kindness of *i* toward *j*:

[7] This theory belongs to the more general framework of psychological game theory; see Genakopolos et al. (1989) and Battigalli and Dufwenberg (2009).

$$k_{ij}(a_i(h),(b_{ij}(h))_{j\neq i}) = \pi_j(a_i(h),(b_{ij}(h))_{j\neq i}) - \pi_j^e((b_{ij}(h))_{j\neq i}) \quad (3)$$

A further term (analogous to (3)) can be constructed that measures whether player *i* thinks that *j* was kind with *i* (λ_{iji}). Through the material payoff of *i* and the kindness functions, we can finally define the following utility function:

$$U_i\left(a_i(h),\left(b_{ij}(h),\left(c_{ijk}(h)\right)_{k\neq j}\right)_{j\neq i}\right) = \pi_i\left(a_i(h),\left(b_{ij}(h)\right)_{j\neq i}\right) + \sum_{j\in N\setminus\{i\}}\left(Y_{ij}\cdot k_{ij}\left(a_i(h),\left(b_{ij}(h)\right)_{j\neq i}\right)\lambda_{iji}\left(b_{ij}(h),\left(c_{ijk}(h)\right)_{k\neq j}\right)\right) \quad (4)$$

The first term stands for *i*'s material payoff. The second term is of particular interest here. Notice that k_{ij} is negative when *i* behaved unkindly with *j*, while λ_{iji} is negative when *j* was unkind with *i*. When only one of the terms is negative, the value of U_i decreases; this captures the fundamental intuition underlying reciprocity: a mismatch between kind and unkind behavior triggers a violation of reciprocity. Finally, Y_{ij} is an exogenous term, which models how much player *i* cares about reciprocity (when it equals zero, the utility function reduces to the material payoff component).[8]

The relation between this approach to reciprocity and voluntary giving is indirect, but particularly relevant. In particular, Dufwenberg and Kirchsteiger's theory is one of the most developed ways of formally imposing constraints on other-regarding preferences. This psychological game theoretic approach can, in principle, accommodate the imposition of further constraints based on the *properties* that a recipient should have from the point of view of the donator in order to be eligible for a charitable act. This would constitute a novel and important addition. What we have in mind, specifically, is that moral considerations are relevant, considerations that are related to (but subtly different from) desert, or the idea that a person deserves something based upon his or her actions.

As an example of data motivating this hypothesis, consider the debate surrounding the system of welfare in America. This is an ideal test case, for the welfare system constitutes an egalitarian redistribution of income among total strangers, and to the extent that it is supported through votes it is a voluntary system – it is, in other words, an act of charity. Nonetheless, popular support for welfare is mixed, with many individuals adamantly opposed to its continuation. Why might people oppose welfare or, to put it differently, refuse to be charitable in this case? Surprisingly, income level, education, and a variety of demographic variables are incapable of adequately discriminating between supporters and detractors; voters are not really concerned about the cost of welfare or fraud; and even agreeing with the idea that income should be evenly redistributed is not a very good predictor. The overwhelmingly significant predictor is instead the answer to the following question: does bad luck cause poverty? Those who believe that poor people are somehow responsible for their poverty are far more likely to oppose welfare (see Fong et al., 2005, for a full discussion). This form of responsibility, or rather irresponsibility, can take several forms. Fong et al. (2005), for example, report survey results indicating that, by more than a five-to-one margin, respondents

[8] The model as presented here is only a partial representation. For the full structure, see Dufwenberg and Kirchsteiger's original paper (2004).

believe that welfare recipients could obtain employment if they tried, 70% of respondents believe it is more financially rewarding to stay on welfare than get a job, 57% believe welfare encourages laziness, and 60% believe welfare encourage out-of-wedlock childbirth.

The welfare survey data reveal a refusal to financially support people who, it is believed, are engaging in socially unacceptable behaviors, such as refusing to contribute to the common good via employment or having children outside of marriage. To some extent the opposition to welfare can be interpreted in terms of reciprocity. What is meant by reciprocity requires specification. Most authors use reciprocity to mean reciprocal altruism in particular, or a sort of tit-for-tat in which the benefactor expects the recipient to somehow repay the favor, even if unknowingly. This is not the most parsimonious explanation, however. After all, welfare supports anonymous strangers that any given individual is unlikely to ever meet, and even many very wealthy persons – who presumably stand little to gain – support welfare. More pertinent is what is referred to as *strong reciprocity* (Gintis, 2000; Gintis et al., 2005). The theory is that people tend to behave pro-socially and punish antisocial behavior, at a cost to themselves, even when the probability of future interactions is extremely low, or even zero. This is fundamentally different from the *weak reciprocity* associated with reciprocal altruism, for it remains a motivating factor even when future compensation is unlikely or impossible.

Strong reciprocity is different from pure altruism, however, for it is still conditional. It is just that the condition is not of equivalent worth as the beneficent act. What this means is that the possible recipient must satisfy certain criteria, for instance (to continue with the welfare example) be willing to work. But it is *not* the same as saying the potential recipient has a legitimate claim to compensation in the standard sense of reciprocity. In other words, the recipient has not actually provided an offsetting benefit to the benefactor, nor does the benefactor *ever expect* the recipient to return the favor. Rather, the benefactor has a true other-regarding preference, one in which he or she is genuinely concerned with the payoff of another individual that is independent of his or her own payoff – but that concern, importantly, only exists for those who satisfy the relevant moral criteria.

Selfishness, warm-glow, preferences related to fairness, reciprocity, and deservedness are therefore all crucial components in the decision to act charitably. Other factors might be important as well, but these motivations, incorporated into economic models such as those outlined above, provide a generative starting point for the neuropsychological investigation of charitable giving. Indeed, it is difficult to even begin research into the neurobiology of charitable giving without the aid of precisely formulated models. Having outlined such models, we now turn to the very new and very small neuroscientific literature on charitable giving (see Mayr et al., 2009, for a review).

3. NEURAL EVIDENCE

Neuroeconomics is an experimental approach that couples the mathematical precision and simplifying assumptions of economic models with the methods and data from cognitive neuroscience (Glimcher et al., 2009). Neural data is useful for the study of charitable giving, as well as other economic behaviors, for several reasons. To begin with, neural evidence places a constraint on the range of viable psychological theories, since proposed cognitive mechanisms must align with possible neurophysiological function. In essence this works by enhancing the epistemic criteria of internal consistency: not only must the proposed cognitive mechanism align with behavioral data, but now it must also mesh with theories about different brain regions and the functions they perform. Second, brain activity during charitable giving is likely to overlap with brain activity observed in other contexts and for which there is already a good deal of theoretical understanding. This means that our understanding of brain function in

general can be used to help interpret the neural signals observed during charitable behavior and, potentially, stimulate hypotheses for testing at cognitive and behavioral levels. Finally, neural evidence might reveal motivations for charitable giving that are not accessible behaviorally, either because subjects are unwilling to report or unconscious of the actual impetus for their behavior.

This final point is especially relevant for uncovering the psychological motivations driving charitable behavior. As mentioned previously, egoism can be broadly construed to entail non-material selfish benefits, such as warm-glow. This is a legitimate interpretation, but the consequence is that egoism becomes seemingly impossible to refute on the basis of behavioral data alone. One problem is that self-reported reasons claiming genuine altruism can almost always be reinterpreted as an instrumental desire toward a selfish, ultimate desire, one that subjects may be unable to access. And it is difficult to bypass this obstacle with behavioral evidence other than self-report because a decision maker, whether motivated by warm-glow or a true other-regarding preference, will likely *act* the same. Helping other people for their own sake and doing it only to generate warm-glow feelings both require *actually* helping. This is *not* to say that behavioral data are irrelevant, only that it is difficult to disentangle motivational theories that make similar behavioral predictions.

But what if we could more directly observe the motivational systems at work? Sober and Wilson (1998, p. 205-208) discuss the example of marine bacterium that are obligate anaerobes (they cannot survive in the presence of oxygen) as an analogy for the problem of teasing apart psychological egoism from psychological altruism. They note that the bacterium could avoid oxygen by utilizing a device that directly detects oxygen or, alternatively, it could use a magnetosome sensitive to the gravitational field of the earth (oxygen is more abundant near the water surface, so gravity could be used to as a guide in swimming away from the surface). But bacteria with either an oxygen detector or a magnetosome would behave the same – they would swim downward. However, assuming we knew enough about bacterial anatomy, we could resolve this dilemma by dissecting the bacterium and observing what devices are inside. Does it have an oxygen detector or does it have a magnetosome? Sober and Wilson note that this is possible, in principle, for the question of how human altruism operates in the brain but, writing a decade ago, rightly concluded that "even if neurobiology will answer this question one day, it offers little guidance now" (p. 207). This, fortunately, is beginning to change.

The neural investigation of charitable giving specifically is only just beginning. Jorge Moll and colleagues (2006) conducted one of the first functional magnetic resonance imaging (fMRI) studies. Nineteen subjects were endowed with U.S. $128. Each person was then presented with a series of real charitable organizations, including a brief mission statement for each, and given the opportunity to either donate or oppose a donation to each organization from his or her endowment. Decisions were strictly anonymous, and subjects were aware that real money was at stake, both for themselves and the stated charitable organization. The design entailed several possible payoff conditions for each decision: pure monetary reward (YOU: $+2, ORG: $0), non-costly donation/opposition (YOU: $0, ORG: $5), costly donation/opposition (accepting YOU: $-2, ORG: $5; refusing YOU: $2, ORG $5). Note that the pure monetary reward condition has no consequences for the charitable organization, and therefore the decision should reflect solely egotistical preferences. The non-costly condition, on the other hand, has no personal consequences, and therefore the decision should reflect solely other-regarding preferences for the charity. The costly conditions entail a conflict between egotistical and other-regarding motivations.

There were two main results. First, the ventral tegmental area (VTA) and striatum, components of the brain's reward system, were activated by both pure monetary rewards and decisions to donate. This finding suggests that charitable giving and personal gains share the anatomical systems underlying reward reinforcement and expectancy. Second, a direct contrast between the pure monetary reward and donation conditions revealed that neural activity was greater in the ventral striatum for donations and, most intriguingly, that the subgenual area was uniquely activated during donations. To the extent that the donation condition elicits the experience of reward above and beyond the monetary gain condition, the interpretation of warm-glow motivation is reasonable. In other words, the enhanced ventral striatum activity might represent the experience of utility from the act of giving *per se*.

But the presence of warm-glow does not rule out the existence of other motivations as well. The subgenual area has been repeatedly implicated in the expression of social attachment (e.g., Bartels & Zeki, 2004; Aron et al., 2005) and the release of neuromodulators, such as oxytocin and vasopressin, which are likewise thought to have important social functions (e.g., Zak et al., 2005). It is therefore tempting to speculate that subgenual activity reflects neural evidence of some form of conditional altruism or strong reciprocity. At the very least, the subgenual area is a neuroanatomical region outside of the putative reward system, and thus charitable decisions entail a type of processing that is perhaps meaningfully different from the utility processing of personal payoffs. Such a conclusion really is speculative though. Aside from awaiting replication, the interpretation of subgenual activity is somewhat complicated by the fact that the charitable organizations used in the study were deliberately chosen on the basis of their support or opposition to socially controversial issues, for instance euthanasia, gun control, and abortion. The observed activity might therefore reflect a reaction to the social issue that is independent of the decision to donate. This seems unlikely, however, and as we argued above social values and moral assessments are important considerations in the decision of whether or not to donate to a particular public good. To put it differently, the subgenual activity might very well reflect a reaction to stimulating socially values, but that reaction exists as a relevant factor in the decision of whether or not to behave charitably.

Harbaugh et al. (2007) conducted a second neuroimaging study, this time without the additional variable of moral assessment and with a design that more clearly tested the possibility of other-regarding preferences. Nineteen subjects were endowed with US $100 and given a series of payoff situations which they could either accept or reject. The charity was a food bank. There were also, however, three unique conditions that did not require a decision: a tax condition, in which subject money was involuntarily transferred to the food bank; a pure mandatory payment to the subject, at no cost to the charity; and a pure mandatory payment to the charity, at no cost to the subject. Activation of the reward system was again found for both personal gains and gains to the charity. There was also greater striatum activity during voluntary rather than mandatory donations, again supporting the existence of warm-glow. However, in a contrast between the mandatory treatments, activation in the ventral striatum was found for both personal monetary gains and the mandatory tax. This latter activation cannot possibly be interpreted as a warm-glow effect. The subject, after all, did not volunteer to donate the money.

The most impressive finding, however, was the predictive model that Harbaugh et al. tested. Recall the impure altruism utility function, $U_i = U_i(x_i,G,g_i)$. They used neural activity during the pure mandatory payments to each subject as an indicator of his or her marginal utility of money (x_i), and activity during pure mandatory payments to the charity as an indicator of the marginal utility derived from increases in the public good (G); the transfers were mandatory, so g_i (personal donation) was irrelevant. Those with a larger neural response to the mandatory personal payoff were labeled "egoists," and those with a larger response to

the mandatory payoff to the charity "altruists." As predicted, the labeled altruists were more charitable, giving nearly twice as much. The authors reasonably interpreted this as evidence of a purely altruistic motive. The larger the neural response to increases in *G*, no matter the source (that is, regardless of g_i), the more likely one is to give voluntarily.

4. CONCLUSIONS

We have argued in this paper that selfish motivations are only one component of decisions related to charitable behavior. People also value warm-glow feelings and have genuine other-regarding preferences, such as concerns about how resources are distributed and strong reciprocity. Such other-regarding preferences are not blind, however, and are instead constrained by a variety of criteria related to the deservedness of the potential recipient. These constraints are not related to material compensation but are rather, in an important sense, moral considerations.

The Moll et al. (2006) and Harbaugh et al. (2007) studies are merely a taste of what is to come from the neuroeconomic approach. These two studies demonstrated that it is possible to identify neuroanatomical regions uniquely activated during conditions of selfishness, pure-altruism, and warm-glow. The data at present do not conclusively resolve the psychological egoism versus altruism debate by any means, but this work is promising. Dissecting the brain, as we might to see if a bacterium has an oxygen detector or magnetosome, is not as far-fetched as it once seemed. For example, the evidence that reward areas of the brain are activated by merely witnessing a charitable act is compelling evidence that humans do, in fact, donate for other-regarding reasons, contrary to the egoism hypothesis. To appreciate the unique force of these data, consider someone who claims to enjoy the fact that another group is benefitted by charity as an end in itself. An advocate of egoism could dismiss this claim as disingenuous; perhaps the person merely *says* such things to earn the approval of the experimenter, or is somehow deluding him or herself. But to also dismiss the neuroimaging data, the critic would have to argue that the subject is somehow capable of self-generating striatal activity despite not actually experiencing a reward, or perhaps postulate a source of striatal activation other than reward. There might be such explanations, but this is a harder case to make, and now the burden of proof would lie with the proponent of egoism.

BIBLIOGRAPHY

Andreoni J. (2008) Charitable giving. In S. N. Durlauf, & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd Edition ed.) Palgrave Macmillan.

Andreoni J. (1995) Cooperation in Public-Goods Experiments: Kindness or Confusion? *The American Economics Review , 83* (4), 891-904.

Andreoni J. (1989) Giving with impure altruism: applications to charity and Ricardian equivalence. *Journal of Political Economy , 97*, 1447-1458.

Andreoni J. (1990) Impure altruism and donations to public good: a theory of warm-glow giving. *Economic Journal , 100*, 464-477.

Andreoni J. (2006) Philanthropy. In S.-C. Kolm, & J. M. Ythier (Eds.), *Handbook of Giving, Reciprocity, and Altruism* (pp. 1201-1269) Amsterdam: North Holland.

Andreoni J. (1988) Why Free Ride - Strategies and Learning in Public-Goods Experiments. *Journal of Public Economics , 37*, 291-304.

Andreoni J., & Miller, J. (2002) Giving According to GARP: An Experimental Test of Consistency of Preferences for Altruism. *Econometrica , LXX*, 737-53.

Aron A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005) Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology , 94*, 327-337.

Bartels A., & Zeki S. (2004) The neural correlates of maternal and romantic love. *NeuroImage , 21*, 1155-1166.

Battigalli P. & Dufwenberg M. (2009) Dynamic Psychological Games. *Journal of Economic Theory , 144*, 1-35.

Bolton G. & Ockenfels A. (2000) ERC: A Theory of Equity, Reciprocity and Competition. *American Economic Review , 90*, 166-93.

Camerer C. *Behavioral Game Theory. Experiments in Strategic Interaction.* Princeton, New Jersey: Princeton University Press.

Charness G. & Rabin M. (2001) Understending Social Preferences with Simple Tests. *The Quarterly Journal of Economics , 117* (3), 817-69.

Cookson R. (2000) Framing Effects in Public Goods Experiments. *Experimental Economics , 3*, 55-79.

Dawes R. & Thaler, R. (1988) Anomalies: Cooperation. *The Journal of Economics Perspectives , 2* (3), 187-97.

Dufwenberg M. & Kirchsteiger G. (2004) A theory of sequential reciprocity. *Games and Economic Behavior , 47*, 268-298.

Fehr E. & Fischbacher U. (2002) Why Social Preferences Matter: The Impact of Non-Selfish Motives in Competition. *The Economic Journal , 112* (478), C1-C33.

Fehr E. & Gächter S. (2000) Cooperation and Punishment in Public Good Experiments. *The American Economic Review , 90* (4), 980-94.

Fehr E., & Schmidt K. M. (1999) Theories of fairness, competition, and cooperation. *Quarterly Journal of Economics , 114*, 817-868.

Fong C. M., Bowles S. & Gintis H. (2005) Reciprocity and the Welfare State. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral Sentiments and Material Interests* (pp. 277-302) Cambridge, MA: MIT Press.

Gächter S. & Herrmann B. (2009) Reciprocity, Culture and Human Cooperation: Previous Insights and New Cross-Cultural Experiment. *Philosophical Transactions of the Royal Society B* (364), 791-806.

Genakopolos J., Pearce D., & Stacchetti E. (1989) Psychological Games and Sequential Rationality. *Games and Economic Behavior , 1*, 60-79.

Gintis H. (2000) Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology , 206*, 169-79.

Gintis H., Bowles S., Boyd R., & Fehr E. (Eds.) (2005) *Moral Sentiments and Material Interests.* Cambridge, MA: MIT Press.

Glimcher P. W., Camerer C. F., Fehr E. & Poldrack R. A. (Eds.) (2009) *Neuroeconomics: Decision Making and the Brain.* San Diego: Elsevier Inc.

Goeree J., Holt C. & Laury S. (2002) Private Costs and Public Benefits: Unraveling the Effects of altruism and Noisy Behavior. *Journal of Public Economics , 83*, 255-276.

Harbaugh W. T., Mayr U., & Burghart D. R. (2007) Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science , 316*, 1622-1625.

Herrmann B. & Gätcher S. (2008) Antisocial Punishment Across Societies. *Science , 319*, 1362-7.

Independent Sector. (2001) *Giving and volunteering in the United States 2001.* Washington, DC: Independent Sector.

Isaac R. M. & Walker J. M. (1988) Group size effects in public goods provision: the voluntary contributions mechanism. *Quarterly Journal of Economics 103*, 179-199.

Kim O. & Walker, M. (1984) The Free Rider Problem: Experimental Evidence. *Piblic Choice , 43*, 3-24.

Leydard J. (1995) Public goods: a survey of experimental evidence. In J. H. Kagel, & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 111-194) Princeton, NJ: Princeton University Press.

Marwell G. & Ames, R. (1981) Economists Free Ride, Does Anyone Else? *Journal of Public Economics , 15*, 295-310.

Mayr U., Harbaugh, W. T., & Tankersley, D. (2009) Neuroeconomics of charitable giving and philanthropy. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision Making and the Brain* (pp. 303-320) San Diego, CA: Elsevier Inc.

Moll J., Krueger, F. Zahn, R. Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences , 103* (42), 15623-15628.

Osborne M., & Rubinstein, A. *A Course in Game Theory.* Cambridge, MA: MIT Press.

Ostrom E., Walker, J., & Gardner, R. (1992) Convenants With and Without a Sword: Self-Governance is Possible. *American Political Science Review , 86* (2), 404-17.

Palfrey T., & Prisbrey, J. (1997) Anomalous Behavior in Public Goods Experiment: How Much and Why? *American Economic Review , 87*, 829-46.

Rabin M. (1993) Incorporating Fairness into Game Theory and Economics. *American Economic Review , 83*, 1281-1302.

Sober E., & Wilson, D. S. (1998) *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, MA: Harvard University Press.

Tversky A., & Kahneman, D. (1991) Loss Aversion in Riskless Choices: A Refernces-Dependent Model. *The Quarterly Journal of Economics , CVI*, 1039-62.

Von Neumann, J., & Morgenstern, O. (1944) *Theories of Games and Economic Behavior.* Princeton: Princeton University Press.

Yaari M., & Bar-Hillel, M. (1984) On Dividing Justly. *Social Choice and Welfare , 1*, 1-24.

Zak P. J., Kurzban R. & Matzner W. T. (2005) Oxytocin is associated with human trustworthiness. *Hormones and Behavior , 48*, 522-527.

Why Cognitivists About Practical Reason can Only be Semi-Cognitivists[+]

Giacomo Mollo[*]
gmollo@umich.edu

ABSTRACT

Debates about the nature and the role of rationality are part of the core of philosophy of economics. Holding rational beliefs and being instrumentally rational are some of the primary desiderata that economic theory demands when modelling the behavior of an agent. Philosophy itself, and especially philosophy of action, has also explored analogous issues and can certainly provide substantial insights about the nature of practical rationality. An agent is demanded to hold rational belief about the world: theoretical rationality requires the agent to hold beliefs that are consistent and coherent. On a practical level, the agent is capable of intending: this activity brings about the state of affairs that he desires. Rationality urges the agent to hold intentions consistently with the set of beliefs that he is accepting. Practical rationality in particular demands the agent to respect the consistency and coherence requirements for intentions.
In the following paper I am going to explore and criticize cognitivism about practical reason, one of the most disputed position in action theory. Cognitivists claim that we can make sense of the rational requirements for consistency and coherence of intentions by appealing exclusively to the requirements of consistency and coherence of beliefs. Practical rationality is in some sense reducible to theoretical rationality. I will explore this claim and assess whether this is an acceptable possibility. And of course, I will point out that there is space for some criticism.

PROLOGUE

Cognitivism about practical reason is the position according to which we can explain the requirements of consistency and coherence of intentions in terms of the requirements of consistency and coherence of beliefs. My main interest is to analyze the validity of the cognitivist's claim that demands of theoretical rationality (TR) can fully explain all demands of practical rationality (PR) and in particular to test whether the cognitivist account of belief-involvement of intention can be successful in explaining both consistency and coherence of intentions.

I will start by considering what are assumed to be the requirements of practical and theoretical rationality. I will use examples to illustrate how exactly these requirements work and I will make clear what they are supposed to accomplish. I will then consider the relationship between requirements of TR and requirements of PR. This will be followed by an explanation of how cognitivists about intentions have attempted to claim that rational requirements of practical reason are grounded in (or derive from) rational requirements for theoretical reason. I will consider the general case where cognitivists attempt to show that (1) *inconsistency of intention can be explained in terms of inconsistency of beliefs*, and (2) *incoherence of intentions*

[+] I would like to thank Professor Sarah Buss (University of Michigan, Ann Arbor) for the long and exciting discussions about this topic and the referees from Humana.Mente for their helpful criticisms.
[*] Department of Philosophy, University of Michigan

can be explained by incoherence of beliefs. I will consider cases (1) and (2) and assess their plausibility.

As for claim (1), I will *take* into account an objection inspired by Brunero's kidnapping scenario[1]. I will make some interpretive clarification of the case – I will distinguish between intending not to A and not intending to A - and demonstrate that the cognitivist is able to provide a strong case in favour of his position no matter which interpretation we adopt.

Secondly, I will consider claim (2) and explain how Bratman[2] has objected to this position. He shows how there is a possibility to hold false beliefs about intentions: this will create a situation in which we have coherence of beliefs in TR and incoherence of intentions in PR. This occurrence challenges the cognitivist 's idea that coherence of intentions is grounded in coherence of beliefs. I will object to Bratman by assessing whether holding a false belief about intention is a genuine possibility.

The rest of the paper will present an alternative attack against cognitivism. If we assume that (a) the standard formulation of coherence of intention is correct and (b) the Strong Belief Thesis holds, I show that we have good ground to dispute the claim that incoherence of intentions can be reduced to incoherence of beliefs.

My conclusion is that it is not possible to be a complete cognitivist: consistency of intentions is systematically explained from a cognitivist's perspective and the account is able to address a number of objections. If however we consider coherence of intentions, we can see how no claim about belief-involvement that a cognitivist supports can possibly reduce incoherence of intentions to incoherence of beliefs.

THEORETICAL RATIONALITY AND PRACTICAL RATIONALITY: EACH HAS ITS OWN DEMANDS

If an agent wants to be rational about his own beliefs, he needs to respect the two main requirements of theoretical rationality: consistency of beliefs (COB) and coherence of beliefs (COHB). If a set of belief honours both demands of theoretical rationality then the set is considered rational.

Belief Consistency requires that all beliefs are not mutually exclusive if considered together. One possible formulation is:

> *COB: Theoretical rationality demands that if one believes p, and believes p→-q , then one must not believe q.*

We can illustrate this requirement by using an example.

> *Theoretical rationality requires that IF you hold beliefs:*
> *Belief 1: [It's raining]*
> *Belief 2: If [it's raining] then NOT [Streets are dry] (i.e. Streets are wet)*
>
> *THEN*
>
> *(if you want to respect the COB requirement) you MUST NOT hold the belief:*
> *Belief 3: [Streets are dry]*

[1] Brunero, J. Against Cognitivism about Practical Rationality. *Philosophical Studies*, August 2008.
[2] Bratman, M. Intentions, Beliefs, Practical, Theoretical. In J. Timmerman, J. Sorupski, S. Robertson, eds. *Spheres of Reason*, OUP.

I take this requirement as a wide scope requirement: if I hold beliefs 1, 2, 3 at the same time - and therefore my beliefs are not consistent- then in order to be consistent I must discard either belief 1, belief 2 or belief 3.[3]

Coherence of beliefs (COHB or Belief Closure) is the requirement that you need to believe the logical consequences of your beliefs. Bratman et al.[4] refer to it as explanatory coherence: the idea is that each belief X needs to have an explanation (i.e. a series of other beliefs Y,Z,W) that are able to account for belief X. One possible formulation is:

> *COHB: Theoretical rationality requires that if one believes p, and believes that p→-q, then one must believe –q.*

Again, we can use the same example:

> *Theoretical rationality requires that IF you hold beliefs:*
> *Belief 1: [It's raining]*
> *Belief 2: if [It's raining] then NOT[Streets are dry] (i.e. Streets are wet)*
>
> *THEN*
>
> *(if you want to respect the COHB requirement) you MUST hold the belief:*
> *Belief 3: NOT[Streets are dry]*

On the other hand, practical rationality defines the requirements for a rational system of intention and of intentions and beliefs. If an agents wants to be rational about his own beliefs and intentions, he needs to respect the two main requirements of practical rationality: consistency of intentions (COI) and coherence of intentions (COHI). If a set of intentions and beliefs honours the demands of practical rationality then the set is considered rational.

Consistency of intentions explains how intentions and beliefs are compatible if taken together:

> *COI: Practical rationality requires that if one intends to x, and believes that if one will x then one will not y, then one must not intend to y.*

We can use an example to clarify the notion:

> *Practical rationality requires that IF you hold:*
> *1. Intention : [I intend to travel to Prague this afternoon]*
> *2.Belief : [If I travel to Prague this afternoon, I will not travel to New York this afternoon]*

[3] Generally speaking, being a wide scope requirement simply means that there are three ways to remove yourself from an irrational state: as mentioned you can discard either 1, 2 or 3. The scope of the requirements is not attached to a specific attitude – this would be a narrow scope requirement- but ranges over a disjunction, in this case, of beliefs. For a more detailed treatment of the subject see Broome 1999, 2007 and Kolodny 2005.

[4] Bratman, M. Intentions, Beliefs, Practical, Theoretical. In J. Timmerman, J. Sorupski, S. Robertson, eds. *Spheres of Reason*, OUP.

THEN

(if you want to respect the COI requirement) you MUST NOT hold the intention:
3.Intention: [I intend to travel to New York this afternoon]

I take this requirement also as a wide scope requirement: if I hold 1, 2, 3 at the same time (and therefore my intentions and beliefs are not consistent) then in order to be consistent I must discard either 1, 2 or 3.

Coherence of intentions (COHI or Means-End coherence) explains what intentions we should hold given our beliefs and our other intentions:

COHI: Practical rationality requires that if one intends to x, and believes that if one will x, then one must intend to y, then one must intend to y.

Again we can use an example to clarify the position:

Practical rationality requires that IF you hold:
1.Intention: [I intend to travel to Prague this afternoon]
2. Belief: [If I travel to Prague this afternoon, then I must intend to buy a ticket to Prague]

THEN

(if you wan to respect the COHI requirement) you MUST hold the intention:
3.Intention: [I intend to buy a ticket to Prague]

Again I take COHI to be a wide scope requirement: If I hold 1 and 2 and fail to hold 3, then I am in a situation of incoherence of intentions. In order to hold coherent intentions I can either abandon 1, abandon 2 or form intention 3.

COGNITIVISM ABOUT PRACTICAL REASON

Is there a relationship between the requirements of theoretical rationality and the requirements of practical rationality? Cognitivists generally think that we can appeal to theoretical rationality to explain the requirements of practical rationality.

Their approach includes two different moments: in the first place they formulate a position that is able to relate beliefs and intentions. Secondly, they explain why given this position it is possible to claim that the rational requirements of practical rationality are grounded in rational requirements for theoretical rationality.

There are three different ways of considering the relation between beliefs and intentions. Each view has a different idea of what is defined as belief-involvement i.e. the explicit relation between intention and the belief that is involved when the intention is formed.

According to the Identity Thesis (or Pure Cognitivism) the intention "I intend to A" *is* a belief that "I WILL do A". Intentions are some sort of beliefs: they coincide with a special kind of

belief. This is taken to be the strongest position of the three. This view can be found in the early works of Harman (1976) and in the work of Velleman[5].

A different position is the Strong Belief Thesis: this claims that the intention "I intend to A" *involves* the belief that "I WILL do A". The idea is that if I am in the state of intending to A, then I am also in the state of believing that I will A. Unlike the Identity Thesis, an intention is a separate entity from a belief. This position figures in the later work of Harman (1987).

Lastly, the Weak Belief thesis claims that the intention "I intend to A" *involves* the belief that "it is *possible* that I WILL do A". Holding an intention requires holding the belief that it will be possible that I will perform the action that I am intending. This is clearly the weakest position.

In general we can say that the relation between intention and belief pointed out by one theory or the other will tell you how to derive a certain belief from the corresponding intention. For instance if you hold the intention "I intend to A" then, according to the Strong Belief Thesis, you do hold the belief that "I will A".

As pointed out earlier, a definition of the belief-involvement is the first step toward explaining the requirements of practical rationality by appealing to the requirements of theoretical rationality. It is now time to see exactly how this movement is accomplished.

1. CONSISTENCY OF INTENTIONS IS EXPLAINED BY CONSISTENCY OF BELIEFS

Consider first of all the general case of inconsistency of intentions by using the example I presented earlier in the text. For the sake of clarity I will distinguish beliefs and intentions in two separate columns.

BELIEFS	INTENTIONS
(b) Belief: [If I travel to Prague this afternoon, I will not travel to New York this afternoon]	(a) Intention: [I intend to go to Prague this afternoon] (c) Intention: [I intend to travel to New York this afternoon]

This is clearly a case of inconsistency in practical reason: If I intend to go to Prague AND I hold the belief that if I travel to Prague this afternoon then I won't travel to New York, then I cannot rationally hold the intention to travel to NY this afternoon.

Let's now assume the Strong Belief Thesis[6] and see how cognitivists relate the requirement of consistency of intention in practical reason with the requirement of consistency of beliefs in

[5] Velleman, D. *Practical Reflection*, University of Chicago Press.

[6] The entire paper will focus on the Strong Belief Thesis version of cognitivism. This is the position in between the radical and the weak positions and, as the reader will see later in the text, it seems the most promising when it comes to successfully address certain objections against cognitivism. I will briefly show how the Identity Thesis can be problematic but I will not take into consideration the Weak Belief Thesis.

theoretical reason. The belief-involvement principle in the case of the Strong Belief thesis is: *the intention "I intend to A" involves a belief that "I will do A"*. If we consider the example, then the intention [I intend to go to Prague this afternoon] (a) *includes* in some sense the belief [I will go to Prague this afternoon] (a*) by the Strong Belief Thesis. By the same token [I intend to travel to New York this afternoon] (c) *includes* the belief [I will travel to New York this afternoon] (c*). In a schematic form:

BELIEFS		INTENTIONS
(a) Belief:* *[I will travel to Prague]*	← by Strong Belief Thesis	(a) Intention: [I intend to travel to Prague this afternoon]
(b) Belief: [If I travel to Prague this afternoon, I will not travel to New York this afternoon]		
(c) Belief:* [*I will travel to New York this Afternoon*]	← by Strong Belief Thesis	(c) Intention: [I intend to travel to New York this afternoon]

It is at this point clear how the inconsistency of intention that occurred between (a), (b) and (c) can be explained by the inconsistency of beliefs (a*), (b) and (c*).

THE KIDNAPPING CASE

Let me now take into consideration one specific objection to cognitivism. Brunero[7] points out that the Strong Belief Thesis entails that believing that one will X is *necessary* for intending to X. He then claims that this is problematic: it is in fact possible to show that believing one will X is not sufficient for intending to X. [8] This means that Brunero interprets the Strong Belief Thesis as saying that believing that one will X is *sufficient* for intending to X. This latter position is vulnerable to a number of different counterexamples. Brunero describes cases in which it is possible to believe that one will X without him intending to X. Suppose a gangster is now kidnapping me. As he ties me up I hear from his driver that we are going to the gangster's hideout. In this case I do not intend to go to the gangster's hideout but I believe that I will go to the gangster's hideout[9]. We hold a belief that I will X without intending to X. This is contrary to the claim that believing that one will X is sufficient for intending to X. Brunero is right in claiming that believing one will X is *not sufficient* for intending to X. But is it really true in the

[7] Brunero, J. Against Cognitivism about Practical Rationality. *Philosophical Studies*, August 2008.
[8] Ibid. *"According to the Strong Belief Thesis, believing that one will x is necessary for intending to x. But problems for the above attempt to use BC [belief consistency] to explain IC [intention consistency] arise from the fact that believing one will x is not sufficient for intending to x. I'll argue here that (a) believing one will x is not sufficient for intending to x […]".*
[9] Ibid. *"Surely I do not intend to go to the gangster's hideout, even though I believe I will".*

first place that - as Brunero seems to interpret it - the Strong Belief Thesis entails that holding a belief *is sufficient* for intending?

I believe that the Strong Belief Thesis cannot be understood as Brunero interprets it: There is no element in its definition that would suggest how holding a belief would be *sufficient* to produce an intention. All it claims is that *the intention "I intend to A" involves the belief that "I will do A"*. From this formulation it is clear that there is in fact a necessity relation, namely that "Intending to A" necessarily requires a belief that "I will A". There is however no sufficiency relation that transpires from the original definition: nothing tells us that forming a belief is sufficient to produce a corresponding intention. In his objection, Brunero seems to interpret the definition of the Strong Belief Thesis as *the intention "I intend to A" ONLY involves a belief that "I will do A".* If we consider his interpretation, then we can deduce a sufficiency relation between belief and intention: if an intention to X only involves a belief to X, then a belief to X can be said to be sufficient to produce the intention to X. This last formulation is equivalent to the Identity Thesis but it is inaccurate as an interpretation of the Strong Belief Thesis.

In conclusion, Brunero's kidnapping case – as it is presented- is certainly a good counterexample to the Identity Thesis but it is inadequate to challenge the Strong Belief Thesis.

I still believe that the kidnapping case is an interesting example and I want to propose an alternative way in which it can still challenge cognitivism. As we have seen, cognitivists claim that it is possible to explain inconsistency of intentions by appealing to inconsistency of beliefs. In the kidnapping case we have a situation in which I do NOT intend to go to the gangster's hideout but I do believe that I will go to the gangster's hideout.

If we assume the Strong Belief Thesis and attempt to see which belief is involved in (a) then we will have a situation as the following:

BELIEFS		INTENTIONS
(a) Belief:* *[I will NOT go to the gangster's hideout*	← by Strong Belief Thesis	(a) Intention: [I do not intend to go to the gangster's hideout]
(b) Belief: [I will go to the gangster's hideout]		

This is problematic in that we can observe consistency in practical reason that is NOT reflected in consistency in theoretical reason since (a*) and (b) are inconsistent. We have no reason to say that intentions are inconsistent, but if we apply the Strong Belief Thesis then we cannot doubt that we are holding inconsistent beliefs. This is clearly a problem for cognitivism.

"INTENDING NOT TO A"

It is possible to respond to this objection on different grounds. Let's now focus on how we can interpret intention (a) in the example. Prima facie, by "I do not intend to go to the gangster's

hideout" I might mean "I intend not to go to the gangster's hideout" or "I intend to abstain from going to the gangster's hideout". I have formed my intention and my intention is to not go to the gangster's hideout. Since we do have an intention, a cognitivist would legitimately apply the Strong Belief Thesis to understand what belief is involved by that intention. In a schematic form:

BELIEFS		INTENTIONS
(a) Belief:* *[I will NOT go to the gangster's hideout*	← by Strong Belief Thesis	(a) Intention: [I INTEND NOT to go to the gangster's hideout]
(b) Belief: [I will go to the gangster's hideout]		

As pointed before, we see how this interpretation presents a problem for the cognitivist's account: we are presented with inconsistent beliefs (a*) and (b) and with intention (a). We have thus a case in which inconsistency of beliefs is not reflected in inconsistent intentions.

At this point I believe the cognitivist can still provide a response to this counterexample. This would require to analyze the exact *kind* of belief that is involved by intention (a).

Velleman[10] suggests an interesting distinction between how we understand intentions and how we understand predictions. Understanding someone's intention to do A requires understanding the evidences that shows that he wants to do A. On the other hand, understanding a prediction that he will do A involves understanding the evidences that he will do A *independently* of the will of the agent to do A.

Let's now see how this distinction can help cognitivists to respond to the kidnapping case. The question is whether belief (a*) is the correct formulation of the kind of belief involved in the intention [I INTEND NOT to go to the gangster's hideout]. Following Velleman's distinction we would need to be clear on the will of the agent when attempting to understand his intentions. Does (a*) really helps understanding the evidences that shows whether the agents wants (or doesn't want) to go to the gangster's hideout or not? I believe that we can produce an alternative formulation of the belief involved in (a):

[10] Velleman, D. *" Perhaps the difference is that an agent's intending to do something is normally to be explained by his wanting to do it, whereas his predicting that he will do something is to be explained by his having evidences that he will". Practical Reflection*, University of Chicago Press. Pages 94-95.

BELIEFS		INTENTIONS
*(a**) Belief:* *[I will NOT go to the gangster's hideout <u>BY MY OWN WILL</u>]*	← <u>by Strong Belief Thesis</u>	(a) Intention: [I do NOT intend to go to the gangster's hideout]
(*b) Belief:* *[I will go to the gangster's hideout]*		

We have here applied the Strong Belief Thesis and formulated the correct kind of belief that is involved in intention (a). Belief (a**) expresses the wants of the agent holding intention (a): if the agent could choose, then he will not go to the gangster's hideout i.e. he will not go to the gangster's hideout by his own will. From a cognitivist perspective, this is the exact belief that is involved by intention (a): it helps us understanding the evidences that shows whether or not the agent is willing to go to the gangster's hideout. Belief (b) on the contrary is not derived from the agent's decisions or wants but it is a prediction given a certain state of the world (namely the fact that the agent is tied up sitting in the gangster's car). It involves understanding the evidences that he will go to the gangster's hideout *independently* of his will to go to the gangster's hideout.

It is now finally clear how (a**) and (b) are consistent. By distinguishing between kinds of beliefs the cognitivist has one way to successfully respond to the kidnapping case and invalidate its role as a counterexample.

"NOT INTENDING TO A"

There is however a second possible interpretation of "I do not intend to go to the gangster's hideout" (a) in the kidnapping example. This might in fact mean that the agent *failed* to form the intention to A. In other words the agent is incapable to produce the intention "I intend to A". He does NOT embark in the process of intention formation at all and there is no intention whatsoever that appears in his practical reasoning.

In the light of the latter formulation of (a), we can consider its implications for cognitivism. By "I do not intend to go to the gangster's hideout" we mean is that I do not form the intention to go to the gangster's hideout. There is no intention whatsoever in the picture:

BELIEFS	INTENTIONS
(b) Belief: [I will go to the gangster's hideout]	(a) Intention: […]

It is evident that the kidnapping case does not pose any problems for cognitivism: we lack an intention that could be reduced to a belief through a belief-involvement principle. Under this interpretation the issue about consistency does not arise in the first place. The cognitivist is therefore immune to the counterexample.

We have considered the cognitivist's claim that consistency of intentions can be explained by consistency of beliefs. We have tested the claim by providing a counterexample to it. We have also seen how the cognitivist has different ways to respond to the challenges and firmly defend his main claim. It seem therefore plausible to conclude that there are solid cases that the cognitivist can employ to show that we can explain the consistency of intentions in terms of consistency of belief. I am not here claiming that this is the right way of explaining consistency of intentions in practical reason. What I am trying to convey is that the cognitivist seems to have the tools to defend his explanation of consistency of intentions in practical reason.

2. COHERENCE OF INTENTIONS IS EXPLAINED BY COHERENCE OF BELIEFS

Cognitivists claim that incoherence in practical reason can be reduced to incoherence in theoretical reason: if we perceive an incoherence of intentions then there must be a parallel incoherence in the system of beliefs that the agent is holding. This latter will be able to explain the former.

Let's consider the general case of incoherence of intentions as Bratman[11] presents it. We have seen that practical rationality demands that if "I intend to E" and I hold the belief that "If one will E, then one will intend to M" (M is the means I must take to accomplish end E), then I am required to form the intention " I intend to M". If you do not intend to M then your intentions are incoherent. Take the initial example of incoherence of intentions and rephrase it to fit Bratman's formulation. If I hold the intention [I intend to go to Prague this afternoon] (a); hold the belief that [buying a ticket is necessary to go to Prague] (b) and [intending to buy a ticket is necessary to buy a ticket](c) and still hold [I do not intend to buy a ticket] (d), then my intention-belief system is incoherent. Let me represent the situation by highlighting its two main elements, beliefs and intentions. I will use again the left column for beliefs and the right column for intentions:

BELIEFS	INTENTIONS
(b) Belief: [Buying a ticket is necessary to travel to Prague]	(a) Intention: [I intend to travel to Prague this afternoon]
(c) Belief: [Intending to buy a ticket is necessary to buy a ticket][12]	(d) Intention: [I do not intend to buy a ticket]

[11] Bratman,M. Intention, Belief, Practical, Theoretical. In J. Timmerman, J. Sorupski, and S. Robertson, eds. *Spheres of Reason*, OUP.

[12] I am including belief (c) to be faithful to the characterization that Bratman provides. It can be pointed out that this belief is by no means required to make the system of beliefs/intentions coherent as it does not appear in the original formulation of COHI presented earlier in the text.

Given the formulation of coherence of intentions illustrated above, it is now relatively straightforward to see how this is a case of incoherent intentions. I do not intend to buy a ticket when the demands of practical rationality require me to do so.

Let's now assume the Strong Belief Thesis and see how cognitivists relate the requirement of coherence of intention in practical reason with the requirement of coherence of beliefs in theoretical reason. The position claims that: the intention "I intend to A" involves the belief that "I will do A". If we consider the example mentioned above, then we can see how by the Strong Belief Thesis the intention [I intend to travel to Prague this afternoon] (a) involves the belief [I will travel to Prague this afternoon] (a*). By the same token [I do not intend to buy the ticket] (d) involves the belief [I will not buy the ticket] (d*). In a schematic form:

BELIEFS		INTENTIONS
(a) Belief:* *[I will travel to Prague]*	← by Strong Belief Thesis	(a) Intention: [I intend to travel to Prague this afternoon]
(b) Belief: [Buying a ticket is necessary to travel to Prague]		
(c) Belief: [Intending to buy a ticket is necessary to buy a ticket]		
(d) Belief:* *[I will NOT buy a ticket]*	← by Strong Belief Thesis	(d) Intention: [I do not intend to buy a ticket]

From a cognitivist's standpoint, we can therefore observe how the incoherence of intentions that do not respect the demands of practical rationality – set (a); (b); (c); (d) – can be seen as incoherence of beliefs that do not respect the demands of theoretical rationality – set (a*); (b); (c); (d*).

Bratman challenges this account by presenting a case in which we observe coherence of beliefs and, at the same time, incoherence of intentions.

BRATMAN AGAINST COGNITIVISM

In "*Intention, Belief, Practical, Theoretical* "[13] Bratman argues that it is not possible to reduce a case of incoherence in practical rationality to a case of incoherence in theoretical rationality. Cognitivism seems to allow the possibility of holding a false belief about one's own intention:

[13] Bratman, M. Intentions, Beliefs, Practical, Theoretical. In J. Timmerman, J. Sorupski, S. Robertson, eds. *Spheres of Reason*, OUP.

this will generate a situation in which even though someone's beliefs are coherent, her intentions will be incoherent. This is how Bratman argues to support this claim:

> Suppose I intend E and know that E requires both M and that I intend to M. If I still do not intend M my intentions suffer from means-end incoherence. But suppose that I in fact do not intend M, I nevertheless falsely believe that I intend M. So my beliefs are that E, that E requires both M and that I intend M, and that M. There is no incoherence (though there is falsity) in this structure of beliefs. So means-end coherence is not belief coherence.[14]

If cognitivism allows the possibility of holding a false belief regarding one's intentions (i.e. I falsely believe I intend M when in fact I don't) then we have a situation where we have *coherence in our system of beliefs and at the same time we hold incoherent intentions* ('I intend to E' and 'I do not intend M'). Since coherence of intentions derives (according to cognitivism) from coherence of beliefs then coherence in one's beliefs system should be *reflected* in one's system of intentions. According to Bratman, the possibility of holding a false belief precludes this parallelism. In a schematic form:

BELIEFS		INTENTIONS
(a*) Belief: [I will travel to Prague (E)]	← by Strong Belief Thesis	(a) Intention: [I intend to travel to Prague this afternoon]
(b) Belief: [Buying a ticket (M) is necessary to travel to Prague]		
(c) Belief: [Intending to buy a ticket is necessary to buy a ticket]		
		d) Intention: [I do not intend to buy a ticket]
(e) [FALSE] Belief: *[I intend to buy the ticket]*		
(f) Belief: *[I will buy the ticket]*		

As we can observe, by introducing false belief (e) we are faced with a perfectly coherent system of beliefs and at the same time with a system of incoherent intentions. This would count as a counterexample to the explanation that cognitivists provide to reduce incoherence of intentions to incoherence of beliefs.

[14] Ibid.

AGAINST BRATMAN

First of all it is necessary to establish what is the interpretation of Bratman's claim "in fact I do not intend M" or, in my example, " I do not intend to buy the ticket" (d). Two are the possibilities if we follow the distinction that I presented earlier between holding the intention "I intend not to A" and failing to form the intention "I intend to A".

On one hand Bratman might mean 'I intend NOT-M' or 'I intend to NOT buy a ticket'. If this is the case, then his position cannot possibly hold. Since the agent is holding an intention to not perform an action, the cognitivist might claim that (given that intention involves belief) the agent needs to have a belief of the kind "I will not buy the ticket". This will produce incoherence also in theoretical reasoning and it would dismiss Bratman's counterexample. The interpretation however misses the very idea of what we have described as incoherence.

On a second interpretation the sentence " I do not intend to buy the ticket" (d) might be paraphrazed as "I did not form the intention to buy the ticket". This is certainly more plausible and is able to sustain Bratman's line of reasoning. The relevance given to the fact that incoherence of intentions entails my failing to form the right intention will be later helpful to support the critique that I will move against cognitivism. As for now let's consider how we can criticise Bratman from a different perspective even if we allow this second interpretation. Following the second interpretation, the final picture would be:

BELIEFS		INTENTIONS
(a*) Belief: [I will travel to Prague (E)]	← by Strong Belief Thesis	(a) Intention: [I intend to travel to Prague this afternoon]
(b) Belief: [Buying a ticket (M) is necessary to travel to Prague]		
(c) Belief: [Intending to buy a ticket is necessary to buy a ticket]		
		(d) Intention: […] {should be: I intend to buy a ticket}
(e) [FALSE] Belief: *[I intend to buy the ticket]*		
(f) Belief: *[I will buy the ticket]*		

Your intentions are incoherent in that you FAIL to form the intention that you must form according to the demand of practical rationality. At the same time your beliefs are coherent in that you DO FORM the belief that you must form according to the demand of theoretical

rationality. If we accept this interpretation, then we can appreciate how Bratman's case is a counterexample to cognitivism.

There are two main lines of reasoning that can be used to attack Bratman's counterexample.

On one hand, a cognitivist might show that a closer analysis of the example presented will reveal coherence in practical reason together with coherence in theoretical reason. Earlier in this paper we considered different versions of cognitivism: Identity Thesis, Strong Belief Thesis and Weak Belief Thesis. If Bratman's picture is a counterexample to cognitivism then it needs to challenge all three versions of cognitivism. It is clear that the counterexample challenges the Strong Belief Thesis and the Weak Belief Thesis. But what about the Identity Thesis? As we have seen, according to the Identity Thesis (or Pure Cognitivism) the intention "I intend to A" *is* a belief that "I WILL do A". Intentions are some sort of beliefs: they coincide with a special kind of belief. The belief "I will buy the ticket" (f) *is* the same as the intention "I intend to buy the ticket". We are thus holding the intention "I intend to buy the ticket" in practical reason as the coherence requirement would command us to do. This is precisely the case because my intention "I intend to buy the ticket" is the same as my belief "I will buy the ticket". If this is true, then we are presented with coherence of beliefs and *at the same time* coherence of intentions. This is contrary to what Bratman is claiming. There is therefore at least one version of cognitivism that is not effectively challenged by Bratman's counterexample.

A second way to attack Bratman is to show that we are actually presented with incoherence of beliefs in theoretical reason as well as incoherence of intentions in practical reason. Bratman claims that what makes beliefs coherent – even when intentions are incoherent – is the possibility of holding a false belief about one's intention. In the following part I will attack this claim and show that there is no genuine possibility for the agent to hold false beliefs about future intentions. If my argument is correct then Bratman's counterexample to cognitivism will lose its appeal.

Let's begin by breaking the idea of 'falsely believing that I intend to buy the ticket' into its main components. I am now going to make some basic assumptions that will help supporting the rest of the argument. First of all it is crucial to distinguishing believing and intending and determine which is the functional role of each element in our mental economy. Believing is a cognitive state: I am in a mental state in which I have reasons to believe something about the world. I have evidences that I take as supportive for my belief and I accept this belief as true. The role of a cognitive state is to accept or reject information about the world on the basis of evidences. Intending, on the other hand, is a motivational state: it affects my immediate behavior and, generally speaking, it initiates the action that is the content of my intention. I take intention to be the first step toward a conscious action. Its role is to move my behavior toward a specific direction. This is an extremely broad distinction but it is helpful to understand exactly the kind of belief that we hold when we hold a belief about our own intention. From this characterization it follows that I cannot be in a motivational state that applies to the future exclusively: since intention is immediately motivational, then my intentions can only concern my present behavior.[15] I can also be in a cognitive state that tells

[15] I can intend to perform an action in the future but this intention is not disconnected from my present actions. I can intend to boil an egg in the future but this intention has an immediate motivational impact on the means that I intend to use now to reach my goal. I take the intention to perform a future action to be necessarily connected with the present motivational state of the agent. This is compatible with Bratman's ideal of future-directed intentions but it is not the focus of the discussion. See Bratman 1987.

me something about my future motivational state. This is precisely the case in which I believe that I intend to buy the ticket. I am now in a cognitive state: I believe something about the world; in this special case I believe something about my future self. I am in a cognitive state whose content is a future motivational state. I am holding the belief that I intend to buy the ticket: this is a particular kind of belief, namely a belief *about* my own future intentions.

Let's now examine how a cognitive state – belief – about a motivational state can be made true or false. In other words what I am asking is how exactly can I truly or falsely believe that I intend to buy the ticket. Bratman argues that the main problem with cognitivism is that it allows cases in which we can hold false beliefs about intentions i.e. cases where an agent falsely believe he intends to X when he in fact he does not intend to X. Unfortunately he does not fully explain how cognitivism can allow such a possibility and in particular how exactly belief about intentions can be made true or false. We can however infer the underlying picture that Bratman seems to imply when he claims that we can hold false beliefs about intentions. Consider the example in which I hold the belief that I intend to drink the coffee that is in front of me in one minute. As the time comes I fail to intend to drink the coffee. I might claim that I am falsely believing that I intend to drink the coffee in one minute: I am lacking that very motivational state when the time comes. The presence of the intention to drink the coffee when the proper time comes is what makes the belief about my intention true or false. If, in fact, I formed the intention to drink the coffee, then I would have truthfully held the belief that I intended to drink the coffee in one minute. The belief about my intention is made true or false by the presence of that very intention. The belief "I believe I intend to drink the coffee in one minute" is made true or false by the presence or the absence of the intention "I intend to drink the coffee" when the proper time comes. Hence the possibility of holding a false belief about one's intention i.e. that *"I in fact do not intend M"* when at the same time *"I nevertheless falsely believe that I intend M".*[16]

I believe that this is a faulty characterization and that there is no genuine possibility to hold a false belief about one's intention if we consider this picture. To show this I will consider two main cases, the synchronic case and the diachronic case. Let's start with the synchronic case: the belief about intention to do X and the intention to do X itself occur at the same time. Think of an example in which I initiate an action by my intending and at the same time I am fully aware of it. We can say that the agent is in the cognitive state "I believe at t_1 that I intend to buy the ticket at t_1". Let's say that t_1 is exactly now: I am in the motivational state "I intend to buy the ticket at t_1". At the same time, by observing myself as being moved by that intention, I have good evidence for the presence of my intention and therefore good grounds to support the belief – i.e. be in the cognitive state - "I believe at t_1 that I intend to buy the ticket at t_1". If at t_1 I held the belief "I believe at t_1 that I intend to buy the ticket at t_1" *without* at the same time the presence of the intention "I intend to buy the ticket at t_1", then I would have a situation in which I falsely believed about my intention. Is this last case a genuine possibility? I am inclined to say no: if I do not hold the intention to do X at t_1, then I do not have any reason to form a belief that I hold the intention to do X at t_1. This would misrepresent what is going on in my practical reason. In general, to rationally hold a belief about an intention, one requires the presence of the intention that the belief is specifying. In order to see whether that intention is present the agent needs evidences. I believe that many elements can count as evidences that you are/are not holding an intention: these include being spontaneously aware

[16] Bratman, M. Intentions, Beliefs, Practical, Theoretical. In J. Timmerman, J. Sorupski, S. Robertson, eds. *Spheres of Reason*, OUP.

of your own intention, consciously observing you own system of intentions by careful introspection, or seeing yourself initiating an action right now. If by introspection I cannot find any intention to lift the cup in front of me now – or I do not see myself initiating now the action of lifting the cup in front of me - it would be irrational to hold the belief " I believe now that I intend to lift the cup in front of me now". In the synchronic case we are in a situation in which the presence of the intention "I intend to do X now" supports the belief " I believe now that I intend to do X now". The absence of such intention does not allow the belief about that intention to emerge in the first place. In this case rationality demands that beliefs about intention are either supported by actual intention or are not formed at all. Therefore in the synchronic case there is no possibility to hold a false belief about intention in the sense that Bratman emphasizes.

This is the simplest case and Bratman might reply that it does not reflect the scenario that he was presenting. What he is concerned with, as we have seen, are cases of beliefs about future intentions. In those cases, he wants to claim that we can hold false beliefs about future intentions. Let's call this category diachronic cases i.e. situation in which the beliefs about intention and the intention itself occur in separate times. Take for example the belief: "I believe now that I intend to buy the ticket at t_2 " – where t_2 is in the future. I am in a cognitive state now that expresses something about my motivational state at t_2. In order for Bratman's counterexample to work we need the possibility to falsely believe now that I intend to buy the ticket at t_2. In Bratman's picture, the absence or presence of the intention at t_2 is what makes the belief true or false. But is this a real possibility?

There are only two moments at which the belief " I believe now that I intend to buy the ticket at t_2 " could be made true or false: these are when the 'now' is in the past – call it t_1 - and when then 'now' coincides with t_2. The first question is: Can I hold a false belief at t_1? I would say that given the very nature of the belief at t_1, that belief cannot be true or false at t_1. The truth making condition of the belief at t_1 has not yet happened: the occurrence or not occurrence of the intention at t_2 will determine whether the belief is true or false. At t_1 we are clearly not in a position to assign a truth value to that belief. If this argument is correct then at t_1 we cannot possibly claim that we are holding a false (or true) belief about our future intention. The belief is neither true nor false in that the truth making condition has not happened yet. If we consider the behavior of the agent at t_1 we see that he acts at t_1 as if he will actually have the intention at t_2. This would not make the belief itself true or false: all we can say is that there is no way to distinguish the behavior of an agent at t_1 that will actually form the intention at t_2 from the behavior of a second agent at t_1 that will fail to form the intention at t_2. Both agents are holding at t_1 a belief that is neither true nor false.

Let's say that t_2 comes – i.e. 'now' is t_2 – and that I do NOT form the intention to buy the ticket. Can I hold a false belief at t_2? Can I falsely believe now –i.e. at t_2 - that I intend to buy the ticket at t_2? If Bratman is right, since I did not form an intention then my belief is false. But is this a real possibility? As the reader can notice, this is the same as a synchronic case i.e. a case in which belief about intention and intention itself happen at the same time. The absence of the intention to buy the ticket will give no evidences to rationally hold a belief about that intention. In other words if I do not intend to buy the ticket at t_2 then there is no possibility that I can rationally hold the belief "I believe at t_2 that I intend to buy the ticket at t_2".

Imagine a case in which you are in love with a perfect stranger that you happen to see everyday at the coffee shop where you go each morning. I might consciously form the intention to talk to the stranger the next day. I am aware of my mental process and I might come to think that I believe now that I intend to talk to her tomorrow. The following day I

enter the coffee shop, see the person and my emotional reaction prevents me to accomplish what I did intend to do the previous day: my voice disappears and I cannot think of anything to say. Does that mean that yesterday – or today – I falsely believed that I intended to talk to her when I actually did not intend to do so? Would today's failing to accomplish what I intended to do make yesterday's belief false? I do not think so: yesterday my belief could not be either true or false: the truth making condition – whether I will or won't talk to her tomorrow – did not happen yet. Today as I walk in the cafeteria and miserably fail to talk to that person, I do not have any evidence to accept the belief "I believe now that I intend to talk to her now" anymore.

This is much in line with Setya's claim that "the mind is not transparent to itself "[17]: we might be just wrong about assessing our future psychological states. What I am claiming is that we either are not in the position to say whether the assessment is correct – we are holding now a belief about a future intention- or whenever we attain such position then it would be irrational to hold that belief without proper evidences.

In conclusion there is no space to hold a false belief about intention: it is in fact the case that either your belief about intention cannot assume any truth value or that you do not have any ground to form a certain belief about intention. Contrary to what Bratman claims, I do not believe that holding a false belief about intention is a genuine possibility. If the argument is correct, then the counterexample that Bratman present against the cognitivist loses its appeal. If this is the case then the cognitivist is safe: it is in fact the case that we have incoherence in practical reason – we fail to form the intention to buy the ticket – and incoherence in theoretical reason – we fail to form the belief that we will buy the ticket.

Suppose now for the sake of the argument that the previous reasoning is completely mistaken: let's grant that holding a false beliefs about intention is a genuine possibility. If this is the case, I believe that Bratman will still have to solve one important problem.

First of all, the cognitivist might claim that the mere introduction of belief (e) does not make the system of beliefs coherent or incoherent. Belief (e) is so to speak neutral: the belief "I believe that I intend to do X" is not the kind of belief that we need to form to produce a coherent set of beliefs in theoretical reason. The bare introduction of (e) in the set of beliefs does not make beliefs coherent. Bratman seems to be aware of this: what makes the system of beliefs coherent is the presence of belief (f) " I will buy the ticket". But how is belief (f) produced?

One can argue that beliefs "I will travel to Prague this afternoon" (a*) and "Buying the ticket is necessary to travel to Prague" (b) are sufficient to produce the belief "I will buy the ticket" (f) and make the set of beliefs coherent. Bratman however claims that it is (e) - the possibility to form false beliefs about one's intention - the key element that makes the system of beliefs coherent. The challenge for Bratman is to explain what is the exact role of (e) to make the set of beliefs coherent: from the previous considerations it seems that we can easily hold a coherent set of beliefs even without the presence of (e). There must be therefore some theoretical movement that will make the possibility of holding a false belief about intention a crucial element for making beliefs coherent. Unless this issue is addressed, I do not think that the counterexample that Bratman presents is successful in challenging cognitivism.

[17] Setiya, K. Cognitivism About Instrumental Reason. *Ethics* 117, 2007.

The idea presented in the first critique to Bratman is what led me to believe that there might be an alternative way to object to cognitivism. It seems to me that the problematic move is done when connecting theoretical reason with practical reason. I do not see particular problem in reducing (a) to (a*) but there is certainly an ambiguity in interpreting (d) " I do not intend to buy the ticket". The interpretations we can possibly provide to (d) might prevent the possibility of its translation in the sphere of beliefs.

The general case for incoherence presented earlier in the text, shows how intention (d) [I do not intend to buy the ticket] – and interpreted as "I intend NOT to buy the ticket" - is reduced by the Strong Belief Thesis to the belief (d*) [I will not buy the ticket]. This is the case that the cognitivist endorses. In the earlier example, incoherence of intentions in practical reason is reduced to the incoherence of belief (a*); (b); (c) and (d*). We have a situation in which incoherence of intentions in practical reason is explained by incoherence of belief in theoretical reason.

I want to claim that this is not the correct interpretation. Incoherence of intentions occurs because I *fail* to form the intention "I intend to buy the ticket". Being incoherent is certainly not the same as intending –or to *form* an intention to do- the opposite of what I was demanded to intend by the coherence requirements of practical reason. Incoherence, on the contrary, pertains the fact that we *fail to form* an intention for the means that our beliefs would require us to intend in order to accomplish the end. A situation of incoherence of intentions is a situation in which I hold (a); (b); (c) and fail to produce the correct intention "I intend to buy the ticket".[18] So if we want to see which intentions I am holding when I am incoherent we will find that there is *no intention whatsoever* regarding the action "buying the tickets". In a schematic form:

BELIEFS		INTENTIONS
(a) Belief:* *"I will go to Prague"*	← by Strong Belief Thesis	(a) Intention: [I intend to go to Prague this afternoon]
(b) Belief: [Buying a ticket is necessary to go to Prague]		
(c) Belief: [Intending to buy a ticket is necessary to buy a ticket]		

[18] At this point some critics would challenge the idea that incoherence of intentions *only* amounts to fail to form the right intention which is demanded by the coherence requirement of rationality. They can argue that one can in fact be incoherent also by forming an intention with the wrong content . This is not an issue: even if we form an intention with the wrong content, at the same time we are still failing to form the right intention. This latter is the key to incoherence. When we consider incoherence, it seems to me that there is no difference between the situation in which (a) I fail to form the required intention and (b) I fail to form the required intention *and* I form an intention with the wrong content. In both circumstance we are incoherent because we fail to form the correct intention.

(d)* *?*	(d) Intention: […] (I <u>do not form</u> the intention to buy the ticket)

In (d) we observe an absence of intention. The agent, even though holding beliefs/intentions (a), (b) and (c), failed to form the correct intention in conformity with the requirement of coherence of intention.

Since there is no intention upon which we can apply the Strong Belief Thesis then there can't be a parallel belief that will be able to produce incoherence of beliefs. Cognitivists in fact lack a principle that connects non-formed-intentions with beliefs: there is no relation that is made evident regarding the belief involvement of intentions that are failed to be produced. If this is correct then it is not the case, given the classical formulation of belief involvement in any cognitivist account, that incoherence of intentions in practical reason can be possibly explained with incoherence of beliefs in theoretical reason.

The cognitivist might respond that in this case you still have incoherence of beliefs: the demand of theoretical rationality requires you to derive the belief that you will buy the ticket (d*) if you hold beliefs (a), (b) and (c). There is no belief (d*) and therefore we have incoherence in theoretical reasoning. At the same time no intention (d) has been formed even though its production was required by the coherence requirement of practical reason. You can see how the absence of my intention to buy the ticket occurs together with the absence of my belief that I will buy the ticket. The cognitivist might go further and claim that whenever we fail to form the belief required by the demand of theoretical rationality, then we also fail to form the intention required by the demand of practical rationality. He might insist on the fact that there is a situation of theoretical incoherence every time we encounter practical incoherence. We can therefore observe incoherence of beliefs and incoherence of intentions.
The problem with this response is that it does not support the original cognitivist claim that we can understand incoherence in practical reason by reducing it to incoherence in theoretical reason. Even if we concede that the cognitivist is able to demonstrate the constant coexistence of theoretical incoherence with practical incoherence, then it does not follow that one is explicable by the other. He did not show how what went wrong in cases of practical incoherence is reducible to what went wrong in cases of theoretical incoherence. The cognitivist still needs to explain how exactly the failure in practical reason is connected with the failure in theoretical reason or how exactly incoherence in practical reason just is incoherence in theoretical reason.

As I understand the cognitivist picture, the only available instrument to connect intention and belief is a belief-involvement principle: depending on the version of cognitivism – Identity Thesis, Strong Belief Thesis or Weak Belief Thesis – this principle will provide a bridge between practical reason and theoretical reason. Intentions will be therefore related to their corresponding belief. When we consider cases of practical inconsistency, the cognitivist has the proper instruments to explain how exactly inconsistency in practical reason can be explained by inconsistence in theoretical reason. If however we consider cases of practical incoherence, the cognitivist does not have those theoretical tools that can provide a satisfactory account in favour of the claim that we can explain incoherence in practical reason by incoherence in theoretical reason. As I pointed out earlier in the text, a belief-involvement principle cannot possibly be applied when an intention is absent.

There is in fact a full theoretical gap: we would require the cognitivists to provide a 'non-belief involvement principle" that would explain how the absence of an intention somehow *involves* the absence of a corresponding belief. In its current formulation, the cognitivist account cannot explain how the failure to form an intention is connected with the failure to form a belief. It seems therefore that cognitivism cannot possibly provide an account for incoherence of intention in terms of incoherence of beliefs.

CONCLUSION

Can you now possibly be fully cognitivist regarding all the requirements of practical rationality? Given the idea of belief-involvement, can you explain the rational requirements of consistency and coherence in practical reason by appealing to the rational requirements of theoretical reason?

My answer is no. The cognitivist's picture *can* provide a sound and structured explanation on how to derive inconsistency of intentions from inconsistency of beliefs. However, as it is formulated, it *cannot* explain how to reduce incoherence of intentions to incoherence of beliefs. Given the present debate, a reasonable position to hold is to be a semi-cognitivist: this view endorses cognitivism regarding the consistency requirement but cannot accept cognitivism when it comes to explain the coherence requirement. We have seen how the two rational requirements presuppose different structural elements, a fact that is particularly evident when we consider failure of respecting those demands. If we assume the belief-involvement principle alone, then we can see how it can be applied on the consistency requirement and it cannot possibly be applied to the coherence requirement.

BIBLIOGRAPHY

Bratman M. (forthcoming) Intentions, Beliefs, Practical, Theoretical. In J. Timmerman, J. Sorupski, S. Robertson, eds. *Spheres of Reason*, OUP.

Bratman M. (1987) *Intention, Plans and Practical Reason*. Harvard University Press.

Broome J. (1999) Normative Requirements. *Ratio*, 12, 398-419.

Broome J. (2007) Wide or Narrow Scope? *Mind*, 116, 359-370.

Brunero J. (2008) Against Cognitivism about Practical Rationality. *Philosophical Studies,* August.

Kolodny N. (2005) Why be rational? *Mind*, 114, 509-563.

Setiya K. (2007) Cognitivism About Instrumental Reason. *Ethics* 117.

Velleman D. (2007) *Practical Reflection*, University of Chicago Press.

Comparing Preferences

Mauro Rossi[*]
mauro.rossi@umontreal.ca

ABSTRACT

The orthodox view in economics is that interpersonal comparisons (ICs) of preferences present insurmountable epistemic difficulties and, thereby, have no scientific legitimacy. A recent line of thought argues against this position by investigating how ordinary people make ICs of preferences. The underlying idea is that the problem of ICs can be solved if we can find scientific evidence showing that ICs can be reliably made in everyday life. In this paper, I provide an assessment of this strategy. I consider four arguments attempting to show that the conditions for having reliable ICs can be satisfied. I argue that all these arguments fail and reject this strategy as unsuccessful.

> The susceptibility of one mind may, for what we know, be a thousand times greater than that of another. But, provided that the susceptibility was different in a like ratio in all directions, we should never be able to discover the difference. Every mind is thus inscrutable to every other mind, and no common denominator of feeling seems to be possible.[1]

1. INTRODUCTION

It is commonplace that, in everyday life, we compare preferences belonging to different people with respect to their intensity. We typically make such comparisons with relative ease. Moreover, we often do not find inter-personal comparisons of preferences more difficult than intra-personal comparisons, that is, comparisons involving our own preferences.[2] Things change as soon as we consider the matter from a theoretical point of view. Several authors claim that interpersonal comparisons (ICs henceforth) of preferences are either impossible,[3] or meaningless[4] or, at least, that they are not factual claims, but rather normative statements.[5] The main reason for theoretical scepticism is that, while the empirical evidence is sufficient for establishing comparisons about the intensity of a single individual's preferences, it appears to

---

[*] Postdoctoral Fellow, Centre de Recherche en Éthique de l'Université de Montréal (CRÉUM).

[1] S. Jevons, *Theory of Political Economy*, 4th edition, Macmillan, London 1911 (1871, 1st ed.), p. 14.

[2] See D. Davidson, *Judging interpersonal interests*, in *Foundations of social choice theory*, ed. by J. Elster – A. Hylland, Cambridge University Press, Cambridge 1986, pp. 195-211, reprinted as *Interpersonal Comparisons of Values*, in D. Davidson *Problems of Rationality*, Oxford University Press, Oxford 2004, p. 59.

[3] See S. Jevons, *op. cit*, p. 14. For an early reaction against the impossibility claim, see I.D.M. Little, *A Critique of Welfare Economics*, 2nd ed., Clarendon Press, Oxford 1957 (1950, 1st ed.), Chapter IV.

[4] See K. Arrow, *Social Choice and Individual Values*, 2nd ed., Wiley, New York 1963 (1951, 1st ed.), p. 9.

[5] See L. Robbins, *An Essay on the Nature and Significance of Economic Science*, Macmillan, London 1932, p. 139.

be insufficient for establishing comparisons about the intensity of different individuals' preferences. In turn, this has led various areas of scientific research – most notably, economics – to dismiss ICs of preferences as scientifically illegitimate.

This rejection has not been without repercussions. It is true that economic theory does not need ICs of preferences to explain the behaviour of the main economic agents, i.e. consumers and firms. On the other hand, if ICs of preference strengths are not allowed, welfare economics is unable to settle distributive conflicts by considering the relative importance that different individuals attach to competing states of affairs. In social choice theory, it is impossible to aggregate individual preferences in order to obtain a social ranking of alternative states of affairs, which satisfies, amongst the others, a condition of non-dictatorship. Even worse, if ICs turn out to be impossible, the meaningfulness of various ethical doctrines e.g. preference utilitarianism and at least some versions of objective list theories of well-being, is entirely compromised. Finally, the rejection of ICs has a considerable impact on more applied spheres, such as health care and policy making, which require the use of normatively significant measures indicating how different people's preferences compare in terms of strength.

In the course of the years, the constraints posed by the rejection of ICs have started to appear intolerable. As a consequence, several strategies have been explored to give ICs new scientific legitimacy. In this paper, I want to examine one of them in particular. Broadly speaking, the main idea underlying this strategy is that, by investigating how ordinary people make ICs of preferences in everyday life, it is possible to find a solution to the main theoretical problems concerning ICs. Consider the common view of our comparative practice. It is generally held that ICs of preferences are based on the ascription of preferences with specific content and intensity to other people and to ourselves. If this picture is correct, the suggestion is that the solution to the problem of ICs should be based on the analysis of the conditions that ought to be satisfied in order for us to make reliable ascription of preference strengths to other people and to ourselves.

This approach invites us to examine two different kinds of question: the question of mental ascription, that is, the question of how ordinary people assigns mental terms to other people and to themselves; and the question of the meaning of mental states, that is, the question of what ordinary people mean when they employ mental terms. Broadly speaking, there are two main theories of mental ascription: Theory Theory (TT) and Simulation Theory (ST). According to the former, ordinary people ascribe mental states by means of a 'Theory of Mind' that they, more or less tacitly, possess. According to the latter, ordinary people ascribe mental states to others by trying to simulate, or replicate, their mental activity. On the other hand, there are two main theories of the meaning of mental states in contemporary philosophy of mind: (commonsense) functionalism and experientialism. According to functionalism, the meaning of a mental state is given by the set of causal laws in which such a mental state figures and which relate it to inputs, other mental states and behavioural outputs. Instead, according to experientialism, the meaning of a mental state is given by the more or less conscious experiences that the subject has of it.

Given that the problem of comparing different people's preferences concerns a specific type of mental states (i.e. preferences) and one of their properties (i.e. strength), one would expect the existence of a large literature in philosophy of mind connecting the problem of ICs to these fields of research. Instead, and quite surprisingly, philosophers of mind have almost completely ignored the issue of ICs. Alvin Goldman constitutes the only significant exception. Indeed, in his "Simulation and Interpersonal Utility", Goldman attempts to bring the problem

of ICs in line with current debates in philosophy of mind and epistemology.[6] For the purpose at stake, Goldman's approach has two limitations: it focuses mainly on ICs of happiness and it is too specific, as it considers only ST as a theory of mindreading and experientialism as a theory of the meaning of mental states.

In this paper, I want to extend Goldman's approach by focusing explicitly on ICs of preference strength and by considering the main theories of mindreading and of the meaning of mental states. I shall pursue a twofold goal. First, I shall individuate the conditions that ought to be satisfied in order to have reliable ICs of preference strength with respect to all the main accounts of how ordinary people make ICs of preference strength. Second, I shall examine some arguments attempting to show that these conditions can, at least in principle, be satisfied. By so doing, I believe it is possible to have a full assessment of the strategy under consideration. Ultimately, I shall claim that this strategy is unsuccessful. I shall offer an argument by elimination. No matter which account we adopt of how ordinary people make ICs of preference strength, all the arguments proposed in the literature fail to show that we can make reliable ICs of preference strength. Obviously, my analysis does not entail that no such argument exists. Nevertheless, the current state of research about ICs legitimates a moderate form of scepticism: given that no solution explored so far proves to be successful, it might as well be the case that ICs of preferences present an unsolvable problem.

I shall proceed as follows. In section 2, I shall illustrate the problem of ICs of preferences in more detail. Following Goldman, I shall specifically focus on its epistemological dimension, according to which the problem is whether or not we can have scientific knowledge of, or, at least, scientifically justified beliefs about, how different people's preferences compare in terms of strength. In section 3, I shall present, respectively, the main theories of mindreading and the main theories of the meaning of mental states present in contemporary philosophy of mind. In section 4, I shall consider the issue of scientific justification. Although scientific justification may require the satisfaction of several requirements (e.g. publicity, replicability, measurability), for simplicity here I shall focus only on one of them, namely, the requirement that the relevant mechanisms on which ordinary people's comparative practice is based be reliable for the purpose of making ICs. In section 6, I shall discuss five different arguments attempting to show that this condition can be satisfied. I shall argue that all these arguments fail. I shall summarise my findings in the conclusion.

2. THE PROBLEM

The problem of ICs of preferences presents several dimensions. However, two of them are especially important for the previously mentioned areas of scientific research. The first is the metaphysical dimension. The relevant question is whether or not there are facts about ICs of preference strength. The second is the epistemological dimension. The relevant question is whether or not we can have epistemic access to these (alleged) facts about ICs of preference strength. These dimensions are often confused in the literature. However, they differ in important respects and they should be distinguished. The metaphysical question is clearly prior to the epistemological question. If there is no fact of the matter about ICs of preference strength, then no question of epistemic access arises. Although the former is a contentious issue, in what follows, I shall simply presuppose that we can positively address the metaphysical question and focus instead on the epistemological question about ICs. Broadly speaking, this is the question of whether or not we can have knowledge of or, at least, justified, ICs of preference strength. More narrowly, it is the question of whether or not we

[6] See A. Goldman, *Simulation and Interpersonal Utility*, «Ethics», 4, 1995, pp. 709-726.

can have 'scientific' knowledge of, or 'scientifically' justified beliefs about, how different people's preferences compare in terms of strength. This issue is particularly important. Indeed, if we cannot have scientific knowledge of, or scientifically justified beliefs, about how different people's preferences compare in terms of strength, welfare economics, social choice theory, normative and applied ethics remain in trouble even if there is a fact of the matter about ICs of preferences. It is thus important to see why ICs of preference strength raise specific epistemological worries.

As the problem of ICs of preferences often arises in the context of economics, it is useful to start our illustration from this area of analysis. In economics, preferences are defined as binary relations R, that is, relations between two items. The items included in the preference domain vary according to different decision theories.[7] For the purpose at stake, it is not necessary to commit to any specific ontology. In what follows, I shall refer to the object of preference as the option that an individual faces. We can then say that an individual prefers an option *x* to another option *y* and express this in the following way: *x* R *y*. If preferences satisfy certain axioms, they can be represented by a numerical function. More specifically, if they satisfy the ordering axioms, i.e. completeness and transitivity, they can be represented by an ordinal utility function, unique up to a monotone increasing transformation. If preferences satisfy the expected utility axioms, i.e. the ordering axioms, the independence axiom and the Archimedean axiom, they can be represented by an interval utility function, unique up to a positive affine transformation. An interval function is a cardinal function. While an ordinal function is supposed to preserve simply the order of the individual's preferences, a cardinal function is supposed to preserve additional information. More specifically, a cardinal function is supposed to capture the degree to which an individual prefers one option rather than another, or, in other words, the intensity of the individual's preferences for alternative options. Typically, in order to fix a cardinal scale of measurement, it is (necessary and) sufficient to fix two points on the scale, namely, the zero and the unit. One feature of an interval scale of measurement is that both the zero and the unit are 'arbitrary'. Thus, assigning the utility value 0 to one option does not mean that the individual prefers that option with zero intensity. It is simply an arbitrary choice, which fixes one of the relevant points on the cardinal scale of measurement.

Let us now consider an example. Suppose there are two individuals, *i* and *j*, and four options *x*, *y*, *w*, *z* ∈ A. Individual *i* ranks the options in the following way: $x\ R_i\ y\ R_i\ w\ R_i\ z$. On the other hand, individual *j* ranks the options in the following way: $w\ R_j\ z\ R_j\ x\ R_j\ y$. Suppose we represent their preferences on an interval scale. In particular, we assign the value 1 to the most preferred option and the value 0 to the worst option in their preference rankings. We can then assign a value that represents the intensity of their preferences for the other options, where these values are relative to the best and the worst options in each individual's ranking. Suppose it is the case that $u_i(y) = u_j(x) = 0.6$. Suppose it is also the case that $u_i(x) - u_i(y) = u_j(w) - u_j(x) = 0.4$. In the first case, can we conclude that individual *i* prefers option *y* with the same strength with which individual *j* prefers option *x*?[8] In the second case, can we conclude that

[7] More specifically, preferences range over either acts, or propositions, or prospects. For acts, see L. Savage, *The Foundations of Statistics*, Wiley, New York 1954. For propositions, see R. Jeffrey, *The Logic of Decision*, 2nd ed., University of Chicago Press, Chicago 1983 (1965, 1st ed.). For prospects, or lotteries, see J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton 1944.

[8] This is the case of an interpersonal comparison of utility levels. ICs of utility levels are judgments of the form: $u_i(x) \geq u_j(y)$.

the difference in strength of individual *i*'s preference for option *x* over *y* is the same as the difference in strength of individual *j*'s preference for option *w* over *x*?[9]

The answer to both questions is negative: the empirical evidence is not sufficient to conclude that identical utility values represent identical preference strengths. The reason is the following. As we have seen before, in order to fix an interval scale of measurement it is (necessary and) sufficient to fix an arbitrary zero and an arbitrary unit. In the case of preferences, we fix the scale of measurement by assigning the value 0 to the worst option and the value 1 to the best option. However, this is only sufficient to fix the scale of measurement for 'each' individual. It is not sufficient to fix a 'common' scale for both individuals. In order for the preference scale to be a common one, it must be the case that both individuals prefer their best option and their worst option with the same strength. The problem is that the evidence does not tell us anything at all about how different individuals' preferences for their best and worst options compare in terms of strength. As it is typically put, the evidence is consistent with the case where individual *i* prefers the best (the worst) option with intensity ten times greater than *j*.[10]

The previous example shows that ICs of preferences are underdetermined by the empirical evidence. At first sight, this poses a threat to the possibility of having scientifically justified ICs. At least, this seems to be the case if we adopt an evidentialist theory of epistemic justification.[11] The argument is straightforward. If all the possible empirical evidence is insufficient to determine ICs of preference strength, then, if the empirical evidence is what makes ICs of preference strength justified, it follows that ICs cannot be justified. On the other hand, if this is the only source of the problem, the epistemological problem of ICs might as well have a positive solution.[12] One reason is that evidentialism is not the only theory of epistemic justification. According to reliabilism, for instance, what makes a belief justified is not the empirical evidence, but the reliability of the processes by means of which the belief in question is formed.[13] Thus, even if all the possible empirical evidence is insufficient to determine ICs of preference strength, these can nonetheless be justified, provided that they are acquired through reliable processes, that is, through processes that tend to produce true beliefs. The result is that, if we have independent grounds to prefer reliabilism to evidentialism as a theory

[9] This is the case of an interpersonal comparison of utility differences. ICs of utility differences are judgments of the form: $u_i(x) - u_i(y) / u_j(w) - u_j(z) = \lambda$, for some $\lambda \in \Re$.

[10] It is worth noticing that the problem is independent from which empirical evidence one considers. Economists typically take choice behaviour, under conditions of both certainty and uncertainty, to be the only admissible evidence for the ascription of individual preferences. However, as List has recently shown (See C. List, *Are Interpersonal Comparisons of Utility Indeterminate?*, «Erkenntnis», 58, 2003, pp. 229-260), it is not possible to compare different individuals' preference strengths in a meaningful way even if we extend the set of admissible evidence by including the individuals' latency of choice, i.e. the time delay between the presentation of the option and the actual choice-making, the probability of choice, i.e. the probability of choosing one option rather than another the probability of choice (See I. Waldner, *The Empirical Meaningfulness of Interpersonal Utility Comparisons*, «The Journal of Philosophy», 4, 1972, pp. 87-103), their verbal expressions (See J. Harsanyi, *Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility*, «The Journal of Political Economy», 63, 1955, pp. 309-321), their expressive reactions (See R. Weintraub, *Do Utility Comparisons Pose a Problem?*, «Philosophical Studies», 92, 1998, pp. 307-319), their facial expressions, body temperature and other proxies (See C. List, *op. cit*).

[11] For a paradigmatic statement of this position, see R. Feldman and E. Conee, *Evidentialism*, «Philosophical Studies, 48, 1985, pp. 15-34.

[12] See A. Goldman, *Simulation and Interpersonal Utility*, *op. cit*.

[13] For a paradigmatic statement of this position, see A. Goldman, *What is Justified Belief?*, in *Justification and Knowledge*, ed. by G. Pappas, Kluwer Academic Publisher, Reidel 1979, pp. 1-23.

of epistemic justification, we might be able to solve the epistemological problem of ICs, even if ICs are underdetermined by all the possible empirical evidence.

Let us go back to the goal of this paper. I said in the introduction that I am interested in assessing a specific strategy, which attempts to find a solution to the main theoretical problems concerning ICs of preference strength by investigating how ordinary people make ICs of preferences in everyday life. It is now clear that the problem about ICs of preference strength with which I am concerned is the problem of whether or not we can have scientific knowledge of, or scientifically justified beliefs about, how different people's preferences compare in terms of strength. In this paper, I shall adopt reliabilism as my theory of epistemic justification. According to reliabilism, in order for a belief to be justified, it is not necessary that we know that it is reliably acquired. It is only sufficient that, as a matter of fact, such a belief is reliably acquired. However, for other purposes, this may not be enough. For instance, if we want to give ICs of preference strength new scientific legitimacy, it may indeed be important to know whether or not ICs of preferences can be reliably made. The relevant question is thus whether or not there is scientific evidence that ICs of preference strength can be reliably made. As the strategy under consideration specifically focuses on how ordinary people make ICs of preferences in everyday life, this question can be further specified as the question of whether or not there is scientific evidence that ordinary people can reliably make ICs of preference strength in everyday life.[14] This will be the object of my paper.

3. THEORIES OF MINDREADING

This section is divided in two parts. In the first, I shall briefly illustrate how the main theories of mindreading presented in the literature explain the ascription of mental states to a target. This involves considering what ordinary people mean when they use mental terms. In the second part, I shall attempt to clarify how these theories might explain the ascription to a target, and the interpersonal comparison, of preference strengths.

Let us start with the first part. Different disciplines formulate their explanatory accounts of people's mindreading capacity at different levels of description, i.e. the personal, the sub-personal and the physical level. Most of the philosophical literature is concerned with the sub-personal level of description. The common strategy is to conceive the mind as a system, where mental states and processes are characterized functionally. The goal is then to identify the underlying information-processing mechanisms that need to be postulated in order to explain our mindreading capacity. In the course of the years, two main approaches have emerged: Theory Theory (TT) and Simulation Theory (ST). TT characteristically accounts for the mindreading capacity by positing cognitive processes that exploit "an internally represented "knowledge structure" - typically a body of rules or principles or propositions - which serves to guide the execution of the capacity to be explained".[15] In short, TT explains mental ascription

[14] See A. Goldman, *Simulation and Interpersonal Utility*, *op. cit.*

[15] See S. Stich and S. Nichols, *Folk Psychology: Simulation or Tacit Theory?*, «Mind and Language», 7, 1992, pp. 35-36. See also S. Stich and S. Nichols, *Second Thoughts on Simulation*, in *Folk Psychology: The Theory of Mind Debate*, ed. by M. Davies – T. Stone, Blackwell, Oxford 1995, pp. 87-108, S. Stich and S. Nichols, *Cognitive Penetrability, Rationality and Restricted Simulation*, «Mind and Language», 12, 1997, pp. 297-326, S. Nichols *et al*., *Varieties of Off-Line Simulation*, in *Theories of Theories of Mind*, ed. by P. Carruthers – P. Smith, Cambridge University Press, Cambridge 1996, pp. 39-74, and S. Nichols and S. Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford University Press, Oxford 2003.

by arguing that the folks possess a 'Theory of Mind' (ToM), to which they have a more or less conscious access.[16]

As far as the meaning of mental states is concerned, TT is generally associated with analytic functionalism. According to functionalism, the meaning of a mental state is given by the set of causal laws in which that mental state figures. Such causal laws specify how each mental state is related to environmental inputs, other mental states and behavioural outputs. It may be the case that the agent employing mental concepts is incapable of specifying all these constitutive causal relations. Indeed, this may require a sophisticated analysis. If we think of the defining causal relations as part of the ToM that the agent possesses, then we can say that such a theory operates tacitly, or, equivalently, that the theory is tacit. As far as mental ascription is concerned, then, TT assumes that the folks ascribe mental states to other people by observing external events (i.e. inputs and outputs) and inferring the relevant mental states by reference to the causal relations postulated by the ToM that they possess.

ST offers an alternative account of mental ascription. The basic idea is that mental ascription involves individuating another individual's targeted mental states by imagining being subject to the same mental states to which she is subject. The ST approach to mindreading comes in different forms. In this paper, however, I shall primarily focus on the account advocated by Alvin Goldman.[17] According to Goldman, first, the simulator asks herself what mental states she would have if she were subject to the initial mental states of the simulated agent. By so doing, she feeds her own information-processing mechanisms with pretend inputs, which supposedly correspond to the other person's initial mental states. These

[16] There are two variants of the TT approach to mindreading, namely, the scientific-theory theory (STT) and the modularity theory (MT). According to the former, the ToM that the folks use for mindreading is both learnt and stored in the mind in the same way as scientific theories are. In the course of their development, children proceed as little scientists, formulating hypotheses on the basis of the information available and revising them in the light of new data. According to the latter, the ToM is neither learnt nor stored in the same way as scientific theories are, but it is rather included in one or more innate modules. For the current purpose, however, we can ignore the distinction between the two approaches. See H. Wellman, *The Child's Theory of Mind*, MIT Press, Cambridge, Mass. 1990, J. Perner, *Understanding the Representational Mind*, MIT Press, Cambridge, Mass. 1991, J. Gopnik and H. Wellman, *Why the child's theory of mind really* is *a theory of mind*, «Mind and Language», 7, 1992, pp. 145-171, J. Gopnik and H. Wellman, *The theory theory*, in *Mapping the Mind: Domain Specificity in Cognition and Culture*, ed. by L. Hirschfiled – S. Gelman, Cambridge University Press, New York 1994, J. Gopnik and A. N. Meltzoff, *Words, Thoughts and Theories*, MIT Press, Cambridge, Mass. 1997, for a defence of the STT approach. See A. Leslie, *Pretence and representation: The origins of "theory of mind"*, «Psychological Review», 94, 1987, pp. 412-426, A. Leslie, *Some implications of pretense for mechanisms underlying the child's theory of mind*, in *Developing Theories of Minds*, ed. by J. Astington – P. Harris – D. Olson, Cambridge University Press, Cambridge 1988, pp. 19-46, A. Leslie, *Pretending and Believing: Issues in the theory of ToMM*, «Cogniton», 50, 1994, pp. 211-238, A. Leslie, *How to acquire a representational theory of mind*, in *Metarepresentation: A Multidisciplinary Perspective*, ed. by D. Sperber, Oxford University Press, New York 2000, pp. 197-223, A. Leslie and T. German, *Knowledge and ability in "theory of mind": One-eyed overview of a debate*, *Mental Simulation*, in ed. by M. Davies – T. Stone, Blackwell, Oxford 1995, pp. 123-150, and S. Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind*, MIT Press, Cambridge, Mass. 1995, for a defence of the MT approach.

[17] See, in particular, A. Goldman, *Interpretation Psychologized*, «Mind and Language», 4, 1989, pp. 161-185, A. Goldman, *In defense of the simulation theory*, «Mind and Language», 7, 1992, pp. 104-119, A. Goldman, *The mentalizing folk*, in *Metarepresentations*, ed. by D. Sperber, Oxford University Press, Oxford 2000, A. Goldman, *Simulation theory and mental concepts*, in *Simulation and Knowledge of Action*, ed. by J. Dokic – J. Proust, John Benjamins, Amsterdam 2002, pp. 1-20, A. Goldman, *Simulating Minds*, Oxford University Press, Oxford 2006.

mechanisms run 'off-line' and produce pretend outputs. The simulator then introspects these mental states and recognises them as belonging to a certain type. Finally, she ascribes them – by analogy – to the simulated agent.[18]

One crucial feature of Goldman's account is its emphasis on introspection. Indeed, third-personal mental ascription presupposes that the simulator is capable of introspecting her own mental states in order to ascribe mental states to another individual. If this is the case, the ST explanation of mindreading requires an account of ordinary people's introspective capacity. One suggestion is that the simulator recognises her own mental states with respect to the function that they occupy in her mind-system. Goldman gives two reasons to reject this option. The first is that functionalism characterises mental states as dispositions, which may involve relationships with events to which the person has not yet epistemic access (e.g. future events) or will never have epistemic access (e.g. in the case of subjunctive relationships). The second is that, in order for an individual to recognise her own occurrent mental states, functionalism requires that she be able to identify the indefinitely large number of other attitudes to which her occurrent mental states are related as a matter of definition. In turn, this seems to burden self-ascription with excessive computational requirements.[19]

In the light of these problems, Goldman suggests that a more plausible account explains self-ascription in terms of the capacity of internally detecting mental properties that are both categorical and non relational (or at least non-massively relational). There are two candidate types of properties satisfying these criteria: phenomenological and non-phenomenological properties. In his "Simulation and Interpersonal Utility", Goldman opts for the former candidate. Accordingly, the simulator recognises her mental states on the basis of their phenomenology, that is, on the basis of 'what it is like' to have them in specific circumstances. Effectively, this means embracing an experientialist view of the meaning of mental states, that is "the traditional view that mental language gets its meaning, primarily and in the first instance, from episodes of conscious experience of which the agent is more or less directly aware".[20] It follows that the corresponding ST account of mental ascription is associated with a view of mental states as phenomenologically real states, to which the agent has introspective – privileged, although not infallible – access.[21]

We can now examine more closely how the two main mindreading approaches explain the folks capacity to ascribe preferences to other individuals and to compare them in terms of strength. Let us consider TT first. As seen above, TT is typically associated with a functionalist understanding of mental states. Preferences can be defined in functionalist terms as mental states that are causally related to certain inputs, and that, in combination with other mental states, produce certain behavioural outputs. According to some authors, decision theory is the

[18] See A. Goldman, *Interpretation Psychologized*, *op. cit*.

[19] See A. Goldman, *The Psychology of Folk Psychology*, «Behavioral and Brain Sciences», 16, 1993, pp. 15-28, and A. Goldman, *Simulation theory and mental concepts*, *op. cit*.

[20] See A. Goldman, *Simulation and Interpersonal Utility*, *op. cit*, p. 712.

[21] See A. Goldman, *The Psychology of Folk Psychology*, *op. cit*. It is worth noticing that Goldman has recently changed his mind about the meaning of mental states. At present, he defends the view that mental concepts pick out categorical properties of mental states that are non-phenomenal properties. See A. Goldman, *Simulating Minds*, *op. cit*. In what follows, I shall still focus on the version of ST associated with experientialism for two reasons. The first is that this is the most developed version of ST. The second is that Goldman's most recent account of the meaning of mental states is not yet elaborated in sufficient details to provide a basis for an accurate discussion of the problem of ICs of preference strength.

research programme that attempts to specify some of these relevant relations.[22] In particular, decision theory conceives preferences as mental states that lead to choices, in combination with beliefs and desires.[23] Thus, if we define preferences in functionalist terms, the property of preferential strength can be conceived as the causally efficacious property, which leads an individual to behave in a certain way, when subject to specific circumstances and in the presence of other beliefs and desires, in accordance with the causal laws defining the notion of preference.

Suppose now that an observer, e.g. a judge, wants to compare another individual's preferences with her own in terms of strength. The first step concerns third-person mental ascription. The judge observes the relevant external events (i.e. instances of the input-types and output-types that are included in the definition of preference) and infers both the content and the strength of the other individual's relevant preferences, by reference to the causal relations postulated by the ToM that she possesses. The second step concerns first-person mental ascription. Orthodox TT suggests that first-person mental ascription entirely parallels third-person mental ascription. This means that self-ascription is based on inferences mediated by the ToM that the judge possesses. Less orthodox TT approaches relax this position by conjecturing that first-person mental ascription involves the use of recognitional devices or mechanisms – which either make the use of the ToM invisible, but not completely irrelevant, or confine it to certain specific purposes – and ends up with the self-ascription of both a specific content and a specific strength to one's own preferences. The last stage concerns the interpersonal comparison of preferences. The judge has formed a belief about the intensity of the other individual's preferences and a belief about the intensity of her own preferences. Straightforwardly, she can now combine those beliefs to make an interpersonal comparison of preference strengths.

Let us now consider ST. As seen above, Goldman's ST version is associated with an experientialist understanding of mental states.[24] Preferences can be defined in experientialist terms as mental states that give raise to certain experiences in a subject. It may be the case that there is no unique phenomenal experience that different individuals have in common when they are in a preference-state. However, it is enough that there is a family of experiences that are sufficiently similar to constitute a preference-type. According to an experientialist understanding, then, preference strength is a felt property, a qualitative experience of the individual that has preferences. The subject has introspective access and can discriminate the strengths of his preferences. As such, preference strength is a real psychic magnitude, whose

[22] See D. Lewis, *Philosophical Papers*. *Vol. 1*, Oxford University Press, Oxford 1986, P. Pettit, *Decision Theory and Folk Psychology*, in *Foundations of Decision Theory*, ed. by M. Bacharach – S. Hurley, Basil Blackwell, Oxford 1991, pp. 147-175, and P. Pettit, *Preference, Deliberation and Satisfaction*, «Royal Institute of Philosophy Supplement», 81, 2006, pp. 131-154.

[23] Roughly speaking, there are three possible ways to conceive the relationship between desires and preferences. First, one can be eliminativist about preferences and claim that the notion of preferences is syncategoramatic. It is simply a way to conveniently describe an individual's desires and their relations. However, there are no real mental states corresponding to preferences. Second, one can be reductivist and claim that preferences are real mental states but mental states that reduce to desires in one sense or another, e.g. they constitute a specific, e.g. relational, class of desires. Finally, one can maintain that preferences are derivative on desires, in the sense that they are related to, and determined by, them; but they do not reduce to desires, except in the loose sense that they are both pro-attitudes of some sort. I think that the functionalist position fits more comfortably with the latter position, which I shall thereby adopt in what follows.

[24] See particularly the account offered by A. Goldman, *Simulation and Interpersonal Utility*, *op. cit*.

meaning arises "from points or intervals on the experiential scale"[25] that the term denotes, and which the subject experiences and can introspectively discriminate.

Once again, suppose that one individual, e.g. the simulator, wants to compare another individual's preferences with her own in terms of strength. The first step concerns third-person mental ascription. The simulator asks herself which content and intensity her preferences would have if she had the simulated agent's initial mental states. This involves recreating in imagination the same qualitative experiences of the individual whose preference she wants to compare. Then, the simulator discriminates the intensity of these experiences through introspection and classifies them as experiences of preferences with a specific intensity. Lastly, she ascribes such preference strengths to the other agent by analogy. The second step concerns first-person mental ascription. Self-ascription proceeds by direct introspection. The simulator detects her own preferences, discriminates their intensities and ascribes them to herself. Finally, the last stage concerns the interpersonal comparison of preferences. Once again, the simulator has formed a belief about the intensity of the other individual's preferences and a belief about the intensity of her own preferences. She can then combine those beliefs to make an interpersonal comparison of preference strengths.

It is worth noticing that both TT and ST set only minimal conditions for the 'possibility' of forming beliefs about how different people's preferences compare in terms of strength. In both cases, the explanation of how ordinary people make ICs of preference strength is consistent with the possibility that these beliefs are systematically mistaken. Two questions arise. First, what conditions should be satisfied in order for ordinary people to make reliable ICs of preference strength, within a TT and a ST account of their mindreading capacity? Second, can these conditions, at least in principle, be satisfied?

4. CONDITIONS FOR RELIABLE ICS OF PREFERENCE STRENGTH

In this section I want to consider the first of the previous questions. Let us start with TT first. If ordinary people make ICs of preference strength as the TT approach suggests, there are two requirements that must be satisfied for such ICs to be reliably made. First, the judge's inferences about preference strengths must be based on the correct inputs. Second, they must be based on the correct theory about the compared individuals' mind.

The first requirement is straightforward. If the evidence that the judge uses is not correct, then she is likely to reach wrong conclusions about the observed agents' preference strengths. The second requirement holds that the judge's inferences must be based on the correct theory about the compared agents' mind. This means that the causal laws that form the ToM that the judge possesses must correctly represent, or at least very closely approximate, the way in which the relevant information-processing mechanisms of the targeted agents work. For simplicity, let us refer to this requirement as the condition of ToM-to-mind similarity.

Let us now move to ST. Summarising Goldman's own position and a large literature on ST, we can distinguish three requirements that must be satisfied for ICs to be reliably made. First, simulation must be based on the correct inputs. Second, the simulator and the simulated agents must be similar at the level of the relevant information-processing mechanisms. Third, the simulator's relevant information-processing mechanisms must operate in the same way in imagination as in reality.

The rationale underlying the first requirement is identical to the TT case. If the simulator feeds his information-processing mechanisms with incorrect inputs, then she is likely to reach

[25] See A. Goldman, *Simulation and Interpersonal Utility*, *op. cit*, p. 713.

wrong conclusions about the simulated agents' preference strengths. One problem arises. In Goldman's account, the inputs are pretend mental states corresponding to the agent's actual mental states. Thus, in order for simulation to be reliable, it must be the case that the simulator can correctly individuate the intensity of the agent's actual mental states. However, this task presents the same difficulties associated with the comparison of different individuals' preference strengths. Thus, the assumption that the simulator can feed the 'correct' inputs into his 'off-line' system simply begs the question.

In the light of this problem, it seems better to take the simulated agent's environmental circumstances, rather than pretend mental states, as inputs of simulation. The simulator does not begin by asking herself what preferences she would have if she were to have another individual's initial mental states. Instead, she begins by asking herself what preferences she would have if she were in the other individual's initial circumstances. ST must be thus complemented with causal knowledge about the relations between environment and mental states such as beliefs and desires. Moreover, it must be complemented with knowledge about the history of the simulated agent, which should be used to identify which environmental circumstances constitute relevant inputs in specific situations, amongst the infinite ones that the mere observation of the simulated agent's situation allows one to consider. Clearly, this moves ST towards a more hybrid formulation.

Consider now the second requirement, according to which the simulator and the simulated agent must be similar at the level of the relevant information-processing mechanisms. Let us refer to it as the assumption of interpersonal psychological similarity. The rationale of this requirement is intuitive. Even if the simulator feeds her information-processing mechanisms with the correct inputs, she will reach the wrong conclusions about the agent's mental states unless they are psychologically similar in the respects that matter for forming preference strengths.

Finally, consider the third requirement. Even if the assumption of interpersonal psychological similarity is satisfied, so that the simulator forms preferences that are similar to those of the target individual on the basis of 'actual' causal circumstances, it might still be the case that she forms preferences that are radically different from those of the other individual on the basis of 'imagined' causal circumstances. After all, simulation works 'off-line', whereas interpersonal psychological similarity is a thesis about 'on-line' information-processing mechanisms. In order for simulation to be reliable for ICs, the previous requirement must be complemented with the requirement that the off-line working of the individual's mind-system approximates its online working.

5. FIVE ARGUMENTS

In this section I want to examine whether or not the conditions for having reliable ICs of preference strength can be met. There is now considerable evidence that the third requirement for the reliability of simulation can be satisfied.[26] Thus, here, I shall mainly focus on the first and the second requirements for the reliability of both TT and ST. Following and extending Goldman's analysis, I shall consider five arguments, which I shall call, respectively, the argument from mindreading predictive success, the argument from neuroscience, the argument from evolution, the argument from scientific practice and the argument from nativism.

[26] See, for instance, G. Currie and I. Ravenscroft, *Mental Simulation and Motor Imagery*, «Philosophy of Science», 64, 1997, pp. 161-180.

The argument from mindreading predictive success claims that the fact that mindreading is reliable for predictive purposes provides *prima facie* evidence that mindreading is reliable also for the purpose of making IC judgments. The reason is that IC judgments are based on the same mental ascriptions that lead us to reliable behavioural predictions. The objection against this argument is that, even if we grant that mindreading is reliable for predicting an agent's behaviour, we cannot conclude that mindreading is also reliable for making ICs of preference strength. Let us see why by considering TT and ST in more detail.

Consider TT first. The judge ascribes preference strengths to herself and to the target, which best predict their respective behaviour on the basis of the available evidence. The problem is that it is possible for the judge to make both correct behavioural predictions and incorrect ICs of preference strength. For instance, the following is one possible reason. As preferences are characterised in functionalist terms, their intensity is relative to the intensity of other mental states, e.g. other preferences. However, such mental states may never become manifest in overt behaviour. It follows that two individuals' preference strengths may appear identical, from the point of view of all the possible empirical evidence, and yet be different, insofar as they are relative to the intensity of mental states that never generate behavioural outputs. Clearly, the judge can make successful behavioural predictions even if not all of the agent's mental states become manifest in overt behaviour. On the other hand, her mindreading activity may be unreliable for making ICs, because there is no way to show that the inputs that are relevant for ICs are really correct.

Consider ST now. Once again, reliable behavioural predictions are consistent with different and incompatible IC judgments. This means that although it may be true that "empirically observed success at empathy-based predictions of behaviour does go some distance toward supporting psychological isomorphism",[27] it is not true that predictive success goes far enough in showing that such a psychological isomorphism is high enough to lead to correct ICs of preference strength. One problem is that the simulator and the simulated agent may differ with respect to one of the relevant information-processing mechanisms, despite the fact that this difference never becomes manifest at the predictive level. For instance, it may be possible that the simulated agent responds to the environmental inputs by forming desires with intensity ten times greater than the simulator's. If the evidence about the two individuals is perfectly identical, the simulator can make successful predictions of the other agent's behaviour and yet incorrect ICs of preference strength.

To summarise, success at predicting an agent's behaviour requires both a less fine-grained individuation of the relevant inputs and a looser degree of similarity than those required for having reliable ICs of preference strength. At best, predictive success shows that mindreading is reliable for predictive purposes. However, it does not offer a reason to think that mindreading is reliable also for making ICs of preference strength.

The argument from neuroscience claims that, if it is possible to establish well-defined correlations between the intensity of the judge's neural activation during preference ascription and the intensity of the agent's neural activation during preference formation, then it is possible to claim that the mechanisms underlying mindreading are isomorphic to the mechanisms underlying preference formation. This argument specifically fits the ST mindreading approach. Indeed, as simulation employs 'off-line' the same mechanisms that are activated during 'on-line' preference formation, the hypothesised neural correlation should figure amongst the predictions of the theory. If a correlation of that sort can be robustly

[27] See A. Goldman, *Simulation and Interpersonal Utility*, *op. cit*, p. 724.

established, it may thus be possible to vindicate the assumption of interpersonal psychological similarity.

This argument faces some problems. To begin with, although there is evidence that some mental states, e.g. disgust, are located in specific brain regions and, thereby, that different individuals undergoing those states share common neural properties, the same is not true for other mental states, like preferences. Perhaps, this is simply a problem of limited empirical evidence. It might as well be the case that one day scientific research will discover the neural correlates of preferences. If so, the argument from neuroscience may enjoy a good fate. Even if we grant this possibility, however, I believe that the prospects of success are dim. The existence of a common neural region dedicated to preference formation does not imply, *per se*, that an identical neural activation corresponds to the formation of preferences with identical strength across individuals.

Consider the experientialist understanding of preferences associated with ST. If we grant the possibility that the qualitative character of experiences is not fully accounted by their neurophysiological character, it is clear that identical neuronal activation across individuals may correspond to preference experiences that are significantly different at the level of strength. This is the same as admitting that interpersonal isomorphy at the physical level does not necessarily imply interpersonal isomorphy at the subjective level.

Things do not change if we adopt a functionalist understanding of preferences. According to it, preference strengths are individuated not only with respect to external inputs and outputs, but also with respect to other mental states. Crucially, these states can be both occurrent and non-occurrent. The problem is that neural activation registers only occurrent mental states. In order to conclude that different individuals' preference strengths are the same when their neurons fire with the same intensity, we need to assume that they are identical with respect to all those non-occurrent mental states which might impact on their occurrent preference strengths. However, we have no epistemic reason to accept this *ceteris paribus* assumption. Once again, the result is that interpersonal isomorphy at the physical level does not imply interpersonal isomorphy at the functional level.

If the previous points are correct, the hypothesised correlation between the judge's mindreading mechanisms and the agent's preference formation mechanisms does not prove that ICs can be reliably made. The reason is the following. Although the activation of the neural region dedicated to mindreading may co-vary with the activation of the neural region dedicated to preference formation, it may still be the case that the judge does not get the other individual's preference strengths right. In other words, interpersonal neural correlation may be consistent with systematic errors about preference strength attribution. Therefore, the argument from neuroscience does not support the hypothesis that ICs of preference strength can be reliably made.

The argument from evolution claims that evolutionary pressure might have favoured the development of a close isomorphism between the observer's ToM and the target's information-processing mechanisms, in the case of TT, or between the simulator's and the simulated agent's information-processing mechanisms, in the case of ST. The reason is that this would have maximised the expected fitness of the members of a relevant group by endowing them with competitively advantageous features for the typical environment encountered by the group. Roughly speaking, individual fitness is assessed with respect to the (probabilistic) propensity of spreading one's own genes into the next generation. Clearly, this propensity may be enhanced by the individual's capacity to correctly predict, explain or interpret other group members' behaviour. If this is the case, evolutionary pressure might have favoured the development of a strong convergence in the working of different individuals' mindreading mechanisms and, thereby, a high level of intersubjective agreement about the outcomes of

mindreading. The problem is that, as I shall illustrate below, intersubjective agreement presupposes a less demanding degree of similarity than the one required for having reliable IUC judgments. Given that a higher degree of similarity would be unnecessarily costly, it follows that the argument from evolution does not support the claim that ICs of preference strength can be reliably made.

Let us consider TT first. Suppose there are two individuals trying to mindread each other. Suppose that the ToM that each of them uses closely represents the working of the other individual's mind. Finally, suppose that the empirical evidence about their preferences is identical. On the basis of the ToM in their possession, they will both conclude that they have the same preference strengths. However, it may as well be the case that the first individual's occurrent preference strengths are relative to a different set of non-occurrent mental states, which never become manifest. The consequence is that intersubjective agreement is consistent with the case where ICs of preferences are unreliably made. Intersubjective agreement merely requires that the individuals make similar assumptions about a suitable set of non-occurrent mental states, i.e. those that have a chance of becoming manifest in overt behaviour.

Let us now consider ST. Suppose that there are two individuals simulating each other's mental life. Suppose that both individuals are completely identical at the level of the relevant information-processing mechanisms, except for the fact that the first individual forms desires with intensity ten times greater than the second individual's, when responding to the same environmental stimuli. If the relevant evidence is the same, both individuals will conclude that they have identical preference strengths. Indeed, both individuals ascribe preferences to the other on the basis of their own cognitive machinery, under the assumption that the other individual is similar to them in the relevant respects. However, by so doing, they make incorrect ICs of preference strength, because, by stipulation, the intensity of the first individual's desires is ten times greater than the intensity of the second individual's desires. Intersubjective agreement requires a looser degree of interpersonal psychological similarity than the one required for making reliable ICs.

The argument from scientific practice starts from the observations that, on the one hand, most scientific theories are underdetermined by the empirical evidence, like ICs of preference strength, and yet that, on the other hand, there are often reasons to prefer one theory rather than another on grounds of simplicity, parsimony and non-arbitrariness. These are pragmatic virtues that characterise scientific practice and that help break the underdetermination of scientific theories by the empirical evidence. Applied to ICs of preference strength, the argument works in the following way. Suppose that two different judges, in the case of TT, or simulators, in the case of ST, reach intersubjective agreement about other individuals' preference strengths. It is true that preference strengths may be relative to other mental states that never become manifest or that there may be hidden interpersonal differences concerning the relevant information-processing mechanisms, but if all the empirical evidence is otherwise the same, the best explanation of the mindreaders' intersubjective agreement is that the other individuals' preference strengths are really the same. In both cases, the recommended conclusion looks like the simplest, the most parsimonious and the least arbitrary hypothesis.

The objection against this line of thought is that it contrasts with the very explanatory practice in the mindreading literature. In the case of TT, the best explanation of intersubjective agreement in mindreading is that different judges 'take' their theories about other people's mind to be correct, on the one hand, and the *ceteris paribus* assumption about non-occurrent mental states to be satisfied, on the other hand. This does not mean that the ToM in their

possession or the inputs that they consider are really correct. Likewise, in the case of ST, the best explanation of intersubjective agreement is that different simulators 'take' other individuals to be just like them at the level of the relevant information-processing mechanisms. Once again, this does not mean that simulators and simulated agents are really alike. In other words, the mindreading literature explains intersubjective agreement by holding that the judges or the simulators merely 'take' the reliability requirements to be satisfied, without holding that they really are. If this the case, the argument from scientific practice does not show that ICs of preference strength can be reliably made.

The argument from nativism is specifically advocated by Goldman and pursues an analogy with Chomsky's nativist approach in linguistics.[28] Chomsky's analysis starts from the observation that children belonging to the same community end up acquiring the same grammar. This fact is particularly striking because grammar acquisition is radically underdetermined by the empirical evidence. According to Chomsky, it is not plausible to assume that children use purely pragmatic criteria, such as simplicity and parsimony, in order to learn a common grammar amongst the infinitely many possible ones that are consistent with the available empirical evidence. Instead, Chomsky suggests that children possess an innate and universal body of knowledge, which guides them in the process of language learning. Such a body of knowledge is not only important during the development process. Indeed, it is the very body of knowledge on which the grammaticality judgments of adult competent speakers are based.

Goldman invites us to conceive the problem of ICs in analogy with linguistics. The starting point is the observation that different observers reach frequent intersubjective agreement about ICs of preference strength. This fact is particularly striking because, as we have seen, ICs are radically underdetermined by the empirical evidence. As the analogy with linguistics suggests, it is not plausible to assume that different observers form the same beliefs, amongst the infinite ones licensed by the empirical evidence, on the basis of purely pragmatic considerations.[29] Rather, it is more plausible to hold that they form the same beliefs on the basis of the possession an innate and highly representative ToM, in the case of TT, and innate and highly similar information-processing mechanisms, in the case of ST. According to Goldman, if the nativist hypothesis gives linguistics "epistemic respectability", so does it with ICs of preference strength.[30]

The crucial concept is that of 'innateness'. The question of what innateness is has generated a particularly intense philosophical debate in the past few years.[31] Although there is an evident lack of agreement in the literature, the most recent positions suggest taking 'nativism' as equivalent to 'psychological primitivism'.[32] Roughly speaking, innate cognitive capacities are psychological primitives. In turn, psychological primitives are entities or processes that, on the one hand, are mentioned in the best psychological explanations of human behaviour; and, on the other hand, whose acquisition cannot be explained by any psychological theories, but only by a theory at a lower level. The important issue is to see

[28] See N. Chomsky, *Rules and representations*, Basil Blackwell, Oxford 1980.
[29] Cfr. J. Harsanyi, *Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility*, op. cit, and J. Harsanyi, *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press, Cambridge 1977.
[30] See A. Goldman, [1995], pp. 725-726.
[31] See F. Cowie, *What's Within? Nativism Reconsidered*, Oxford University Press, Oxford 1999, P. Griffiths, *What is innateness?*, «Monist», 85, 2002, pp. 70 – 85, R. Samuels, *Nativism in cognitive science*, «Mind & Language», 17, 2002, pp. 233 – 265, M. A. Khalidi, *Innate Cognitive Capacities*, «Mind and Language», 22, 2007, pp. 92-115.
[32] See specially F. Cowie, *op. cit*, and R. Samuels, *op. cit*.

whether or not there is scientific evidence supporting the nativist hypothesis in the case of ICs of preference strength.

The first objection is that, even if such evidence exists, it does not show that ICs of preference strength can be reliably made. After all, nativism is a hypothesis about the acquisition of the mindreading capacity rather than a hypothesis about its reliability. In order to turn the argument from nativism into an argument of the right kind, we should reformulate it as claiming that the reliability of the relevant information-processing mechanisms constitutes a 'primitive' in a theory of how ordinary people make ICs of preference strength, in the sense that the best explanation of their comparative capacity and of their frequent intersubjective agreement requires such property. In the light of the previous discussion, however, we can easily reject even this alternative version of Goldman's argument. As we have seen above, the best explanation of people's comparative practice and intersubjective agreement merely requires the assumption that ordinary people 'take' the reliability requirements to be satisfied, not they really are. Therefore, even the argument from nativism fails.

6. CONCLUSION

Solving the problem of ICs of preference strength is of vital importance for welfare economics, social choice theory and ethics. It is therefore not surprising that several attempts have been made in the course of the years. In this paper, I examined one strategy in particular, which tries to solve the problem of ICs by looking more closely at how ordinary people make ICs of preference strength in everyday life. The question that I specifically considered is whether or not there is scientific evidence that ordinary people make ICs of preference strength in a reliable way. I discussed five arguments, which offer various reasons to think that the answer to this question is affirmative. In this paper, I argued that they all fail. This does not mean that no positive argument exists, which can vindicate the strategy under scrutiny or, more generally, solve the problem of ICs. However, given that all the solutions explored so far turn out to be unsuccessful, this paper provides *prima facie* reason to think that scepticism about ICs may be here to stay.

BIBLIOGRAPHY

Arrow K. (1995) *Social Choice and Individual Values*, 2nd ed., Wiley, New York 1963 (1951, 1st ed.).

Baron-Cohen, S. *Mindblindness: An Essay on Autism and Theory of Mind*, MIT Press, Cambridge, Mass., for a defence of the MT approach.

Chomsky N. (2002) *Rules and representations*, Basil Blackwell, Oxford 1980.

Cowie, F. *What's Within? Nativism Reconsidered*, Oxford University Press, Oxford 1999, P. Griffiths, *What is innateness?*, *Monist,* 85, pp. 70 – 85.

Currie G. and Ravenscroft I. (1997) Mental Simulation and Motor Imagery, *Philosophy of Science*, 64, pp. 161-180.

Davidson D. (1986) Judging interpersonal interests, in *Foundations of social choice theory*, ed. by J. Elster – A. Hylland, Cambridge University Press, Cambridge, pp. 195-211.

Davidson D. (2004) *Problems of Rationality*, Oxford University Press, Oxford.

Feldman R. and Conee E. (1985) *Evidentialism*, *Philosophical Studies*, 48, 1985, pp. 15-34.

Goldman A. (1979) *What is Justified Belief?*, in *Justification and Knowledge*, ed. by G. Pappas, Kluwer Academic Publisher, Reidel, pp. 1-23.

Goldman A. (1989) Interpretation Psychologized, *Mind and Language*, 4, pp. 161-185.

Goldman A. (1992) In defense of the simulation theory, *Mind and Language*, 7, pp. 104-119.

Goldman A. (1993) The Psychology of Folk Psychology, *Behavioral and Brain Sciences*, 16, pp. 15-28.

Goldman A. (1995) *Simulation and Interpersonal Utility*, *Ethics*, 4, 1995, pp. 709-726.

Goldman A. (2000) *The mentalizing folk*, in *Metarepresentations*, ed. by D. Sperber, Oxford University Press, Oxford.

Goldman A. (2002) Simulation theory and mental concepts, in Simulation and Knowledge of Action, ed. by J. Dokic – J. Proust, John Benjamins, Amsterdam, pp. 1-20.

Goldman A. (2006) *Simulating Minds*, Oxford University Press, Oxford.

Gopnik,J. and Wellman H. (1992) *Why the child's theory of mind really* is *a theory of mind*, Mind and Language, 7, pp. 145-171.

Gopnik J. and Wellman H.(1994) *The theory theory*, in *Mapping the Mind: Domain Specificity in Cognition and Culture*, ed. by L. Hirschfiled – S. Gelman, Cambridge University Press, New York.

Gopnik J. and Meltzoff A. N. (1997) *Words, Thoughts and Theories*, MIT Press, Cambridge, Mass.

Harsanyi J. (1955) Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, *The Journal of Political Economy*, 63, pp. 309-321.

Harsanyi J. (1977) *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press, Cambridge.

Jeffrey R. (1983) *The Logic of Decision*, 2nd ed., University of Chicago Press, Chicago (1965, 1st ed.).

Jevons S. (1911) *Theory of Political Economy*, 4th edition, Macmillan, London (1871, 1st ed.).

Khalidi, M. A. (2007) Innate Cognitive Capacities, *Mind and Language*, 22, 2007, pp. 92-115.

Leslie A. (1987) Pretence and representation: The origins of "theory of mind", *Psychological Review*, 94, 1987, pp. 412-426.

Leslie A. (1988) *Some implications of pretense for mechanisms underlying the child's theory of mind*, in *Developing Theories of Minds*, ed. by J. Astington – P. Harris – D. Olson, Cambridge University Press, Cambridge, pp. 19-46.

Leslie A. (1994) Pretending and Believing: Issues in the theory of ToMM, *Cogniton*, 50, , pp. 211-238.

Leslie A. (2000) How to acquire a representational theory of mind, in *Metarepresentation: A Multidisciplinary Perspective*, ed. by D. Sperber, Oxford University Press, New York, pp. 197-223.

Leslie A. and German T. (1995) *Knowledge and ability in "theory of mind": One-eyed overview of a debate*, *Mental Simulation*, in ed. by M. Davies – T. Stone, Blackwell, Oxford, pp. 123-150.

Lewis D. (1986) *Philosophical Papers*. *Vol. 1*, Oxford University Press, Oxford.

List L. (2003) Are Interpersonal Comparisons of Utility Indeterminate?, *Erkenntnis*, 58, pp. 229-260.

Little I.D.M. (1957) *A Critique of Welfare Economics*, 2nd ed., Clarendon Press, Oxford (1950, 1st ed.).

Nichols S. *et al*. (1996) *Varieties of Off-Line Simulation*, in *Theories of Theories of Mind*, ed. by P. Carruthers – P. Smith, Cambridge University Press, Cambridge, pp. 39-74.

Nichols, S. and Stich, S. (2003) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford University Press, Oxford.

Perner J. (1991) *Understanding the Representational Mind*, MIT Press, Cambridge, Mass.

Pettit P. (1991) *Decision Theory and Folk Psychology*, in *Foundations of Decision Theory*, ed. by M. Bacharach – S. Hurley, Basil Blackwell, Oxford, pp. 147-175.

Pettit P. (2006) Preference, Deliberation and Satisfaction, Royal *Institute of Philosophy Supplement*, 81, 2006, pp. 131-154.

Robbins L. (1932) *An Essay on the Nature and Significance of Economic Science*, Macmillan, London.

Samuels R. (2002) Nativism in cognitive science, *Mind & Language*, 17, pp. 233 – 265.

Savage L.(1954) *The Foundations of Statistics*, Wiley, New York.

Stich S. and Nichols S. (1992) Folk Psychology: Simulation or Tacit Theory?, *Mind and Language*, 7, pp. 35-36.

Stich S. and Nichols S. (1995) *Second Thoughts on Simulation*, in *Folk Psychology: The Theory of Mind Debate*, ed. by M. Davies – T. Stone, Blackwell, Oxford, pp. 87-108.

Stich S. and Nichols S. (1997) Cognitive Penetrability, Rationality and Restricted Simulation, *Mind and Language*, 12, 1997, pp. 297-326.

von Neumann J. and Morgenstern O. (1944) *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.

Waldner I. (1972) *The Empirical Meaningfulness of Interpersonal Utility Comparisons*, The Journal of Philosophy, 4, pp. 87-103.

Weintraub R. (1998) Do Utility Comparisons Pose a Problem?, *Philosophical Studies*, 92, pp. 307-319.

Wellman H. (1990) *The Child's Theory of Mind*, MIT Press, Cambridge, Mass.

Does Freedom of Choice cause Satisfaction?

Margherita Bottero[*^]
Margherita.Bottero@hhs.se

ABSTRACT

This paper investigates the relation between freedom of choice and individual well being, as measured by satisfaction. I look at analyse the hypothesis of a positive correlation between the two variables, controlling for other relevant covariates and unobserved heterogeneity. The results show that freedom is significantly correlated with satisfaction, a finding robust to several alternative model specifications.

INTRODUCTION

Well being has recently encountered renewed interest as a policy objective, motivating the flourishing of a new branch of economics, known as economics of happiness. Part of this literature is focused on uncovering the determinants of well being, often understood as satisfaction over one's life or feeling of happiness. By now, however, virtually no study has investigated the relatively intuitive question of whether freedom matters for individual well-being. If one understands freedom as freedom of choice[1] (FoC), the existing theory endorses a positive answer (Pattainx, Xu, 1990, Sen, 1999 and Sugden, 1998). Providing empirical evidence in support of this claim encounters a twofold problem. First, it is not yet clear how to correctly measure freedom, since there is no unique definition of it. Second, freedom is most likely to be endogenous with other measures of well being, and the relation is widely blurred by the presence of unobserved heterogeneity. Does more freedom cause higher satisfaction or is being satisfied with one's life the cause of higher self reported freedom? To what extent are unobservable ethics or value systems affecting the relationship under study?

In this paper I will mostly concentrate on the problem of unobservables. As for the theoretical part, I will be measuring FoC as Autonomy Freedom (Sugden, 1998, Bavetta and Peragine, 2006, Bavetta and Guala, 2003). Autonomy Freedom is a metric of freedom with a formal characterization (Bavetta and Peragine, 2006), maintaining that the essence of freedom is having opportunities to choose from in a conscious way. More importantly in this context, data on Autonomy Freedom are collected in a publicly available dataset[2] (for a justification, see Bavetta et al., 2008). Some preliminary research (Bavetta et al., 2008) reports that FoC is correlated with several important socio-demographic variables, as income, employment, marital status and religion. This evidence is suggestive that FoC may be relevant for social studies, further motivating the interest in how it relates to individual well being.

Well being is a broad concept with a wide array of definitions and measures. In what follows, I will measure it as satisfaction. From a philosophical perspective, in fact, being free to choose induces an individual to be "the master of one's own life" (Mill, 1859), by allowing him or her to take the choices more appropriate to his or her lifestyle. Empirically, this means that

---

[*] Department of Economics Stockholm School of Economics
[^] The author thanks the Wallander/Hedelius Foundation for financial support.
[1] Where freedom of choice means the extent to which an individual has opportunities allowing him to take meaningful choices.
[2] The World Values Survey and the European Values Survey.

FoC should be expected to impact measures of individual satisfaction with one's own life, where satisfaction comes from the appraisal of one's conscious choices (Bavetta and Guala, 1998).

Guided by the above considerations, I will investigate the hypothesis that FoC relates positively with the satisfaction that an individual feels with his or her life. While I do not address the problem of endogeneity, I try to deal with the unobserved heterogeneity issue by controlling for fixed effects. The results show that there is a strong, positive correlation between the FoC and satisfaction, holding across all specifications, with and without fixed effect controls.

The remainder of the paper is organized as follows: the next section quickly reviews the previous research. After that, two sections discuss the dataset and the empirical strategy. Section 5 provides the results and the following one comments the empirical evidence. A concluding section follows.

PREVIOUS RESEARCH

The existing literature is relevant for the present work in two respects. First, they will guide the choice of the relevant variables to be included in the regression of FoC on satisfaction. Second, they will provide a benchmark against which to test the consistency of the estimates.

Almost all studies find that income increases with satisfaction (Schnys, 1998, Frey and Stutzer 2002), and a similar correlation holds for being employed, especially self-employed (Oswald, 1997), having a higher education and living in a big city. Satisfaction is also strongly positively related with being married and being male, while it is still debatable whether it increases or not with having children (Frey and Stutzer, 2006). Several works show that satisfaction is a convex function of age, with a peak down around the mid-forties. Moreover, being religious is a determinant of satisfaction (Clark and Lelkes, 2005). For a more comprehensive review of the studies, see Bjornskov et al. (2008). To my knowledge, no study includes a measure of FoC among the covariates.

Establishing a clear axiomatic measure for FoC has been on the agenda of Social Choice theory for many years. Pattanaix and Xu (1990) suggested that FoC should be measured by the number of options available to an individual. Alternatively, Sen proposed to look at the number of options available to and preferred by the decision maker (Sen, 1991).

In this work, I will measure FoC as Autonomy Freedom, a conceptualization due to Sugden (1998) and Bavetta et al. (1997, 2003). They maintain that a metric of freedom should capture the extent to which a choice set allows the decision maker to come to choose an option x rather than to what extent it gives access to (ex post preferred) options, without any requirement on the decision process (Mill, 1898, Bavetta and Guala, 1997, 2003). There are two reasons why I favour this notion. It is an interesting policy variable because it captures the value of having opportunities, which recently has been central for many political goals[3] and it is possible to work with it empirically by using the following question from the European Value survey[4].

[3] See acceptance speech delivered by president-elect Barak Obama, http://www.marketwatch.com/news/story/text-barack-obamas-acceptance-speech/story.aspx?guid=%7B523A921D-6E5F-4103-BA81-23A1ACE29EBE%7D

[4] In the original work, the authors refer to the World Value Survey, but the question is unchanged in the more handy European Value Survey.

Some people feel they have completely free choice and control over their lives, while other people feel that what they do has no real effect on what happens to them. Please use this scale where 1 means "none at all" and 10 means "a great deal" to indicate how much freedom of choice and control you feel you have over the way your life turns out (for a justification, Bavetta et al., 2008). Moreover, preliminary work shows that this measure has some statistical significance in relation with other socio-demographic variables (Bavetta et al., 2008), but the issue of its relation with satisfaction is not studied there[5].

THE DATA

The data that I am using are taken from the worldwide integrated European Values Survey[6], a survey that gathers information on how people perceive themselves and the society around them, similar in spirit to Eurobarometer[7] and the ISSP survey[8]. The sample has a cross section dimension consisting of more than 250,000 individuals from roughly 60 countries worldwide, and a time series of four waves, 1981-1984, 1989-1992, 1994-1997, and 1999-2003.

To measure well being, I will be using data on satisfaction, as gauged by the answer to the question *All things considered, how satisfied are you with your life as a whole these days?*

Respondents provided an answer ranging from 1 to 10, with 1 meaning total dissatisfaction and 10 complete satisfaction. Apart from FoC, I will be using ten other covariates, whose descriptive statistics are displayed in Figure 1.

Covariate	Obs.	Mean	Std. Dev.	Variance	Smallest	Largest
SAT	263097	6.61	2.48	6.18	1	10
FoC	248224	6.64	2.44	5.97	1	10
INC	228938	4.67	2.47	6.13	1	11
EMPL	259689	3.18	2.15	4.64	1	8
TOWN	181216	4.88	2.51	6.32	1	9
EDU	187668	4.44	2.27	5.19	1	8
CHILD	244024	1.93	1.70	2.90	0	20
M.STATUS	263038	2.62	2.17	4.73	1	7
BIRTH	264839	1953	17.46	304.85	1881	1988
SEX	267660	1.5	.49	.24	1	2
REL_0	238328	1.3	.56	.31	1	4
REL_1	204721	56.2	12.81	164.16	1	86
POL	193531	5.56	2.26	5.13	1	10

Fig. 1. Descriptive Statistics

Most of them are self explanatory. Married is a dummy taking value 1 if single and 2 if the respondent is married. Sex is coded with 1 being 'Male' and 2 'Female'. REL_0 stands for the question *Independently of whether you go to church or not, would you say you are a religious person?,* with possible answers: 1 'A religious person', 2 'Not a religious person', 3 'A convinced

[5] The authors regress FoC on a set of covariates (sex, age, marital status, education level, income and employment status, political orientation, trust level, individual's support in market-oriented institutions; religious beliefs), but without accounting for fixed effects.
[6] http://www.europeanvalues.nl/.
[7] http://ec.europa.eu/public_opinion/.
[8] http://www.issp.org/.

atheist' and 4 'Other answer'. Finally, POL corresponds to the question: *In political matters, people talk of "the left" and "the right." How would you place your views on this scale, generally speaking?,* whose answers are coded with 1 meaning 'Left' and 10 'Right'.

THE EMPIRICAL STRATEGY

One way to estimate empirically the relation between FoC and Satisfaction is to try to control as much as possible for the large unobservable heterogeneity lying in the background. More specifically, it is likely that some of the unobservable variables are connected with the effect of being in a certain country (country effect) and being interviewed in a certain wave (time effect). Consider for instance the participants in the wave 1989-1992: the effect produced on their reported answers by the fall of the Berlin wall is per se unobservable, but it is possible to capture it by including time effects. Similarly, country effects account for those ethics or values that are specific to a certain area and affect the relation between FoC and Satisfaction in ways that are intrinsically unobservable.

Given the panel nature of the dataset that I am using, which tracks a cross section of countries across time, the natural technique to employ is to estimate a model with fixed effects. I will investigate the partial effect of the observable independent variables described in Figure (1) on Satisfaction assuming the following model

$$E[SAT \mid \mathbf{x}, \mathbf{c}]$$
$$SAT_{it} = \mathbf{x}_{it}\beta + \mathbf{c}_i + u_{it} \qquad (1)$$

where x_{it} is a vector containing the observable covariates[9], c_i a vector of unobservable random variables and u_{it} a idiosyncratic error term, where i refers to the country and t to the wave.

According to the theory, to consistently estimate models like (1), the explanatory variables x_{it} need to be uncorrelated with the error term, although only once the c_i's are accounted for[10]. Moreover, since the panel data I am using is unbalanced, in the sense that some countries drop off the cross section in some periods, consistency requires also the x_{it} to be uncorrelated with the drop-off pattern.

Formally, the strict exogeneity assumption reads

$$E(u_{it} \mid \mathbf{x}_i, \mathbf{s}_i, c_i) = 0 \Leftrightarrow E(s_{it}, \mathbf{x}_{is}' u_{it}) = 0 \quad s,t = 1,2...,T \qquad (2)$$

where s_{it} is selection indicator vector $\mathbf{s}_i \equiv (s_{i1}, s_{i2}, ..., s_{iT})'$, such that $s_{it}=1$ if $(SAT_{it}, \mathbf{x}_{it})$ obtained, zero otherwise.

Intuitively, this assumption is likely to be satisfied. The independence of the covariates with the error term should be sufficiently warranted by the conditioning on the fixed effects, and the correlation with the attrition pattern should be zero, given that drop-offs in the dataset are intrinsically random, being due to technical constraints in the survey design.

I will estimate model (1) by running four different regressions, whose results are displayed in the next section. First, I will estimate a baseline regression, which captures the relation between Satisfaction and FoC without accounting for unobservable heterogeneity:

[9] FoC, Income, Size of Town, Education, Employment, Marital status, number of Children, Religious personor not, Politcal self positioning, Sex, Year of Birth.
[10] strict exogeneity assumption (see Wooldridge, p. 579).

$$SAT_{it}=\beta_{0it}+FoC\beta_F+\mathbf{x}_{it}\boldsymbol{\beta}_x+du_{it} \quad (3)$$

Then I will add a country fixed effect dC_i

$$SAT_{it}=\beta_{0it}+FoC\beta_F+\mathbf{x}_{it}\boldsymbol{\beta}_x+\mathbf{dC}_i\boldsymbol{\beta}_C+du_{it} \quad (4)$$

where $\mathbf{dC}_i$ is a vector of country dummies, so that $\mathbf{dC}_i$=1 if i=C and zero otherwise. The third model specification accounts for time effects $\mathbf{dT}_t$ in a similar way

$$SAT_{it}=\beta_{0it}+FoC\beta_F+\mathbf{x}_{it}\boldsymbol{\beta}_x+\mathbf{dT}_t\boldsymbol{\beta}_t+du_{it} \quad (5)$$

where $\mathbf{dT}_t$ is a vector of time dummies such that ds_t=1 if t=T and t=1,2,3,4 and zero otherwise. Finally, I have a complete specification including country fixed effects, time fixed effects and country specific time effects, where the latter captures all the unobservables related with being in a certain country in a certain wave

$$SAT_{it}=\beta_{0it}+FoC\beta_F+\mathbf{x}_{it}\boldsymbol{\beta}_x+\mathbf{dC}_i\boldsymbol{\beta}_C+\mathbf{dT}_t\boldsymbol{\beta}_t+\mathbf{d}(\mathbf{C}_i*\mathbf{T}_t)\boldsymbol{\beta}_{it}+du_{it} \quad (6)$$

where obviously $\mathbf{d}(\mathbf{C}_i*\mathbf{T}_t)$ is a dummy such that $\mathbf{d}(\mathbf{C}_i*\mathbf{T}_t)$=1 if i=C and t=T and zero otherwise.

This procedure recovers the partial effect of FoC on the dependent variable, and the estimates will be consistent as long as the exogeneity assumption is satisfied.

RESULTS

The results from regression (3)-(6) are displayed in the following Tables. As the theoretical research on FoC suggests, the estimates imply a large, positive and statistically significant association between FoC and Satisfaction. For each specification, the estimated coefficient varies a little in size according to the estimation technique used (OLS, Probit, Probit with survey option). However, for each technique, estimates do not vary across the different specifications, remaining significant, of the same sign and of about the same size.

Estimation results for (3) are shown in Table (1). As mentioned above, the magnitude of the relative effect of FoC drops from 0.34 to 0.2 when (3) is estimated with a multinomial probit estimation technique rather than with pooled OLS. This was expected, since OLS treats the regressors as if they were continuous variables, which inflates the resulting estimates.

	POOLED OLS	PROBIT	PROBIT SVY
FoC	0.340**	0.200**	0.198**
	(73.93)	(80.88)	(15.59)
INC	0.122**	0.059**	0.060**
	(31.39)	(31.36)	(5.94)
EMPL	-0.007	-0.006**	-0.007
	(1.66)	(2.90)	(1.38)
TOWN	0.002	0.011**	0.012
	(0.45)	(6.05)	(1.19)
EDU	0.034**	0.005*	0.006
	(7.68)	(2.38)	(0.72)
CHILD	-0.005	0.021**	0.020
	(0.69)	(6.50)	(1.71)

M.STATUS	-0.047**	-0.007**	-0.008
	(10.00)	(3.10)	(1.31)
BIRTH	-0.000	-0.001**	-0.001
	(0.43)	(4.37)	(0.79)
SEX	0.108**	0.056**	0.057**
	(6.37)	(6.52)	(2.93)
REL_0	-0.211**	-0.049**	-0.045
	(10.44)	(5.16)	(1.54)
REL_1	-0.004**	0.012**	0.011**
	(3.42)	(28.28)	(3.43)
POL	0.066**	0.041**	0.040**
	(15.56)	(19.39)	(6.20)
CONS	4.290**		
	(3.41)		
Obs	59898	59898	59898
# country	62		
Robust z			
* sign.5%; ** sign. 1%			

Table 1. Baseline regression

Table 2 displays results from model (4). Adding country fixed effects does not alter the estimates that I found in the baseline model. The relation between FoC and Satisfaction remains present, with the same positive sign and basically the same magnitude, with the estimated coefficient varying between 0.17 and 0.34. The coefficients on the country dummies are not all zero and some are significant at 5%, but the standard errors are quite large, suggesting evidence of multicollinearity and warning of overemphasizing the results.

	OLS POOLED	PROBIT	PROBIT SVY
FoC	0.338**	0.175**	0.174**
	(18.22)	(14.86)	(15.06)
INC	0.121**	0.058**	0.056**
	(9.34)	(9.72)	(9.19)
EMPL	-0.007	-0.000	0.000
	(0.85)	(0.03)	(0.06)
TOWN	0.002	-0.001	0.000
	(0.21)	(0.24)	(0.02)
EDU	0.034**	0.014**	0.013*
	(3.79)	(2.81)	(2.61)
CHILD	-0.005	-0.002	-0.003
	(0.53)	(0.53)	(0.57)
M.STATUS	-0.048**	-0.025**	-0.026**
	(6.02)	(5.89)	(6.05)
BIRTH	-0.000	-0.001	-0.001
	(0.13)	(1.21)	(0.85)

SEX	0.107**	0.057**	0.058**
	(3.89)	(4.33)	(4.64)
REL_0	-0.213**	-0.116**	-0.113**
	(4.52)	(5.28)	(5.16)
REL_1	-0.005*	-0.003*	-0.002*
	(2.19)	(2.21)	(2.23)
POL	0.066**	0.035**	0.033**
	(5.85)	(6.27)	(5.82)
Constant	4.082		
	(1.68)		
Obs	59898	59898	59898
# country	62		
R-squared	0.168		
t stat **sign.1%, *sign. 5%			

Table 2. Country Fixed Effects

The same considerations follow by looking at the results obtained accounting for time effects rather than country effects, as can be seen from Table 3. Again, the association between FoC and Satisfaction remains positive and significant, between 0.19 and 0.4. Dummies for wave 1 and 3 are automatically dropped for collinearity reasons, and the results show that only the second wave has had a significant positive impact on Satisfaction. This may be explained by the fact that the second wave is the one carried out right after the fall of the Berlin wall. However, standard errors are large, so the interpretation is not clear.

	POOLED OLS	PROBIT	PROBIT_SVY
FoC	0.419**	0.200**	0.198**
	(92.08)	(15.51)	(15.71)
INC	0.142**	0.061**	0.062**
	(36.19)	(6.57)	(6.58)
EMPL	-0.018**	-0.005	-0.006
	(4.06)	(1.13)	(1.30)
TOWN	0.022**	0.009	0.010
	(5.84)	(0.92)	(1.03)
EDU	0.012**	0.004	0.005
	(2.72)	(0.48)	(0.61)
CHILD	0.034**	0.018	0.017
	(5.11)	(1.48)	(1.43)
M.STATUS	-0.009	-0.005	-0.005
	(1.76)	(0.84)	(0.95)
BIRTH	-0.001*	-0.001	-0.001
	(2.07)	(0.65)	(0.62)
SEX	0.119**	0.058**	0.060**
	(6.54)	(2.79)	(3.04)
REL_0	-0.064**	-0.043	-0.038
	(3.10)	(1.46)	(1.27)

REL_1	0.028**	0.013**	0.013**
	(31.58)	(4.63)	(4.55)
POL	0.086**	0.042**	0.041**
	(19.59)	(6.31)	(6.00)
Constant	3.449**		
	(2.74)		
Obs	59898	59898	59898
Number of wave	3		
R-squared	0.237		
Rob t stat * sign 5%; ** sign 1%			

Table 3. Time Fixed Effects

Finally, model specification (6) does not give rise to significant alterations of the estimation results found in the baseline model. Here the marginal effect of FoC ranges between 0.17 and 0.33.

	POOLED OLS	PROBIT	PR SVY (W)	PR SVY (C)
FoC	0.336**	0.174**	0.173**	0.173**
	(18.07)	(14.76)	(14.96)	(14.96)
INC	0.127**	0.060**	0.059**	0.059**
	(9.66)	(9.98)	(9.42)	(9.42)
EMPL	-0.004	0.001	0.001	0.001
	(0.60)	(0.24)	(0.32)	(0.32)
TOWN	-0.006	-0.004	-0.003	-0.003
	(0.82)	(1.32)	(0.95)	(0.95)
EDU	0.031**	0.012**	0.012*	0.012*
	(3.59)	(2.63)	(2.53)	(2.53)
CHILD	-0.005	-0.002	-0.003	-0.003
	(0.54)	(0.52)	(0.60)	(0.60)
M.STATUS	-0.048**	-0.025**	-0.026**	-0.026**
	(6.00)	(5.84)	(6.04)	(6.04)
BIRTH	0.000	-0.001	-0.000	-0.000
	(0.15)	(1.03)	(0.75)	(0.75)
SEX	0.100**	0.054**	0.056**	0.056**
	(3.55)	(3.97)	(4.33)	(4.33)
REL_0	-0.217**	-0.117**	-0.114**	-0.114**
	(5.12)	(6.04)	(5.88)	(5.88)
REL_1	-0.004	-0.002*	-0.002*	-0.002*
	(1.92)	(2.00)	(2.04)	(2.04)
POL	0.065**	0.034**	0.033**	0.033**
	(5.85)	(6.22)	(5.73)	(5.73)
Cons	3.304			
	(1.52)			
Obs	59898	59898	59898	59898

# country	62
R-squared	0.174
Robust t * sign at 5%; ** sign at 1%	

Table 4. Complete

With regard to the other covariates, the estimated partial effects are in line with the rest of the literature. Income is always significant and positively related to satisfaction. Employment status displays a positive correlation as well[11]. The coefficients on education, marital status, religion and sex are significant and positive across all specifications, as one would have expected. I do not find any effect for the number of children but this covariate is generally considered controversial. I cannot find a significant correlation for age, but this may not be surprising given that I impose a linear specification. Surprisingly, political orientation is highly significant across all specifications, with a positive correlation between being right-winged and being satisfied with one's life.

DISCUSSION

This section discusses some critical issues that may have arisen with the empirical methodology, the data and the interpretation of the results.

The methodology that I am using boils down to investigating the effect of FoC on Satisfaction controlling for how much heterogeneity as possible. This impinges on the implicit assumption that satisfaction levels are interpersonally comparable[12]. While this may be in principle questionable, the assumption is commonly accepted in the economics of happiness literature (Bjornskov et alii, 2008), and even has some psycho-neurological foundations.

With regard to the data, it is important to warrant that there are no systematically missing observations, arising from the respondents' inability or unwillingness to answer some questionnaire items. If missing observations follow a pattern related to some unobservable variable, then the regression estimates will be biased and inconsistent. However, by looking at the dataset the problem does not seem a pervasive one: less than 5,000 respondents did not report an answer to the question on satisfaction, over more than 250,000 interviewed.

The choice of fixed effects as a modelling technique is rather natural to deal with part of the heterogeneity problem. However, it can be the case that the fixed effect dummies will take up part of the effect on satisfaction that instead is to be attributed to other covariates, most importantly to FoC. Unfortunately, there is no way to solve this issue, which loses relevance considering that the estimated coefficient for FoC is always sizeable and significant.

For fixed effects models, the consistency of the results depends crucially on the exogeneity of the covariates with respect to the error term, although allowing positive correlation between the explanatory variables and unobservables. As mentioned in Section 4, assumption (2) is quite likely to hold if we look contemporaneously at the covariates and the error term, being this random in a given time period. Zero correlation between lagged errors and covariates may be trickier to defend: some respondents may systematically report lower FoC in t in face of a negative shock in t-1. However, strict exogeneity requires zero correlation once the c_i's have been accounted for: controlling for the unobservables is likely to wash out this kind of feedbacks and make assumption (2) likely to be met.

[11] Employment is coded such that higher values correspond to increasing levels of unemployment.
[12] Note that this does not imply any assumption about cardinality of the categories.

Considering the actual estimates, there are few general comments to be made. First, it may be difficult to extend the results to samples with different characteristics. The fixed effects control for unobservables but these remain a sort of black box. Also, the effects of the long term levels of the observed variables will no longer be captured in the results, which will rather provide an estimate of the effects of variations in the observed variables.

Looking at the output, the standard errors more than double in the specifications including the country dummies, (4) and (6). This evidence is suggestive that there is multicollinearity among the country dummies. To control for this possibility, I compute the Variance Inflation Factor (VIF) of the variables entering specification (2) and (4)[13]. The VIF shows for each variable how much its variance is inflated by multicollinearity: as long as the VIF is less than 10 there is no need to worry about the growing standard errors. Although it is possible to compute the VIF only for the OLS estimations, the results clearly show that when adding the country dummies the average VIF raises sharply from 1.15 to more than 7. Standard errors, however, take overall reasonable values; thus the pattern discovered above, although calling for caution, does not weaken the main results.

Finally, it is important to point out that while this analysis is suggestive of a causal relation running from FoC to Satisfaction, it is by no means evidence for it. The empirical intuition in favour of causality comes from the robustness of the estimates through the different specification (3)-(6). The preliminary evidence on a positive and significant effect of FoC on Satisfaction comes from the baseline regression, which does not control for any unobservable effect. Adding the fixed effects does not affect the estimates of the coefficient, which keep having the same sign and significance. To further test the robustness of this result, for each model specification I run a pooled OLS estimation, a multinomial probit and a multinomial probit that accounts for the survey-nature of data. While the results again remain unchanged, the causality interpretation again remains only a possibility. Reverse causality from Satisfaction to FoC is an objection difficult to counter at this stage, and I cannot rule out the hypothesis of an external, unaccounted variable that drives both dependent and independent variables. Yet, if one accepts these caveats, the results support the notion that FoC matters for well being, suggesting that further research in this area may be worth pursuing.

CONCLUSIONS

This paper assesses the importance of the relationship between freedom of choice and individual satisfaction, as a measure of well being. I investigate the hypothesis that Freedom of Choice has a positive impact on satisfaction, as it is implied by theoretical studies in the Social Choice literature. The results show that freedom indeed correlates positively with satisfaction, although given the few caveats discussed above. What the results do not show is the direction of causality in the relationship under study, which remains an open issue for future research.

BIBLIOGRAPHY

Bavetta, S. and V. Peragine (2006) Measuring autonomy freedom. *Social Choice and Welfare*, 26:31-45.

[13] VIF is the reciprocal of the tolerance of a variable, which is the proportion of the variance for the variable in question that is not due to other explanatory variables: a tolerance close to 0 indicates risk of multicollinearity.

Bavetta, S. and F. Guala (2003) Autonomy freedom and deliberation. *Journal of Theoretical Politics*, 15:423443.

Bavetta, S., M. Bottero, D. Maimone Ansaldo Patti and P. Navarra (2008) *Better than Free: Why Autonomy Favors Freedom, Success and Self-Esteem*, Mimeo.

Bjørnskov, C., A. Dreher and J. A.V. Fischer (2006) Cross-Country Determinants of Life Satisfaction: Exploring Different Determinants across Groups in Society, *Working papers 06-145*, KOF Swiss Economic Institute, ETH Zurich.

Clark, A. and O. Lelkes (2005) Deliver us from evil: religion as insurance. *PER Working Paper 06/03*, European Center for Social Welfare Policy and Research.

Frey, B. and A. Stutzer (2002) *Happiness and Economics: How the Economy and Institutions A ect Well-being*, Princeton University Press, Princeton.

Frey, Bruno and Alois Stutzer (2006) Mispredicting Utility and the Political Process, in Edward J. McCaffery and Joel Slemrod (eds) *Behavioral Public Finance*. New York: Russell Sage Foundation, pp. 113-140.

Mill, J.S. (1859) *On Liberty*, John W. Parker and Son, Oxford University Press, 1991.

Oswald, A.J. (1997) Happiness and Economic Performance, *The Economic Journal* 107, 1815-1831.

Pattanaik, P. and Y. Xu (1990) On Ranking Opportunity Sets in Terms of Freedom of Choice. *Recherches Economiques de Louvain*, 44:173-98.

Schyns, P. (1998) Crossnational differences in happiness: economic and cultural factors explored. *Social Indicator Research* 43, 3-26.

Sen, A. K. (1988) Freedom of choice: Concept and content. *European Economics Review* 32:269-294.

Sen, A. K. (1991) Welfare, Preference and Freedom. *Journal of Econometrics* 50:15-29.

Sen, A.K. (1999) *Development as Freedom*, Oxford: Oxford University Press.

Sugden, R. (1998) The metric of opportunity. *Econ Philos* 14:307-337.

Sugden, R. (2004) The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences, *AER* 94, 4: 1014-1033.

Wooldridge, J.M. (2001) *Econometric analysis of cross section and panel data*, MIT Press, Cambridge Massachusset Introduction.

Understanding Marx's Normative Critique of Capitalism: Reification revisited[+]

Volkan Çıdam[*]
volkancidam@hotmail.com

ABSTRACT

Marxist traditions of thought barely pay any attention to the normative dimension of Marx's critique of capitalism in *Capital*. In this respect Lukács' seminal essay on reification is an exception. Situating Marx within German Idealism, Lukács defines the fundamental normative concern of Marx's writings as an investigation of the conditions of radical freedom. Given its central role in Marx's critique, it is necessary to find an answer to the question of how to conceptualize freedom. Lukács finds that answer in a reinterpretation of the introductory chapters of *Capital*, which seemingly explores only the economic logic of capitalism. Although this is a fruitful theoretical move, I argue that Lukács' conceptualization of Marx's understanding of freedom, implicit in his concept of reification, is problematic. Still, Lukács' interpretive approach provides conceptual resources to revise the concept of reification with a view to develop a better understanding of Marx's normative critique.

It has always been controversial to look for Marx's normative critique of capitalism in his Magnum Opus *Capital*. For such a reading challenges the scientific-image of Marx, who has analyzed the laws of development and dynamics of capitalism, and foreseen its necessary demise. This image of a positivist scientist[1] was firmly established by Engels and his both social-democratic and Leninist followers, who built their case on Marx's seemingly 'objectivist' language in *Capital*. This particular reading of Marx became the dominant discourse in Marxist debates. The term, 'orthodox Marxism' has generally been used to denote this discourse, even though it is difficult to give an exact definition of the term. This is due to the fact that within the tradition of Marxism 'orthodox Marxism' has often been utilized as the most convenient ideological weapon to discredit the opposing fraction. Thus despite their radically different interpretations both Lukács in *Geschichte und Klassenbewußtsein* (London: Red Star Press, 2000) [1923] and Althusser and Balibar, at least implicitly, in their *Reading Capital* (London: Verso, 1979) [1968] declare their own reading of

---

[+] I would like to thank Çidem Çıdam for her very helpful comments, suggestions, and corrections. My work on this paper owes a great deal to conversations with her. Also, I am grateful to Marco Solinas for his encouragement. Finally I would like to thank Deniz Ercan for smiling at the world with his big bright eyes.

[*] Phd Candidate in Philosophy, Johann Wolfgang Goethe-Universität

[1] With 'positivist Image' I don't refer to any particular school of thought within Marxism. Any Marxist approach that adopts the opinion that social sciences are best to be modeled after 'natural sciences' and search for an unanimous method is however essentially positivist. The classical positivist approach can be found in Friedrich Engels, "Dialectics of Nature" in *Marx/Engels Collected Works*, (New York: International Publishers, 2005), Volume 25, p. 313-590 and "Socialism Utopian and Scientific" in *Marx/Engels Selected Works*, (Moscow: Progress Publishers, 1970), Volume 3, p. 95-151. Most prominent Among those who read (accuse) Marx as a positivist are G.A Cohen, *Karl Marx's Theory of History: A Defence.* (New Jersey: Princeton Univ. Press, 1978), and Albrecht Wellmer, *Critical Theory of Society*, (New York: The Continuum Publ. Corp, 1971). For a critical overview of positivism-debate and a successful refutation of this claim see, James Farr, "Marx and positivism" in Ball / Farr (ed), *After Marx*, p. 217-235 (Cambridge, Cambridge University Press, 1984)

Marx as orthodox. In this paper, I call both socialist (and eventually social-democratic) movement of the Second-international and Leninism as orthodox positions. Both socialists of the Second International and Leninists have claimed authority on the *true* interpretation of Marx's work and even though their view on how to overcome capitalism differs considerably, they share some basic interpretative assumptions that draw on Engels' writings rather than that of Marx.

The problem with orthodox Marxism is twofold. On the one hand, because of its misguided understanding of science that condemns any normative judgment as ideological, orthodox Marxism tends to cover up the normative foundations upon which its own project of emancipation is based – as a result, its interpretative approach to *Capital* lacks any normative interest. On the other hand, since the total absence of normative arguments is impossible within a *critical* – let alone a *revolutionary* – theory of society, orthodox Marxism finds it necessary to import a critical normative perspective to its analysis. This perspective is either borrowed from the earlier writings of 'humanist' Marx or, at its worst, is drawn out from a common-sense moral intuition, which has been criticized by Marx in all his writings. Thus, for instance, the ideal-typical orthodox-Marxist legitimates his/her political agenda with the following rhetorical question: Is it not evident that the capitalist form of exploitation is 'unjust' and 'inhuman'? Marx's own answer to this question in *Capital* however would not exactly serve to orthodox Marxist's interests. For, according to Marx, capitalist form of exploitation is both *just*[2] and, as a historical phenomenon, all too human.

The structuralist reading of Althusser can be best understood as a critic of this double strain in orthodox Marxism; scientism, on the one hand, humanism or common sense morality on the other. And yet, Althusser's strategy of setting a radical break between the 'humanist' early writings and the scientific later writings of Marx, perhaps inadvertently, strengthened the discourse of orthodox Marxism. This is the case despite the fact that Althusser's understanding of science significantly differs from that of orthodox Marxism.[3]

In many ways, it is possible to suggest that an earlier attempt to challenge the ascendancy of an amoral, "scientific" reading of *Capital* had proved to be more successful than Althusser's approach because, unlike Althusser's, this approach dealt with the problem of orthodox Marxism directly. According to this view, the normative foundations of Marx's critic of capitalism had to be uncovered and this could only be done, if we read *Capital* with

[2] See the Chapter on "Working Day", where he writes on the antinomy of 'rights' in Volume I (Karl Marx, *Das Kapital* in Marx/Engels Werke, (Berlin: Dietz Verlag) Volume 23, p. 209) and his definition of justice in Volume III (Karl Marx, *Das Kapital* in Marx/Engels Werke (hereafter MEW) (Berlin: Dietz Verlag) Volume 25, p. 351). In order to explain an oxymoron such as a ‚just exploitation' that occupies a central place within Marx' critic of capitalism, it is necessary to unearth the normative foundations of the critique. For a plausible account that argues against this necessity see, Allen Wood, *Karl Marx*, (Oxon: Routledge, 1981). According to Wood it is possible to differentiate between 'moral' and 'amoral' goods. He suggests that the term 'exploitation' refers to a state that is detrimental to a natural, amoral good, to well-being. Thus he concludes, just exploitation is not an oxymoron at all. It seems however that Wood here presents a very Aristotelian argument, in order to define ‚amoral' goods and thus points to a normative foundation without naming it.

[3] See Althusser, *Reading Capital*. According to Althusser, Marx is the initiator of a *new* science and methodology and not a positivist scientist who analyses the eternal laws of society. A brief look at the neo-Althusserian texts would show, however, the similar normative conclusions drawn by this line of thought and orthodox Marxism. See Jan Hoff (among others ed.), *Das Kapital neu lesen*, (Münster: Westfälisches Dampfboot, 2006).

different philosophical lenses than those of Engels.[4] In order to substantiate this approach Lukács developed his concept of reification[5] that primarily rests on the interpretation of Marx's critique of the capitalist mode of social mediation[6] presented in the opening chapters of *Capital*. Lukács convincingly argued that any attempt to understand Marx's critique of capitalism in *Capital*, which disregards his philosophical stance within the tradition of German Idealism, is inadequate. Read within the context of Marxist debates, Lukács' analysis of reification stands out not only by its philosophical rigor and cogency, but also by its unique attempt to search for a normative standard within *Capital* with a view to highlight Marx's normative critique of capitalism.[7]

In this paper, I argue that although Lukács was principally right in his interpretive approach to *Capital*, he was unsuccessful in capturing the essence of Marx's normative critique of capitalism. It is due to this failure that his account of revolutionary politics, which is based on the analysis of reification, succumbs to a politics of macro-subject – a form of politics that has been the object of legitimate concern and has raised doubts about the critical potential of the concept of reification.[8] Although Lukács' particular account of Marx's relation to German idealism is responsible for this outcome, I suggest that a certain Hegelian reading of *Capital*, which Lukács occasionally seems to pursue, can not only help us to save the revolutionary core of Lukács' analysis of reification but also contribute to a better understanding of Marx's normative critique of capitalism. Thus, in what follows, I propose a rereading of Marx's opening arguments in *Capital* to rethink reification and to bring to light its critical potential.

[4] Non-Marxist writers discovered very early that Marx' analyses of political economy has a normative critical core to it. Georg Simmel's analysis in *Philosophie des Geldes* (Georg Simmel, *Philosophy des Geldes* (Frankfurt am Main: Suhrkamp, 1989) [1900]) can be interpreted as a polemic against Marx (See below). Similarly, it would not be far-fetched to claim that Max Weber's renowned analysis of Protestant Ethics takes it inspiration directly from *Capital*. The first analysis of *Capital*, aiming to uncover the normative foundations of the Work belongs to Franz Petry, *Der Soziale Gehalt der Marxchen Werttheorie* (Jena: G. Fischer, 1916).

[5] Marx coins the term 'reification', *Verdinglichung*, in Volume III of *Capital* in reference to his analysis of the ‚fetish charakter of commodity'. In this context the term literally denotes the conversion of social relations to a relation between things. More accurately social relations acquire a reified, *thingly*, quality, while at the same time the relations between things, commodities, take the form of a social relation. As Marx calls it: «It is an enchanted, perverted (verkehrte), topsy-turvy world, in which Monsieur le Capital and Madame la Terre do their ghost-walking as social characters and at the same time directly as mere things.» (MEW 25, p. 838). As I will demonstrate below Lukács takes this term and develops his theory of reification in his seminal essay "Reification and the Consciousnes of the Proletariat" (Georg Lukács, "Die Verdinglichung und das Bewusstsein des Proletariats" in *Geschichte und Klassenbewusstsein*).

[6] By 'social mediation' I refer to a form of social relation that plays a constitutive role for subject positions. A constituted subject position denotes a particular attitude towards other subjects and the outside world. This attitude in return depends upon a particular conceptualization of freedom.

[7] Only recently innovative attempts have been made to unearth the foundations of Marx' normative critic of capitalism in *Capital* that go beyond the analysis of Lukács. See, Georg Lohmann, *Indifferenz und Gesellschaft* (Frankfurt am Main: Suhrkamp, 1991) and Moishe Postone, *Time Labour and Social Domination* (Cambridge, Camobridge University Press, 1993).

[8] Habermas is among the most eminent thinkers that pointed to this problem. See Habermas, *Der philosophische Diskurs der Moderne,* (Frankfurt am Main: Suhrkamp, 1988), p. 95-103 and Habermas, *Theorie des Kommunikativen Handelns Bd.2*, (Frankfurt am Main: Suhrkamp, 1981), Chapters 6 and 8.

This reading draws parallels between the unfolding of Marx's critique and Hegel's discussion of dialectically unfolding knowledge models in his *Phenomenology of Spirit*.[9] My argument is that when read through the lenses of *Phenomenology*, the opening chapters of *Capital* do provide a criterion for the normative critique of capitalist society. This criterion, I claim, is to be found in the dialectic of recognition: for Hegel, as well as for Marx, the dialectic of recognition serves as a model for the intersubjective mediation that is necessary for the subsistence of a true, i.e. free subjectivity. Thus, like Lukács', my reading of *Capital* attributes neither an amoral, positivist understanding of scientific stance to Marx's critique of capitalism, nor searches for external normative standards in the earlier writings of Marx that will supposedly make sense of the objectivist analysis of his late work. I also share Lukács' view that Marx's normative critique of capitalism in *Capital* targets primarily the unfreedom specific to capitalist society and that in Marx's account the conception of freedom plays a fundamental role in defining other (moral and ethical) values such as justice, equality and a conception of good life.[10] I part ways with Lukács, however, on the issue of how to conceptualize the understanding of freedom that is to serve as the normative standard /criterion of Marx's critique of Capitalism. According to my account, to fully grasp Marx's understanding of freedom, it is essential to rethink the concept of reification through the Hegelian dialectics of recognition.

To elaborate these points, in the first part of this paper I turn to two interpretative accounts of Lukács' concept of reification that underline Lukács' indebtedness to Hegelian philosophy. Herbert Marcuse and Axel Honneth, both members of Frankfurt School, find it necessary to have recourse to the concept of reification in search for a normative standard that enables a substantial critique of modern capitalist society on philosophical grounds. My own interpretation of Lucács draws on the insights provided by these thinkers. In the second

[9] This Hegelian reading is not to be confused with a form of Hegelianism particularly prevalent in Germany that reads *Capital* in its relation to Hegel's *Science of Logic*, which takes its origins from what can be called as 'Value-Form-Analysis School' of Hans-Georg Bachhaus, *Dialektik der Wertform. Untersuchungen zur Marxchen Ökonomiekritik* [1974] (Freiburg: Ça Ira, 1997); Helmut Reichelt, *Zur logischen Struktur des Kapitalbegriffs bei Karl Marx* (Frankfurt am Main: Europ. Verl.-Anst., 1970) and Michail Heinrich, *Die Wissenschaft vom Wert. Die Marxche Kritik der politischen Ökonomie zwischen wissenschaftlicher Revolution und klassischer Tradition.* (Münster: Westfälisches Dampfboot, 1991). Despite very valuable contributions to the question of late Marx' methodological considerations in their relation to Hegel's *Science of Logic*, the focus of these works remain to be a better understanding of the **scientific** critique of political economy by Marx. In other words the underlying assumption is absolutely in line with orthodox Marxism's (especially with Leninism's) claim that late Marx use of dialectical method of presentation in *Capital* is yet another proof of his critique of capitalism on purely scientific / amoral grounds. For recent examples of this approach see Leo Èeìerko, "Das 'wirkliche Triebwerk des Kapitals' und seine Beziehung zu Hegels Logik" in Hoff (ed.), *Das Kapital neu Lesen* and Važjulin, Viktor A., "Entwicklung systematisch denken. Ein Vergleich der dialektischen Logik bei Hegel und Marx" in: *Deutsche Zeitschrift für Philosophie*, 53:2, S. 203-18, (2005).

[10] A recent debate between Nancy Fraser and Axel Honneth can be taken as the proof of how fundamentally important the issue at stake here is for contemporary emancipatory politics. See Nancy Fraser / Axel Honneth, *Umverteilung oder Anerkennung? Eine politisch-philosophische Kontroverse*. (Frankfurt am Main: Suhrkamp, 2003) While Fraser opts for a liberal/social democratic double-track politics of redistribution *and* recognition in order to maintain a more *equal* terms of political participation, Honneth in his more Marxian approach calls for a rethinking of the conception of *freedom* in modern capitalist societies, which would only than make emancipatory politics possible. Then as for Marx also for Honneth the ideal dialectic of recognition presents a mode of social mediation that can serve as an orientation for contemporary political actors.

part of the paper, I give a brief analysis of Lukács' concept of reification presented in his seminal essay "Reification and the Consciousness of Proletariat". Here I demonstrate that Lukács' text carries within it the conceptual resources for an alternative understanding of reification. In my concluding remarks, drawing on these resources, I provide a brief account of what a Hegelian reading of the introductory chapters of *Capital* might look like. Such an account is crucial because only a Hegelian reading of this kind can both save the revolutionary core of Lukács' analysis of reification and contribute to a better understanding of Marx's normative critique of capitalism.

I

Writing right after the Bolshevik Revolution, Lukács, in his famous essay "Reification and the Consciousness of Proletariat", challenges the philosophical foundations of the economic determinist Marxism of the Second International. While making philosophical arguments to lead revolutionary Marxism away from Engels' scientific dialectical materialism, Lukács also tries *not* to counter the latter simply by putting forward a naïve anthropological-humanist interpretation of Marx's writings. In other words, he does not want to offer a theory of alienation that is inevitably based on an essentialist understanding of human nature. Lukács develops his concept of reification mainly to address these concerns.

As I will demonstrate below, Lukács ends up failing to accomplish this goal due to the philosophical choices he makes. Lukács' failure, however, is productive because, even though his answers to them are inadequate, he gives a central status to two interdependent and crucial questions in his analysis: First, what is the relation of Marx's philosophy to German Idealism in general? And second, what is Marx's position with respect to the central ethical pursuit of this tradition, namely the investigation of the conditions of existence of radical individual freedom? In many ways, it is possible to suggest that Western Marxism in general, and Critical Theory in particular has emerged and developed in response to these questions formulated by Lukács. Because Lukács' answer to these questions is to be found in his conception of reification, a new assessment of this concept has attained a central role in the self-understanding of Marxist traditions of thought.[11]

It is crucial to note that those who directly engage with the concept of reification, in one way or another turn to Hegel's philosophy. Thus, for instance, Herbert Marcuse in his reconstruction of the *Phenomenology of Spirit* uses the concept to characterize the first three sections of Hegel's work.[12] According to Marcuse's interpretation, all truth claims made by the succeeding models of knowledge in these introductory sections fail on the same grounds: All these models claim to secure their presumable knowledge independently of any historical definite social practice.[13] Therefore, Marcuse argues, «Hegel attempts to show that man can know the truth only if he breaks through his 'reified' world. We borrow the term 'reification' from the Marxist theory, where it denotes the fact that all relations between men in the world of capitalism appear as relations between things […]. The commodity, for instance, embodies in all its qualities the *social* relations of labor […]» (Ibid, p. 112). Hereby Marcuse sets the concept of reification rightly in relation to Hegel's *Phenomenology* as well as to

[11] See the classical work of Martin Jay, *Marxism and Totality. The Adventures of a concept from Lukács to Habermas* (Cambridge: Polity Press, 1984).

[12] Herbert Marcuse, *Reason and Revolution,* (1941) [Oxford: Routledge & Kegan Paul, 1986].

[13] See also the more recent study of Terry Pinkard, who builds on this insight of Marcuse. Terry Pinkard, *Hegel's Phenomenology. The Sociality of Reason*, (Cambridge: Cambridge University Press 1994).

Marx's analyses of commodity fetishism. A few pages later, however, commenting on Hegel's master and slave Dialectic, Marcuse equates the position taken by Marx in his 1844 Manuscripts with the later critique formulated in the first chapter of Capitalism, with commodity fetishism or in Lukács' words with reification: «In 1844 Marx sharpened the basic concepts of his own theory through a critical analysis of Hegel's *Phenomenology of Mind*. He described the 'alienation' of labor in the terms of Hegel's discussion of master and servant […] Marx makes reference to Hegel's definitive insight, which disclosed to him that lordship and bondage result of necessity from certain relations of labor, which are in turn, relationships in a 'reified' world» (Ibid. 115). This conflation of two distinct concepts, that is, of 'alienation' and 'reification' by Marcuse points to an inner tension in Lukács' own analyses.

When he wrote his essay, Lukács did not yet know about the existence of 1844 Manuscripts of Marx. And yet, as I argue below, in his account of revolutionary overcoming of capitalism, building up on *Thesis on Feuerbach* and Johann Gottlieb Fichte's philosophy, Lukács reached to similar conclusions with young Marx's theory of alienation. According to Lukács, the 'proletarian standpoint' is, like the standpoint of the 'slave' in Hegel's account (presumably, for Hegel does not argue exactly in this line), inclined to freedom. For, through his work – in Lukács case also political work – slave builds /educates himself to be the *true* subject of history.[14]

With this line of argument Lukács privileges, as noted by Jürgen Habermas,[15] a certain productivist mode of action over others, and in doing so, he comes very close to a position occupied by 'humanist' criticisms of capitalism, which attach a central importance to the role played by labor / productive action in human flourishing. Humanist critics of capitalism argue that modern (capitalist) division of labor and work processes is responsible for human being's *alienation* from (their own) nature. Because Lukács is aware of the ahistorical and essentialist nature of such a critical account, he tries to disavow this position through lengthy discussions in his essay. Despite all his efforts, however, as we will see, he cannot avoid to fall back into this perspective.[16] Nevertheless, this tension in his essay does not prevent Lukács from trying to establish his own theory of *reification* in opposition to a theory of *alienation*.

By pointing to the different accent put on the normative grounding of the Marxian critique of capitalism in a theory of *alienation* in contrast to a theory of *reification*, a difference, which, as we have seen, was deliberately evaded by Marcuse, I do not want to suggest that there is 'a radical break' between 'young Marx' and 'mature Marx.' As with every thinker, however, it is possible to observe gradual shifts between different works of a given author.[17] Here it suffices to note that the central role attached to 'alienated labor' as the normative ground for the critique of capitalism in *1844 Manuscripts*, leaves its place to a critique of a historically determined form of a *social mediation* in *Capital*, informed by the Hegelian theory of intersubjectivity. Although I am critical of Marcuse's erasure of the difference between a theory of alienation and a theory reification in his reconstruction of Marx's work in relation to *Phenomenology of the Spirit,* I share the fundamental premise of his interpretation: Reification is best understood as a conceptual innovation that shows and

[14] For a detailed account of Lukács' usage of Hegelian Master and Slave Dialectic in his essay see Andrew Arato, "Lukács' Theory of Reification" in *Telos 11*, Spring 1972, p. 58f.

[15] Habermas, *Ibidem*.

[16] For the discussion of Humanism See Lukács, *ibid.*, p. 148f and 208f. Especially the explanation of the privileged status of the 'standpoint of Proletariat' (ascribed class-consciousness of proletariat) is by contrast made from the perspective of a theory of alienation. See 175f.

[17] By far the best account of the positions that Marx takes in his earlier writings is given by Daniel Brudney, *Marx' attempt to leave Philosophy*, (Cambridge: Harvard University Press, 1988).

highlights the indebtedness of Marx's analysis of commodity fetishism to the opening sections of *Phenomenology*, specifically, I suggest, to the chapter on 'Force and Understanding'.

Recently, Axel Honneth also revisited Lukács' concept in his work. Unlike traditional interpretations, in his book entitled "Reification",[18] Honneth refuses to analyze the concept in terms of its relation to German Idealism, in order not to hazard the consequences of the metaphysical assumptions that then have to be made. Thus, he successfully avoids falling back to a productivist model of subjectivity that frames Lukács' problematic discussion of the 'ascribed class-consciousness' of proletariat, which has burdened the Marxist theory and praxis ever since. As a result, instead of conjuring Lukács' eschatological imagery of the macro- subject and object of history that will overcome the state of reification once again, Honneth focuses on social phenomena that helped Lukács to characterize the object of his criticism from a sociological and action-theoretical perspective. Honneth argues that even stripped from its idealist presumptions, on a purely sociological level, Lukács' concept goes beyond the current utilizations of similar concepts, such as Nussbaum's 'objectification', which either describe certain types of human behavior that violate ethical principles or point to a form of instrumental rationality that colonizes the emotional life of the actor.[19] The radical core of Lukács' argument, Honneth suggests, rests on the observation that in modern societies actors are compelled to take a contemplative attitude towards the reality, which they constitute, and in which they participate and interact. In this reading, reification refers to a distorted human praxis, which is to be contrasted with an intersubjective attitude on the part of the subject, and as such, can be reinterpreted as losing sight of a primordial form of relation to the world, as "forgetfulness of recognition".

Although I share Honneth's view that the concept of recognition constitutes the normative horizon of a non-reified social world, in what follows, I refuse to adopt his sociological and action-theoretical approach that sets, in Honneth's own words, a more 'existential mode' of recognition as the standard of critique of a reified world.[20] Instead, I argue that Lukács' own philosophical assumptions are adequate and sufficient to understand the concept of reification within the framework of a theory of recognition. For the philosophical assumptions in Lukács' analysis of commodity fetishism provide us with the conceptual resources to develop an alternative reading, which unlike Lukács' own account, sets the Hegelian dialectic of recognition as the standard of the critique of capitalism. When interpreted in this manner, the concept of reification continues to provide a powerful critique of a mode of social mediation *specific to* capitalist society, as intended by Lukács, rather than referring to a distorted (alienated) human praxis.

II

In order to utilize these insights provided by Marcuse and Honneth to develop an alternative interpretation of Lukács' concept of reification, it is first necessary to give an account of his renowned essay, "Reification and the Consciousness of Proletariat". Let me begin that account with a brief discussion on the argumentative structure of the text. Drawing on Marx's method outlined in his "Introduction to (Grundrisse)",[21] Lukács divides his essay in

[18] Axel Honneth, *Verdinglichung*, (Frankfurt am Main: Suhrkamp, 2005).
[19] Martha Nussbaum, "Objectification„ in Alan Soble (ed.) *The Philosophy of Sex* (Maryland: Rowman & Littlefield, 2002).
[20] See Axel Honneth, *Ibid*. p. 46.
[21] See Karl Marx, MEW 42, p. 19-45.

three major parts. According to Marx, a scientific analysis begins with a naïve pre-conceptual perception of an empirical whole (his example is 'population'). From this perception of a chaotic gathering of facts one has to move analytically to simple abstract concepts. A scientific theory of society brings these abstract concepts together so as to reconstruct a conceptually informed concrete totality. A conscious use of processes of thinking, abstraction and reconstruction are essential for the scientific method. For, only if scientific effort in these processes can be raised to consciousness, the theory can become immune to ahistorical generalizations.

Following this line of thinking, in the first part of his essay, following Marx and drawing on Max Weber and Georg Simmel, Lukács provides a sociological analysis of everyday life within capitalist society to give a picture of the totality that is to be overcome through revolution. This sociological analysis sets the ground for a conceptual, that is, *philosophical* discussion on the foundations of the diverse phenomena brought under the term *reification*. Hence, in the second part of his essay Lukács' discussion is firmly situated within the tradition of German Idealism. Here Lukács puts forward his radical claim and introduces the central discussion of the subject-theoretical conditions of the existence of individual freedom: «Modern critical philosophy springs from the reified structure of consciousness.» (Lukács, Ibid., p. 122) Finally, in the third and final part of his essay, following the scientific method outlined in *Grundrisse,* Lukács draws direct political consequences from his philosophical analysis and attempts to construct an account of a *true* totality. Thus, he argues that the reified world of capitalism can only be dissolved and *appropriated*, if the proletariat sets itself as the subject of socio-historical reality instead of simply being the product of capitalist social order. The moment of revolution is the moment, in which 'the reified structure of consciousness' gives way to Self-consciousness.

Just like its argumentative structure, the substantive content of "Reification and the Consciousness of Proletariat" is also indebted to Marx's thinking. Hence, Lukács begins his essay with a brief analysis of Marx's commodity fetishism. He correctly points out that the term commodity for Marx is not only, or even primarily, an economic category. Instead, Lukács suggest, 'commodity' refers to a social mediation specific and central to capitalism. He, then, sets out to elucidate Marx's somewhat obscure criticism of this social mediation by focusing on the discussion on 'commodity fetishism,' which even today presents the key to diverse interpretations of *Capital*.

It is of theoretical significance to underline that one of the major goals of Lukács' account of commodity fetishism is to challenge the orthodox Marxist view, which interprets commodity fetishism simply as an inverted state of social relations. According to the orthodox Marxist view, social relations between human beings *appear to be* relations between things. Thus, it is argued, this inverted state of affairs, which conceals the true relationship between human beings, has to be corrected to give way to a genuine and conscious, a *decommodified*, relationship between human beings that defy a social relation mediated by commodities. This interpretation, however, is a simplification of Marx's argument and consequently misleading. For, Marx's critique based on the fetish character of commodity is twofold. On the one hand, Marx criticizes the political economists, who present a historical and social relationship as a natural order, for falling victim to the ideology emanating from the commodity form. On the other hand, Marx directs his main critique towards a form of social mediation that sets the subjects of this relationship as atomistic individuals. The existence of two worlds, that is the world of human beings and the world of commodities, which are inversions of each other, is a precondition that holds this

problematic social mediation together. According to Marx, a two-world perspective is symptomatic of such an individuation process that undermines the freedom of the subject fundamentally. In other words, Marx does not directly criticize the *fact* that social mediation takes place behind the back of the producers *via* commodities – for, this is yet another *symptom* of the fundamental unfreedom that emanates from the individuation process –, but the very form of this social mediation.[22]

Building on this more complex understanding of commodity fetishism Lukács defines *reification* as a specific problem of modern capitalism in contrast to *alienation*, which claims validity for a much broader historical period and is usually understood in terms of a fall from some golden age. Instead, Lukács uses the concept of reification to give an expression to the effects of commodity fetishism in all spheres of capitalist society. Thus, looking for conceptual means to explain a broader range of phenomena in modern everyday life, Lukács brings Max Weber and Georg Simmel into the discussion.

The principle of "formal equality", on which Lukács builds his theory of reification, is an abstraction from Marx's commodity-form analysis. The concept of "formal equality" not only helps Lukács to identify an essential feature of the social mediation specific to bourgeois society but also enables him to incorporate the rationalization theory of Weber into his critique of capitalism. Through the increasing rationalization that is peculiar to modern world, Lukács argues, instrumental-rationality is institutionalized in every sphere of life. Parallel to the commodity producing abstract labor, which generates a formal equality between the products so that they can be exchanged, instrumental rationality, too, sets everything into a relationship of formal equality. In order to function smoothly, modern society relies on instrumental rationality to reduce everything to a certain form of objectivity, which can, then, be *calculated*. Therefore, Lukács concludes, not only in the assembly line, but virtually in every sphere of life, where, in Habermas' words, instrumental rationality colonizes the life-world, the individuals are subjected to an objective scheme of rationalized activity.

The institutionalization of instrumental rationality compels individuals to a *contemplative stance*, because, to use Lukács' own words, 'the activity of the agent becomes less and less active' as the agent acts simply in conformity with the given laws of the particular sphere in which he finds himself. The notion of contemplative stance is the key to understand Lukács' concept of reification. To explain this point it is necessary to quote Lukács at some length:

> The contemplative stance adopted towards a process mechanically conforming to fixed laws and enacted independently of man's consciousness and impervious to human intervention, i.e. a perfectly closed system, must likewise transform the basic categories of man's immediate attitude to the world: it reduces space and time to a common denominator and degrades time to the dimension of space […] Thus time sheds its qualitative, variable, flowing nature; it freezes into an exactly delimited, quantifiable continuum filled with quantifiable 'things'; it becomes space. (Lukács, Ibid., p. 101)

Already with this thesis the conceptual transition to the philosophical discussion of reification is provided for, which will situate the problem within German Idealism. For here, without trying to conceal it, Lukács introduces the idea that the modern capitalist rationalization distorts even 'time', that is, to put it in Kantian terminology, 'the subjective condition under

[22] Hence, Marx does not propose a political strategy against the ills of 'commodification' as proposed by critics inspired by the orthodox marxist interpretation, but rather demands the overcoming of this form of individuation process through the constitution of an intersubjective one.

which all our intuitions take place'.[23] In other words, depriving the transcendental Subject of its inward intuition, capitalist society reverses the achievement of Kant's "Copernican Revolution" in philosophy, which gives the primary role in understanding, or more correctly, in constructing the reality, to the knowing Subject.

Before going into Lukács' discussion of the concept of reification in the philosophical part of his essay, it is necessary to take into account a significant problem that arises at this juncture of his argument. The key position, which the notion of contemplative stance attains in Lukács' analysis, reveals how deeply he is influenced by Georg Simmel's inquiry into the modern forms of subjectivity in his book *Philosophy of Money.*[24] Written explicitly as a polemic against Marx, Simmel's study is also one of the first works to appreciate the normative content of Marx's analysis of the social mediation in commodity-form.[25] According to Simmel's reading, the social mediation inherent in the workings of money economy brings into being an atomistic form of individual freedom that manifests itself as an attitude of *indifference* towards others. This attitude of indifference however, Simmel argues, is first to be understood on purely descriptive terms and for that reason cannot be judged as presumably Marx does, only in negative terms. For Simmel, it is possible to have a more balanced view of the modern form of freedom, if the enabling and empowering qualities generating from the attitude of indifference are taken into consideration. It can be argued that while appropriating Simmel's analysis Lukács rejects this balancing act. For although he doesn't explain why, in Lukács' account the contemplative stance is necessarily the undesirable outcome of capitalist socialization. And yet, Lukács' strictly negative evaluation of the contemplative stance requires a normative standard. Since Lukács identifies Kant's philosophy as a part of the problem of reification, it would be self-defeating for his argument to build his normative critique on the grounds that capitalist society presents a distortion of the capacities of transcendental subject as defined by Kant. Thus, at times Lukács tries to solve this problem of normativity by presenting the contemplative stance as a pathological state, as an outcome of dehumanization brought by capitalism: «The worker's fate is typical of society as a whole in that this self-objectification, this transformation of a human function into a commodity reveals in all its starkness the dehumanized and dehumanizing function of the commodity relation.» (Ibid., p. 104) However, this is a similarly self-defeating argument. As his lengthy discussion in the third part of his essay, which presents *his* 'orthodox Marxist' critique of reification as an alternative to what he takes to be the problematic account of the humanist notion of alienation makes it clear, Lukács is aware of this problem. There remains only one other option to substantiate Lukács' negative normative judgment of contemplative stance. Lukács tries to pursue this option in his philosophical discussion of reification, where he enters the realm of German Idealism.

Within the scope of this paper it is not possible to give a detailed account of Lukàcs' engagement with German Idealism. Suffice it to say that in a bold attempt he brings together

[23] Kant, *Kritik der reinen Vernunft*, (Hamburg: Meiner Verlag, 1998), p. 109f.

[24] Georg Simmel, *Philosophie des Geldes*, [Frankfurt am Main: Suhrkamp, 1989] (1900).

[25] In fact Simmel does not conceal his concern in writing his study, even though the main opponent Marx's name is mentioned only three times in the text. In Introduction he writes: «Methodologically the basic intention can be expressed in the following manner. The attempt is made to construct a new story beneath historical materialism such that the explanatory value of incorporation of economic life into the causes of intellectual culture is preserved, while these economic forms themselves are recognized as the result of more profound valuations and currents of psychological or even metaphysical pre-conditions.» (Ibid, p. 12) In other words Simmel wants to hold primordial facts against Marx's critique of capitalism based on value-form analysis.

the philosophical agendas of several philosophers, among them primarily Kant, Schiller, Fichte, Hegel and Marx, around a relatively simple thesis. I will conclude this part of the paper by elucidating this thesis and pointing out its unexplored potential.

On most general terms Lukács builds his argument on the fundamental assumption that our attitude towards the world depends upon how we perceive, comprehend and/or grasp (*begreifen*) it. Since our attitude *towards* the world determines how we act *in* the world, the question of perception is of normative value. On this assumption, Lukács begins the second part of his essay with two interdependent claims. First, commenting on Kant's second preface in his first *Critique*, where he brings forward the metaphor of Copernican Revolution,[26] Lukács argues as follows: «Modern philosophy sets itself the following problem: it refuses to accept the world as something that has arisen (or e.g. has been created by God) independently of the knowing subject, and prefers to conceive of it instead as *its own product*» (Ibid, p. 123). The defining feature of Kant's understanding of subjectivity is a particular understanding of freedom, that is, autonomy. Lukács broadens this conception of freedom to include a notion of self-determination, by highlighting the contribution of the subject in constructing the reality. With the notion of knowledge as 'product', here Lukács emphasizes his genuine Marxist and arguably 'materialist' concern. Second, a couple of pages later Lukács puts forward a complementary claim: «What is novel about modern rationalism is its increasingly insistent claim that it has discovered the *principle* which connects up all phenomena which in nature and society are found to confront mankind. Compared with this, every previous type of rationalism is no more than a *partial system*.» (ibid, p. 125) To cut a long discussion short, this principle is to be found in the notion of Subject, or to put it more technically in the concept of *Self-Consciousness*. The way we perceive and/or construct the world as well as our attitude towards it depends upon our understanding and construction of this very constituent Subject. Thus, Lukács reformulates the fundamental question of German Idealism: How are we to understand the Subject, so that it does not contradict its definition, namely that it is free? Or to put it in Marxian terms: What are the conditions of existence of individual freedom?

In answering these questions Lukács' text provides conceptual resources for an alternative understanding of reification, which he does not pursue himself. In order to distinguish this alternative better, it is necessary to challenge the apparent linearity of Lukács' account of German Idealism. For though it may seem that here Lukács is simply mapping the philosophical path leading to Marx, he actually draws different conceptual resources from different philosophical systems within German Idealism. This explains the central role given to Fichte as opposed the marginal role of Hegel in his subject-theoretical attempt to substantiate his negative normative judgment of contemplative stance. Since it is the definition of the Subject and its normative horizon that determines the ultimate definition of reification as a state that *ought* to be overcome, the central role attached to Fichte is crucial.

Lukács' reliance on Fichte is, if partly, the result of his zealous attempt to situate Marx in German Idealist tradition. It can also be seen as a backlash of his romantic anti-capitalism.[27] In any case, Lukács in this part of his essay draws mainly on an earlier text of Marx rather than on the section on 'commodity fetishism' in *Capital*, from which, as we saw, he originally derives his concept of reification. In the first thesis on Feuerbach, young Marx coins the concept of Praxis. Lukács brings this concept in a fruitful exchange with Fichte's 'Subject of Action' and concludes that at the moment the subject-object of history, that is, the

[26] See Kant, *Ibid*, p. 15-40.

[27] On this point see the detailed analysis of Michael Löwy, *From Romanticism to Bolshevism*, (London: NLB, 1979).

proletariat becomes conscious of its constituting force, the conditions of existence of individual freedom will be established.

The eschatological tone of such an account is deeply problematic and has been criticized by many others. The same account can also be criticized for its political implications, for it brings with it a series of questions that remain relevant for a certain tradition of Marxist politics. To name a few: What is the difference between a Jacobin "We" and a Fichtean-Proletarian Subject? Can the self-consciousness of the proletariat be delegated? Who decides when the proletariat is mature for a revolution? And finally, how do we know, whether the proletariat has reached to self-consciousness, if our answer is not simply to say that the proof of pudding is in eating it? All these questions refer to a normative deficit in Lukács' account, despite his attempt to provide for a powerful normative critique of capitalism. If this is the case, then it is Lukács' definition of reification that needs to be criticized. For it is this definition, which is responsible for the normative deficit in question.

To put it simply, with his definition of reification, Lukács targets a *state* of a Subject, which is to be exchanged with another state. The active *Subject* is the proper Subject, since it is capable of keeping in control of the reality that it posits, whereas the contemplative stance lacks this capacity and *therefore* ends up with a reified reality beyond its reach. Understood in these terms, reification turns out to be the loss of a fantasy of omnipotence. This state of omnipotence, however, is arguably itself a reified state. For this fantasy defies any mediation between subject and object, positing an immediate identity between the two. In other places, however, Lukács seems to point to a different understanding, particularly when he gives a more Hegelian account of the superceding (Aufhebung) of reification. Thus, for instance, despite the fact that Lukács still reads Hegel through the lenses of Fichte and consequently assigns a crucial role to the master and slave dialectics, at certain points in his account, he seems to suggest that the overcoming of reification is possible through a form of social mediation, which would constitute the subjects as constituents of the reality that they take part in. This social mediation involves a reflective self-consciousness that requires an intersubjective relationship through the construction of a plural identity, a 'We' that would not undermine the individual differences of its members. Lukács refers not only to Fichte, but also repeatedly to the *Phenomenology of Spirit* in support of this argument and specifically to the transition from the chapter on 'Force and Understanding', where the ahistorical and monological conceptualization of the subject is criticized, to the chapter on 'Self-consciousness'.[28] Thus Lukács sets the problem as that of finding the appropriate Subject that would strip off the limitations put on the monological Subject. Because of his essentially Fichtean approach, however, he puts his emphasis on how this appropriate Subject would overcome the separation between subject and object, instead of clarifying the intersubjective mediation required for this overcoming. Thus bringing Hegel and Fichte together in one paragraph Lukács writes:

> Here in our newly-won knowledge where, as Hegel puts it in the *Phenomenology*, «the true becomes a Bacchantic orgy in which no one escapes being drunk», reason seems to have lifted the veil concealing the sacred mystery […] and discovers […] that it is itself the solution to the riddle. But here, we find […] the decisive problem of this line of thought: ***the problem of the subject of the action, the subject of the genesis***. For the unity of subject and object, of thought and existence, which the 'action' undertook to prove and to exhibit, finds both its fulfillment and its substratum in the unity of the genesis of the determinants of thought and of the history of the evolution of reality. But to comprehend this unity it is necessary both to discover the site

[28] Compare Lukács, ibid., p.157f, 161f., 191f., 211f.

> from which to resolve all these problems and also to exhibit ***concretely*** the 'we' which is the subject of history, that 'we' whose action is in fact history. (Ibid, p. 161)

Lukács refers above first to Hegel in pointing out the significance of the conceptualization of self-consciousness in realizing what he deems to be the radical freedom of subject. Instead of following Hegel's definition of self-consciousness that presupposes an intersubjective mediation between subjects, however, he turns to Fichte and defines the essential feature of a free subjectivity with the notion of 'Subject of Action'. And yet, he reintroduces the Hegelian perspective to account for the historical genesis of the 'Subject of Action'. This historical subject, 'we', is the 'proletariat as a class'. Again referring to Hegel, Lukács writes:

> The individual can never become the measure of all things. For, when the individual confronts objective reality he is faced by a complex of ready-made and unalterable objects which allow him only the subjective responses of recognition or rejection. Only the class can relate to the whole of reality in a practical revolutionary way. (The 'species' cannot do this as it is no more than an individual that has been mythologized and stylized in a spirit of contemplation) (Ibid., p. 211).

As Hegel argues, and Lukács agrees, a plural identity, a 'we', cannot be constituted in reference to a pre-given nature of the 'species', or in reference to human nature. The revolutionary subject that will overcome the division between subject and object, has to come into being through a process of social and historical mediation. Instead of reflecting upon the process of this mediation, the necessity of which is thus grounded in his theory of reification, Lukács chooses, however, to focus on the *capacity* of this revolutionary subject to appropriate the reality. As a result both reification and its overcoming are presented as *states* of a singular and/or a plural subject rather than referring to the process of its constitution in and through social and historical relations. Thus instead of following Hegel's perspective and Marx's analysis of commodity and critique of commodity fetishism in defining the mode of social mediation that would overcome the atomistic subjectivity of individuals living in capitalist society, Lukács puts his hopes in a surrogate of a plural 'Identity', that is, a revolutionary party that has to play the role of a 'Subject of Action'. Lukács does not clarify what form of individuation process has to replace the reified social relations. Thus, at the end of his essay, in his response to the question of how to organize the proletariat to constitute it as the subject-object of history, Lukács ends up with no choice other than to quote Engels affirmatively: «The proof of the pudding is in the eating […] This pudding, however, is the making of the proletariat into a class: the process by which its class consciousness becomes real in practice» (ibid, p. 217). With this turn taken, Lukács' concept of reification fails to capture the essence of Marx's normative critique of capitalism, to which we now turn.

III

The first chapter of *Capital* gives a very tightly knit account of how we are to understand the commodity as the most abstract category and commodity-form as the most elementary form of capitalist relations of production.[29] As Marx argues, the key to understanding this abstract

[29] Marx points to the centrality of the first chapter in *Preface to the first edition of Capital* as follows: «The value-form, whose fully developed shape is the money-form, is very elementary and simple. Nevertheless, the human mind has for more than 2,000 years sought in vain to get to the bottom of it all, whilst on the other hand, to the successful analysis of much more composite and complex forms,

category lies in its *form*, that is, in the form of social mediation, which transforms labor into *Value* (MEW 23, p. 94f.). Marx begins his analysis with an individual commodity, then moves on to the analysis of a relation initially between two commodities, which gives way to an analysis of the *world of commodities*. Two difficulties arise in interpreting this chapter, if Marx's method of exposition is not taken into consideration.

First, in Marx's account there is an apparent inconsistency with respect to what constitutes value. For Marx suggests that value-creating labor is both an abstract energy quantum and *abstract labor*, understood as a product of the social mediation necessary in order to produce commodities. Second, it becomes extremely difficult to make sense of Marx's very detailed expository conceptual account that move from a single commodity to the world of commodities. With respect to the second point, many commentators find refuge in Marx's statement at the very beginning of his analysis of the commodity-form. Since in the beginning of this section Marx claims to have already solved the riddle of commodity-form, the remaining fifteen pages that constitute the value-form analysis is frequently ignored by the readers of Marx.[30] I suggest that a reading which draws a parallel between this chapter and the chapters of 'Perception' and 'Force and Understanding' of the *Phenomenology of the Spirit* solves these difficulties.

The explanatory power of such a reading can be summarized as follows. First, through this reading it is possible to track Marx's methodological interest in his exposition (Darstellung) and account the seeming inconsistencies in his discussion. In the first expository step of this chapter, a single commodity, as a *thing*, is presented as the object of analysis, whereas in the value-form analysis that follows this account a *rational-relation* is at stake. Marx searches, I argue, for the adequate model of knowledge to comprehend and fully develop the object of his analysis. In doing so, Marx, just like what Hegel does, in the *Phenomenology of Spirit*, ascribes the inconsistencies and contradictions within his exposition to the insufficiencies of the models of knowledge. Understood in this way, the insufficiency of the first model is superceded in the second model. This is why, for instance, the first definition of *value-creating labor* as energy quantum differs from the second definition of *abstract labor* as product of social relations. For this second definition follows from the first definition's insufficient understanding of the abstractness of labor.

Second, when read through the lenses of the *Phenomenology of the Spirit,* it becomes possible to discern the substantive content of the expository development of the object of analysis in *Capital* from a single commodity to the world of commodities. Marx conceptualizes the first, empirical object of his analysis as a *thing* with *special qualities*, that is, both as a thing to be consumed and as a thing of value. In doing so, he defines 'commodity' in terms of the object of Hegel's chapter, 'Perception'. The insufficiency in conceptualizing this thing with its double nature is superceded by the consideration of a *relation* between commodities, which can only be the object of a *rational* conceptualization.

---

there has been at least an approximation. Why? Because the body, as an organic whole, is easier to study than are the cells of that body. In the analysis of economic forms, moreover, neither microscopes nor chemical reagents are of use. The force of abstraction must replace both. But in bourgeois society, the commodity-form of the product of labour — or value-form of the commodity — is the economic cell-form. To the superficial observer, the analysis of these forms seems to turn upon minutiae. It does in fact deal with minutiae, but they are of the same order as those dealt with in microscopic anatomy». (MEW 23, p. 12)

[30] See among many such interpretations Ernst Michael Lange's otherwise insightful commentary. Ernst Michael Lange, "Wertformanalyse, Geldkritik und die Konstruktion des Fetischismus bei Marx" in *Neue Hefte für Philosophie*, Heft 13, p. 1-46.

Following the Hegelian discussion of 'Force and Understanding,' Marx characterizes commodities in this relation as realizations of an immaterial *force*, so that they attain a phenomenal existence, which is a "form". It is only after establishing this that Marx begins to analyze the *commodity-form*. Parallel to Hegel's model of knowledge in "Force and Understanding", the relation among commodities imprints them with their *value-form*, and embodies a law, *law of value*, which regulates the world of commodities.[31] This world presupposes its *inverted* image, the world of human beings, who exchange commodities in order to make use of their material, not *phenomenal,* qualities. The mediation between these two worlds constitutes the social relation called *capital* in the proceeding chapters.

Most importantly drawing this parallel between the opening chapters of *Capital* and the *Phenomenology of Spirit* makes it possible to account for the normative core of Marx's critique of capitalism. Within the scope of this paper, it is impossible to fully present the detailed interpretative analysis, which is necessary to substantiate this point. With this in mind, in my concluding remarks, I will limit my discussion to providing a brief account of the conclusions drawn from such an analysis.

In both Hegel's account of *force* and Marx's analysis of *commodity-form* a distinct form of subject-relation is thematized. The normative standard of Marx's critique of capitalism is contained within the critique of this form of subject-relation. In the analysis of *value-form* Marx describes a social mediation, a society, which is based on a radically individualistic, atomized form of subjectivity. This is not just a descriptive account that denotes, say, the *indifference* of Subjects against each other. Marx, following Hegel's account of *force,* describes a world of *monads* as proposed by Leibniz. Both for Hegel and Marx the form of Subjectivity described by Leibniz entails an understanding of freedom and individuality that is contradictory within itself. A monadological conception of freedom is deficient because it proves to be a fully determined relation of necessity between subjects as the monadic individuality gives way to an abstract universal form of equality, which obliterates all forms of difference. In Marx's analysis of *value-form*, the detailed exposition of this deficient form of Subjectivity – deficient, because the self-reflection of commodities presuppose an absolute Subject (i.e. Money), which annihilates the subject status of commodities – contains, at the same time, its immanent, *determinant negation*, that is, the proper form of subjectivity. This proper form of subjectivity and the conception of freedom and individuality that lie at its core can be defined as a true, non-deficient, relationship of intersubjectivity. In *Phenomenology of Spirit,* Hegel defines this reciprocal relationship of intersubjectivity with the help of dialectics of recognition. In modeling capitalist mode of social mediation after a dialectical relation that necessarily leads to the dialectic of recognition, Marx sets the standard of his normative critique of capitalism as recognition.

This briefly sketched alternative reading of the opening arguments of *Capital,* thus, contributes to a better understanding of Marx's normative critique of capitalism. It also enables us to rethink Lukács' concept of reification along the lines argued by Marcuse and Honneth. Both Hegel in his chapter 'Force and Understanding' in the *Phenomenology,* and Marx in his analysis of commodity-form, provide a critique of a *reified* social world based on a specific social mediation. They both emphasize that the reality perceived by a Leibnizian monad-Subject, which is the outcome of such a deficient mediation, is a reality beyond the control of the subject by definition. Because, Hegel and Marx argue, insofar as the monad-Subject claims to appropriate the reality as its own, it has to posit it independently of any historically definite social practice, that is, as a reified thing in itself. Thus, the reality attains a

[31] The parallel concept Hegel uses at this juncture is "ruhiges Reich der Gesetze", 'tranquil kingdom of laws'. See Hegel *Phänomenologie des Geistes* (Hamburg: Meiner Verlag 1988), p. 111.

detached quality, and becomes a power over the monad-Subject rendering its relation to other subjects inessential. For, the monad-Subject acquires its individuality / Identity through the mediation of an omnipotent reality, namely through God. In capitalist society, Marx argues, the constituents of society are imprisoned in such a form of individuality. The social relation among atomized individuals can be characterized by an attitude of *indifference*, for their individuality / identity is *given* through the mediation of an omnipotent subject, i.e. money. In this sense, this social mediation criticized by Marx can be called **reification** proper. This implies that the revolutionary praxis against reification *is* to reject all social relations that succumb to a deficient mode of intersubjectivity.

For the revolutionary praxis, then, the dialectics of recognition, which expresses the ideal relationship of intersubjectivity, serves as the norm for a free individuality. In other words, all social mediations that fail to adhere to the normative ideal of recognition are reified. Understood as such, this revised conception of reification retains the revolutionary impetus of Lukács' criticism of capitalism, while, at the same time, providing a critical standpoint for the constitution of a new society, in which in Hegels words «the unity of the different independent self-consciousness which, in their opposition enjoy perfect freedom and independence: 'I' that is 'We' that is 'I'» (Ibid,. p.127) is possible.

BIBLIOGRAPHY

Althusser L. and Balibar É. (1979) *Reading Capital*, London: Verso [1968].

Arato A. (1972) Lukács' Theory of Reification in *Telos 11*, Spring 1972.

Backhaus H. (1997) *Dialektik der Wertform. Untersuchungen zur Marxchen Ökonomiekritik*, Freiburg: Ça Ira [1974].

Brudney D. (1988) *Marx's attempt to leave Philosophy*, Cambridge: Harvard University Press.

Cohen G.A (1978) *Karl Marx's Theory of History: A Defense*, New Jersey: Princeton Univ. Press.

Engels F. (2005) Dialectics of Nature in *Marx/Engels Collected Works*, Volume 25, New York: International Publishers, p. 313-590.

Engels F. (1970) Socialism Utopian and Scientific in *Marx/Engels Collected Works*, Volume 3, Moscow: Progress Publishers, p. 95-151.

Farr J. (1984) Marx and positivism in Ball / Farr (ed) *After Marx*, Cambridge: Cambridge University Press, 1984, p. 217-235.

Fraser N. and Honneth A. (2003) *Umverteilung oder Anerkennung? Eine politisch-philosophische Kontroverse*, Frankfurt am Main: Suhrkamp.

Habermas J. (1981) *Theorie des kommunikativen Handelns*, Bd. 2, Frankfurt am Main: Suhrkamp.

Habermas J. (1988) *Der philosophische Diskurs der Moderne*, Frankfurt am Main: Suhrkamp.

Hegel G.W.F (1988) *Phänomenologie des Geistes*, Hamburg: Meiner Verlag.

Heinrich M. (1991) *Die Wissenschaft vom Wert. Die Marxche Kritik der politischen Ökonomie zwischen wissenschaftlicher Revolution und klassischer Tradition*, Münster: Westfälisches Dampfboot.

Hoff J. (among others ed.) (2006) *Das Kapital neu lesen*, Münster: Westfälisches Dampfboot .

Honneth A. (2005) *Verdinglichung*, Frankfurt am Main: Suhrkamp.

Jay M. (1984) *Marxism and Totality. The Adventures of a concept from Lukács to Habermas* Cambridge: Polity Press.

Kant I. (1998) *Kritik der reinen Vernunft*, Hamburg: Meiner Verlag.

Lange E. M. (1978) Wertformanalyse, Geldkritik und die Konstruktion des Fetischismus bei Marx in *Neue Hefte für Philosophie*, Heft 13, p. 1-46.

Leibniz G.W. (1998) *Monadologie*, Stuttgart: Reclam.

Löwy M. (1979) *From Romanticism to Bolshevism*, London: NLB.

Lohmann G. (1991) *Indifferenz und Gesellschaft*, Frankfurt am Main: Suhrkamp.

Lukács G. (2000) Die Verdinglichung und das Bewusstsein des Proletariats in *Geschichte und Klassenbewußtsein*, London: Red Star Press [1923].

Marcuse H. (1986) *Reason and Revolution,* Oxford: Routledge & Kegan Paul [1941].

Marx K. (2001) *Das Kapital* in *Marx/Engels Werke* (MEW) Band 23-25, Berlin: Dietz Verlag [1890].

Marx K. (2005) *Marx/Engels Werke* (MEW) Band 42, Berlin: Dietz Verlag.

Nussbaum M. (2002) Objectification„ in Alan Soble (ed.) *The Philosophy of Sex*, Maryland: Rowman & Littlefield.

Petry F. (1916) *Der Soziale Gehalt der Marxschen Werttheorie*, Jena: G. Fischer.

Pinkard T. (1994) *Hegel's Phenomenology. The Sociality of Reason*, Cambridge: Cambridge University Press.

Postone M. (1993) *Time Labour and Social Domination*, Cambridge, Cambridge University Press.

Reichelt H. (1970) *Zur logischen Struktur des Kapitalbegriffs bei Karl Marx*, Frankfurt am Main: Europ. Verl.-Anst.

Šešerko L. (2006) Das 'wirkliche Triebwerk des Kapitals' und seine Beziehung zu Hegels Logik in Hoff(ed.) *Das Kapital neu Lesen*, Münster: Westfälisches Dampfboot.

Simmel G. (1989) *Philosophie des Geldes*, Frankfurt am Main: Suhrkamp [1900].

Važjulin V. A. (2005) Entwicklung systematisch denken . Ein Vergleich der dialektischen Logik bei Hegel und Marx in: *Deutsche Zeitschrift für Philosophie*, 53:2, S. 203-18.

Wellmer A. (1971) *Critical Theory of Society*, New York: The Continuum Publ. Corp.

Wood A. (1981) *Karl Marx*, Oxon: Routledge.

Recensione

Filosofia dell'economia

Daniel M. Hausman

AAVV, *Filosofia delle scienze*, a cura di Nicla Vassallo, Einaudi, Torino, 2003

Alberto Binazzi
alberto.binazzi@humana-mente.it

Nel 2003 la casa editrice Einaudi pubblicava *Filosofia delle scienze*, un'antologia curata da Nicla Vassallo, nota filosofa italiana che, col suo lavoro, tesseva un metatesto sui problemi epistemologici sollevati da ogni scienza. Nel libro, che comprende articoli monografici dedicati alla filosofia della biologia, alla filosofia della chimica e alla filosofia della scienze cognitive, compariva anche la traduzione in lingua italiana dell'articolo di Daniel Hausman *Philosophy of Economics* (1998). Il testo in questione, per quanto datato, è ancora oggi una delle più chiare introduzioni alla materia. Hausman infatti risulta un eccellente didatta, non per niente è curatore a sua volta di diversi *handbooks* e manuali usati con profitto dagli studenti e da coloro che si avvicinano per la prima volta a questa pur giovane disciplina.

L'autore spiega con molta efficacia come le questioni d'interesse filosofico offerte dal pensiero economico siano molteplici: il rapporto tra condotta economica, welfare e diritti, la natura della razionalità, le basi epistemologiche e metodologiche delle scienze sociali, il legame tra le varie scuole economiche in relazione al più ampio problema dello statuto delle scienze naturali. Numerose sono state, soprattutto nel secolo appena trascorso, le proposte di soluzione al problema sollevato da John Stuart Mill sulla natura dei principi fondanti della scienza economica, suscettibili, secondo il filosofo britannico, di essere stabiliti legittimamente tramite l'introspezione. Molte di queste proposte, secondo Hausman, tra cui le tesi di Popper, Lakatos, Friedman, non sono state in grado di affrontare correttamente il tema della inadeguatezza empirica delle teorie economiche. Negli ultimi decenni, poi, sono state avanzate nuove letture epistemologiche e metodologiche, come quelle di McCloskey, Baugh, Caldwell e Rosenberg, anche molto critiche nei confronti della pretesa da parte dell'economia di costituirsi a tutti gli effetti come una scienza empirica.

L'economia contemporanea appare oggi come una disciplina assai differenziata al suo interno. In parte complementari alla visione 'ortodossa' o neoclassica, si sono sviluppate numerose scuole che testimoniano, secondo Hausman, un'evidente non omogeneità tra le varie posizioni: si osservano varie correnti, tra cui la scuola austriaca, quella istituzionalista, quella marxiana, le dottrine socio-economiche, quelle comportamentiste, post-Keynesiane, neo-Ricardiane, la scuola di Chicago, la neoistituzionalista. Se tale frammentazione ha un interesse specifico per il filosofo, le principali implicazioni epistemologiche che lo riguardano concernono, in particolare, il dibattito sulla supposta neutralità dell'economia nei confronti della politica, il problema della plausibilità del vocabolario mentalistico mutuato dalla *folk psychology* per le spiegazioni di tipo causale riguardanti le scelte individuali, il problema del naturalismo (l'economia desidera diventare scienza naturale, ma, allo stesso tempo, i testi economici sono pieni di teoremi dimostrati deduttivamente) l'eccessiva idealizzazione e astrazione che non tiene conto delle interferenze di natura sociale e culturale, e, infine, il rapporto che la nozione di causalità intrattiene con la statistica e la metodologia.

Il rifiuto della proposta di Mill di legittimare l'autonomia dell'economia attraverso il ricorso all'introspezione e al metodo deduttivo, ha stimolato numerose ricerche che hanno evidenziato come i fenomeni economici siano influenzati da molteplici fattori (sociali, culturali,

comportamentali, politici) estrinseci alle teorie economiche stesse e non ancora adeguatamente considerati. In questa luce, le teorie economiche non sono suscettibili di essere messe alla prova dell'esperienza come qualsiasi altra teoria scientifica, a causa di fenomeni di interferenza che comportano numerose assunzioni di sostegno necessarie per derivare implicazioni opportunamente 'falsificabili'. I tentativi, infine, da parte di Rosenberg e McCloskey, di rinunciare al potere predittivo delle teorie economiche insieme al rifiuto delle questioni metodologiche, più che offrire concrete alternative alle tesi di Mill, hanno affermato, negandola, l'insolubilità del problema posto dal filosofo britannico. Celato dietro nuove declinazioni teoretiche, resta, dunque, ancora irrisolto il "Mill's problem", e, cioè, la difficoltà di conciliare le premesse dell'economia con i metodi e gli strumenti delle scienze empiriche.

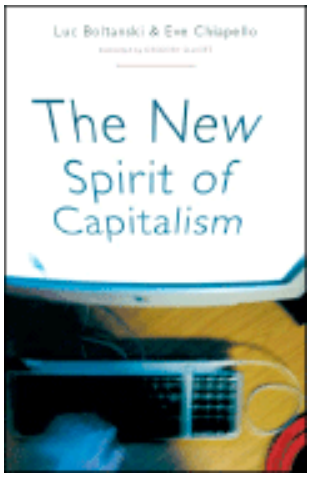
Book Review

The New Spirit of Capitalism

Luc Boltansk, Ève Chiapello

Verso, London & New York, 2007, 656 pp.

George Hull



This historical, sociological and conceptual study, first published in France in 1999, takes as inspiration the social situation of the industrialised West in the mid-1990s. Capitalism was flourishing: returns on invested capital were considerably higher than in the early 80s. On the other hand, unemployment, inequalities and homelessness were also on the rise. Yet, whereas comparable circumstances in the inter-war years and in the 60s had provoked a concerted and redoubtable criticism of the prevailing economic order, in the 90s criticism had largely given way to an economic or Darwinistic fatalism – a rueful (or enthusiastic) acceptance that 'the laws of the market' were indomitable. This diagnosis of a crisis of criticism is, despite various knee-jerk reactions to the present financial difficulties, as pertinent to the situation in Western Europe and the U.S. today as it was ten years ago to the French case on which Boltanski and Chiapello concentrate their attention.

The present crisis of criticism, the authors argue, must be understood in connection with a previous crisis of capitalism, or, more precisely, a crisis of a previous *spirit* of capitalism, which erupted most visibly in the protest movements of 1968. Capitalism, understood as 'an *imperative to unlimited accumulation of capital by formally peaceful means*' (p. 4), although it includes a normative, or justifying, component in its definition, cannot by itself motivate people to participate actively – whether as investors or as wage-workers – in the process of accumulation. On the contrary, Boltanski and Chiapello argue that the rise in the suicide rate from 22.9/100,000 in 1977 to 31.6/100,000 in 1994 (p. 423) can be attributed to the 'anomie' which results when people are directly exposed to the capitalist imperative without an intermediary 'spirit'. Desire for essential goods, or even luxuries, is also not enough, the authors claim, to motivate the kind of engagement with one's work which is required in modern production and services. So capitalism needs a motivating spirit, and must import this spirit from the area of social reality outside the sphere of accumulation.

The spirit of capitalism at a given point in time can be called an ideology, but this should not be taken to imply that it is an inevitable epiphenomenal accompaniment of the economic 'basis', that it is necessarily a false representation of reality, or that one social group consciously uses it in order to manipulate another. Rather, the spirit of capitalism must draw on normative orders, or– in the terminology of Boltanski and Thèvenot's earlier book, *On Justification* – 'cities', which are not only endorsed as correct standards of justice across different groups in the society in question, but furthermore form the criteria for success in institutionalised 'tests', which as a consequence count as 'legitimate tests' rather than bare 'tests of strength' (p. 31). In turn, because these tests take a concrete, institutional form, they have a – partially constraining – impact on the process of accumulation.

This book's intriguing thesis is that capitalism has evolved over the last century not just in response to internal pressures (efficiency, competition) and technological advances, but also

by incorporating large strands of the criticism to which it has been subject into its own motivating spirit. Time and again, criticism of the capitalist imperative itself for its unacceptable consequences, aiming to overthrow or abolish the capitalist economic system, is reinterpreted as a justification for a new form of... capitalism.

To substantiate their claims the authors turn in the first place to the management literature of the 1960s and 1990s, which, in prescribing ways in which a firm should be run, provides managers not only with their own motivations but also with justifications they can use when explaining features of organisation to the workforce. In subsequent historical chapters it is shown that the 'tests' introduced by firms and by government legislation progressively embodied the 'spirits' advocated by the literature directed at managers in these two periods.

The spirit of capitalism in the 60s defined itself in opposition to the capitalism of the first half of the twentieth century. Its defining features can be understood as an incorporation of the substantial criticism directed at capitalism in the inter-war years by European socialist and communist parties, as well as the trade unions. This criticism was largely what Boltanski and Chiapello call 'social critique' – criticism of poverty and inequalities in society, on the one hand, and, on the other hand, criticism of egotistic exploitation of one social group by another. According to this spirit of capitalism, profit should be distributed fairly according to qualification and position in the firm, as opposed to nepotistically, as it often was in earlier, family-based capitalism, and a job in a large firm should provide a worker with lifelong security. The spirit of capitalism in these years had much in common with the ethos of the welfare state.

But it is in opposition to this (second) spirit of capitalism that the new (third) spirit of capitalism defines itself, and whereas the former draws above all on the social critique of the inter-war years, the latter draws predominantly on what the authors call 'artistic critique', which acquired a formidable public voice for the first time in 1968. Artistic critique is criticism of 'inauthentic' activities, emotions and forms of life, on the one hand, and of oppression restricting the individuals' freedom, autonomy and creativity, on the other. Incorporating criticism of industrial capitalism as hierarchical and dehumanising, transforming workers into mere machines for carrying out instructions, the new spirit of capitalism advocates self-management and versatility in the workplace and a new ideal of the manager as an inspiring 'coach' or 'leader', who gets the workers to do what is needed, not by issuing orders, but by converting them to the cause, so that the firm's interests become *their* interests.

This latest spirit of capitalism can, argue Boltanski and Chiapello, be best understood as embodying the values of a new normative order, 'the projectual city', not seen before, and not yet fully articulated – let alone institutionalised. This city, drawing indirectly on neo-Nietzschean philosophy (Deleuze), French and American sociology and the language of information technology, acknowledges that the world of work is now a 'connexionist' world, a mesh of interlinking networks in which the successful actor is the one who has enough transferable skills, contacts and mobility repeatedly to come out on top in this city's chief test: ending one 'project' and beginning another.

While the flexibility prized in the new connexionist world may – in the spirit of a disarmed artistic critique – enhance the autonomy of roving managers, and certainly enhances the revenue of investors on the international financial markets, things are not so rosy for much of the workforce on new flexible contracts with their concomitant insecurity. In 1995 67% of French workers on part-time contracts had accepted these 'for want of something better' (p. 258).

But the same artistic critique that contributed to forming the projectual city also helped to generate the crisis of criticism, which is the book's starting point. For one thing, criticism of

authoritarian hierarchy was from 1968 onwards directed as much at the trade unions and the P.C.F. as at industry. For another, measures required to further autonomy are intrinsically likely to diminish security, and vice versa, so that there is a conceptual tendency for the artistic critique and the social critique to undermine one another.

However, in the most fascinating part of their study the authors outline a model for how the traditional social criticism of exploitation can be adapted to the connexionist world, criticising it from within the normative framework of the projectual city itself. It can be shown that actors with more flexibility and mobility (multinationals, financial firms) rely for their profitability on the fact that other actors (states, local workforces) have less. This fact can then ground claims for more recognition and remuneration on the part of less mobile and versatile actors. The authors perhaps do not make clear how much the traditional model of exploitation by a proprietor class of a working class can still be brought to bear on a connexionist world, as they tend to blur the distinction between what kind of criticism is justified and what kind of criticism is, or can most easily be, institutionalised (unions, laws).

This original and imaginative book, which goes a long way towards making contemporary life intelligible, ends with a subtle and perceptive discussion of new ways in which features of work and consumption can be experienced as inauthentic.

Recensione

Critica della ragione economica. Tre saggi di Kahneman, McFadden, Smith

A cura di Matteo Motterlini e Massimo Piattelli Palmarini

Il Saggiatore, Milano, 2005

Alberto Binazzi

alberto.binazzi@humana-mente.it

I recenti sviluppi di ricerca in economia cognitiva e sperimentale, ponendo come loro principale oggetto di studi il comportamento effettivamente osservato dell'attore economico, hanno incrinato la fiducia in una concezione di razionalità astratta e formale a favore di una visione problematica dei processi decisionali umani in condizioni di razionalità limitata. Gli articoli raccolti in questo volume - la versione in italiano di un lavoro di McFadden apparso su Journal of Risk and Uncertainty intitolato *Razionalità per economisti* e le traduzioni delle *Nobel lectures* di Kahneman e Smith - raccolgono i contributi di tre studiosi impegnati in questo nuovo settore di ricerca che integra i metodi classici dell'economia con quelli tradizionali della psicologia cognitiva e delle neuroscienze. Come ben sintetizzano Motterlini e Palmarini nella postfazione del volume, "gli economisti classici assumono (implicitamente ed esplicitamente) l'esistenza di leggi dell'economia rispetto alle quali le microdeviazioni individuali dovute a istanze psicologiche" sono considerate rumore di fondo diretto ad autocompensarsi a livello macroeconomico.[1] È proprio l'evidenza sperimentale con i suoi numerosi risultati che ha messo in crisi i principi tacitamente assunti dagli economisti. Scopo dei curatori è evidenziare l'importanza della relazione di complementarità tra psicologia ed economia – capace di integrare le metodologie tradizionali in economia attraverso l'uso dell'esperimento controllato in laboratorio - e accogliere, così, la valenza pragmatica e antidogmatica di queste ricerche interdisciplinari, pur rimanendo fedeli all'impostazione originaria di Kahneman e Tversky che concepisce le teorie cognitive non necessariamente alternative a quelle classiche.

La teoria comportamentale delle decisioni (behavioral decision theory) – osserva McFadden – ha origine con il lavoro di von Neumann e Morgenstern sulle scelte in condizioni di incertezza e con la teoria dei giochi. L'approccio matematico ha avuto il pregio di "rendere appetibile l'analisi formale e assiomatica in economia e psicologia" incoraggiando la sperimentazione in laboratorio per controllare la validità descrittiva degli assiomi.[2] Gli importanti contributi delle scienze cognitive hanno recentemente evidenziato come i processi decisionali umani siano influenzati da elementi di natura psicologica quali percezioni, credenze, atteggiamenti, giudizi. Tali risultati hanno messo in crisi il modello dell'uomo di Chicago, un agente che sceglie con

[1] *Critica della ragione economica*. *Tre saggi: Kahneman, McFadden, Smith*, a cura di Matteo Palmarini e Massimo Piattelli Palmarini, Il Saggiatore, Milano 2005, pp.236.

[2] Ivi, pp.27.

l'unico fine di massimizzare preferenze razionali, primitive e coerenti, uniformandosi a rigorosi principi statistici bayesiani (una specie in via di estinzione, come l'ha definita l'autore).

Tra i lavori degli scienziati cognitivi, le ricerche di Kahneman e Tversky hanno affascinato e sgomentato allo stesso tempo gli economisti. I loro risultati evidenziano delle incongruenze tra il comportamento osservato e il modello dell'uomo di Chicago: le principali discrepanze concernono un'elevata variabilità nell'utilizzo dell'informazione, l'uso di euristiche, l'incapacità di massimizzare le preferenze, la sensibilità al contesto, la violazione degli assiomi della scelta razionale, deformazioni percettive, effetti di dissonanza cognitiva. Le nuove evidenze sperimentali, osserva McFadden, convergono verso una conclusione molto severa: la falsità del modello dell'uomo di Chicago a causa di irregolarità cognitive dovute a effetti di contesto, di punto di riferimento, di disponibilità, di superstizione, di processo, di proiezione. Questi risultati dimostrano come "gli esseri umani non riescano a reperire e a elaborare le informazioni in modo coerente e ciò genera una varietà di anomalie, compreso il comportamento che rende i consumatori vulnerabili a essere sfruttati dai mercati".[3] Agli economisti - conclude McFadden - non resta che "far evolvere l'uomo di Chicago nella direzione dell'uomo di Kahneman-Tversky" adottando, di conseguenza, ipotesi di razionalità più blande di quelle classiche e modificando l'analisi economica in funzione di queste.[4]

Il saggio di Kahneman intitolato *Mappe di Razionalità limitata: indagine sui giudizi e le scelte intuitive* illustra i risultati di una lunga collaborazione con Amos Tversky (1937-1996) su temi di psicologia delle credenze e analisi delle scelte in condizione di incertezza. Tra gli argomenti affrontati, euristiche e distorsioni del giudizio, effetti di *framing*, di salienza e ancoraggio, teoria del prospetto, euristiche prototipiche, illusioni cognitive. L'idea centrale sostenuta dallo psicologo israeliano è che i pensieri divergono in una dimensione di accessibilità manifesta nel differente funzionamento dei processi mentali di tipo intuitivo e deliberato. Un ruolo privilegiato viene attribuito alle intuizioni: secondo Kahneman, le intuizioni rappresentano una modalità di elaborazione dell'informazione intermedia tra i due poli costituiti dalle operazioni automatiche della percezione e quelle deliberate del ragionamento. Tale generale distinzione individua due tipi di sistemi (1, 2) di funzionamento globale della mente umana.

A un estremo troviamo il sistema percettivo e intuitivo (sistema 1) le cui operazioni sono automatiche, rapide e poco costose da un punto di vista computazionale. All'opposto, le risposte del sistema 2 sono deliberate, lente, seriali, costose in termini di carico cognitivo e, di norma, "le persone le intraprendono solo se hanno una qualche precisa ragione per farlo".[5] È compito dello scienziato cognitivo, secondo Kahneman, mostrare come molteplici condizioni di scelta siano influenzate notevolmente dal sistema 1. Gli effetti gestaltici osservati nell'ambito della psicologia della percezione, infatti, sono in azione anche in presenza di situazioni cognitive che includono l'elaborazione di stimoli più astratti. Ecco, sinteticamente, alcuni risultati sperimentali: gli effetti di *framing* nel processo decisionale e nella risoluzione dei problemi si manifestano quando differenti descrizioni dello stesso compito evidenziano una notevole discrepanza tra i risultati, la teoria del prospetto indica che le variazioni e le differenze sono molto più accessibili dei livelli assoluti, le euristiche del giudizio offrono una cornice interpretativa per molti errori sistematici nelle credenze e nelle preferenze. Tali effetti documentano un forte allontanamento dalla logica estensionale-normativa della credenza e della scelta: la sostituzione di attributi prototipici ad attributi estensionali sembra essere una

[3] Ivi, pp.64.
[4] Ivi, pp.66.
[5] Ivi, pp.84.

caratteristica incompatibile sia con i principi bayesiani, sia con l'assioma della massimizzazione dell'utilità.

Vernon Smith, nel suo intervento intitolato *Razionalità costruttivista e razionalità ecologica* osserva come la ricerca nel campo della psicologia economica abbia portato alla luce molti esempi in cui venivano violate le ipotesi di razionalità del modello standard o costruttivista. Ereditato da Descartes, Hobbes, Bacone, questo assume che gli agenti possiedano una completa informazione sui guadagni e dimentica l'importanza del ruolo del contesto e della memoria autobiografica degli attori economici. Secondo Smith, appare più opportuno muoversi verso il concetto di razionalità ecologica, una razionalità che emerge da processi evoluzionistici, culturali e biologici. Il modello costruttivista, infatti, per mezzo della teoria razionale, rappresenta una situazione economica tramite un albero di gioco interattivo astratto. Al contrario, il concetto ecologico di razionalità si interroga sulla natura della struttura formale delle situazioni rappresentate. Quali pratiche sociali si prestano a essere spiegate da questi modelli e non altre? Smith accoglie, inoltre, le potenzialità offerte dagli esperimenti in laboratorio, dove è possibile "costruire razionalmente dei controfattuali, come nella storia economica" per verificare ed esaminare la validità di quelli che sono considerati teoremi dimostrati per via deduttiva.[6] Il suo giudizio finale sul modello standard di razionalità è assai critico: l'adesione alla tesi costruttivista, oltre a essere priva di validità empirica, implica, in maniera immotivata, che "ogni agente sia un costruttivista esattamente nello stesso senso in cui lo siamo noi in quanto teorici".[7] Da qui, allora, occorre ripartire integrando la ricerca classica in teoria economica con i contributi offerti dall'economia cognitiva e sperimentale.

[6] Ivi, pp.150.
[7] Ivi, pp.159.

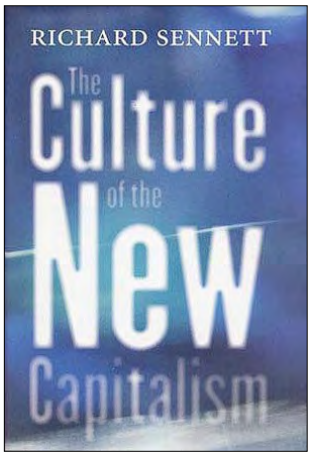
Book Review

The Culture of the New Capitalism

Richard Sennett

Yale University Press, New Haven and London, 2006

Marco Solinas
marcosolinas@tin.it



"But the flexibility they celebrate does not give, it cannot give, any guidance for the conduct of an ordinary life. The new masters have rejected careers in the old English sense of the word, as pathways along which people can travel; durable and sustained paths of action are foreign territories." This was the bitter conclusion of the nearly classic *The Corrosion of Character*, printed in 1998 (*The Corrosion*, p. 147). After the more eminently ethical-moral interlude of *Respect in an Age of Inequality* (2003), Richard Sennett took up again where he had left off in *The Culture of the New Capitalism* (which was given as the Castle Lectures at Yale University in 2004); he again brings up for discussion the consequences of the new institutions' flexibility: "The fragmenting of big institutions has left many people's lives in a fragmented state: the places they work more resembling train stations than villages, as family life is disoriented by the demands of work" (*The Culture*, p. 2). He also begins to oppose this culture with the "ideal of craftsmanship" against which the emerging social order militates, as celebrated in *Craftsman*, published last year (*The Culture*, pp. 4, 194-197, 103 ff.).

But let us get back to the effect of the new institutions. The first aim of *The Corrosion* was to show "the personal consequences of work in the new capitalism", i.e. that "the conditions of the new economy feed instead on experience which drifts in time, from place to place, from job to job. […] short-term capitalism threatens to corrode […] character, particularly those qualities of character which bind human beings to one another and furnishes each with a sense of sustainable self" (*The Corrosion*, pp. 26-27). Sennett takes up his discourse again from this point: "the cultural ideal required in new institutions damages many of the people who inhabit them" (*The Culture*, p. 5). Playing a major part in both books is the fact that the flexible institutions, first of all enterprises and the market, no longer act in the long but the short term, and this transformation involves both self-employed professionals that work in the sphere of the new economy as well as employees. If, as a result, some analyses come to overlap, the sphere is nonetheless extended as now the welfare institutions and the public realm are also considered. This extension corresponds to the adoption of an enlarged perspective: the analysis aims to sketch out the cultural ideal and the embodiment of the ideal man, demanded and promoted by the institutions of new capitalism. In short, at the core is the "new cultural model" or the "culture of the new capitalism" and its effect on society (pp. 3-8). And this provides the structure for the volume: analysis of the transformations in the institutions brought about by the shortened framework of time, developed in the first long chapter ("Bureaucracy"). The relation between fears about being made redundant or left behind and the talent required by the "skilled society", i.e. the relation between work, unemployment and *Bildung* in the second chapter ("Talent and the Specter of Uselessness"). The relation between

consumer behaviour and political attitudes in the third chapter ("Consuming Politics"). Finally, a brief conclusion that aims to outline some public remedies for the ills of the new institutional model ("Social Capitalism in our Time").

Beyond a certain disjointedness, one of the threads that seems to me to unify the analysis of the above mentioned spheres is the critical state of new capitalism's cultural model: its negative effect on different levels. The analysis indeed aims to show that the new model not only produces anxiety and stress in the worker of the flexible enterprises (pp. 52, 65-68, 127, 181), but also presents some deficits and malfunctions in the productiveness and organization of those enterprises, which often prove their "undoing" (pp. 43, 64-71, 156), as well as a scourge on the welfare institutions and the political models (pp. 163 ff., 173-178). And it is this transversal negativity which could perhaps be given further examination. In the sense that if Sennett clearly writes that there is a net "class divide between those who profit from the new economy and those in the middle who do not" (p. 131), and that the new system brings a more and more radical social and economical inequality (54 ff.), it nevertheless seems to me that in the analysis of the functioning of the new system more emphasis could be put on the difference and even struggle between the different interests, needs and social expectations at stake. First of all between the perspective of the workers "below elite levels", increasingly subjugated to the new labour market mechanism, i.e. to the process of "casualization" of the labour force (pp. 48-51, 61-62, 76-77), and the perspective of the top executives, consultants and more generally the elite. While the new mechanisms could in some way have an anxiety-inducing effect on these workers, at the same time they guarantee high incomes and relatively good future expectations (pp. 54, 80-81).

And so different interests come to interact in various ways with the interests of the owners and the old- style stockholders, with their aims to increase dividends and reduce labour costs by recruiting deunionized and cheap young workers (pp. 48-49, 95-97, 134), while also sacrificing quality to achieve "high profits" (p. 106). In its turn a perspective different from that of the "new" stockholders, who instead are looking to make a profit in a short time (pp. 38-40). In brief, if the innermost conflicts of these perspectives were pointed out in a clearer and more articulate way, the negative effect of the new model could perhaps be better defined, more specific, and less transversal. From here one could then also further examine the crucial correlated question of the assertion of the new model. With regard to this, Sennett writes that the "new cultural model" is that of the new economy and that, although it "is still only a small part of the whole economy […] it does exert a profound moral and normative force as a cutting-edge standard for how the larger economy should evolve"; it has also "a cultural influence far beyond its numbers"; from a complementary viewpoint, and in simple terms: "the avatars of a particular kind of capitalism have persuaded so many people that their way is the way of the future" (pp. 10-12). However, no answers are given to the questions about the way in which this "persuasion" has taken place, nor are there any reasons why and how "the extension of the new values is [so] broad", despite their negative effect (p. 182).

To conclude, in order to try to better highlight the conflictual character of the different perspectives at stake and at the same time to try to illuminate the process leading to the assertion of the new model, it seems to me that it may be useful to interpret the working of the new order as the exercising of a new form of hegemony. An approach that could contribute both to examining the different power relations between the social partners at stake and to analysing the processes of obtaining consent. And an approach that could integrate in that "critical mind-set of ethnographers" through which the "ideological proposals", writes Sennett, reveal their elusive character (p. 11). In other words, I believe that what could perhaps arise from this conceptual-methodological integration is a renewed form

of criticism that can unmask those power mechanisms underlying the hegemony exercised by those “new masters” that celebrate flexibility and “have rejected careers in the old English sense of the word” and which, in spite of the negative effect of the new order on the mass of citizens-workers, nevertheless ends up achieving widespread consent in public life.

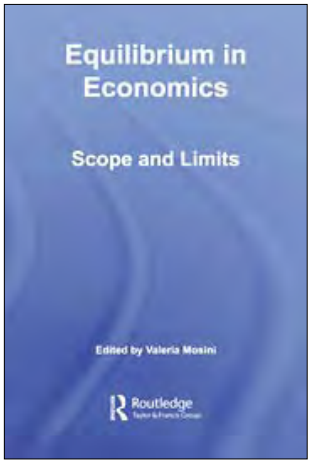
Book Review

Equilibrium in Economics: Scope and Limits

Valeria Mosini

Routledge, London, 2007

Giovanni Casini
giovanni.casini@humana-mente.it



The notion of equilibrium plays a fundamental role in economic science. Yet, scholars often adopt or refer to it uncritically. This situation led to the flourishing of an ample debate in contemporary economics about the value of models based on the concept of equilibrium. The book under review offers an in-depth analysis of this notion by investigating some important questions: How did the notion of equilibrium enter into economics? Which meaning(s) has it been used with? What kind of (past, present, and, possibly, future) role does and can it play in this field?

The volume consists of a collection of papers by authors coming from different academic backgrounds (economists, mainly, but also historians, sociologists and philosophers), who thereby address the issue from different points of view. The work is divided into three parts. The first two of them, 'The interplay of equilibrium notions between the natural sciences and economics' and 'Equilibrium in pre-neoclassical economics', have an historical orientation. They investigate, respectively, the emergence of both theoretical and methodological interchanges between economics and natural sciences and the issue of how the notion of equilibrium have been used and characterized by some of the founders of economics. The last and richest part, 'Equilibrium in present-day economic theory and practice', examines the use of the notion of equilibrium in present economic theory and practice. In what follows, I shall discuss each of these parts in more detail.

The three papers in the first part address the issue of the influence of other scientific fields in the development of the notion of equilibrium in economics. The paper by Grattan-Guinness deals with the theoretical interchanges between mechanics and economics. More specifically, it analyses how the notions of equilibrium, defined in mechanics, have influenced the work of some of the fathers of marginalism, like Jevons, Walras and Pareto. The paper shows how economics inherited many different notions of equilibrium from mechanics and how disproportionate the influence of such notions has been, given the very different scientific nature of disciplines such economics, on the one hand, and mechanics/physics, on the other hand. The second paper, by Bensaude-Vincent and Mosini, focuses on the theoretical interplay between chemistry and economics and analyses the two main notions of equilibrium developed in chemistry by Lavoisier (a static notion) and Le Chatelier (a dynamic notion) , respectively. First, the authors stress that Lavoisier held that the same notions of balance mechanisms and equilibrium were generally at play in nature, and hence present both in chemistry and in economics. Secondly, they analyze how the notion of equilibrium developed by Le Chatelier has been translated into economy by Samuelson – an example of a haphazard use of analogy between two radically different disciplines. The third paper, by Jarvis and Mosini, looks at the interchanges between biology and economy. It analyses, on the one hand, the notion of equilibrium as homeostasis and, on the other hand, a theoretical economic-

derived notion of equilibrium associated with evolutionism and connected to cost-benefit analyses. To summarize, this first part is centered on the theoretical 'exchanges' between economics and other disciplines. We are faced with a rich interplay, which shows how the multifarious has been the notion of equilibrium introduced, since the beginning, in economics. All these three papers raise interesting questions about the value and the correct use of analogy between different fields.

The second part, also constituted by three papers, deepens the analysis of how equilibrium-connected notions have been treated by pre-neoclassical economists. The first paper, by Dixon and Wilson, shows that the selfish, Hobbesian, neoclassical economic agent cannot be equated to the Smithian economic agent. Indeed, Smith's notion of 'simpathy', together with Chalmers' notion of 'character', and with the more recent work of Mead, suggests a notion of economic agency that denies the neoclassical reduction of human agency to 'particle-like behaviour'. The second paper, by van den Berg, dedicated to Isnard's work, shows how the tendency towards an utility-maximization approach was already mitigated, in the thought of this eighteen-century author, by the appeal to dimensions of human agency other than pure egoism, through the emphasis placed on the role of morality and 'virtuous habits' in the characterization of rational behavior. Finally, the paper by Vatin revises the work of Cournot, stressing the fundamental influence of Liebniz's thought over his ideas.

The third and last part, composed by seven papers, is devoted to the analysis of the role of the notion of equilibrium in contemporary economic theory. A big portion of these papers is dedicated to highlighting the ambiguities associated with the notion of equilibrium in economics and to proposing possible paths of disambiguation in order to restore clarity in the discussion about equilibrium. The papers by Lawson, by Samuels and by Chick are clear examples. Lawson starts by analyzing the ambiguous role of the notion of equilibrium in economic theorizing. He distinguishes two possible approaches towards equilibrium, according to which such concept refers, respectively, to a real property of the economic systems (the 'ontic' approach), or simply to a useful tool in the investigation of such systems (the 'epistemic' approach). Lawson believes that much of the confusion in the debate about equilibrium stems from conflating this distinction and confounding 'claims about models and their properties with properties of the reality that the models putatively aim to represent'.

In their papers, Samuels and Chick try to disentangle the various aspects of the notion of equilibrium that are mixed in contemporary economics. Samuels highlights various dichotomies, which different characterizations of equilibrium move from, focusing on the approach adopted towards equilibrium (seen either as an element of reality or as a useful modeling tool) or on its role in argumentative structures. Chick proposes a classification of the many different notions of equilibrium at play in economical theorizing and argues that the equilibrium-controversy should be reconsidered with respect to a better defined notion of equilibrium.

Backhouse starts from a strong instrumentalist point of view about equilibrium: economics is first of all a matter of problem-solving, and the debate about the role of equilibrium has to be treated case by case. The question whether the notion of equilibrium is useful or not depends primarily on its role within models that successfully target particular problems.

The paper by Thomas discusses how econometrics deals with the concept of equilibrium, concluding that this specific field can elaborate interesting economic models that do not employ such a notion.

Freeman points to the multifarious use of the concept of equilibrium in economics, which contrasts with the use of such concept in the natural sciences. Freeman also suggests that, in the former field, the concept of equilibrium is used in a dogmatic, quite religious, way.

Finally, starting from a representative instantiation of the concept of equilibrium in neoclassical economics, i.e., the notion proposed by Lucas, Denis examines the central role that equilibrium gained in the neoclassical approach and argues that this led to overestimating the static analysis of the system at the expense of its dynamic analysis. Denis concludes by stressing the need for the future recovery of a dialectical approach to the study of the economic system.

This book presents varied and rich contributions, which look at the notion of equilibrium from very different points of views, moving from an historical to an epistemological approach. Surely, this collection achieves at least one result: it convinces the reader that the debate about equilibrium in economics needs to be re-assessed from its foundations. Equilibrium is a concept that has been introduced in economics under the urge of various and different external leverages. Such manifold characterization keeps vitiating the current usage of the notion. In order to address properly the question about the role of equilibrium in economics, we need therefore to address the preliminary question about the meaning of the notion of equilibrium itself. The definition and classification of the various notions at play, of their epistemological and functional role in theoretical economics, is the first step for a good grasp of the issue.

At the same time, the richness and variety of the contributions appear also as the main drawback of the book. Despite an evident effort of coordinating and integrating the papers, the presence of so many different perspectives makes it difficult for the reader to gain a consistent overview on the issue and to selecting a research route appropriate for facing the many unlocked problems.

Table of Contents

Recensione

Etica ed economia: il rapporto possibile

A cura di B. Giovanola e F. Totaro

EMP Edizioni, Padova, 2008

Chiara Erbosi

chiara.erbosi@humana-mente.it

Etica ed economia: il rapporto possibile è un libro che raccoglie scritti di vari autori. Il filo conduttore del volume è l'indagine del rapporto esistente tra etica ed economia ai nostri giorni. L'interesse per questo rapporto nasce dalla volontà di interrogarsi su alcuni aspetti degli attuali meccanismi economici, nati soprattutto dalle dinamiche connesse alla logica della globalizzazione e dalla volontà di provare a comprenderli. In particolare tutti i contributi sono accomunati dal tentativo di individuazione di modelli dell'economia rispettosi della dignità della persona, dignità che spesso l'economia dimentica. Come infatti sostiene Totaro nel suo articolo di apertura del libro, la colpa dell'economia è non solo quella di essersi isolata da tutti gli ambiti dello spazio dell'umano ma anche di pretendere di contenerli tutti, con il costante obiettivo di assorbire e inglobare ogni aspetto della persona. E in buona parte si può dire che ci sia riuscita. Così infatti l'autore descrive il comportamento dell'uomo occidentale, corroso a tal punto dalle vorticose dinamiche economiche da essere incapace di modificare comportamenti e atteggiamenti che stanno diventando sempre più pericolosi e dannosi per il nostro pianeta:

> Perché il nostro modo di vivere come donne e uomini dell'Occidente ci conduce a ritenerci degni di considerazione e di autostima solo se siamo in grado di essere sempre più produttori e consumatori di beni materiali ad alto investimento energetico? Perché abbiamo affidato la nostra immagine di umani al possesso crescente di beni individuali che, a fianco di opportunità di scelta che dipendono esclusivamente da noi, comportano lo spreco o l'uso abnorme di risorse che appartengono anche ad altri o sono di rilevanza comune?[1]

Il libro dunque si apre con l'articolo di Francesco Totaro, *Per una misura etico-antropologica dell'economia*, dedicato in particolare alla questione energetica ed ecologica, diventata così urgente e pressante da chiamare l'economia ad una vera e rapida inversione di tendenza e ad una correzione. Totaro si dichiara a favore di un contenimento dell'economia da attuarsi però in modo differenziato a seconda che riguardi paesi dell'ipersviluppo e quelli del sottosviluppo. Se infatti è vero che, in mancanza di illimitate fonti energetiche, «uno sviluppo dei paesi emergenti il quale ricalcasse il modello di sviluppo perseguito dai paesi ricchi condurrebbe al collasso il pianeta terra»[2] è vero anche che è molto difficile domandare ai paesi che vogliono uscire dalla povertà una condotta virtuosa in quanto essi si sentono doppiamente defraudati

[1] F. Totaro, B. Giovanola (a cura di), *Etica ed economia: il rapporto possibile*, Edizioni Messaggero Padova, Padova, 2008, p. 45.

[2] Ivi, p. 44.

dai paesi ricchi «sia per la sottrazione di materie prime subita nel passato, sia per l'eventuale rinuncia forzata ad uno sviluppo che ormai bussa anche alla loro porta»[3].

Il secondo articolo di Sebastiano Maffettone, *Responsabilità di impresa e sostenibilità*, si incentra sul concetto di sostenibilità come il parametro più significativo della responsabilità sociale e come unica possibilità contro l'ossessiva ricerca delle imprese del profitto a breve termine.

Il tema della sostenibilità è centrale anche nell'articolo di Gian Luigi Brena, *Etica d'impresa tra economia ed ecologia*, in cui si sostiene che la responsabilità per l'ambiente non è mai solo dell'impresa ma dell'intera società che è chiamata ad ottenere la cooperazione di tutti i popoli. Questa però non può che esistere su basi di giustizia, riducendo le drastiche disuguaglianze economiche. Sono esse infatti che

> offrono la possibilità di continuare ad inquinare alle imprese dei paesi sviluppati, semplicemente spostando i loro impianti in paesi che non possono permettersi il lusso di occuparsi dell'ambiente. E la situazione di emergenza dei paesi ancora in via di sviluppo li autorizza anzitutto ad industrializzarsi, senza curarsi dell'inquinamento: le persone vengono prima dell'ambiente e devono essere anzitutto superate le situazione di povertà.[4]

Il quarto saggio di Helen Alford, *Impostazione del lavoro nella prospettiva della Laborem exercens*, è dedicata alla visione del lavoro nella prospettiva dell'enciclica *Laborem exercens*. L'autrice mette in evidenza l'importanza della dimensione soggettiva del lavoro che include gli aspetti morali e spirituali della persona.

Il saggio di Stefano Zamagni, *Democrazia, libertà positiva, sviluppo*, è dedicato al modello di democrazia deliberativa, la sola in grado di favorire un processo di democrazia economica grazie all'attuazione di un mercato plurale e ad una fondamentale campagna culturale incentrata sulla figura del consumatore eticamente responsabile.

Il libro si chiude con il saggio di Benedetta Giovanola, *Razionalità etica, razionalità economica e ricchezza antropologica*, che indaga i fondamenti etici ed antropologici dell'economia.

Pur dunque nella diversità dei temi affrontati, il libro cerca di mostrare l'insostenibilità dei modelli economici attuali che continuano ad accentuare disuguaglianze e conflitti e che stanno diventando sempre più deleteri per l'ambiente e per il pianeta, incapace di far fronte al vortice economico che porta intrinsecamente con sé inquinamento dell'atmosfera, rarefazione delle risorse idriche, distruzione delle foreste e del patrimonio naturale. Da una parte, infatti, aumenta la popolazione mondiale, aumentano i consumi anche da parte dei paesi in via di sviluppo e «si diffonde una tecnologia la cui potenza ha già portato i paesi più avanzati a uno sfruttamento invasivo e distruttivo delle risorse naturali»[5]; dall'altra invece la natura non riesce a smaltire e a riciclare in tempo le risorse consumate. Questa situazione che diventerà nel tempo sempre più drammatica richiede l'intervento e la mobilitazione di tutti. E infatti gli autori concordano tutti nel sostenere che un'inversione di tendenza dell'economia si può avere solo con un intervento programmato e condiviso a raggio mondiale e con più consapevolezza e coscienza dei comportamenti quotidiani di tutti gli individui, in particolare di quelli dei paesi ricchi.

[3] Ivi, p. 43.
[4] Ivi, p. 88.
[5] Ivi, p. 80.

Indice

Recensione

Economia e concezione dell'uomo

AA.VV. - a cura di P. Grasselli

Franco Angeli, Milano, 2007

Scilla Bellucci

scilla.bellucci@humana-mente.it

Il volume curato da Pierluigi Grasselli raccoglie le relazioni presentate da alcuni docenti[1] durante una Giornata di Studio tenutasi presso la Facoltà di Economia dell'Università degli Studi di Perugia il 22 Novembre 2006. Suddiviso tematicamente in tre parti[2], offre un ampio ventaglio di prospettive riguardo al pensiero economico contemporaneo. Queste vengono affrontate, così come è possibile intuire dalla partizione del volume, sia da un punto di vista tecnico (aziendale e giuridico) sia da una posizione più speculativa e teorica, prettamente economica, da cui provengono la maggior parte dei contributi.

Immediatamente evidenti sono due caratteristiche che appartengono alla molteplicità dei saggi: l'urgenza di ripensare i fondamenti della scuola economica *mainstream,* ossia quella neoclassica, e la presa in esame della mutata realtà contemporanea che rende necessario tale ripensamento. La corrente *mainstream* si fonda, infatti, su una concezione dell'attore economico, che viene usualmente chiamato *homo oeconomicus*[3], le cui peculiarità consistono in un atteggiamento egoista ed edonista volto alla massimizzazione dell'utilità o del profitto, a seconda che si tratti di un consumatore o di un produttore. Negli Atti qui pubblicati si hanno, invece, molte proposte di allontanamento da questa configurazione che si è rivelata fallace come base di elaborazione di teorie e strategie. Essa, infatti, non tiene conto dei comportamenti effettivi che vengono condizionati dall'emotività, dalla paura, dalla morale individuale e da altri fattori 'devianti' che orientano le scelte in una quantità di direzioni non sempre, o meglio, quasi mai, razionalmente giustificabili dalla logica del mero profitto (Grasselli, Signorelli, Visaggio)[4].

Ma non soltanto sul comportamento del singolo individuo si concentra la critica alla corrente neoclassica delle dottrine economiche. Essa investe la sfera aziendale, analizzando le possibili dinamiche di partecipazione dei lavoratori all'impresa (Uvalic, Montrone); la sfera sociale, contemplando la crescente richiesta di comportamenti 'etici' nella finanza (Moroni, Gigliotti) come nella gestione delle risorse pubbliche (Giannoni e Arcelli) e inserendo nel

[1] Per l'elenco completo degli autori si veda l'indice del volume riportato in calce.

[2] Parte prima: *Homo oeconomicus,* razionalità, reciprocità. Parte seconda: Responsabilità sociale, equità, partecipazione. Parte terza: Responsabilità d'impresa, identità individuale, ordinamento giuridico.

[3] "È probabilmente Vilfredo Pareto nel suo Manuale del 1906 ad avere utilizzato per primo l'espressione latina *homo oeconomicus*; la formulazione del concetto corrispondente è invece più antica dal momento che appare nel saggio di John Stuart Mill *On the Definition of Political Economy and on the Method of Investigation Proper to it* del 1836". Marco Boccaccio in *Economia e concezione dell'uomo*, Franco Angeli, 2007, p. 33.

[4] Tra parentesi sono indicati gli autori di alcuni testi presenti nel volume al cui contenuto si fa riferimento nell'indicazione degli argomenti trattati.

dibattito elementi assiologici e simbolici quali la gratuità e il dono (Montesi). Infine si approda anche al dibattito giuridico rispetto ai diritti e ai doveri (Budelli, Giampieri), ai problemi legati alla globalizzazione dei mercati per i paesi in via di sviluppo (Cerreta) e alla gestione dei beni culturali (Cosi).

In conclusione può essere opportuno riportare un'osservazione in merito al ruolo della formazione cui sono deputati i possibili sviluppi di una nuova dottrina economica:

> L'Università rappresenta quindi il luogo privilegiato per creare culture professionali più avanzate, ancorate ed alimentate da una consapevole dimensione etica in grado di gestire un sistema di responsabilità sociale.[5]

INDICE

[5] Pierluigi Daddi, in *Economia e concezione dell'uomo*, Franco Angeli, 2007, p.13.

19. Il consumo di prodotti "etici" e il comportamento del consumatore responsabile – M. Gigliotti
20. La natura della conoscenza – M. Paoli
21. Parte terza – Responsabilità d'impresa, identità individuale, ordinamento giuridico
22. *Homo oeconomicus* e individuo. Contestazioni e minacce – R.Chiarelli
23. Alla ricerca della "quadratura del cerchio": il principio di sussidiarietà – S. Budelli
24. Condizione degli immigrati ed asimmetrie
25. nella sfera dei diritti e degli obblighi - B. Giampieri
26. Equità contrattuale e mercato concorrenziale – F. Scaglione
27. I codici etici: profili storici e funzione – F. Pascucci
28. La nuova frontiera dei codici etici. Rilevanza giuridica attuale – M. Cerreta
29. Il nuovo accordo sul capitale di Basilea 2, l'etica degli imprenditori e i "rating" nella concessione del credito – G. Gatti
30. Sull' "uomo giuridico tributario" - M. Versiglioni
31. Dalla "conservazione" pubblica dei beni storico – artistici alla "regolazione" e "valorizzazione" dei beni ed eventi culturali – D. Cosi

Recensione

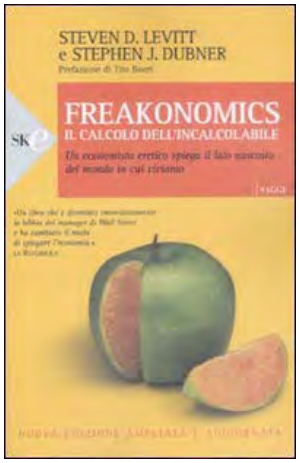
Freakonomics. Il calcolo dell'incalcolabile

Steven Levitt, Stephen J. Dubner

Sperling & Kupfer, Milano, 2008

Chiara Erbosi

chiara.erbosi@humana-mente.it



Nell'estate 2003 Stephen J. Dubner, giornalista e scrittore, fu incaricato dal New York Times di intervistare Stephen D. Levitt, giovane e brillante economista dell'Università di Chicago e vincitore del John Bates Clark Medal, premio assegnato al miglior giovane economista degli Stati Uniti sotto i quarant'anni. Da questo incontro e dalla loro collaborazione nasce *Freakonomics*, un libro insolito e inusuale che si pone l'obiettivo di «raschiare la superficie della vita contemporanea in cerca delle sue dinamiche più profonde»[1].

È un libro che parla di economia, ma il metodo di indagine utilizzato dagli autori è assolutamente estraneo a quello accademico di cui si avvalgono gli economisti: si servono infatti del metodo induttivo, ma basano le loro analisi scientifiche su una vasta raccolta di dati statistici, apparentemente svincolati gli uni dagli altri e privi di relazione tra di loro. Dubner e Levitt applicano il loro metodo a realtà molto lontane dalle questioni economiche e spesso ne derivano risposte sorprendenti, sconcertanti e molto originali. Gli autori, infatti, cercano di trovare motivazioni che spieghino perché la maggioranza degli spacciatori di droga viva ancora in casa con i genitori, perché il livello di criminalità negli Stati Uniti sia diminuito in seguito all'introduzione della legge sull'aborto o come il nome di battesimo possa influire sulla vita sociale di una persona o esporla alla discriminazione razziale. Le argomentazioni portate a sostegno delle loro ipotesi si allontanano quindi dagli ambiti tradizionali dell'economia, per sfociare nella sociologia, nella psicologia e nella criminologia.

Freakonomics affronta quindi temi tra loro diversissimi, partendo però da alcuni punti fermi: gli incentivi sono la pietra angolare della vita moderna e saperli capire è il punto essenziale per comprendere qualsiasi problematica; la saggezza convenzionale è spesso fatta di luoghi comuni che pur sembrando incontrovertibili possono essere confutati; effetti anche spettacolari hanno spesso cause sottili e lontane nel tempo; i cosiddetti esperti, dal criminologo all'agente immobiliare, sfruttano a proprio uso e consumo il vantaggio informativo di cui godono; sapere che cosa misurare rende il mondo meno complesso di quanto non appaia e imparando a studiare i dati nel giusto contesto si possono spiegare fenomeni altrimenti incomprensibili.

Dietro queste premesse c'è l'idea che ogni cosa abbia un volto nascosto, che può essere esplorato solo mettendo in crisi il senso comune e rifiutando le spiegazioni più ovvie. Come già detto, i casi che si soffermano ad analizzare non rientrano nei temi tipici delle scienze economiche, bensì riguardano fenomeni sociali e culturali: dai meccanismi immobiliari al

[1] S. D. Levitt, e S. J. Dubner, *Freakonomics. Il calcolo dell'incalcolabile*, Sperling e Kupfer, Milano, 2005, p. XXXIX.

rapporto genitori-figli. Gli autori si servono «dei più sofisticati strumenti analitici che le scienze economiche potessero offrire»[2] per capire dinamiche sociali che non sono mai state analizzate, per studiare nessi tra cose che nessuno avrebbe mai potuto immaginare, per calcolare l'incalcolabile - come dice appunto il sottotitolo del libro.

Già nell'*Introduzione* gli autori mostrano il loro particolare modo di procedere, cercando di offrirci il nesso tra la riduzione della criminalità minorile negli U.S.A degli anni '90 e l'introduzione della legge sull'aborto. Negli anni '90, infatti, la criminalità minorile avrebbe dovuto subire una grande impennata, secondo il calcolo dei maggiori criminologi e politologi. In realtà, anziché registrare la tanto temuta *escalation* di violenza, la criminalità cominciò a diminuire. «L'entità di quell'inversione di tendenza era impressionante. L'incidenza degli omicidi commessi da adolescenti, anziché registrare quell'impennata del 100 per cento o anche soltanto del 15, secondo lo scenario meno catastrofico di James Alan Fox, nel giro di 5 anni era crollata del 50 per cento»[3]. Tutti cercarono le motivazioni di questa inversione di tendenza e le trovarono nelle cause più svariate: dalla grande espansione economica al moltiplicarsi di leggi sul possesso di armi, dalle nuove strategie di pubblica sicurezza al miglioramento delle tecniche investigative, dall'aumento del numero di poliziotti all'introduzione di pene più severe.

Ma per gli autori la vera motivazione sta altrove e riguarda quello che era successo vent'anni prima ad una giovane donna di Dallas, Norma McCorvey. Priva di istruzione, dedita all'uso di alcol e di stupefacenti, dopo avere già dovuto dare in adozione due bambini, nel 1970 si ritrovò di nuovo incinta e voleva abortire. Ma nel Texas, come in quasi n tutti gli stati del Paese all'epoca, l'aborto era illegale. La donna divenne così lo strumento di un processo collettivo finalizzato a chiedere la legalizzazione dell'aborto. Il 22 gennaio 1973 la Corte Suprema si pronunciò a favore della donna, permettendo l'interruzione volontaria di gravidanza su tutto il territorio degli Stati Uniti. Ovviamente per lei era troppo tardi: aveva già messo al mondo il bimbo e l'aveva dato in adozione ma, senza volerlo, con la sua storia, finì con l'alterare in modo sostanziale il corso degli eventi.

Ma cosa c'entra tutto questo con il crollo della criminalità? Secondo gli autori, è evidente che non tutti i bambini sono uguali: un bimbo nato in un contesto sociale disagiato ha molte più probabilità di diventare un delinquente rispetto a coloro che nascono in condizioni più vantaggiose e che sono bambini voluti e amati. Proprio a seguito dell'introduzione della legge sull'aborto, moltissime donne in condizioni di disagio sociale hanno potuto accedere all'interruzione volontaria di gravidanza impedendo così la nascita di quei bambini che avrebbero avuto maggior probabilità di finire nelle maglie del crimine. Quei bambini non nacquero mai e questo ebbe un potente effetto: «ad anni di distanza, proprio quando questi ragazzi avrebbero potuto compiere il loro debutto come delinquenti in erba, il tasso di criminalità cominciò a diminuire»[4].

Gli autori sottolineano come una tesi di questo tipo possa disturbare molto e come sia, da un punto di vista etico, sconvolgente accettare che «il crollo della delinquenza è stato un beneficio collaterale della legalizzazione dell'aborto. Non occorre chiamare in causa motivi religiosi o morali per trovare agghiacciante che una tragedia privata divenga un beneficio per la collettività»[5]. Gli autori sono ben consapevoli che il tema dell'aborto è estremamente difficile da trattare, ma non possono fare altro che concludere che esiste uno stretto legame tra

[2] Ivi, p. XLII.
[3] Ivi, p. XXIX.
[4] Ivi, p. XXXI.
[5] Ivi, p. 143.

legalizzazione dell'interruzione volontaria di gravidanza e riduzione della delinquenza. D'altra parte, come sostengono nell'introduzione del loro libro, «la morale rappresenta il modo in cui si vorrebbe veder girare il mondo, mentre l'economia rappresenta il modo in cui gira davvero»[6]. Se quindi la loro ipotesi può sembrare inaccettabile da un punto di vista etico, diventa assolutamente congruente sul piano della pura analisi dei dati. Questo è solo uno dei molteplici esempi che si possono fare per mostrare il modo di procedere dei due autori e cioè il tentativo di trovare nessi tra questioni apparentemente svincolate le une dalle altre, andando oltre il buon senso, prendendo le distanze dalla saggezza convenzionale e diffidando sempre delle spiegazioni più ovvie e preconfezionate. Infatti dietro l'apparenza del pensare comune si celano convinzioni superficiali ed errori che solo un'indagine accurata e approfondita può smascherare.

Gli autori non seguono un filo conduttore ben preciso ma spaziano nei più svariati ambiti. Leggendo il libro, si scoprirà così che, a giudicare dal numero di minori che muoiono ogni anno, è più pericoloso avere in casa una piscina che un'arma da fuoco. Annualmente infatti muoiono 500 bambini al di sotto dei 10 anni per annegamento in piscina, mentre sono 100 quelli che rimangono uccisi da un'arma da fuoco, in un paese che dove ne circolano più di 200 milioni. Oppure si scoprirà che la maggioranza degli spacciatori di crack vive ancora con i genitori perché si tratta di un'attività assai poco redditizia e che la struttura societaria su cui si basa lo spaccio è identica a quella delle grandi aziende statunitensi. Levitt sostiene infatti che «lo spaccio di crack funziona in larga misura come qualsiasi altra attività di impresa in senso capitalista: per fare soldi bisogna essere vicino al vertice»[7]. Non ci si deve quindi stupire se uno spacciatore di droga vive ancora in casa con la mamma.

Quando un giovane dottorando in sociologia, Sin Venkadesh, allievo di J.W. Wilson, riesce ad entrare in possesso dei libri contabili di una gang, nei quali sono riportati quattro anni di attività di spaccio di droga, da questi libri emerge che i 120 top manager della gang guadagnano circa 500.000 dollari l'anno, mentre per tutti gli altri membri la retribuzione oraria è di 7 dollari all'ora per gli ufficiali e di 3 dollari e 30 centesimi per i soldati semplici, cioè ben al di sotto di un salario minimo. Rimanere in casa con i genitori è quindi una scelta obbligata. In più le condizioni di lavoro degli spacciatori sono molto pericolose con una probabilità di venire uccisi di 1 su 4. Perché dunque accettare un simile lavoro? La risposta di Levitt è che «la molla è sempre il desiderio di affermarsi in un campo estremamente competitivo, ma che al vertice, fa guadagnare un sacco di soldi. Per non parlare del compenso sotto forma di gloria e di potere»[8]. In quartieri dove il 56% dei minori vive al di sotto della soglia di povertà, il 78% vive in famiglie monoparentali, solo il 5% è in possesso di un titolo di studio superiore, il sentiero che conduce ad un lavoro onesto è quasi invisibile. L'alternativa è lo spaccio, le cui modalità sono analoghe a quelle di ogni altra professione-mito: «una gran massa di concorrenti per un numero limitatissimo di premi»[9]. Anche gli spacciatori si scontrano quindi con le rigide leggi del mercato del lavoro: se l'offerta di manodopera è abbondante, di solito la paga è bassa.

Come per tutte le professioni-mito che hanno molta richiesta, per cercare di arrivare ai vertici, bisogna partire dal basso, con tutti i rischi e sacrifici che questo comporta. Anche in questo caso quindi l'intento degli autori è di mostrare come dietro convinzioni apparentemente fondate e dietro la cosiddetta saggezza convenzionale si nascondano realtà

[6] Ivi, p. XL.
[7] Ivi, p. 96.
[8] Ivi, p. 98.
[9] Ivi, p. 99.

diverse da indagare secondo criteri anticonvenzionali - ma sicuramente più veritieri - capaci di mostrare davvero le dinamiche sociali e di comprendere fenomeni altrimenti inspiegabili.

INDICE

Recensione

Capitalismo totale

Perché la finanza minaccia la democrazia

Jean Peyrelevade

Isbn edizioni, Milano, 2008

Riccardo Furi

riccardo.furi@humana-mente.it

In tempi di crisi economica i sistemi d'informazione di massa ci parlano di 'crollo dei mercati', 'svalutazione delle azioni' e 'bolle speculative', descrivendo con alcuni termini tecnici un mondo che nella più banale delle ipotesi appare distante ed ininfluente dal susseguirsi quotidiano della vita di noi cittadini. Si capisce che la situazione economica non è delle migliori ma raramente vengono fornite spiegazioni esaustive sulle cause e gli effetti delle crisi o su quali sono i meccanismi e principi che regolano le attività finanziarie. La notizia che più spesso salta all'occhio riguarda la tal impresa che ha gonfiato il bilancio o il manager truffatore di azionisti; così che il danno viene circoscritto all'interno di un evento particolare e contingente dovuto alla sprovvedutezza o malafede di qualcuno dei protagonisti della finanza mondiale evitando di interrogarsi sulla natura di questi eventi: se, invero, siano una inevitabile conseguenza della struttura del *capitalismo finanziario*.

In *Capitalilsmo totale* di Jean Peyrelevade possiamo trovare una interessante ed accesa analisi riguardo al dorato mondo della finanza, i meccanismi del sistema ed il ruolo dei suoi attori principali. Poco più di 100 pagine dense e piacevolmente scorrevoli che, senza olismi o sofisticate argomentazioni, rendono espliciti concetti troppo spesso celati da ostici tecnicismi. Il libro di Peyrelevade, già presidente e amministratore delegato di Crédit Lyonnais, nonché consigliere del governo Mauroy (1981-83), coinvolge in una duplice accusa sia il capitalismo finanziario che intellettuali ed esperti di finanza: il primo reo, con tutti i protagonisti che lo animano, di essere causa delle sue stesse crisi e della accentuata diseguaglianza e mal distribuzione delle ricchezze mondiali; i secondi per tacere, nella migliore delle ipotesi - o mistificare, nella peggiore -, alla comunità mondiale, quella che è la realtà economica globale. Contenuti politicamente scorretti, ma obiettivi perché confortati da incontestabili dati, rendono il testo uno strumento che, usato con la dovuta attenzione, permette di capire quanto la previsione di crescita di un'azione possa influenzare direttamente la nostra vita, il nostro posto di lavoro.

La tesi dell'economista francese è chiara e diretta: siamo entrati in un era di capitalismo totale, completamente slegato dalla società ed in cui il fine unico è l'incremento di valore.

Il legame tra il mercato azionario e le imprese, già intrinsecamente tenue, non è mai stato così univoco nei confronti del primo.

Nei primi capitoli, il nostro autore descrive il contesto storico e culturale della nascita del mercato azionario, la fondazione teoretica e la funzione pratica cui doveva assolvere.

Esso era un luogo virtuale in cui si monetizzava il valore delle imprese che poteva essere liberamente scambiato: in tal modo le imprese avevano sempre a disposizione una certa quantità di capitale fresco da poter reinvestire, in ricerca o occupazione, a seconda del prodotto e delle proprie strategie. Si trattava di un sistema strutturato in modo tale che potesse, idealmente, generare un circolo virtuoso di denaro liquido.

È doveroso precisare che Peyrelevade si riferisce principalmente al modello liberista degli Stati Uniti, culla del mercato e origine del nuovo modello di liberismo totale.

Dunque il mercato nasce in funzione dell'impresa. Gli azionisti necessitano dell'intervento intermediario degli istituti di credito, che garantiscono le rendite e decidono gli investimenti da fare per le imprese. Una volta descritti i pregi e i difetti di questo modello, detto *renano*, l'autore ce ne racconta il declino e il passaggio ad una forma monopolizzata dal potere azionario. Dopo la crisi del 1929 si è innescato un processo che ha destituito di valore l'istituto bancario; attratti dalle lusinghe di maggiori rendite, i risparmiatori si sono accasati a istituti di gestione del capitale specializzati che non interpretano più il ruolo di intermediari ma che sono alle loro dirette dipendenze. Il loro unico scopo è l'incremento delle rendite per i loro clienti. Paradossalmente il mercato, oggi, non è più fonte di sostentamento per le imprese: queste sono smembrate, trasferite o chiuse su decisioni delle assemblee degli azionisti, talvolta sulla base di qualche punto in più o in meno di incremento apparso su un display luminoso.

L'accusa dell'economo francese si manifesta nei dati riportati nel capitolo centrale, in cui l'autore orienta la lente d'ingrandimento sugli ingranaggi del sistema; chi sono gli azionisti? Come viene spartita la ricchezza mondiale? Un dato su tutti: il 95% delle ricchezze del pianeta viene controllato dal mercato e spartito in circa 300 milioni di azionisti. Considerando che siamo 6 miliardi, fare le debite proporzioni diventa imbarazzante. Tuttavia, ancora, il nostro autore non è soddisfatto; vuole vederci chiaro sulle declamate regolamentazioni del mercato, sulla funzione degli istituti di analisi d'impresa, i consigli di amministrazione e la struttura della *corporate governante*.

Le critiche di Peyrelevade sono ammonimenti verso questo sistema iniquo che arricchisce pochi a scapito anche degli altri: cos'è un fallimento, se alla fine si può vendere tutto? La provocazione, in questi tempi di globalizzazione, spera di raggiungere anche gli intellettuali, troppo assenti dal dibattito che dovrebbe stabilire la rotta economica mondiale; ed i governi, sempre più schiavi dell'economia finanziaria - se non conniventi nello svendere i diritti e la libertà dei propri cittadini.

INDICE

Recensione

Oikonomia

Edoardo Ferrario e AA. VV.

Lithos editrice, Roma, 2009

Riccardo Furi
riccardo.furi@humana-mente.it

Oikonomia è il titolo del nuovo lavoro di Edoardo Ferrario per la collana "Filosofia" della Lithos Editrice per la quale è autore di *Testimoniare* (2006) e *Voci della fenomenologia* (2007). Quale sua terza opera della collana, *Oikonomia* si rivela essere un progetto ambizioso ed originale che fin dalle prime pagine mostra il capillare e minuzioso lavoro critico, ermeneutico e semantico del nostro autore.

Ferrario introduce l'argomento definendo cosa si intenda con il termine introdotto in filosofia da Aristotele e composto di due parole: *oikos* e *nomos.* Si tratta delle regole o norme di amministrazione della casa, della 'proprietà', del 'fare' – un 'fare' così distinto dall'agire (*praxis*) politico.

Amministrare, trarre profitto, governare, ci fa notare Ferrario, sono ancora oggi concetti chiave che costituiscono l'essenza dell'economia (termine in cui è traslato l'originale aristotelico) e l'economia, a sua volta, non è che l'originario 'fare' (che riguarda la moltitudine delle attività umane, poiché l''economico' sempre si dà nel 'fare') in cui l'uomo produce se stesso, ovvero: l'oikonomia come ontologia dell'essere-nel-mondo.

Il volume si costituisce di due parti: la prima, frutto del lavoro del nostro autore, ha come scopo quello di dare un senso alla domanda se l'oikonomia possa sostituire l'ontologia o si possa costituire come onto-economia.

Nel cercare una risposta, attraverso un'analisi fenomenologica di pregio, Ferrario trova alleati illustri, del calibro di Heidegger e Derrida, che forniranno gli strumenti ed i primi indizi all'infaticabile ricerca dell'autore, una ricerca che si sviluppa attraverso i più importanti concetti e modelli filosofici.

La strategia attuata nel testo di questa prima parte, "Metafisica ed economia", prende le mosse dalla definizione di due caratteristiche dell'economico: la sua indefinibilità e la sua pervasività; esso è dunque indefinibile, perché nel suo senso più ampio si dissolve nella miriade delle pratiche umane; e pervasivo, perché la miriade di pratiche umane si declina molto spesso e in molte parti degli elementi dell'economico.

"Ovunque e in qualche modo si manifesti un motivo o un momento di regolazione, amministrazione, gestione, disposizione, disciplina – sono le parole stesse a dircelo – un procedimento economico è già in corso."[1]

L'obiettivo è dimostrare come sia difficile, se non impossibile, definire cosa non è economico, cosa c'è 'oltre' l'oikonomia, come l''oltre' dell'ontologia e quindi 'oltre' la metafisica heideggeriana.

Oikonomia dunque non come portato dell'ontologia, ma costitutiva di essa: precedente anche alla prospettiva teologica, la quale potrebbe essere interpretata come originaria forma d'economia 'sacralizzata'. Per Ferrraio devono così essere indagati i nessi tra oikonomia ed ontologia, oikonomia e metafisica. Avvalendosi degli strumenti della ricerca di Heidegger si

[1] E. Ferrario, *Oikonomia*, Lithos Editrice, Roma 2009, p. 14.

arriva alla conclusione, altrimenti raggiunta anche dal filosofo tedesco, che è nel pricipio di ragione di Leibniz che va maturandosi il concetto di un onto-oikonomia.

La seconda parte, costituita da una raccolta di saggi di vari autori, si occupa delle "declinazioni dell'economico nella filosofia del nostro tempo" e si avvale di contributi che fanno da approfondimento e corollario alla prima parte. Per brevità di spazio non mi è possibile citarli tutti pur se tutti molto interessanti e meritevoli di studio; vorrei però mettere in evidenza alcune tematiche, in particolare quelle trattate in "Economia e teo-logia negativa", di Giorgia Bordoni; "Etica/Economia" di Emiliano Camertoni; "Giorgio Agamben: economia e nuda vita" di Andrea De Santis.

In conclusione è doveroso precisare che il testo che si ha davanti non è per neofiti ma presuppone una profonda conoscenza delle tematiche filosofiche che riguardano la metafisica e l'ontologia; la passione suscitata da certi interrogativi può tuttavia supplire alcune mancanze e, anche se rimangono necessarie attenzione e concentrazione, ne vale la pena. Buona lettura.

INDICE

Commentary

Karl Marx's *Capital* revisited

Martin Sattler[*]

Rereading Chapter 26 (*The so called original accumulation*) in section VII (*The accumulation process of the capital*) from Karl Marx's *Capital* it became clear to me that his analysis of the historic and economic events that led to the separation between the disposal over the means of production and the disposal over the labour power was characterized by metaphoric notions and rather vague language altogether. It also became clear to me that these pages nevertheless contain a deep insight in the structure of the capitalist system and a kind of basic common sense position. This position is hidden in his analysis from the "child period" of manufacturing industry on to the year 1867, the year in which he finalizes the *Capital*. This position can be seen as a premonition and a foresight of the financial situation of our days and it also opens an insight into the structure of contemporary capitalism.

Marx derives his argument in his analysis first from book II of Adam Smith's *An Inquiry into the nature and causes of the wealth of Nations* where the liberal theoretician treats the "previous accumulation". But there is only an unclear allusion to Smith's theorem of the nature, accumulation and employment of stock. Marx, all *homo symbolicus*, is overcome subsequently by an economic-theological comparison. This comparison shows not only that Marx puts his "political economy" within the great flow of the tradition of historical ideas; it may also suggest a radical departure from this tradition. Yet the metaphors remain ambiguous. Notionally Marx introduces a term by which he points out the meaning of the separation between means of production and labour that has theological connotations.

> This primitive accumulation plays in Political Economy about the same part as the original sin in theology. Adam bit the apple, and thereupon sin fell on the human race. Its origin is supposed to be explained when it is told as an anecdote of the past. In times long gone-by there were two sorts of people; one, the diligent, intelligent, and, above all, frugal elite; the other, lazy rascals, spending their substance, and more, in riotous living. The legend of theological original sin tells us certainly how man came to be condemned to eat his bread in the sweat of his brow; but the history of economic original sin reveals to us that there are people to whom this is by no means essential. (Beginning of Chapter 26)

The separation that creates capital out of money, he continues, must proceed in the following way:

> On the one hand *(there are, M.S.),* the owners of money, means of production, means of subsistence, who are eager to increase the sum of values they possess, by buying other people's labour-power; on the other hand, free labourers, the sellers of their own labour-power, and therefore the sellers of labour. (next page)

[*] Emeritus Professor (Mannheim), director of the Heidelberg School "Hans-Georg Gadamer"

Marx continues and he states the systematic change towards the capitalist system in a somewhat dogmatic way:

> The capitalist system pre-supposes the complete separation of the labourers from all property in the means by which they can realize their labour. As soon as capitalist production is once on its own legs, it not only maintains this separation, but reproduces it on a continually extending scale. (same page)

Continually extending is the scale in which the capitalist production progresses. There are no setbacks and there is no rupture in the scale of the development of capitalist production.

So far the exposition of Marx. There exist of course craftsmen today who own their means of production. In his book "The Craftsman" Richard Sennett, 2008, deals with this phenomenon and shows clearly that the separation between means of production and labour power is not at all radically realized in our days. Maybe we shall look upon the free-lance home computer labourer as working in the "child period" with his own disposal over means of production and the disposal over his own labour power, working late at night or whenever he pleases to do so. The "extending scale" toward the capitalist system is illustrated by Marx very opulently in the subsequent sections of Chapter 26: The colonial system and state debts, the repressive load of taxes and the commercial wars characterize the rise of the capitalist system from "child period" to "large industry". He states this development in 1867. Since then the colonial exploitation (through the World Bank), the state debts in their volume, the tax load have all risen "from the sprout of the manufacturing period" or even from 1867 to our late capitalism of today in a gigantic unforeseen measure.

Later on Marx as a prophetic author touches casually the core of the capitalist accumulation process as it emerges in the original accumulation (*ursprüngliche Akkumulation*):

> It *(the colonial system, M.S.)* was «the strange God» who perched himself on the altar cheek by jowl with the old Gods of Europe, and one fine day with a shove and a kick overthrew them all. It proclaimed surplus-value making as the sole end and aim of humanity. (page 782)

The finality of all human life was, according to Marx, reduced to surplus-value making, and no other cultural activity was henceforth pursued in the capitalist system. This historical breach was seen by Marx in the most radical way and, of course, this observation is not coherent with the facts of life of the society, not at the time of the division between means of production and labour power, not at the time in which the *Capital* was written (1867), nor in the time thereafter. We know many non-surplus oriented activities in our culture: the arts, the crafts, education, etc.

If this is an exaggeration, there is still another observation of primordial character added to the surplus-value orientation. Our author analysing the "primitive accumulation" has another argument at hand to explain the history of economic politics.

> The system of public credit, i.e. of national debts, whose origin we discover in Genoa and Venice as early as the Middle Ages, took possession of Europe generally during the manufacturing period. The colonial system with its maritime trade and commercial wars served as a forcing-house for it. Thus it took root in Holland. National debts, i.e., the alienation of the State – whether despotic, constitutional or republican – marked with its stamp the capitalist era. The only part of the so-called national wealth that actually enters into the collective

> possessions of modern peoples is their national debt. Public debts become the Creed of the capital. And the emerging debts of the state stands for the «sin against the Holy Spirit», which cannot be pardoned […]. (page 783)

A notion with theological implications is used by Marx to illustrate the irrefutability of the obligation to pay back a debt. Yet the metaphor is somewhat problematic. The sin "against the Holy Spirit" is mentioned here. The biblical context of this notion is interesting. The Pharisees accuse Jesus that he wants to exorcize sin by using Belzebub, the super devil. Jesus answers, that he represents the truth and the Holy Spirit and those who oppose the Holy Spirit cannot be pardoned. It is a kind of an oath or curse that is uttered by Jesus against the doubting Pharisees (Mathew 12, 31-32). What does this notion mean when used for characterizing the state debts? First of all the implication of so strong a biblical notion means that Marx wants to state that he is absolutely sure in raising the argument of the state debts and the irrefutability of the obligation to pay them back. Yet since the middle ages the creditors believed that one could consolidate state debts by state credits, by lending out money from the treasury of the state as a stop-gap loan. For Marx that is a breech of the basic principle of logic which is underlined by his recourse to the metaphor of the "sin against the Holy Spirit". In his argument state debts form only a structural element in the primitive accumulation and not any more in the capitalist system. What he could foresee is that they do play a major role also today in our fully grown industrial societies.

> The public debt becomes one of the powerful levers of primitive accumulation. Colonial system, public debts, heavy taxes, protection, commercial wars etc., these children of the true manufacturing period increase gigantically during the infancy of modern industry. (page 783)

Not only do they increase gigantically during the infancy of modern industry as seen in the year 1867, but the "infancy" of the capitalist system lasted until our days and the elements Marx saw in the "child period" are still with us and they are growing into a super-gigantic scale in our days. The basis of the colonial system are for Marx the credit given by the state to the colonial companies, the debt-making by the state in order to provide this credit, the surety warrant or the "bail-out", as the state debt is euphemistically called today.

The structure of the economic development knows many forms that Marx would put in his systematic approach to history into the "child period", in the manufacturing phase or into the original primitive accumulation phase. The core of his "political economy" is revealed by this argument, namely that one cannot spend, what hasn't been earned beforehand. If there is a breach against this rule the "scale" of the "capitalist system" emerges "continually", thereby producing "alienation" which is the "*fall of Man*" or the "*fall of Adam*".

There are of course many breaches of the obligation to pay back state debts. There is inflation and there are "financial reforms" with the issue of a new currency that make state debts shrink and they have been adopted ever so often in economic history since 1867.

Marx appears in his analysis of the "child period of capitalism" as a surprisingly conservative economic thinker, who holds fast to the principle that one cannot spend, what one hasn't earned beforehand. *"The fall of man"* in times of *"primitive accumulation"* and ever since is a persisting drama which takes place in the breach of this obligation.

Commentary

La grande trasformazione.
Le origini economiche e politiche della nostra epoca

Karl Polanyi

Einaudi Editore, Torino, 1974

Giovanni Pancani

giovanni.pancani@humana-mente.it

Se il pensiero di Karl Polanyi desta ancor oggi un vivo interesse lo dobbiamo senza dubbio in buona parte a *La grande trasformazione:* elaborata sul finire degli anni '30, quest'opera della maturità di Polanyi è, come noto, costituita da un suggestivo e complesso intreccio di sociologia, storia, economia, politica e, sopratutto, antropologia.

Il compito di chi scrive è quindi reso particolarmente complesso dall'elaborazione dell'opera stessa, che peraltro non distingue sempre in maniera esplicita i diversi piani del discorso; questo però niente toglie alla chiarezza del volume, come alla lucidità di sguardo di Polanyi, che in quest'opera è affilatissimo.

La trattazione de *La grande trasformazione* riguarda l'individuazione del meccanismo istituzionale che ha permesso alla civiltà del XIX secolo di crollare: questo meccanismo viene individuato nel libero mercato.

Polanyi espone acutamente la nascita del capitalismo attraverso lo sviluppo e le modifiche istituzionali e legislative avvenute in Inghilterra nel secolo XVII mostrando come, da un contrasto d'interessi iniziale che opponeva parti distinte della società, si sia originata quella spinta al cambiamento che ha determinato un'inarrestabile reazione a catena. Elementi costitutivi ne sono stati la legislazione sulla povertà (le cosidette *Poor's Law*), l'*Act of Settlement* (disposizioni mediante le quali si poneva un limite al mercato del lavoro) e il loro necessario superamento, che portò alla creazione di un equilibrio instabile, del tutto nuovo e peculiare, che finì per diffondersi su scala globale.

Il libero mercato, secondo Polanyi, non avrebbe comunque potuto resistere indefinitamente all'interno del resto della società, senza che questa prendesse delle misure per difendersi; tali misure avrebbero dovuto essere una limitazione al libero mercato e la disarticolazione della vita industriale.

Anche così, però, sotto un diverso profilo, esso avrebbe messo in pericolo il resto della società: una siffatta contraddizione è quella che porterà lo sviluppo del sistema di mercato ad arrestarsi, trascinandosi dietro le società che si basavano su di esso.

La forte linearità del discorso di Polanyi a questo punto comincia ad aprirsi ad analisi storico-antropologiche più particolareggiate: se è vero, e lo è, che alcune tesi di Polanyi appaiono superate nella misura in cui sono stati superati i testi degli storici cui fa riferimento (esempio calzante è il caso delle opere di Pirenne, che restano comunque a tutt'oggi suggestive ed importanti per chiunque studi storia medievale), non è invece facile superare la lezione che Polanyi trae dalle sue vastissime letture.

Una di queste lezioni, esposta ne *La grande trasformazione,* consiste nella tesi di fondo secondo cui vi è un tratto di unicità e di innaturalità nelle società di mercato: dopo un'analisi

delle più remote origini sociali di quello che può essere considerato l'inizio del processo storico che in Inghilterra determinerà lo sviluppo della prima Rivoluzione Industriale, Polanyi sposta la sua visuale dimostrando abilmente ed in maniera stringente come i fenomeni economici poggino sempre su basi valoriali extra-economiche.

Questa idea, che non si può dire essere né facilmente contestabile né banale - o almeno non lo è *all'interno* di una società di mercato -, ci permette oggi un confronto con quelle categorie di lungo periodo che dominano i rapporti economici rendendoli, per così dire, immersi (*embedded:* letteralmente "sprofondati") nell'intera vita sociale.

Queste idee, come molte altre, danno al volume di Polanyi un valore che, come risulterà evidente, è largamente indipendente dall'analisi sociologica della dinamica degli scambi di beni da parte degli isolani delle Trobriand o dalla visione di Aristotele sul rapporto tra economia domestica ed economia di vendita, finalizzata ad un guadagno.

Convincente come ricostruzione concettuale è anche la successiva esposizione di come sia avvenuto il predominio dei rapporti economici sulla società, di cui erano parte speciale: questa assimilazione da parte del mercato, portata alle sue estreme conseguenze, rende evidentemente possibile quella contraddizione secolare che trova la sua più piena espressione nel rapidissimo e violento mutamento che le istituzioni liberali europee subirono negli anni '30.

Ed è qui il cuore delle analisi di Polanyi: l'intera struttura del volume gravita attorno al fascismo come possibile e necessario sbocco sociale di quella contraddizione secolare tra una "sostanza umana e naturale" da una parte e l'assimilazione dell'intera società da parte della sfera economica dall'altra.

Tutto ciò sarebbe stato possibile proprio a seguito del secolo della Pace dei Cento Anni, l'Ottocento, durante il quale il mercato autoregolato era divenuto la fonte e la matrice dell'intero sistema: su di esso, infatti, poggiava lo stato liberale. La base aurea, inoltre, altro non era che il tentativo di espandere il mercato su scala internazionale. A sua volta, l'equilibrio tra le potenze era diventato una sovrastruttura che si sosteneva sulla base aurea - ed in parte operante su di essa.

Il complesso equilibrio nato dalla interazione tra queste quattro istituzioni viene a erodersi ed infine a crollare con la dissoluzione del terzo polo della Triplice Alleanza e, con il concorso dell'alta finanza, cominciano quelle fluttuazioni distruttive che porteranno prima ad una serie di crisi economiche nel trentennio 1879-1909, poi al generarsi delle precondizioni alla Grande Guerra. Questa, che è ancora una guerra tra potenze ottocentesche, apre, con la sua fine, qualcosa di nuovo, un mutamento che è anche la definitiva affermazione della potenza e del dominio sociale incontrastato della sfera economica sul resto della società, per mezzo dell'abolizione della partecipazione democratica nella vita pubblica: questo, per Polanyi (che comunque ne dà varie e precise definizioni) è il fascismo.

Proprio su questo punto l'analisi di Polanyi si fa incredibilmente avvincente e credibile: del fascismo vengono colte varie sfumature - e chi volesse approfondire questo aspetto troverà un antecedente a *La grande trasformazione* in *The essence of fascism*, pubblicato ben undici anni prima, nel 1935: in quel testo, si afferma concisamente che "dopo l'abolizione della sfera politica democratica resta solo la vita economica; il capitalismo organizzato nei diversi settori dell'industria diventa l'intera società: questa è la soluzione fascista" (pp. 392).

Il fascismo si configurerebbe quindi come il prodotto di una società di mercato che si rifiuta di funzionare e che si rifiuta di farlo in base alla scoperta della stridente contraddizione che l'idea stessa di libero mercato autoregolato conteneva già *in nuce.*

Questa visione, che può apparire semplicistica, è arricchita da un'integrazione che agli occhi di un lettore italiano ha una pregnanza particolare: oltre a quanto già detto, infatti, il fascismo è anche una sorta di reazione emotiva, quasi uno sfogo isterico e momentaneo, radicato in

ogni comunità industriale. Da questo punto di vista il fascismo, contrariamente a come si è autodefinito, non è tanto un *movimento*, quanto una *mossa* o una *tendenza latente* di una società e di tutto un mondo che vede crollare i pilastri della propria vita e che vede risucchiare, insieme ad essa, anche la propria identità.

In questo senso il fascismo è, con buone probabilità, la più riuscita combinazione di nichilismo e dogmatismo mai apparsa all'interno della società occidentale nel suo complesso.

Straordinariamente penetrante, articolato e suggestivo il volume di Polanyi riesce a superare l'inadeguatezza storica di alcune sue parti, con un affresco dello sviluppo della società occidentale convincente ed addirittura profetico, come in quelle parti dove accenna alla distanza crescente tra l'uomo e la natura, ovvero a come la logica del libero scambio vada ad intaccare l'equilibrio dell'ecosistema "trasformando il suo ambiente in un deserto".

Commentary

Problemi epistemologici dell'economia

Ludwig von Mises

Armando Editore, Roma, 1988

Alberto Binazzi
alberto.binazzi@humana-mente.it

I saggi raccolti in questo volume "rappresentano il necessario e preliminare studio per quel corretto esame dei problemi che ho tentato di portare a termine in *Human Action. A Treatise on Economics*." Queste le parole introduttive di Ludwig von Mises in un testo che è considerato un classico dell'epistemologia dell'economia e dell'epistemologia in generale. Molteplici le questioni analizzate: la natura dell'induzione, il rapporto tra teoria ed esperienza, la problematica della comprensione dei fatti storico-sociali, il legame tra sociologia e storia, alcuni argomenti fondamentali di economia politica come la teoria soggettiva del valore e la convertibilità dei beni capitali. Dietro le dispute di politica economica, infatti, vi sono problemi epistemologici e metodologici non eludibili: "Non è fuori luogo far notare che le questioni astratte di logica e metodologia hanno una stretta influenza sulla vita di ogni individuo e sul destino di tutta la nostra cultura".[1] Allievo diretto di Carl Menger ed Eugen von Böhm-Bawerk, Von Mises considera compito dell'economia la formulazione di ciò che egli chiama 'leggi esatte', leggi, cioè, che non necessitano di esperienza alcuna per essere confermate o refutate. La sua è una posizione aristotelica, secondo cui si può arrivare a leggi valide semplicemente per via deduttiva. Dell'impostazione mengeriana, oltre al primato del metodo deduttivo, von Mises recupera l'idea di scienza esatta caratterizzata dal rifiuto di ogni filosofia della storia, come di posizioni positivistiche, empiristiche, collettivistiche, insieme alla primarietà del procedimento compositivo e dell'individualismo metodologico (concezione che vede nelle azioni dei singoli individui e non in entità metafisiche l'origine intenzionale o non intenzionale dei fenomeni sociali). Compito principale della sua analisi "è quello di distinguere la teoria a priori dalla storia e dalla scienza empirica, dimostrando l'assurdità dei tentativi della scuola storica e istituzionalista di riconciliare ciò che è logicamente incompatibile".[2] Rivendicando l'autonomia della scienza economica sia da approcci storicistici che sperimentali, von Mises afferma "la legittimità logica della scienza che ha come suo oggetto le leggi universalmente valide dell'azione umana, leggi indipendenti dal tempo, dal luogo, dalla razza, dalla nazionalità, dalla classe di appartenenza dell'attore economico".[3]

L'azione per Mises consiste nell'economizzare i mezzi disponibili per la realizzazione dei fini prescelti. Le scienze naturali, invece, nel loro processo di individuazione di regolarità dell'esperienza chiamate leggi, non espongono un disegno o cause finali. Allo stesso modo, i limiti degli approcci storicistici in economia si palesano nell'incapacità di fare previsioni sul futuro, così come per lo storico dell'arte è impossibile prevedere, sulla base del passato, gli stili di moda futuri. Come precisa Lorenzo Infantino nella postfazione, le teorie scientifiche sono, in opposizione a qualsiasi approccio storicista, a-teleologiche, non hanno pretese di realismo e si sottraggono alla tentazione di considerare gli eventi umani come diretti verso un destino. Le riflessioni epistemologiche di von Mises, osserva Infantino, oltre a ribadire gli errori commessi

[1] L. von Mises, *Problemi epistemologici dell'economia*, Armando Editore, Roma 1988, pp.30.
[2] Ivi, pp.27.
[3] Ivi, pp.26.

dagli approcci storicistici e neopositivisti in economia, affermano la fallibilità delle teorie scientifiche, il rifiuto della logica induttiva e il primato del dato teorico per l'indagine scientifica. Inoltre, è possibile rintracciare nella dottrina misesiana alcuni risultati anticipatori delle tesi epistemologiche popperiane.[4]

Che ruolo assegnare, oggi, alla teoria dell'azione se confrontata con i recenti sviluppi di ricerca in economia cognitiva e sperimentale? Tra le attuali linee di ricerca di questo nuovo settore di studi, troviamo il problema del rapporto tra matematica e realtà, l'analisi dei modelli di razionalità, l'applicazione di rappresentazioni mentali e mappe cognitive ai processi decisionali, i fenomeni di categorizzazione e biases, il ruolo delle simulazioni al calcolatore, l'utilizzo di tecniche di brain imaging. Settori di ricerca, alcuni, molto lontani dalla concezione aristotelica di von Mises, che, sì, enucleava l'elemento prasseologico dell'agire economico, ma, contemporaneamente, lo depurava da qualsiasi componente di natura cognitiva. In un punto le lucide argomentazioni di von Mises mantengono un rapporto di grande attualità con i recenti sviluppi cognitivi in economia: i ricercatori nel campo dell'economia sperimentale e cognitiva contemporanea sarebbero concordi nel ritenere i loro risultati più come appartenenti a una scienza cognitiva interessata alle questioni economiche, piuttosto che come una vera e propria specializzazione della scienza economica. Benché sia connotato come un approccio fortemente critico nei confronti del modello neoclassico, quello alla Kahneman-Tversky non viene considerato dagli autori stessi come indipendente dagli approcci normativi classici e né riducibile a questi. Mises è chiaro su questo punto: "L'economia comincia laddove la psicologia finisce".[5] Quest'ultima, infatti, occupandosi di fatti psichici, non ha a che vedere con la teoria dell'azione umana, priva di elementi riconducibili alla soggettività. L'economia, concepita come ramo della prasseologia, è, sì, teoria dell'azione umana, ma ne è, per definizione, azione razionale. E il concetto di razionalità cui von Mises aderisce è quello classico: non esistono azioni *a*-razionali in economia, l'azione razionale coincide con l'atto stesso di economizzare. Tutto ciò che possiamo considerare azione umana, osserva Mises, perché va oltre il comportamento meramente reattivo degli organi del corpo umano, è razionale. E ancora, "il confine che separa l'economico dal non economico, non si può cercare nell'ambito dell'azione razionale, ma coincide con la linea che separa l'azione dalla non azione".[6] A un primo sguardo, dunque, le linee di ricerca sviluppate dai cognitivisti appaiono logicamente incompatibili con la teoria dell'azione misesiana, autonoma, allo stesso tempo, sia da impostazioni matematizzanti (nota è la sua concezione dell'equilibrio come espediente teorico e non come tentativo di descrivere la realtà) che psicologiste (rifiuto di spiegazioni inerenti a dati psichici).

I rapporti tra economia e psicologia hanno caratterizzato da sempre la ricerca economica.[7] Anche Schumpeter ha osservato come gli economisti si siano sempre fabbricate a loro piacimento le osservazioni sui processi psichici che ritenevano opportune nonostante riconoscessero la presenza nell'analisi economica di problemi che avrebbero potuto beneficiare dei metodi elaborati dagli psicologi.[8] A patto, però, di definire chiaramente il confine tra le due discipline, evitando utilizzi strumentali. Per esempio, la legge dei compensi decrescenti della terra si riferisce a un fatto che si potrebbe chiamare fisico, ma ciò non significa che siamo obbligati a considerarla un'asserzione relativa alla fisica. Similmente, se affermiamo che il desiderio di cibo diminuisce in maniera inversamente proporzionale alle

[4] L.von Mises, *La mentalità anticapitalistica*, Armando Editore, Roma 1988, pp.9.
[5] Ivi, pp.32.
[6] Ivi, pp.154.
[7] M.Novarese, S.Rizzello, *Economia Sperimentale*, Mondadori, Milano 2004, pp.8.
[8] J.Schumpeter, *Storia dell'analisi economica*, Universale Scientifica Boringhieri, Torino 1976, pp.33.

porzioni successivamente ingerite, con ciò non stiamo prendendo nulla in prestito dalla psicologia, bensì formuliamo un fatto di esperienza comune. In questo senso, per Schumpeter, non c'è molta psicologia nelle proposizioni economiche. Ciò non implica che l'economia debba disinteressarsi della ricerca in psicologia, ma parlare di leggi psicologiche in economia costituirebbe un flagrante abuso. L'economista classico fa a meno di teorie psicologiche e afferma che "i teoremi dell'economia hanno il carattere di leggi scientifiche".[9]
La teoria dell'azione misesiana poiché "prende in considerazione solo l'essenziale, è formale, assiomatica e le azioni sono concepite non nella loro forma concreta, ma come costruzioni formali".[10] Essa "è a-priori, non empirica. Così come la logica e la matematica, non deriva dall'esperienza, è antecedente all'esperienza. È, per così dire, la logica dell'azione e dell'atto".[11] È evidente qui l'influenza kantiana nella concezione di attore economico che conosce "ciò di cui la natura e la struttura dei nostri sensi e delle nostre menti ci rendono conto" e nella fede in una forma logica immutabile e senza tempo della ragione. In che modo, però, dobbiamo recepire il tentativo misesiano di enucleare l'elemento *in*tenzionale, e ritenerlo, allo stesso tempo, 'a priori', ovvero, precedente a qualsiasi esperienza umana? Nemmeno gli assiomi della matematica e della logica sono, ad oggi, considerati a-priori, figuriamoci quelli dell'economia. Si pensi a Quine, per esempio, filosofo che ha affermato come la distinzione tra verità analitiche e sintetiche, se applicata senza restrizioni a tutti gli enunciati teorici, sia, in realtà, una scomposizione invalida da un punto di vista logico. E fanno bene, evidentemente, a dubitarne gli economisti cognitivi e sperimentali che hanno avuto il merito di mettere in primo piano la complessità della nozione di razionalità, termine, questo, fondamentale per tutto il pensiero economico. È utile notare come le varie linee di ricerca, soprattutto in ambito microeconomico, siano riconducibili a differenti impieghi della nozione di razionalità. La concezione classica di razionalità assume, infatti, la forma della soluzione di un problema di ottimizzazione in relazione alla ricerca dell'interesse personale. Tale modello ha subito notevoli trasformazioni individuabili in due deviazioni interne al paradigma neoclassico: il concetto di razionalità del rapporto mezzi-fini, che utilizza modelli probabilistici (sia soggettivistici che oggettivistici) e il concetto di razionalità limitata, per cui la razionalità - procedurale e agita in situazioni di informazione non disponibile senza costo - non può essere definita in termini di ottimalità. Infine, un terzo filone, che indaga la natura strategica del concetto di razionalità opponendosi alla visione tradizionale o parametrica, si serve, in massima parte, della teoria dei giochi. Anche von Mises critica la concezione classica dell'*homo oeconomicus*.[12] L'economia moderna non comincia dall'azione dell'uomo d'affari, ma da quella dei consumatori, dall'azione, cioè, "di tutti". In questa posizione siamo forse autorizzati a individuare una qualche affinità con i recenti sviluppi cognitivi? Se prestiamo ascolto al recente premio Nobel Vernon L. Smith e alla sua severa critica al modello dell'uomo di Chicago, probabilmente sì, ma, domandiamoci, von Mises ne avrebbe accolto facilmente le relative implicazioni? Sebbene i contributi di von Mises siano considerati tuttora in gran parte attuali, dato che "molti dei suoi temi principali – i diritti di proprietà, le regole della responsabilità, l'efficacia dei mercati, l'inutilità dell'interventismo, la supremazia dell'individuo – sono divenuti degli elementi importanti nella teoria e nella pratica microeconomica", egli non poteva immaginare come si sarebbe sviluppato cinquanta anni più tardi lo studio delle decisioni

[9] L.von Mises, op.cit., pp.104.
[10] Ivi, pp.41.
[11] Ivi, pp.40.
[12] Ivi, pp.179.

umane.[13] Dopotutto, osserva Smith, la concezione di von Mises riflette l'approccio metodologico universale del suo tempo che considerava l'economia necessariamente una scienza non sperimentale.[14] Tale lettura retrospettiva potrebbe non rispecchiare adeguatamente la posizione di von Mises, in quanto, lo abbiamo già ricordato, egli concepiva l'approccio sperimentale in economia, al pari di quello storicista, logicamente incompatibile con la teoria dell'azione umana. Se consideriamo legittima quest'ultima obiezione, in che modo valutare affermazioni come "l'economia sperimentale è di forte sostegno alla teoria di von Mises dei prezzi di mercato, ma anche alla teoria dell'equilibrio in condizioni stazionarie o in cambiamento dinamico"?[15] Conosciamo la posizione critica di Mises nei confronti del concetto di equilibrio ed egli non poteva certo immaginare che i colleghi sperimentali avrebbero cercato di verificarla in laboratorio; ma è proprio questo l'obiettivo di molta ricerca in economia sperimentale: portare alla luce tramite l'esperimento controllato in laboratorio, l'elemento di "costruzione ideale" di molte teorie economiche, tra cui, appunto, l'equilibrio. Von Mises considerava l'azione umana come consapevolmente intenzionale, e, allo stesso tempo, concepiva i mercati capaci di autoregolarsi indipendentemente dalle scelte consapevoli e deliberate degli individui. Con la sua teoria dell'azione, la dottrina di von Mises ha influenzato profondamente gran parte del pensiero economico del liberismo riformista anglosassone. Basti pensare a Lionel Charles Robbins, economista britannico suo contemporaneo che nel celebre *Saggio sulla natura e sul significato della scienza economica* definirà l'economia come lo studio del comportamento umano caratterizzato da un generale imperativo di economicità poiché dominato dalla privazione. La natura dinamica e volitiva dell'agente economico è un tratto costante del pensiero misesiano. In *La Mentalità anticapitalistica*, saggio del 1956 in cui analizza le radici e le conseguenze del diffuso pregiudizio anticapitalistico, von Mises afferma che gli uomini destinati a prevalere sugli altri sono quelli "che meglio riescono a fare i conti con l'incertezza insita nel mutamento istituzionalizzato" e che non temono di confrontarsi con la preoccupazione derivata dalla imprevedibilità del futuro.[16] Il tratto più caratteristico dell'uomo è che "egli, con un'attività a ciò finalizzata, non cessa mai di cercare il miglioramento del proprio benessere".[17] E per realizzare i propri scopi, oltre alla necessità "di respirare aria libera", depurata, cioè, da pregiudizi ideologici che hanno l'unico obiettivo di introdurre un finalismo coattivo nella vita degli uomini, l'attore economico misesiano deve essere in grado di effettuare scelte razionali. I recenti contributi in economia cognitiva hanno evidenziato come, in realtà, i comportamenti degli attori economici non siano sempre sotto il controllo del pensiero razionale e, soprattutto, intenzionale. Come riconosce Smith, "Mises sottovalutava enormemente il funzionamento dei processi mentali inconsci".[18] La ricerca neuroscientifica ha dimostrato come anche i fenomeni di presa di decisione in situazioni di incertezza siano affrontati dal cervello al di sotto di un livello di accessibilità cosciente. A coloro che si affidano alla supremazia della ragione nella teoria della scelta, risulta arduo accettare il ruolo (evolutivo) esercitato dalle emozioni nell'azione umana. Una buona parte di esseri umani, evidentemente, non sceglie in base a paragoni basati su preferenze, anche se in determinati contesti sarebbe opportuno. Osserva Smith, "i loro cervelli conservano risorse di pensiero attenzionale, concettuale e simbolico

[13] V.L.Smith, *Riflessioni su L'azione umana*, *cinquant'anni dopo*, Istituto Bruno Leoni Occasional Paper, brunoleonimedia.servingfreedom.net/OP/61_Smith.pdf, n.61, Novembre 2008, pp.2.
[14] Ibidem.
[15] Ivi, pp. 3.
[16] L.von Mises, *op.cit*, pp.14.
[17] Ivi, pp.25
[18] V.L. Smith, *op.cit*., pp.4.

perché sono scarse le procedure per delegare la maggior parte delle decisioni a processi autonomi (incluse le emozioni) che non richiedono attenzione conscia".[19] Si potrebbe perfino discutere, alla luce dei risultati delle neuroscienze contemporanee, se le nostre decisioni siano sempre guidate da contenuto mentale, o per dirla con Mises, se siano sempre precedute da pensiero. Purtroppo, non abbiamo oggi una teoria unitaria del pensiero che ci consenta di argomentare adeguatamente sull'attualità di queste posizioni. Ciò che sembra cominciare a farsi strada è una concezione ecologica di razionalità che distingue tra una razionalità teorica e una evolutiva che impara dagli errori. Il dibattito sulla razionalità degli attori economici, suggeriamo, dovrebbe essere esteso anche a quella degli economisti stessi: le recenti ricerche in economia cognitiva, infatti, hanno avuto il merito di portare alla luce i noti errori di introspezione dei teorici che attribuiscono -su base assiomatica- una completa informazione agli agenti economici.

Consapevoli della necessità di un maggiore approfondimento di quello qui sinteticamente preso in esame, possiamo considerare il testo di Mises come portatore di idee e concetti di grande attualità, soprattutto per la sua 'proiettabilità' verso settori di ricerca anche molto lontani da quelli che Mises stesso avrebbe potuto (o voluto) immaginare.

[19] V.L.Smith, Razionalità costruttivista e razionalità ecologica in *Critica della ragione economica*. *Tre saggi: Kahneman, McFadden, Smith*, a cura di Matteo Motterlini e Massimo Piattelli Palmarini, Il Saggiatore, Milano 2005, pp.205.

Commentary

Sulla natura e l'importanza della scienza economica

Lionel Robbins

Torino, UTET, 1947

Scilla Bellucci

scilla.bellucci@humana-mente.it

Lionel Robbins, con quest'opera, propone una tesi per la fondazione della scienza economica che, oltre ad essere stata accolta dalle tradizioni successive con un ampissimo margine di approvazione, intende emancipare la stessa attraverso un riconoscimento alto: di disciplina autonoma e scientificamente fondata.

Alla data di pubblicazione di questo che lui volle chiamare 'saggio', il 1932, il pensiero economico, infatti, aveva ancora forti connotazioni di carattere politico e sociale da cui sentiva la necessità di svincolarsi, non certo perché ne risultasse un processo meno ricco o decisamente arido, bensì per poter acquisire strumenti di valutazione che comportassero un maggior rigore logico e scientifico. Robbins tentò quest'operazione affidandosi al metodo deduttivo, per indicare una via d'indagine che potesse portare nella direzione del risultato desiderato. Per quanto ciò possa apparire in contrasto con le teorie di alcuni economisti, le cui eredità di pensiero sono evidentemente presenti in Robbins (Wieser, Weber, Keynes, tra gli altri), egli effettivamente organizzò in un sistema idee che già nei decenni precedenti erano venute a costituirsi come 'scienza economica', ma che, prive di una reale e coerente sistematizzazione, non potevano dirsi tali.

A fondamento della sua teoria, l'autore pose una definizione di economia che è rimasta famosa fino ad oggi: «L'Economica è la scienza che studia la condotta umana come una relazione tra scopi e mezzi scarsi applicabili ad usi alternativi».[1]

Recuperato il concetto di scarsità dal pensiero di Auguste Walras, l'autore individua i "quattro caratteri fondamentali" del comportamento umano che segnerebbero il confine del campo di interesse dell'economista:

> Gli scopi sono molteplici; il tempo e i mezzi per conseguirli sono limitati e sono capaci di usi alternativi; nello stesso tempo, gli scopi hanno diversa importanza...Ora, la molteplicità degli scopi non ha *in sé* un necessario interesse per l'economista...Né la solo limitazione dei mezzi è *per sé* sufficiente a dare origine a fenomeni economici... Né, ancora, l'applicabilità alternativa di mezzi scarsi è condizione sufficiente...Ma quando il tempo e i mezzi per conseguire gli scopi sono limitati e sono suscettibili di applicazione alternativa, e gli scopi possono essere distinti in ordine d'importanza, allora la condotta assume necessariamente la forma di una scelta...ha un aspetto economico.[2]

[1] Lionel Robbins, *Sulla natura e l'importanza della scienza economica,* Torino, UTET, 1947, p. 20.

[2] Lionel Robbins in *Op.cit*., pp. 16-17.

La semplicità dei postulati che reggono la struttura basilare del ragionamento economico sarebbe tale da assurgere ad una sorta di autoevidenza indiscutibile, non necessitante di prove empiriche, essendo frutto delle quotidiane esperienze di vita. A tal proposito è possibile muovere una critica ai principi che validerebbero tale ragionamento: Robbins, infatti, inserendosi in una prospettiva soggettivista, si affida all'introspezione, trascurando di chiarire con maggior efficacia il ruolo della verifica. Desta curiosità che proprio questo aspetto sia stato trascurato, o meglio osservato da un punto di vista poco 'scientifico', nel complesso di un lavoro che tende ad allontanare l'economia dai lacci che la legano ad altre discipline e dalle implicazioni di ogni pensiero morale e valoriale.

L'autore, pur concentrandosi sulla relazione tra mezzi e fini, infatti, disconosce per entrambi la possibilità di costituirsi ad oggetto d'interesse specifico per l'economia, giacché niente può permettere all'economista di esprimere un giudizio sugli uni o sugli altri. Non è il principio dell'economicità che determina necessariamente il comportamento umano in vista del raggiungimento di uno scopo, anzi, molti altri fattori, quali la politica, la religione e simili, si rivelano ben più influenti nell'operare una scelta piuttosto che un'altra.

Ciò che può risultare incoerente è l'ammissione, da parte di Robbins, di alcuni elementi di valutazione che poco sembrano rispettare un'obiettiva scientificità. Egli sostiene, infatti, che «..benché non sia vero che le proposizioni dell'economia analitica siano fondate su di una qualsiasi particolare psicologia, tuttavia esse implicano incontrovertibilmente elementi di natura psicologica - o per meglio dire, psichica».[3] Le spiegazioni che vengono addotte a giustificazione di questa inclusione concettuale appaiono mancanti di una logica davvero efficace. Come nel caso succitato della superfluità di prove empiriche per determinare la validità dei postulati teorici, anche qui si ha un riferimento forte all'interiorità e addirittura ad elementi psichici che non vengono mai qualificati in modo chiaro.

> Dopo tutto, il nostro compito è di spiegare certi aspetti della condotta individuale, ed è molto disputabile che questo possa farsi in termini che non implichino nessun elemento psichico. È certissimo che, sia o non sia con ciò soddisfatto il desiderio della massima austerità di linguaggio, certi vocaboli come scelta, indifferenza, preferenza, e simili, noi in realtà li *intendiamo* in termini di esperienza interiore.

E ancora

> Non è in realtà possibile intendere i concetti di scelta o delle relazioni fra mezzi e scopi, che sono i concetti centrali della nostra scienza, in termini di osservazioni di dati esterni. Il concetto di condotta rivolta ad uno scopo, nel nostro senso, non implica necessariamente un indeterminismo finale; ma contiene, nella catena della spiegazione causale, anelli che sono psichici, non fisici...Il fatto che questi dati abbiano essi stessi natura di giudizi di valore non rende necessario che debbano essere valutati come tali. Essi non sono giudizi di valore compiuti dall'osservatore. Ciò che importa alle scienze sociali non è se i giudizi individuali di valore siano *corretti* nel senso definitivo della filosofia del valore, ma se essi siano *formati* e se siano anelli essenziali nella catena della spiegazione causale.[4]

Se, da una parte, l'esposizione di questo pensiero è priva di una struttura filosoficamente accettabile, giacché niente giustifica come si possano assumere in qualità di 'dati' elementi interiori, né come si proceda al fine di ottenere da un'iniziale giudizio di valore un elemento neutro di valutazione causale, è altresì sorprendente notare come in un'opera così datata

[3] Ibidem, p.105.
[4] Ibidem, pp. 108,109.

siano presi in considerazione aspetti che oggi costituiscono parte di molta ricerca in campo economico.

Rifiutando di assumere un metodo di indagine meramente matematico e contrapponendosi già al modello classico di *homo oeconomicus*, Robbins si avvicina ad una concezione dell'uomo e dei suoi processi mentali meno astratta che in passato. Sebbene l'idea di scienza economica come disciplina indipendente sia alla base dell'intero lavoro, oggi ci è possibile considerare quest'aspetto in una prospettiva diversa, anacronistica, se si vuole, ma nel senso di una previsione e di una preparazione. I progressi contemporanei delle neuroscienze, infatti, hanno finito con l'interessare direttamente gli studi sul comportamento economico degli individui e i loro risultati vengono applicati alla ricerca economica in virtù di un mutamento concettuale a cui, leggendo le sue parole, si può credere l'autore non si sarebbe trovato impreparato.

Altro tema che emerge dal testo è il rapporto tra Economia e Tecnologia; Robbins rimane famoso, appunto, anche per aver trattato questo aspetto con la solita energia e per aver formulato, in proposito, un pensiero specifico. Egli assume che anche la tecnica debba essere considerata un 'dato', poiché da essa dipendono alcuni metri di valutazione. Tale affermazione dipende ancora da una visione ampia riguardo a ciò che può interessare le scienze economiche: talvolta la tecnica (o tecnologia) produce oscillazioni nei prezzi delle merci, talvolta da essa dipendono occupazione o disoccupazione, ma sempre la «La storia degli strumenti è la storia del genere umano»5. In nota a questa proposizione leggiamo

> Le distinzioni da me adottate in questo punto sono simili a quelle adottate dal dott. STRIGL...[Egli] cerca di presentare la Interpretazione materialistica come una teoria primitiva di ciò ch'egli chiama *Datenänderung*(mutamento dei dati)..tende a nascondere la deficienza insita in quella dottrina in quanto rifiuta di tener conto dei mutamenti nelle valutazioni finali, salvo a considerarli come derivazioni da cambiamenti dal lato dell'offerta. Io, invece, desidero vivamente di mostrare la distinzione fondamentale che passa fra qualsiasi spiegazione della storia sorgente dall'analisi economica, quale noi la conosciamo, e la spiegazione tentata dall'Interpretazione materialistica.[6]

Non è difficile vedere in queste parole un'intenzione che si pone al di là delle dispute accademiche, pur dovendosene occupare diffusamente. Tutto il testo è costellato di osservazioni che si oppongono o seguono le teorie di pensatori contemporanei a Robbins. Questo ci rende misura di quanto allora fosse accesa la disputa intorno alla 'natura delle scienze economiche'; di quanto non si sia trattato di un impulso solo individuale, ma sociale e culturale, riuscire in quest'impresa.

L'altra osservazione che è possibile proporre, riguarda la il pensiero intero dell'autore. Vero è che egli tenta di sfuggire le maglie di queste altre suggestioni, ma, contemporaneamente, è impossibile pensare alle sue affermazioni come slegate dal fermento culturale e politico dell'epoca, cui, peraltro, sono lapalissianamente intrecciate. In risposta alla 'Interpretazione materialistica' che vorrebbe i mutamenti negli scopi e nelle valutazioni dipendere dai mutamenti nelle condizioni tecniche di produzione e offerta, leggiamo:

> Tutto ciò può esser vero o falso, chiacchiericcio pseudo-hegeliano o intuizione profonda ...È una proposizione di ordine generale circa la causazione dei moventi umani, che, dal punto di vista della Scienza economica, è pura metafisica... L'Economica può ben fornire un importante strumento per la spiegazione della storia; ma non v'è nulla nell'analisi economica che ci autorizzi ad affermare che tutta la storia debba essere spiegata in termini 'economici', se il vocabolario 'economico' viene adoperato come equivalente di tecnicamente materiale. L'Interpretazione

[5] Ibidem, p.52.
[6] Ibidem, p.52.

materialistica della storia è giunta ad esser chiamata Interpretazione economica della storia, perché si pensava che oggetto dell'Economica fossero 'le cause del benessere materiale'. Una volta che si è compreso che ciò non è, l'interpretazione materialistica deve reggersi da sé o cadere; la scienza economica non porge nessun sostegno alle sue dottrine, né postula in nessun punto le relazioni ch'essa afferma. Dal punto di vista della Scienza economica, i mutamenti nelle valutazioni relative sono dati primari.[7]

[7] Ibidem, pp. 53 e segg.

Commentary

Storia della economia

John K. Galbraith

Rizzoli Editore, Torino, 1990, 368 pp.

Il grande crollo

John K. Galbraith

Bollati Boringhieri, Torino, 1991, 218 pp.

Giovanni Pancani

giovanni.pancani@humana-mente.it

Nonostante il notevole numero di storie dell'economia pubblicate negli ultimi venti anni, il volume di Galbraith rimane, a suo modo, una lettura utile: a differenza, infatti, di molte opere simili la *Storia della economia* di Galbraith unisce due aspetti diversi che sono solitamente separati, ovvero l'approccio storico e quello teorico.

Come è chiarito nel primo capitolo (*Uno sguardo al paesaggio*) il tentativo di Galbraith è esplicitamente quello di considerare l'economia come un riflesso del mondo in cui si sono sviluppate quelle stesse teorie: in questo modo Adam Smith è letto nel contesto delle prime fasi della rivoluzione industriale, che a sua volta è interpretata attraverso alcune delle idee centrali di Smith.

Questo approccio circolare definisce bene cosa Galbraith intenda per storia dell'economia: non una storia degli economisti e del loro pensiero (o perlomeno non solo), ma una storia integrata che esamini il passato in tutte le sue variabili economiche e storiche principali.

Nello sviluppo del testo sono quindi presi in considerazione aspetti diversissimi: partendo dai primi tre capitoli dedicati rispettivamente al mondo classico, al medioevo ed al nascente mercantilismo, si arrivano ad esaminare parti de *La ricchezza delle nazioni* di Smith. Da questo punto in poi il testo si fa progressivamente più denso: in effetti la costruzione del volume di Galbraith è armonicamente progressiva ed il testo diviene sempre più ricco e denso col passare dei capitoli; una vasta parte è quindi dedicata all'economia del '900, che occupa oltre un terzo del volume, mentre quasi altrettanta è occupata dalla seconda metà dell'800.

Galbraith ci guida attraverso la formazione del pensiero economico liberale classico, con le sue lacune ed ipostatizzazioni, la Legge Bronzea dei salari di Ricardo e la legge di Say, fino all' "energico e penetrante" dissenso di Karl Marx, (cui è dedicato un capitolo, l'undicesimo, breve ma molto istruttivo); tutto questo, assieme all'inevitabile incontro- scontro tra Marx e Keynes, la lucidità caustica di Veblen, il rigore di Von Mises ed Hayek, la struttura e la natura del capitalismo e moltissimi altri temi sono raccontati da Galbraith in modo leggero ma mai vuoto, anche ricorrendo ai ricordi personali di Galbraith stesso, che nella sua lunghissima vita professionale ha conosciuto direttamente molte delle personalità descritte negli ultimi dieci capitoli almeno del volume.

Complessivamente l'opera è molto godibile, la lettura risulta abbastanza scorrevole ed esente da dettagli tecnici, anche se la parte riguardante la moneta e la sua evoluzione (il cap. XII) è piuttosto articolata: la ricostruzione storica, come detto, è anche (e forse sopratutto) concettuale e questo dona al volume una certa compattezza.

Il volume si chiude con due brevi capitoli, in cui si tirano le fila del superamento del pensiero keynesiano e della diffusione delle idee di Milton Friedman.

La distinzione, oggi scontata, tra micro e macroeconomia, la persistenza ortodossa del pensiero classico nella microeconomia e di come questo pensiero sia andato in crisi con la complessità crescente dell'industria stessa, dove l'equivalenza classica del costo marginale dei salari e del ricavo marginale "diventa una caricatura inverosimile": tutti questi temi trovano un abbozzo nella parte conclusiva del volume.

Quella di Galbraith è un'opera che pur essendo stata scritta oltre vent'anni fa, mantiene ancora una freschezza notevolissima, specialmente nei suoi capitoli finali, dove si fa più intensa la riflessione sulla natura dell'economia, del mercato e della burocrazia aziendale.

Un tema particolarmente importante affrontato in questi ultimi capitoli è quello relativo al crac del 1929: viene ricostruita l' opinione e l'atmosfera allora prevalente tra gli economisti e sono descritte le conseguenze sociali prodotte dal crac.

Questo fondamentale tema è però meglio affrontato ne *Il grande crollo:* nella vastissima letteratura sulle cause e lo svolgimento del crollo di Wall Street dell'ottobre 1929 il testo di Galbraith si è imposto fin dall'uscita, alla fine del 1954, come un classico.

Galbraith stesso, infatti, essendo nato nel 1908 e soprattutto avendo avuto un ruolo politico sotto la presidenza Roosevelt disponeva nei giorni del peggior disastro economico della storia moderna di un osservatorio privilegiato: giovanissimo, a soli 32 anni, era stato nominato responsabile del contenimento dei prezzi da parte del governo.

Il testo, la cui composizione fu stimolata nientemeno che da Arthur Schlesinger Jr., si compone di dieci capitoli preceduti da una breve prefazione e da una divertentissima introduzione, scritta nel tipico stile brillante ed affilato di Galbraith che ricostruisce la genesi del libro ed insieme alcune vicende capitategli nello stesso periodo, fra cui un incontro/scontro con Joseph Mc Carthy.

La storia che prelude al crollo di Wall Street è per alcuni aspetti sorprendente simile a quella che gli Stati Uniti – e conseguentemente il mondo intero – si stanno trovando a vivere oggi: come oggi, infatti, all'origine della recessione del 1929 c'è lo scoppio della bolla immobiliare in Florida (una bolla che però oggi coinvolge numerosi paesi europei), e come allora si ripropongono gli stessi metodi, come quello adottato da Charles Ponzi, un imprenditore di Boston che truffò con il boom immobiliare moltissime persone, un metodo portato oggi alle estreme conseguenze nel caso del crac Madoff, in cui sono stati bruciati ben 50 miliardi di dollari.

Altrettanto e forse ancora più interessante è il fatto che l'ultima parte della presidenza Hoover assomigli notevolmente alla fine della presidenza di George W.Bush, un fatto questo recentemente messo in luce tra gli altri anche dal premio Nobel per l'economia Paul Samuelson.

Il boom azionario americano ebbe il suo impulso decisivo probabilmente ad effetto di una rivalutazione della sterlina inglese nel 1927, decisa dal governo Churchill: questa manovra determinò una liquidità improvvisa a disposizione di chiunque volesse investire in Borsa negli USA.

Similmente a quello che era successo in Florida, dove si era riusciti a svincolare il diritto all'investimento dagli oneri del possesso concreto attraverso il meccanismo del compromesso – un sistema che liberava da molti rischi di mercato, almeno inizialmente – così anche nella Borsa di New York accadde qualcosa di simile con i cosiddetti "titoli a riporto". Il mercato fu

così libero di girare a vuoto, senza rischi apparenti: moltissime banche accordarono credito agli speculatori ed ai loro agenti di cambio, fin quando il sistema non cominciò a mostrare delle crepe.

Negli ultimi mesi del 1928 il mercato cominciò a dare segnali di nervosismo e cominciarono a manifestarsi quelle ambiguità e compromissioni che porteranno al crac dopo poco più di un anno.

Il segretario del Tesoro Andrew Mellon perseguì una politica di completo *laissez-faire*, mentre l'unica manovra davvero radicale che avrebbe potuto forse sgonfiare la bolla senza farla esplodere, ovverosia la richiesta al Congresso dell'autorizzazione ad interrompere direttamente le contrattazioni a riporto riponendone le condizioni di legalità nelle mani del Congresso stesso al fine di un regolamentazione non fu attuata da parte dell'organo competente, il Consiglio dei Governatori della Riserva Federale, esso stesso in parte composto da speculatori.

La Borsa era considerata sempre meno come la registrazione di prospettive aziendali di lungo termine e sempre più come il luogo in cui si compivano macchinazioni artificiose, dove si creavano dall'oggi al domani enormi ricchezze sulla base di "voci" e di informazioni riservate.

Questo comportò un distacco crescente tra l'informazione pubblicamente disponibile sull'andamento del mercato (informazione che si scoprì in seguito essere infiltrata dagli stessi poteri economici che erano coinvolti nelle operazioni di Borsa) e la realtà.

Al culmine del nervosismo un gruppo formato dai maggiori finanzieri della nazione decise di comprare in modo massiccio tutti i titoli sottovalutati, salvando il mercato per quello che può essere considerato il canto del cigno della Borsa di New York: soltanto pochi giorni dopo di fronte ad una nuova ondata di panico, lo stesso gruppo di finanzieri si riunì e decise esso stesso di vendere. Fu il panico.

Il prezzo del Martedì Nero di Wall Street, come noto, fu durissimo da pagare sia per gli Stati Uniti che per il resto del mondo: dieci anni di depressione economica, disoccupazione di massa ed il sorgere del nazismo in Germania.

Il testo di Galbraith ha esplicitamente valore come testimonianza storica e non si occupa delle conseguenze sociali del 1929, come invece accade in *Storia della economia*; come testo di storia dell'economia tenta però di metterci in guardia da possibili nuovi casi di tracollo della Borsa e dalle loro nefaste conseguenze; a questo proposito è assai istruttivo leggere le ultime pagine dell'ultimo capitolo, *Cause e conseguenze* per trarne i debiti insegnamenti sulla natura e lo sviluppo del capitalismo.

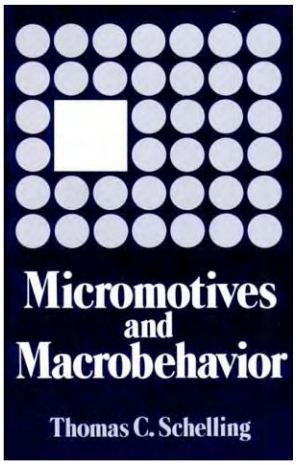
Commentary

Micromotives and Macrobehavior

Thomas Schelling

W.W Norton and Company, 1978

Paolo Maffezioli[*]



My goal is not so much to provide a summary of Thomas C. Schelling's *Micromotives and Macrobehavior*, but rather to try to clarify some of the issues that Schelling deals with from a philosophical point of view. Two questions immediately arise. First, What sort of perspective do we have in mind when we speak of providing a philosophical analysis of *Micromotives and Macrobehavior?* Second, is this perspective adequate for offering an account, as far as possible complete, of a book with such content?

Indeed, *Micromotives and Macrobehavior* seems nothing but an informal introduction to formal methods in social sciences. Such formal methods are the tools used to analyse the main subject matter of the book: the interrelationship between the individuals and the social aggregate they comprise. In particular, Schelling's project is the theoretical explanation of how individual purposive behaviour (micromotives) can determine group equilibrium or a disequilibrium process (macrobehavior). In other words, the book «explores the relation between the behaviour characteristics of the *individuals* who comprise some social aggregate, and the characteristics of the *aggregate*».[1]

Both these concepts, individuals and aggregate, deserve some words of clarification. On the one hand, Schelling's individuals are goal-directed, namely, they behave in a way that we might call purposive: «they have preferences, pursuing goals, minimizing effort or embarrassment or maximizing view or comfort».[2] Rarely do people care about the outcome of the aggregate; typically their own decisions and their own behaviour are motivated by their own interest. In this perspective, the entire book can be read as a sort of generalized Prisoner's Dilemma: behind any action there is a utility-based reason to act. This is a purely game-theoretical assumption. Schelling argues that, even if it is sometimes misleading, it is indispensable for the analysis of the relationship between individual goals and aggregate behaviour.

Once the game's scenario is fixed, some interesting consequences follow. One is that privately optimized decisions might lead in aggregate to outcomes that are undesirable for everyone. As Schelling says at the beginning, «there are several reasons we might interest ourselves in what it is that those people [are] doing, or [think] they [are] doing, or [are] trying to do [...]. One is that we do not like the result».[3]

On the other hand, the aggregate, which results from goal-directed individual behaviour, is not merely an extrapolation from the individual. Of course, there are some easy cases in which

[*] PhD Student, University of Florence

[1] Thomas C. Schelling, *Micromotives and Macrobehavior*, W. W. Norton and Company, 1978. p.13. Italics in original.

[2] Thomas C. Schelling, *cit.*, p. 17.

[3] Thomas C. Schelling, *cit.*, p. 12.

this happens, e.g. the case in which I choose an action no matter what the others decide to do. For instance, if we know that every driver turns his lights on at sunset, it is easy to guess what the features of the aggregate are: all the car lights in a local area will turn on at about the same time. But the most interesting situations are those in which people's behaviour depends on the behaviour of other people, that is, when their behaviour is not (or not only) constrained by some kind of external resource but also by the behaviour of other individuals. For instance, in the example above, it is no longer trivial to derive the features of the aggregate if we know that every driver turns his lights on only when the oncoming cars have already done so. The aggregate is now a sort of system of interactions in which individuals respond to an environment, which include other people's responses to their own environment.

Another important feature of the micromotives underlying individual actions is the fact that they are not market motivations. Schelling writes that quite a number of activities operate almost entirely outside the purview of the market, so they cannot even be expected to receive market-like benefits or treatment. Some examples are the languages we choose to speak, the person we decide to marry, whom we live near, which games we like to play, etc.

Schelling claims that the behaviour he is concerned with is not economical *strictu sensu*. His interest is in aggregate behaviour when market mechanisms are not operating, namely, in all those situations in which those who are involved are not voluntary participants: traffic jams, sex-distribution in campus dining hall, sending of Christmas cards, membership in bridge or tennis clubs, racial composition of neighbourhoods, and so on.

A very illuminating example of this kind of behaviour is the one Schelling begins his book with: the commonly observed seating pattern in a public lecture hall, where the audience is packed into the back rows leaving four, five or six rows in front of the lecture hall completely empty. Any attempt to explain such a situation using the techniques of economics will fail; not because these techniques are inadequate, but because the explanation is, in a sense, incomplete.

The reason for this failure is that the individuals are not involved in that situation in virtue of the fact that they want to be involved, but it simply so happens that they are. Schelling's aim is then to enlarge the scope of social science beyond the analysis of economic behaviour and to show us that «economics [is] a large and important special case, rather than a model for all social phenomena».[4]

Let me sketch in short the sequence of the book's topics.

CHAPTER ONE is a sort of synopsis of the entire monograph and it is definitely the most interesting part for the purpose of giving, as we said above, a philosophical account of Schelling's work. The Chapter presents what is to be considered the main question of the book: if the individuals act purposively but the behaviour of others either constrains what the individuals can do or is part of the individuals' preferences, what kind of aggregate behaviour will result?

CHAPTER TWO develops the models of nine situations where the behaviour of the aggregate is a logical necessity, unaffected by the specific goals or behaviour of the individuals. Indeed, they do not even merit the appellative of "models"; they are merely "definitional statements": their being true does not depend on the way individuals act or on what their goals are; the truth of such statements derives directly from the definition of the terms that occur within such statements.

[4] Thomas C. Schelling, *cit.*, p. 27.

For instance, telephoning is a typical activity through which one can influence the behaviour of other people. Now, surely it is true that none of us ever makes just as many calls as he receives, but it is also true that if the system of calling people on the phone presupposes a sender and a receiver, then the number of incoming calls is equal to the number of outcoming ones. Therefore, in the aggregate, it is true that the total number of received calls equals the total number of made calls, although it could happen (and in a large number of cases it really does happen) that we individually receive more calls than we make, or vice versa.

This example illustrates an important feature of a large class of statements, which the economists are very interested in: «propositions that are true in the aggregate but not in detail, and are true independently of how people behave».[5] The reason of the emphasis on such statements is that economics is manly concerned with exchanges of equivalent values.

Although we don't need any mathematical tool to recognise that the example of telephoning is true, Schelling points out that even in this case one finds surprises in the relationship between the micromotives level and the macrobehavior level. «Considering how banal these propositions sound, it is astonishing how many hours of committee meetings have been spent on proposals to mix men and women in dormitories, or blacks and whites, or freshmen and sophomores, in ways that violated the simple arithmetic principle that no matter how you distribute them, the numbers in all dormitories have to add up to the numbers that there are».[6] When the social situations become even a little more complex, intuition is not enough to realise that there are very simple and inescapable mathematical laws, like the one mentioned above, which underlie these situations.

CHAPTER THREE introduces some models, or families of models, in which individual goals and behaviours do affect aggregate behaviour. Schelling's analysis primarily focuses on such models.

One family is that of the "critical mass" models. Critical mass models describe a large number of phenomena that become self-sustaining once the level of activities passes some minimum level. «What it is common to all [critical mass situations] is the way people's behavior depends on *how many* are behaving a particular way, or how much they are behaving that way».[7] Schelling gives the academic example of a "dying seminar", i.e. the seminar whose popularity depends on its popularity; it will live or die, depending upon whether the sufficient condition is satisfied, namely, upon whether those taking part in it consider the number of other participants sufficient, relative to some standards. In other words, if enough other people are doing something, we will make the same choice. No one wants to be the only person at a club. So we go to a club only if we expect enough other people to be there to make it enjoyable.

The most important subclasses of critical mass models are "lemons" and "tipping" models. The former involve buyers and sellers who respond in different ways to the actual proportion of bad used car, i.e. the "lemons", in the terminology of Akerlof.[8] If lemons are rare, the market will be stable because of exchanges will mutually advantageous for both buyers and sellers. By contrast, when the number of lemons increases and when it exceeds the critical number, the entire market may disappear. The latter models describe the instabilities in the racial composition of neighbourhoods. «It was observed that the entrance of a few members

[5] Thomas C. Schelling, *cit.*, p. 49.
[6] Thomas C. Schelling, *cit.*, p. 59.
[7] Thomas C. Schelling, *cit.*, p. 94. Italics in original.
[8] George A. Akerlof, The market of 'lemons': Quality uncertainty and the market mechanism, *Quarterly Journal of Economics*, 84:3, 1970.

of a minority into a neighborhood often causes some among the formerly homogeneous population to leave, or to sign of leaving. Their departure left openings, so more member of the minority could enter; the increase in new residents induced more of the old to leave [...] Some of the departure might be motivated by the minority entrants who have already arrived, some by the belief that the process, would continue [...]».[9]

Two other classes of models that are strictly related to critical mass models refer to, respectively, the "commons"-type of processes and the expectation-determined processes. More specifically, the former class refers to all those situations in which the individuals act purposively and, by doing so, produce an aggregate that is considered undesirable by everyone. The Prisoner's Dilemma is the most famous and clear example of one situation of this kind. J. Hirshleifer[10] has pointed out that the juxtaposition of commons processes with the critical mass phenomena is quite misleading. Indeed Prisoner's Dilemma situations show cyclical structures that are typical of the class of phenomena described by critical mass models. Perhaps this is the reason why Schelling is interested these situations in this Chapter.

On the other hand, expectation-determined processes refer to a class of processes that has the structure of self-fulfilling prophecies: «the general idea is that certain expectations are of a such character that they induce the kind of behavior that will cause the expectation to be fulfilled. [...] For example, if a particular minority is considered incapable of holding responsible positions, they will not be hired for responsible positions; they will have no opportunity for responsible positions; and lacking any such experience, they may indeed be incapable».[11]

CHAPTER FOUR and CHAPTER FIVE examine the application of the models presented above for two kinds of problems: segregation and integration. In particular, Chapter four deals with segregation and integration with respect to sex and race, which are considered discrete variables. Instead, Chapter five is dedicated to an analysis of segregation and integration when the variables of the system are continuous: age and income.

CHAPTER SIX is quite independent from the other chapters. It was published before then[12] and it could be read independently of the entire book. It considers a possible situation – one that might soon become a reality, in the author's opinion – where parents would have the opportunity of selecting their children's genes.

CHAPTER SEVEN is about Prisoner's Dilemma with coalitions; this Chapter gets rather technical, with the help of a lot of mathematics. As Piergiorgio Odifreddi pointed out in his introductory note to the first Italian edition of *Micromotives and Macrobehavior*[13], the difference between game theory as first pioneered by Nash, Selten and Harsanyi and the very same theory as developed by Schelling and Aumann is that while the first economists limited themselves to the analysis of conflicts in which two or more players have no possibility of making coalitions, Schelling and Aumann studied conflicts with coalitions. This is the reason why Schelling and Aumann received the Nobel Prize in Economics.

The result of a basic Prisoner's Dilemma situation is that privately optimized decisions (micromotives) lead to an aggregate (macrobehavior) that is undesirable for everyone. Is this result inescapable? That is, is it always true that, if we play a Prisoner's Dilemma-based game,

[9] Thomas C. Schelling, *cit.*, p. 101.

[10] J. Hirshleifer, Review of *Micromotives and Macrobehavior* by Thomas C. Schelling, *Journal of Economic Literature*, Sol 18, No. 3 (Sept., 1980), pp. 1092-1094.

[11] Thomas C. Schelling, *cit.*, p. 115.

[12] Choosing our children's genes, in Mack Lipkin and Peter T. Rowley (eds), *Genetic Responsibility*, Plenum Press, 1974.

[13] Piergiorgio Odifreddi, *Nota introduttiva* a Thomas C. Schelling, *Micromotivazioni della vita quotidiana*, Bompiani, 2008.

the cooperative choice is strictly dominated by the non-cooperative one? Schelling's answer is negative. His concern now is to extend the definition of the basic version of the game (i.e. the two-player version), in order to «[catch] the spirit of the Prisoner's Dilemma»[14], when it is played by more than two persons. Schelling makes two hypotheses: first, an individual is always better off, the more other people choose their unpreferred alternative; second, the individual's own preference is constant no matter how many among the others choose one way or the other. These assumptions lead to the definition of Uniform Multi-Persons Prisoner's Dilemma (MPD), namely a situation in which:

1. There are *n* people, each with the same binary choice and the same pay-offs.
2. Each person has a preferred choice regardless of what the others do and everybody prefers to make the same choice.
3. Whichever choice a person makes, the more people among the others choose their unpreferred alternative, the better off that person is.
4. There is a number *k*, greater than 1, such that if the number of the individuals choosing their unpreferred alternative is equal or higher than *k*, then these individuals are better off than if they had all chosen their preferred alternative; but if the number of the individuals choosing their unpreferred alternative is less than *k*, then this is not true anymore.

In this scenario, the central issue is the parameter *k*. What is it, exactly? It represents «the minimum size of any coalitions that can gain by abstaining from the preferred choice. It is the smallest disciplined group that, though resentful of the free riders, can be profitable for those who join (though more profitable for those who stay out)».[15]

The rest of the Chapter is a sort of diagrammatic variation on this basic theme, namely the possibility of coalitions in Prisoner's Dilemma contexts.

Let me now try to offer the philosophical analysis that I have mentioned at the beginning.

The first consideration is a methodological one. Schelling's approach, as he himself states, is not a market approach. The attention to market-free contexts leads immediately to a series of methodological reflections about the consequences of such an analysis, i.e. the divergence of the 'equilibrium' from the 'optimum'.

An equilibrium is a situation in which several things that have been interacting are eventually in balance, at rest. Typically, and in spite of Prisoner's Dilemma situations, from equilibrium analysis economists derive the conclusion that self-interested micromotives often lead to surprising and socially useful coordination in the aggregate; a system in equilibrium is thus a good one. This is the old and fascinating idea by Adam Smith of an invisible hand.

Schelling does not try to assess whether a Smithian approach to economic phenomena is right or wrong; what he does is simply to argue that, when a Smithian approach is applied to non-market contexts of behaviour, it is surely wrong. One can still believe that something in the idea of an invisible hand is true, but one has also to admit that this is true only in situations different from those that Schelling considers.

Schelling insists on the fact that equilibria are simply convenient theoretical results; «there is nothing particularly attractive about an equilibrium».[16] Not every situation that is in equilibrium is a just or an optimal situation; the body of a hanged man is in equilibrium when it

[14] Thomas C. Schelling, *cit.*, p. 217.
[15] Thomas C. Schelling, *cit.*, p. 218.
[16] Thomas C. Schelling, *cit.*, p. 26.

finally stops swinging, but it is hard to acknowledge that someone could say that the man is all right. From this methodological point of view, *Micromotives and Macrobehavior* can be considered as an attempt to provide a large number of formal tools for all the situations where reaching the equilibrium is not the general and definitive purpose that we are trying to achieve. However, in order to recognise the asymmetry between equilibria and desired aggregate results, we need to look beyond a purely market-oriented approach. This is the goal of Schelling's book.

The second consideration concerns the Prisoner's Dilemma. Schelling shows how we can extend this class of games in such a way that to join a coalition is no more the kind of choice we do not prefer to make. The failure of cooperative choice in the basic version of Prisoner's Dilemma is due to the fact that individuals act solely to promote their own interests and to the fact that this behaviour is rational. From this perspective, it is irrational to join a coalition and to keep the agreements. The rational micromotive is to free ride. Even if everyone recognised that the aggregate resulting from rational micromotives is not the desirable aggregate, this acknowledgment would not constitute a sufficient reason for choosing the unpreferred alternative of cooperating.

Nevertheless, in observed real-life situations, coalitions do work. How can we explain that? The sceptic may suggest that the intrinsic limit of game theory is the incapability of giving an adequate definition of rationality. If we adopt a purely game theoretic concept of rationality, coalitions are irrational. The sceptic derives the conclusion that what we need is a sort of "conversion" of rationality.[17] It seems that to make the cooperative choice rational we have to change our concept of rationality. In other words, if we want to capture the logic underlying coalitions we have to give up game theory. Otherwise, the expression "Prisoner's Dilemma with coalition" would be a logical contradiction like Quine's «round square cupola on Berkeley College».[18]

Is the sceptic right? Schelling might reply that it is certainly true that the Prisoner's Dilemma, in its basic version, does not capture the spirit of all those social contexts in which we have a preference to join a coalition rather than staying out. In other words, there is certainly a sense in which the Prisoner's Dilemma is not adequate to the facts. However, this does not mean that cooperative choices are irrational or that their rationality exceeds the resources of game theory. Joining a coalition can be rational not in the light of a different concept of rationality, but in the light of the very same idea of rationality that we find in the Prisoner's Dilemma. Thus, there is no need to abandon the game-theoretic framework. We just have to refine it.

Schelling's extension of the Prisoner's Dilemma in its MPD version has to count as a proof of the fact that it is possible to understand the rationality of coalitions from the perspective of the game theoretic notion of rationality. In MPD situations people act purposively like in the basic version of the game. What is different is the social context in which they act, and not the individual reasons leading them to perform in the way they do. In Schelling's terminology, what is new in MPD situations is the aggregate macrobehavior and not the individual micromotives. Individual micromotives are still self-interested and rationality is still goal-oriented; but the aggregate macrobehavior strictly depends on how many micromotives comprise the aggregate. From this perspective we can understand better the reason why Schelling considers the case of "commons" (typically categorized as Prisoner's Dilemma situations) in Chapter three while dealing with critical mass models. The parameter *k*

[17] Jean Hampton, *Hobbes and the Social Contract Tradition*, Cambridge University Press, 1986.
[18] W. v. O. Quine, On What There Is, in *From a Logical Point of View*, Harvard University Press, 1953.

mentioned above and representing the minimum number of people who have to choose their unpreferred alternative in order to make coalition the preferred choice is considered a point of critical mass. It seems that Schelling's is concerned, among other things, with a sort of new formulation of the Prisoner's Dilemma, that is, with the attempt to integrate the Prisoner's Dilemma within the large family of critical mass models.

Schelling's revisited Prisoner's Dilemma suggests a final consideration concerning the general methodology of game theory after the appearance of *Micromotives and Macrobehavior* in the late '70s and its role within the theory itself. There are several reasons why *Micromotives and Macrobehavior* was considered revolutionary. The friendly approach to the subject, which makes it «a book not just for the bookshelf or the doctoral seminar, but for the bedside table», surely counts as one of these reasons.[19] It also seems to me that behind the author's intention to write a fun book to read, a clear position about the methodology of social sciences is active. Indeed, the real novelty of *Micromotives and Macrobehavior* was a completely different attitude towards the general methods that, in Schelling's perspective, game theory had to employ. We can highlight Schelling's change of perspective by asking the question: does individual behaviour cause systems, or do systems cause individual behaviour?[20] Or, in other words: what is the relationship between facts and model? Which one of these two elements – facts or model – has the priority? Does game theory derive facts from the formal model, or is the model derived from the facts?

Schelling's position is in favour of the priority of facts over the model. As Richard Zeckhauser claims, Schelling «studies a real-world problem and develops a conceptual model. He then takes that conceptual model back to a dozen real-world problems to see how it applies, and then ricochets back to refine the model. He keeps the process going until he is happy with his model, and satisfied with his insights into the problems that most interest him».[21] What we want to do now is clarifying the philosophical motivation leading Schelling to this methodological attitude.

If we look at this topic from a philosophical point of view, we see that we face here the classical problem posed by the dichotomy between deductive and inductive methods. The philosophical debate about the distinction between deduction and induction is vast. Given the limited space available, it will not possible to analyse the matter in great depth. Nevertheless some words of clarification are needed.

The concepts of deduction and induction remind us immediately of an old distinction. I am referring to the distinction that identifies deduction with the inference from statements referring to general concepts to statements referring to particular cases; and, conversely, induction with the inference from particular statements to more general statements. The development of probability theory and statistics allowed contemporary philosophy of science to refine this distinction[22]. According to it, when we talk about deduction and induction we talk about two different types of inference, namely, two ways to derive a conclusion from certain premises. If the conclusion follows from the premises with the force of logical necessity, we

[19] Terry Connolly, Review of *Micromotives and Macrobehavior* by Thomas C. Schelling, Administrative Science Quarterly, Vol. 24, No. 3 (Sep., 1979), p. 504.

[20] Del Taebel and Charles Elder, Does the Individual Behavior Cause Systems or do Systems Cause Individual Behavior? *Urban Affairs Quarterly*, Vol. 15 No. 2, Dec. 1979, pp. 229-235.

[21] Richard Zeckhauser, Thomas Schelling, Ricochet Thinker, in Robert Dodge, *The Strategist: The Life and Times of Thomas Schelling*, Puritan Press, New Hampshire, 2006, pp. vii-xii.

[22] See C.G. Hempel, P. Oppenheim, *Studies in the Logic of Explanation*, in «Philosophy of Science», 15, pp. 135-175, 1948; W.C. Salmon, *Four Decades of Scientific Explanation*, Minnesota 1989.

say that such an inference is deductive; otherwise, if the conclusion follows from the premises with a certain degree of probability, the inference is inductive.

It is my opinion that Schelling's revolutionary contribute to game theory can be appreciated only if we consider the latter way of conceiving the distinction between deduction and induction. Making this distinction in terms of logical deduction and deduction with a certain degree of probability, instead of the classical explanation in term of inference from particular to general (induction) and from general to particular (deduction), gives us the opportunity to reflect on the role of statistics within Schelling's work. Statistics is indeed the science that derives conclusions with a certain degree of probability. If we get an insight into Schelling's method of analysis, we see that conclusions about the macrobehavior in the aggregate are statistically – and not logically – derived from individual micromotives. The presence of statistical inferences in *Micromotives and Macrobehavior* is in fact pervasive. Before *Micromotives and Macrobehavior*, game theory, like any mathematical theory, was a deductive science. After fixing the models, statements concerning facts were thereby derived in a purely deductive way. The employment of statistical tools in *Micromotives and Macrobehavior* radically changed the scenario. The model has now to be adjusted to the facts, and not vice versa. The formal apparatus is subject to a continuous process of refinement in order to adequately describe social phenomena that are essentially variable. Thus, Schelling's strategy is to start with individual micromotives and to derive, with a certain degree of probability, the resulting aggregate macrobehavior.

At this point we start to see the consequences of Schelling's inductive approach. Neither the features of formal models are considered immutable nor social phenomena are derived from them in a merely logical way. Formal models are subject to revisions. If the models are statistically determined, it is clear that the answer to the question about priority, if any, between model and facts is that facts have the priority over the model. But Schelling's preference for the inductive method would be a *petitio principi* if we did not focus on the fact that this preference is grounded on and justified by the use of statistics. The ground for adopting an inductive method is thus the employment of statistics.

In this scenario we can better understand the role of Schelling's examples. These are not only a didactic strategy to make the book fun to read, but they are part of a distinctive methodology. Historically, the endorsement of an inductive method is the element which distinguishes *Micromotives and Macrobehavior* from the other game theoretical approaches to social phenomena. Recent contributes like Malcolm Gladwell's *The Tipping Point*[23] or *Freakeconomics*[24] by Dubner and Lewitt cannot be understood without the intermediate role of Schelling's *Micromotives and Macrobehavior*.

[23] Malcolm Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*, Back Bay Books, 2002.

[24] Steven Levitt and Stephen J. Dubner, *Freakeconomics: A Rogue Economist Explores The Hidden Side of Everything*, William Morrow/Harper Collins, 2005.

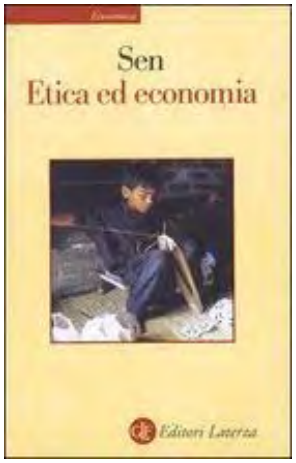
Commentary

Etica ed economia

Amartya Sen

Laterza Editore, Roma-Bari, 2002

Redazione
redazione@humana-mente.it



Basato su una rielaborazione delle *Royer Lectures* tenute dall'autore all'università di Berkeley nel 1986, questo breve volume di Amartya Sen vuole essere il punto di congiunzione tra due campi tradizionalmente affini eppure sempre più distanti quali l'etica e l'economia.

Questi due ambiti del sapere non sono resi distanti l'uno dall'altro solo dalla crescente specializzazione accademica – che domina entrambi –, ma anche, evidentemente, dalle loro concrete dimensioni sociali e civili su scala globale.

Da qui una doppia serie di difficoltà: Sen imbocca coraggiosamente in questo testo l'analisi tecnica delle possibili congiunzioni ed interazioni tra parti specifiche dell'etica contemporanea ed i principali risultati limitativi da egli stesso o da altri dimostrati (ad esempio il Teorema di Arrow), arrivando a conclusioni assieme assai equilibrate e sorprendenti.

Questa analisi tecnica, come noto, ha avuto seguito in vari altri volumi (da *Lo sviluppo è libertà* a *Globalizzazione ed oltre*), nei quali l'autore ha preso posizione anche sull'altro aspetto della distanza tra etica ed economia, quello, per così dire, contingente e concreto.

Nello specifico, in *Etica ed economia*, è scandagliato un punto fondamentale del pensiero di Sen: non solo l' *homo oeconomicus* classico, che è immaginato come un soggetto agente sulla sola base della massimizzazione dell'interesse personale, non rappresenta la miglior approssimazione del comportamento umano effettivo ma, da questa massimizzazione, non è possibile dedurre una teoria che assicuri invariabilmente condizioni economiche ottimali.

A questo proposito, la distanza che Sen pone tra l'approccio "ingegneristico" all'economia e l'approccio "comprensivo" - qui rappresentato da Aristotele -, trova le sue conseguenze nei "risultati di culmine", che contrappone ai "risultati comprensivi" descritti ne *Lo sviluppo è libertà.*

Avvalendosi di questa osservazione fondamentale, Sen scava diritto fino alle origini del liberalismo, arrivando ad Adam Smith: la distanza tra i due diversi approcci viene dunque indagata alla radice e, con questa reinterpretazione del pensiero di Smith, si conclude la prima delle tre conferenze di cui è composto il volume.

Corollario di questa visione – anch'esso poi ampliato ne *Lo sviluppo è libertà* – è una maggiore complementarità tra economia ed etica, come anche il loro reciproco influsso. Se infatti l'uomo non agisce sulla scorta di sole valutazioni personali, allora sarà una più vasta razionalità di gruppo ad agire sulle sue scelte: è questa la via attraverso cui l'etica può influire sull'economia. A sua volta l'etica del diritto non riconosce abbastanza il ruolo valoriale di questi termini e si può quindi giovare dell'analisi economica dell'interdipendenza generale per una migliore definizione dei diritti e della libertà.

La seconda conferenza, dal titolo *Giudizi economici e filosofia morale,* parte dalla peculiare sorte toccata all'economia del benessere vittima, secondo Sen, di un restringimento di visuale, dovuto al progressivo abbandono dell'etica come campo affine di interscambio. L'idea secondo cui il singolo sarebbe mosso unicamente dall'interesse personale – e quindi diretto senza alcun vincolo etico - ha fatto sì che risultasse certa un'identificazione tra efficienza economica e ottimalità paretiana.

Questo criterio (ma in realtà, a proposito di tale questione, si dovrebbe parlare più appropriatamente di "definizione") è, per così dire, vuoto: manca infatti di un principio interno che lo guidi verso la creazione di un ordine realmente "ottimalizzato": il risultato di Pareto non afferma alcunché a proposito di come debba essere distribuita la ricchezza prodotta, mentre l'idea secondo cui uno stato sociale è ottimizzato quando non possa essere aumentata l'utilità di nessuno diminuendo per questo l'utilità di qualcun altro, lascia aperta la possibilità descritta dal calzante esempio di Sen, in cui nessun povero può essere aiutato senza ledere economicamente un ricco.

A questo punto segue una discussione delle manchevolezze del criterio di Pareto in rapporto all'utilitarismo ed al welfarismo: una concezione assai ristretta dell'etica, collegata all'aspetto "ingegneristico" assunto dall'economia, ha anche fatto sì che l'utilitarismo (ed il welferismo come sua parte costitutiva e "pratica") negasse un valore intrinseco ai diritti di per sé presi.

Una considerazione di valore riguardo ai diritti, che riconosca loro un peso al di là del semplice rispetto in quanto vincoli, può invece scuotere alle fondamenta la teoria economica tradizionale.

Nella terza ed ultima conferenza Sen analizza quindi i requisiti di una valutazione etica sistematica ed il ruolo delle conseguenze sulla libertà e sui diritti di tale valutazione.

In questa ampia parte finale, il discorso di Sen si arricchisce e si complica notevolmente.

Vengono sopratutto rigettati i tre punti fondamentali che hanno caratterizzato il pensiero economico nelle sue implicite definizioni del comportamento dell'uomo: l'assunto che questo sia mosso da un interesse personale egoistico; che l'obiettivo finale sia la massimizzazione del benessere personale; che ogni forma di scelta sia fondamentalmente basata su obiettivi personali senza alcuna limitazione o adattamento in ragione del riconoscimento della reciproca interdipendenza dei processi di scelta da parte di altri soggetti agenti.

Rigettare tutto ciò, con Sen, significa non solo riavvicinare l'etica e l'economia, ma anche accogliere una definizione più ampia e più accurata di razionalità: significa, in sostanza, rivedere alla radice la nostra idea di uomo.

Commentary

Il banchiere dei poveri

Muhammad Yunus

Milano, Feltrinelli Editore, 2006, pp. 272

Redazione
redazione@humana-mente.it

"[...] se vogliamo riuscire dobbiamo puntare sulla fiducia. Fin dal primissimo giorno abbiamo stabilito che il nostro sistema avrebbe fatto a meno di polizia e tribunali. Noi partiamo dal principio che dobbiamo essere capaci di far marciare autonomamente i nostri affari; altrimenti faremmo meglio a lasciar perdere la banca e cercarci un altro mestiere. A tutt'oggi per recuperare i nostri crediti non ci siamo mai serviti di avvocati né di altre figure professionali esterne alla banca.

Nella stessa logica non esistono da noi atti giuridici tra la banca e il cliente. Noi stabiliamo rapporti con le persone non con i documenti. Il nostro legame riposa sulla fiducia e il successo o il fallimento della nostra iniziativa dipendono dalla forza del rapporto personale con l'utente."

Questa lunga citazione, al pari di molte altre, è una buona chiave d'accesso per cominciare a capire che cosa sia e cosa possa rappresentare la Grameen Bank (letteralmente, la Banca Rurale) per un paese come il Bangladesh: una banca con uno scopo sociale preciso: la lotta alla povertà in uno dei paesi più poveri del mondo attraverso il microcredito senza garanzie.

Fondata nel 1977 dopo un periodo sperimentale illustrato nel primo capitolo, la Grameen Bank ebbe il suo primo ufficio in una capanna con il tetto di lamiera, senza acqua corrente, senza telefono, senza servizi igienici: dopo venti anni di attività, nel 1997, la Grameen è già presente in 36.000 villaggi del Bangladesh, presta soldi a 12 milioni di persone (il 10% della popolazione del Bangladesh) ed ha aiutato un numero incalcolabile di persone ad uscire dalla miseria.

Il suo fondatore, Muhammad Yunus, oggi internazionalmente noto anche per aver ottenuto il Premio Nobel per la Pace nel 2006, ci racconta con questo volume in modo leggero, suggestivo e quasi *naif* la storia di Grameen Bank ed insieme la storia della sua vita.

E' la storia, densissima di vita e d'incontri, di un economista arrivato negli Usa nel 1965, docente prima a Boulder, nel Colorado, poi alla Vanderbilt University e tornato quindi in Bangladesh nel 1972, dove subisce lo shock di una povertà incredibile, amplificata dalle catastrofi naturali e per niente alleviata dalle azioni della World Bank (che tenterà anche, a più riprese, di assumere il controllo della Grameen).

Questa diffusissima miseria, che spingeva verso l'usura e vere e proprie forme di lavoro schiavista, viene incrinata da un concetto rivoluzionario di credito, chiaramente descritto nell'elenco delle Sedici Risoluzioni. Tali risoluzioni sono le regole fondanti l'azione della Grameen Bank e, al loro interno, la parola "soldi" o la parola "interesse" non appaiono *mai*.

A questi straordinari risultati se ne aggiungono altri, quali ad esempio quelli dovuti alla precisa e volontaria tendenza della Grameen a prestare soldi alle donne anziché agli uomini: superando una forma di ghettizzazione molto diffusa che obbliga ancora oggi molte donne del Bangladesh a non avere quasi proprietà e a dipendere esclusivamente dal marito o da parenti

uomini, la Grameen ha reso possibile una liberazione della donna impensabile fino agli anni '70.

Le sei parti in cui si divide il volume danno conto di una serie lunghissima ed incredibile di successi economici e sociali ottenuti con il microcredito; le prime tre parti trattano in prevalenza della fondazione della Grameen e delle attività sperimentali e collaterali che sono state necessarie per il suo consolidamento, mentre le altre tre riguardano rispettivamente l'esportabilità del modello di microcredito (oggi la Grameen Bank presta soldi anche negli Usa ed è presente in altri 100 paesi del mondo, Italia compresa), alcune riflessioni su libero mercato, la lotta alla povertà e gli obiettivi sociali, ed infine i nuovi orizzonti che si aprono di fronte all'esperienza della Grameen Bank.

Sopratutto in queste tre ultime parti si percepisce chiaramente quanto sia stata profonda e intensa la riflessione sulla povertà all'interno della Grameen Bank, e come questa riflessione si sia legata strettamente a questioni soltanto in apparenza lontane quali l'assistenza sanitaria, la scolarizzazione di base e le telecomunicazioni.

In effetti Grameen è un modello interamente nuovo di pensare l'impresa, la concorrenza ed il libero mercato: è il tentativo, straordinariamente riuscito, di lottare contro la miseria più assoluta e di vincerla dall'interno, facendo leva sulla parte migliore delle dinamiche di mercato.

Dai tempi della prima stampa di questo volume, nel 2001, e ancor più dai tempi della fondazione della Grameen Bank, il Bangladesh ha fatto dei passi avanti incredibili, nonostante la lunghissima serie di problemi sociali ed economici che gravano sul paese.

Interessantissimo e utile a cambiare la nostra visione dei processi economici, il volume si chiude con una postfazione ormai invecchiata ma utile come spaccato storico, sul vertice mondiale del microcredito del 1997 di Washington. e due altri brevi scritti, rispettivamente una *Proposta per l'istituzione di un centro internazionale di tecnologia informatica per l'abolizione della povertà globale* ed il discorso in occasione della premiazione con il Nobel per la pace nel 2006.

Nella postfazione è raccontata la storia dell'organizzazione del vertice sul microcredito, un evento definito da Hilary Clinton, per l'occasione copresidente del convegno assieme alla regina Sofia di Spagna ed all'ex primo ministro giapponese Tsutumo Hada, come “uno degli appuntamenti mondiali più importanti degli ultimi tempi”; ed in effetti il vertice del 1997 con i suoi tremila invitati da 137 paesi, è stato forse il momento più importante per il lancio a livello mondiale dell'idea del microcredito.

Il secondo testo si occupa invece di fare il punto su come deve essere organizzata e funzionare una struttura che operi a vantaggio dei poveri sfruttando le tecnologie informatiche: dai finanziamenti alle partnership, dalle prospettive agli obiettivi.

L'ultimo testo invece consiste, come detto, nel discorso pronunciato da Yunus nel 2006 in occasione del conferimento del Premio Nobel: in poche pagine Yunus fa i conti con la radice ultima del terrorismo, che *in nuce* è la stessa che espresse Kennedy nel 1961, quando affermò che “anche la fame è guerra”, e che quindi all'origine dell'odio vi sia l'ineguaglianza e l'ingiustizia a cui sono sottoposte molte popolazioni nel mondo. La soluzione è ancora una volta nella lotta alla miseria: infatti “la povertà è l'assenza di tutti i diritti umani. Le frustrazioni l'ostilità e la rabbia che nascono dalla miseria non possono offrire un sostegno alla pace in nessuna società. Per costruire una pace stabile dobbiamo trovare il modo di dare opportunità alla gente di vivere vite dignitose.”

Commentary

La Globalizzazione e i suoi oppositori

Joseph E. Stiglitz

Einaudi, Torino, 2002

La globalizzazione che funziona

Joseph E. Stiglitz

Einaudi, Torino, 2006

Giovanni Pancani
giovanni.pancani@humana-mente.it

Pochi sono i libri come *La globalizzazione ed i suoi oppositori* di Joseph Stiglitz che sono riusciti ad animare tanto a lungo e tanto in profondità la discussione sulla globalizzazione esprimendone in modo così nitido gli aspetti salienti: forse soltanto *No Logo* di Naomi Klein ha avuto un impatto paragonabile a questo libro.

Il volume della Klein era però scritto da una giornalista, e quindi da un' outsider dell'ordine economico mondiale mentre Stiglitz, come noto, scrisse questo volume nello stesso periodo in cui ricevette il Nobel e poco dopo aver lasciato la carica di *senior vice president* alla World Bank, carica questa seguita di poco a quella di capo consulente economico sotto la presidenza di Bill Clinton.

Come è ricordato dall'autore all'inizio del primo capitolo, dopo le contestazioni di Seattle del 1999 il dibattito sulla globalizzazione è esploso a livello internazionale con violente rivolte durante i vertici del G8: una parte consistente dell'opinione pubblica occidentale ha finito per comprendere, seppur confusamente, che qualcosa non andava nel processo di globalizzazione.

Buona parte dell'intreccio storico analizzato nel volume di Stiglitz riguarda naturalmente gli interventi di quelle due istituzioni nate a Bretton Woods nell'estate del 1944 che avrebbero dovuto, nelle intenzioni originarie, assicurare la scomparsa della povertà e la stabilità economica globale: rispettivamente queste istituzioni erano la World Bank e l' Fmi.

Una terza istituzione, il Wto, non è particolarmente presente ne *La globalizzazione ed i suoi oppositori* e troverà maggiore spazio negli scritti successivi di Stglitz come *Commercio equo per tutti* e *La globalizzazione che funziona.*

Sia la World Bank che l'Fmi hanno operato secondo l'autore in modo sistematicamente dannoso nei confronti dei paesi cui hanno rivolto l'attenzione: in modo particolare la World Bank ha prima modificato essenzialmente il suo ruolo e poi ha intensificato notevolmente i suoi sforzi dopo il 1981, anno dell'epurazione di Hollis Chenerey e Robert Mc Namara, rispettivamente *chief economist* e presidente dell'istituzione bancaria più importante del mondo.

La sostituzione con Ann Krueger e William Clausen permise alla World Bank la creazione dei Piani di adeguamento strutturali, cosa che trascinò l'Fmi nella stessa direzione: era l'inizio, seppur embrionale, di quel neoliberismo che con la vittoria di Reagan negli USA e della Thatcher in Gran Bretagna diede avvio al *Washington Consensus,* ovverosia a quella sostanziale identità di vedute economico – politiche che allineerà World Bank, Fmi e Tesoro degli Stati Uniti nei confronti dei paesi in via di sviluppo causandone spesso i tragici tracolli economici e

sociali degli anni '80 e '90.

Naturalmente dopo il crollo del Muro di Berlino Fmi e World Bank si impossessarono di un campo enormemente più vasto di quello posseduto in precedenza – un campo che dal 1991 sarebbe stato equivalente a quello del mondo intero – influenzando ancor più pesantemente l'economia mondiale.

Rileggendo a distanza di sette anni *La globalizzazione ed i suoi oppositori* si resta in effetti impressionati dalla abissale distanza che separa oggi la Cina, con uno sviluppo economico impetuoso, dalla Russia: la lunga agonia russa degli anni '90 è dettagliatamente descritta nel quinto e sesto capitolo, che restano ancor oggi interessantissimi per ripercorrere la storia di cosa non è andato per il verso giusto nella più imponente transizione economica della storia del Novecento.

Il volume copre, come è facile immaginare, un vastissimo spettro di argomenti e passa dall'attenta analisi della crisi del Sud-Est asiatico (capitolo quarto) ai problemi che i più diversi paesi dell'Africa sub-sahariana hanno avuto nel fronteggiare l'Fmi (capitolo secondo), fino ad arrivare alla necessaria conclusione illustrata nell'ultimo capitolo del volume, il nono (*La strada da percorrere*): cosa cambiare nelle istituzioni economiche globali e come trasformare democraticamente la globalizzazione per liberarne le potenzialità.

Un libro stimolante e necessario per capire molti aspetti altrimenti oscuri dell'economia internazionale e per afferrare davvero quale sia la posta in gioco quando si parla di globalizzazione: chiunque abbia seguito la progressione di scritti non tecnici che sono stati pubblicati da Stiglitz dopo *La globalizzazione ed i suoi oppositori* ha potuto constatare come il tema della globalizzazione abbia avuto importanti svolgimenti e precisazioni.

Come detto, già ne *La globalizzazione ed i suoi oppositori*, il dibattito veniva incentrato sul funzionamento interno degli organismi monetari internazionali e sulle conseguenze che questi producono, in un resoconto generale e convincente di come interi continenti e singoli paesi avessero subito i diktat di organismi quali la World Trade Organization e l'International Monetary Found.

Osservate da un acuto conoscitore dell'economia, le posizioni adottate dal Wto e dall'Fmi nei confronti dei governi che hanno avuto bisogno di credito da parte di questi organismi internazionali, apparivano a Stiglitz come un mescolanza di cattiva politica, ideologia e dogmatismo.

Una serie apparentemente disconnessa di interventi monetari che avrebbero dovuto assolvere alle funzioni per le quali tanto il Wto quanto il Fmi erano nati, ovverosia garantire liquidità a quei paesi che ne avessero avuto bisogno e creare le condizioni per la stabilità economica globale sono reinterpretati in modo assai convincente e persuasivo da Stiglitz come atti unicamente a beneficiare Wall Street.

Non compariva naturalmente nessuna "pistola fumante" - come del resto si poteva leggere nella prefazione del libro - ma diveniva chiaro come fosse possibile interpretare un insieme apparentemente disordinato di politiche macroeconomiche e monetarie del Fmi *come se* fosse un disegno preciso a tutto vantaggio dei mercati finanziari; se a questo aggiungiamo che tali politiche erano appoggiate dal Tesoro degli Stati Uniti (solo ed unico paese ad avere diritto di veto sulle decisioni prese in seno all'Fmi) diviene chiaro come la globalizzazione sia diventata una promessa mancata e per molti popoli addirittura un peso insostenibile.

Similmente, ne *La globalizzazione che funziona* Stiglitz analizza ciò che non va della

globalizzazione, ma in un contesto addirittura più ampio: è dedicato più spazio al commercio equo, nel cap. III, con una breve ed utile storia dei trattati commerciali internazionali, ed all'ambiente, nel cap. VI; sono affrontati inoltre problemi complessi quali quelli relativi ai diritti di proprietà intellettuale e della regolamentazione dei brevetti (cap. IV), di diritto fallimentare per quei paesi che non possono affrontare il peso del debito (esemplare a questo proposito il caso dell'Argentina) oltre che, come ne *La globalizzazione ed i suoi oppositori*, casi di singoli paesi alle prese con le politiche dell'Fmi (cap. II).

Sono infine toccati, nel IX cap., aspetti strutturali della necessaria riforma del sistema di riserva valutario mondiale, dal quale dipendono in modo sistemico i problemi di instabilità finanziaria globale: questo capitolo, interessantissimo, rivaluta anche alcune vecchie idee keynesiane quale la creazione di una valuta universale (il cosiddetto "bancor" che doveva essere emesso dal Fmi secondo la visione originale di Keynes).

L'ultimo capitolo è riservato a considerazioni politiche: le precedenti analisi economiche vengono a convergere in una visione estremamente ricca e raffinata di cosa la politica internazionale deve fare – *deve* e non *dovrebbe*, si badi bene, perché per l'autore le riforme proposte sono inevitabili, ed anzi sono le sole che possono permettere alla globalizzazione di ricominciare a crescere economicamente e di realizzarsi compiutamente.

In questo ultimo capitolo il volume di Stiglitz supera largamente le conclusioni cui era arrivato l'autore ne *La globalizzazione ed i suoi oppositori*, disegnando dettagliatamente le modalità di riforma necessarie a garantire una nuova e più equa gestione dei processi economici e finanziari globali.

Scritto pochi mesi prima che fallisse la banca inglese Northern Rocks e che si innescasse l'attuale crisi economica e finanziaria mondiale, questo volume resterà con ogni probabilità in futuro all'attenzione degli storici e degli economisti per la sua notevole capacità di penetrare i meccanismi interni con cui si è sviluppata ed è entrata in crisi la globalizzazione.

Commentary

Economie senza denaro.

I sistemi di scambio non monetario nell'economia di mercato

Maurizio Pittau

Editrice Missionaria Italiana, Bologna, 2003

Stefano Liccioli
stefano.liccioli@humana-mente.it

Il saggio di Maurizio Pittau *Economie senza denaro* è un'agile riflessione su tutti quei sistemi di beni e servizi che non si avvalgono dell'intermediazione del denaro. L'assunto di partenza è che la moneta ha da sempre condizionato fortemente i rapporti economici e sociali: oggi come non mai «tutto, o quasi, è denaro; tutto dipende dal denaro; tutto si riconnette al denaro», arrivando a condizionare la cultura e la psicologia delle persone. Esso viene troppo spesso ritenuto uno mero strumento tecnico, quando in realtà influenza anche la fiducia e la natura delle relazioni umane e può causare isolamenti tra gli uomini. «Se si riflette sul fenomeno presto ci si rende conto del ruolo eccessivo che il denaro ricopre nel nostro immaginario. Da strumento d'intermediazione negli scambi esso è divenuto un'entità sempre più potente e imperscrutabile da cui dipende il nostro futuro». In questo senso appare appropriato il paragone che l'autore stabilisce tra fede e denaro, una forza, anche quest'ultima, capace di muovere il mondo.

Come sarebbe la vita umana se si potesse, in alcune occasioni, fare a meno dei soldi? E' la domanda che si pone l'autore. Un interrogativo che parrebbe riportarci in un'epoca superata ed ormai dimenticata, quella del baratto; in realtà non sarebbe proprio così e, soprattutto, non si tratterebbe meramente di un passo indietro. Infatti, analizzando i sistemi di scambio non monetario esistenti, ci possiamo accorgere che l'assenza del denaro comporta un aumento di fiducia, attraverso fattori che l'economia classica non ha mai preso in considerazione: l' altruismo, la cooperazione e la reciprocità. In questa ottica «le persone agiscono non solo in base ad interessi egoistici, ma anche a sentimenti e valori». Inoltre queste esperienze, all'interno dell'economia di mercato con cui convivono, contribuiscono alla ricchezza non solo sociale, ma anche economica delle realtà in cui si trovano.

Pittau definisce economie senza denaro tutte quelle esperienze in cui i partecipanti si «scambiano volontariamente beni e servizi, senza l'intermediazione dei soldi, secondo un rapporto di reciprocità». In generale la moneta tradizionale è sostituita con forme particolari di soldi o tempo: l'idea di base è che il benessere individuale e sociale si persegue più attraverso le relazioni interpersonali, piuttosto che con il consumo dei beni. Non si tratta di economie alternative a quelle di mercato, ma complementari: occorre «concepire l'attività economica in una logica non solo individualistica, ma anche di reciprocità al fine di favorire dinamiche di socializzazione». Queste esperienze, secondo Pittau, non riducono il cittadino ad un lavoratore, né ad un consumatore, ogni individuo mantiene le proprie peculiarità, così come i legami sociali sono salvaguardati e non vengono sottomessi ai meccanismi di produzione dei beni materiali e dei fenomeni commerciali. Leggi, contratti e razionalità economica non bastano a

garantire stabilità e prosperità alle società moderne, c'è bisogno di dinamiche più profonde come la generosità, gli obblighi morali, i doveri verso la comunità. Il dono, la fiducia, l'altruismo sono aspetti che contraddistinguono le economie senza denaro. Per quel che concerne il dono, viene messo in luce l'importanza del dare, ricevere, restituire, tutti gesti importanti per abbattere gli isolamenti e le esclusioni: in una società globalizzata, dove domina l'interconnessione planetaria, è diventato più difficile riconoscere il proprio vicino, cresce la solitudine degli individui. La dinamica del dono invece cerca prima di tutto di stabilire relazioni sia sul piano economico che solidale: il primo è basato «sul riconoscimento formalizzato del valore intrinseco delle prestazioni e delle risorse offerte e ricevute»; il secondo si fonda su una reciprocità tra due soggetti, che ammette la possibilità di dare – ricevere – restituire, una logica che unisce e consolida.

Per Pittau uno delle conseguenze più importanti prodotte dall'assenza del denaro è l'aumento della fiducia nel prossimo ed in particolare dell'altruismo. Infatti quest'ultimo valore, inserito nel sistema economico, ha mostrato che la presenza di soggetti generosi all'interno della società modifica la qualità dei rapporti interpersonali.

L'autore offre anche esempi concreti di forme di scambio non monetario presenti nei Paesi post-industrializzati, dalle prime esperienze in Canada negli anni Settanta fino ad arrivare al 1995 in Italia con la "Banca del Tempo". Soffermandoci brevemente su quest'ultima esperienza, diffusa anche in Spagna ed in Svizzera, possiamo notare che essa si presenta come «un modo per potenziare la rete di reciproco aiuto tipica dei rapporti di buon vicinato […]. Aderendo alla Banca del Tempo si ottengono servizi, e oggi anche oggetti, che permettono di soddisfare piccoli bisogni immediati e al contempo concorrono a potenziare le reti di relazioni e solidarietà sul territorio». Non si tratta dunque di volontariato, in cui persone disponibili offrono il tempo delle loro attività a soggetti che ne usufruiscono, ma di "reciprocità indiretta", cioè «ogni trasferimento accende debiti e crediti di tempo nei confronti di tutti gli altri, non del singolo interessato»: si scambiano così prestazioni, saperi, beni, sviluppando un senso di solidarietà. Il tempo si configura come l'unità di misura, il valore del servizio è determinato dal tempo impiegato nel trasferimento: tutte le attività sono valutate in tempo e non circola denaro se non quello per la copertura delle spese vive. Il campo di azione di una Banca del Tempo deve essere necessariamente ridotto (un quartiere, un piccolo comune, una scuola), per facilitare la socialità ed evitare le difficoltà di spostamento proprie dei territori molto ampi. In questi anni le Banche del Tempo, ove presenti, hanno generalmente avuto un ruolo prioritario nel «creare coesione, integrazione, sviluppo di rapporti intergenerazionali e interculturali, pari opportunità tra donne e uomini».

Pittau sposta poi l'attenzione del lettore sui sistemi di scambio non monetari nelle economie tradizionali del Sud del mondo, in particolare sulla "società informale" diffusa in ampie zone di questi Paesi poveri. «Le società informali sono costituite da *economie popolari*, in cui piccole imprese o artigiani lavorano per la clientela popolare […]. Nella società informale, nonostante la monetarizzazione e l'ambiente mercantile circostante, il denaro non ha lo stesso significato, né lo stesso utilizzo che ha nelle economie occidentali». In queste ultime infatti il denaro sta diventando sempre più un'astrazione (assegni, carte di credito….), invece nelle periferie popolari di molti paesi impoveriti i soldi sono concreti, tangibili, spesso assumono la forma di collane d'oro e d'argento, bestiami. Così in molte aree del Sud del mondo il legame sociale funziona sulla base dello scambio che, verificandosi senza moneta, si fonda sul triplice obbligo di donare, ricevere, restituire. Inoltre, «in ampie zone del Sud del mondo, a differenza dei Paesi post-industrializzati del Nord, esiste una scarsa monetarizzazione e la società è caratterizzata da legami di solidarietà, bassi livelli di spesa, assenza di dinamiche di creazione di nuovi bisogni e insufficienza della produzione per la vendita». Pittau attribuisce la ragione di

molti insuccessi dei tentativi di sviluppo di aree del Sud, alla difficoltà di conciliare una cultura che privilegia il momento comunitario, la solidarietà ed il rapporto fra l'uomo e la natura, con modelli che considerano solo il privato ed il profitto.

La soluzione che l'autore indica come la migliore per questi paesi poveri è il modello d'economia "neoclanica", un sistema in cui il sociale permette di vincere la povertà: esperienze non monetarie (sperimentate in Ecuador, Messico, Thailandia, Venezuela, Senegal…) che non perseguono il profitto, ma hanno come fine lo sviluppo dell'economia sociale, facendo attenzione ai valori della solidarietà, del rispetto dei diritti umani, dell'educazione e dell'ambiente.

In conclusione il libro di Maurizio Pittau ha il merito, a mio parere, di richiamare l'importanza della dimensione etica per ogni sistema economico. In seguito alle recenti speculazioni finanziarie che hanno portato al crollo di grandi complessi bancari, da più parti si è detto che la finanza dovesse riacquisire un'anima etica. Già Aristotele, affrontando il tema dell'economia, distingueva tra una "sana" attività di scambio di beni e la sua degenerazione, cioè l'accumulazione di beni e di denaro, che il filosofo considerava ingiusta: egli probabilmente avvertiva il rischio che, quando i soldi non sono considerati più un mezzo, ma un fine, il sistema sociale può entrare in crisi. Di fatto questo pericolo si è verificato ai nostri giorni, quando l'avidità e la ricerca di una ricchezza non fondata sul lavoro sono stati la causa di tracolli finanziari che hanno finito per danneggiare anche e soprattutto molti incolpevoli cittadini.

La domanda che sorge è quanto queste esperienze di scambi di beni e servizi senza l'intermediazione del denaro possano davvero incidere sul sistema economico dei vari Paesi. Infatti il dubbio è che rimangano degli episodi di nicchia, "stritolati" da una logica finanziaria che mette in primo piano il profitto, a discapito delle relazioni.

Al di là della reale efficacia a livello economico di queste esperienze di scambio non monetario, credo che esse siano dei punti di riferimenti per un economia che non vede le persone solo come potenziali consumatori, ma come uomini e donne. Come conclude Pittau: «In quest'economia senza denaro, o nella quale esso torna ad essere un mezzo governato da chi lo usa, le persone tornano al centro di qualsiasi transazione, scoprendosi esse stesse protagoniste dello scambio e non più merce […] L'essere umano è in primo luogo un essere di relazione e non un essere di produzione».

Commentary

NeuroEconomics: an introductory review

Matteo Baccarini [*,†]
bccmtt@unife.it

Silvano Zipoli Caiani [‡]
silvano.zipoli@unimi.it

1. OVERVIEW

A new intriguing field of research called *Social Neuroscience* has quickly emerged in the past ten years. Social Neuroscience aims at investigating how the brain and the body affect social behaviours by combining a biological and a social approach, with the goal of informing and refining theories of social processes and behaviours (Harmon-Jones & Winkielman, 2007). In just a few years, social neuroscience has reached many results concerning the biological mechanisms underlying social processes such as *emotion* and *action understanding, interpersonal interactions,* as well as the development of *mind reading* and *linguistic abilities,* thus supporting the idea that a new multidisciplinary interaction between social and biological sciences should be further encouraged.

Within the general field of social neuroscience researchers have studied the neurobiological basis of *human decision making*, as well as *risks* and *rewards categorisation*. Such studies have given raise to a new sub-field of scientific research, called *neuroeconomics*. Neuroeconomics investigates the neural mechanisms underlying decision tasks characterized by profitable or unfavourable outcomes. It combines the classical knowledge concerning human economic behaviours and the explanatory power of neuroscience (Schultz, 2008). As for many other new areas of research, the development of neuroeconomics is characterized by contrasting opinions, so that enthusiastic introductions (Glimcher et al., 2008) are coupled with sceptical analysis (Bernheim, 2008; Harrison, 2008) or radical oppositions (Gul, et al., 2008). Despite this fact, a positive stance is actually prevalent in the literature.

Neuroeconomics seeks an explanatory analysis of social and economic behaviours in neural circuitry, using experimental techniques such as fMRI (functional magnetic resonance imaging) TMS (transcranial magnetic stimulation) pharmacological interventions and other methods. The astonishing methodological progress characterizing recent neurobiology makes it reasonable to expect the development of new ambitious programs, changing the boundaries between cognitive neuroscience and traditional social studies. Indeed, the emergence of neuroeconomics is positively correlated with the increasing number of publications linking neurobiological findings with old economic questions (fig. 1) and with the increasing use of the term “neuroeconomics” across many established disciplines such as biology, economics and philosophy (fig. 2). Since 2002 at least four major journals from neuroscience, economics and philosophy, have dedicated a special issue to neuroeconomics (Neuron, 2002; Games and Economic Behavior, 2005; Philosophical Transactions of the Royal Society, 2008; Economics and Philosophy, 2008). Articles about neuroeconomics have appeared in several newspapers

[*] Department of Human Sciences, University of Ferrara
[†] National Institute of Neuroscience and Neuroscience Center, University of Ferrara
[‡] University of Milan

and magazines, such as *The Wall Street Journal, The Economist, Newsweek, Forbes,* and *The New York Times.* Moreover, a research attempting to quantify the use of the term "neuroeconomics" within scientists' personal web pages shows that more than thousand scientists from many different disciplines use this word to describe their research interests (data-base Scirus).

The previous data prove that the rise of neuroeconomics is one of the most interesting scientific phenomena, not only for neuroscientists or traditional economists, but for philosophers interested in the dynamics of contemporary science as well. A great deal of interest about neuroeconomics is also due to the development of one of its sub-branches, i.e. *neuromarketing*. Research in neuromarketing investigates consumers' sensorimotor, cognitive, and affective response to marketing stimuli, measuring changes in their brain activity. In recent years many private companies, especially in US and UK, began to offer market counselling services based on neuroimaging analyses, concerning, for example, car preferences, the relationship between smells and colours of food products and the neural effect of advertising media delivering commercial messages (Lee et al., 2007). An example comes from a famous neuroimagining study applied to market phenomena (McClure et al. 2004) which tried to use an fMRI in order to explain behavioural preferences between two diffuse kinds of beverage.

The goal of this review is to identify the main features characterizing researches in neuroeconomics. More specifically, our attention will be devoted to the introduction of some recent experimental findings, showing how neurobiological research can contribute to economics. For this purpose, we will introduce some findings concerning the role of neurobiology for game theory analysis, focusing our attention on risk decision tasks, which represent a perfect case study for showing how neurobiological findings may cast new light on the role of emotional feelings during decision making. Understanding the neuronal mechanism underlying decision making might indeed allow economics to be built upon a more robust empirical domain and open up novel research directions which are yet unexplored.

In the next section we will start by introducing the main theoretical features characterizing neuroeconomics.

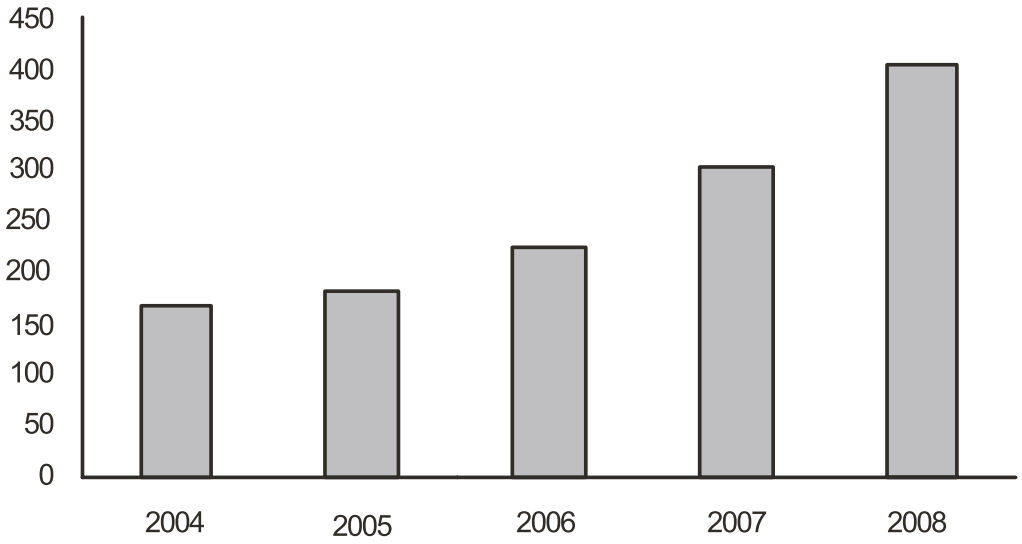


Fig. 1 - Trend of publications that use both the terms "neuroscience" and "economics" (*ScienceDirect* data base).

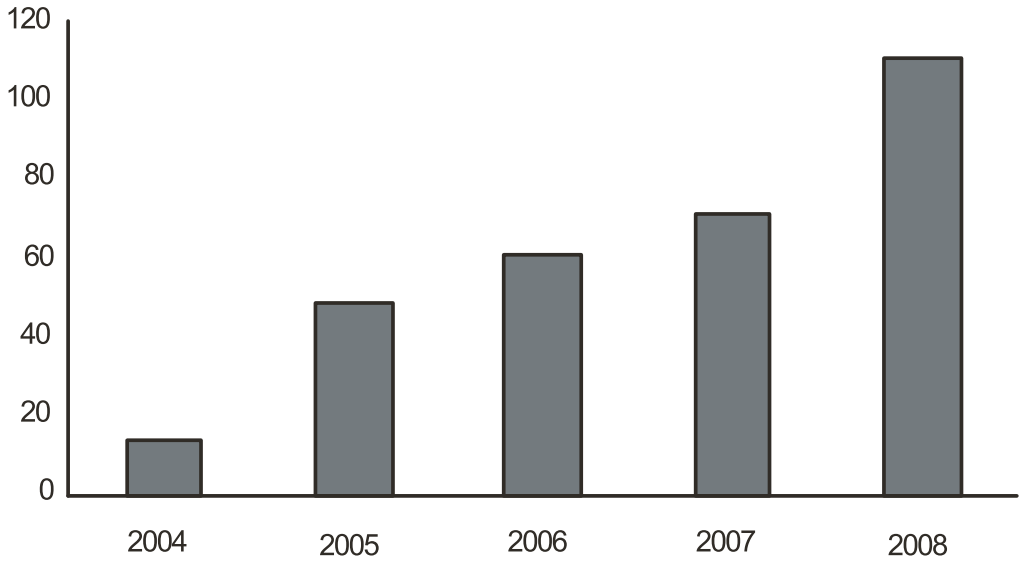


Fig. 2 - Trend of publications that use the single term "neuroeconomics" (*ScienceDirect* data base).

2. INTRODUCTION

For many centuries the study of the human social behaviour has been characterised by the implicit assumption that human beings can be conceived as *dis-embodied* beings. In the light of this assumption, contemporary social sciences have largely developed without paying attention to the underlying bodily mechanisms influencing social decisions. (Chorvat & McCabe, 2004). This is clearly the case for classical economic disciplines, which are based on the *idealistic* concept of *homo economicus*. According to this model, people interact as perfectly rational, unemotional and self-interest guided "maximizers" (Kenning & Plassmann, 2005) instead of situated social actors. Despite the copious insights of classical economical studies, the limitations of the *homo economicus* model are hard to ignore (Kenning & Plassmann, 2005). Many empirical studies have repeatedly shown that subjects frequently manifest *non-opportunistic* or *reciprocal* behaviours, as well as many other anomalies that are not predicted by the classical conceptualization (for an introductory review on anomalies in classical economics see: Camerer & Thaler, 1995). These studies have thus made it compelling the need to explore a radical revision of traditional rationalistic paradigms. In the last two decades economics began to seriously consider the insights coming from psychology. A new wave of economical studies, well represented by the rise of *behavioural economics*, emerged thereby (Thaler 1980; Camerer 2003; Rabin 2002, Camerer, Lowenstein, & Prelec, 2005). *Neurobiological studies* might represent today a new frontier to improve economical theories.

In our view *neuroeconomics* is the attempt to understand economy-related social behaviours by revealing the neurobiological mechanisms by which human decisions are taken. Differently from the two definitions introduced by Don Ross (2008) we conceive neuroeconomics as an extension of the empirical domain adopted by behavioural economics, aiming at *informing* research in economics with new data concerning the relationships between manifest behaviours and underlying neural processes. Thus, neuroeconomics doesn't represent the attempt to replace economics with neuroscience, instead it aims at offering new resources for a more complete understanding of human social interactions based on the adoption of a multidisciplinary stance.

Neuroeconomics can be considered a natural extension of *bioeconomics* (Gheslin & Landa, 1999). Whereas bioeconomics focuses on the ultimate causes of behaviour (evolutionary for example) neuroeconomics seeks to discover proximate causes of choice behaviour at the level of the anatomy and functional physiology of the human brain (Zak, 2004; Vromen, 2007). Following Glimcher & Rustichini (2004) we can say that neurobiological studies represent the ultimate element of a *discipline* which has the goal of providing a *multilevel* theory of human behaviour by relating three different contexts of analysis: the *economical*, the *psychological* and the *biological*. Neuroeconomics holds that the neurobiological analysis of social behaviour helps us understand what kinds of cognitive components are involved in decision making and may reveal unexpected structures of human mind that no behavioural study can show. Thus, understanding behavioural anomalies at the neural level might contribute to the development of a new theory of choice.

One problem is that, while economics is interested in idealised law-like generalisations, psychology is interested in concrete empirical contingencies. This difference is reflected in the fact that these two disciplines give a different meaning to the word "subject" which represents a pivotal element for both fields (Lowenstein et al., 2001). On the one hand, classical economics assumes that a subject is perfectly rational and absolutely impersonal, and that his goal is to obtain what he desires in the most efficient way. On the other hand, psychology is dealing with the empirical dynamics of a fallible subject whose behaviour is radically conditioned by many extra-rational influences. Thus, psychology and economics seem to be very different and hard to combine. But if we look deeper, we can see a strong affinity between them.

The idea that economics and psychology may advance with reciprocal advantages is not totally new. Indeed this insight was already well formed in the economical debate of the early 20's, as demonstrated by these words by Rexford Turgwell:

> Man is equipped with the psychical and physical make-up of his first human ancestors; he is the sort of being who functions best in the exhilarations and the fatigues of the hunt, of primitive warfare, and in the precarious life of nomadism. He rose superbly to the crises of these existences. Strangely and suddenly he now finds himself transported into a different milieu, keeping, however, as he must, the equipment for the old life. Fortunately his power of reflecting (there seems to be an innate tendency to reflect and learn which is a distinguishing characteristic of our species) has enabled him to persist under the new conditions by modifying his responses to stimuli. (Turgwell, 1922).

Economics and psychology have two similar domains of research: they both study human behaviour. The progressive lost of *idealisation* in contemporary mainstream economics is caused by typical psychological results concerning the importance of emotions in decision making (Bechara et al., 2005; Deppe et al., 2005). In this context, the rise of neuroscience, with its abrupt methodological development (Kenning et al., 2005) makes it possible to have new insights about *extra-rational* influences in human behaviour, by introducing a new level of analysis, different from the one offered by a *purely psychological* discipline (Bechara et al., 2005; Smith et al., 2002).

When neuroeconomics started, the explanatory power of classical economics was in serious trouble because of the evidence concerning *conflicting behavioural preferences* and so-called *irrational decisions* (Schultz, 2008). During the late 90's several converging findings set the stage for the rise of neuroeconomics (Glimcher et al., 2008). For instance, Platt & Glimcher (1999) hypothesized that a decision model may provide a powerful framework for studying the neural processes that intervene between sensation and behaviour (Platt & Glimcher, 1999).

Yet, the collaboration between economics and neuroscience is having a more crucial impact on the former field rather than on the latter (Camerer et al., 2003, Camerer, 2008). For example, the typically neuroscientific idea that the brain works in a *parallel* rather than in a *serial* way has suggested that decision making is not a monolithic computation. Several neuroeconomical models have been trying to formalise such a distinction. For example, we can speak about the dualism between "hot" and "cold" ways of reasoning (Bernheim et al., 2004) as well as of the one between "deliberative" and "affective" systems (Lowenstein et al., 2004).

As noted by Camerer et al. (2005) in order to add value to economics, neuroscience has to suggest new insights and useful solutions to old questions. Neurobiological research in economics can point out commonalities between categories traditionally conceived as distinct. For example, mainstream economics considers the pleasure from obtaining money and the pleasure from drugs (or food) as two radically different phenomena, considering the first as a an indirect value for the goods and services that it might produce. However, neural experiments show that the same dopaminergic reward circuitry in the mesolimbic system is activated by both drugs and money-related experiences (Montague & Berns, 2002; Schultz, 2002) suggesting that money provides a direct reinforcement that is independent from what people plan to buy (Camerer et al., 2005). Moreover, using event-related fMRI, Breiter et al. (2002) showed that gains produced predominant activation in the nucleus accumbens and hypothalamus of the right emisphere, whereas monetary losses produce left hemisphere activity (especially in the amygdala). Findings such as these indicate that gains produce neural *rewards*, whereas losses provoke emotional responses associated with *fear* (Zak, 2004).

To add another example, neurobiology can challange some of the most commonly accepted assumptions in economics such as the idea that a subject only strives to obtain what he likes. There is considerable contrary evidence from neuroscience and psychology that the motivation to perform an action is not always tied up to hedonic consequences. Thus, *want* and *like* may diverge (Berridge, 1999). So, as noted by Camerer et al. (2005) if we are not in the position to infer what people like from what they say they want, or from what they choose, then a different method for measuring subjective liking is required in order to avoid incorrect attributions.

3. LOSING RATIONALITY

It is a commonplace that people often take decisions under conditions of uncertainty, where they do not know the final outcome of their actions. Decisions under uncertainty look as if the subjects were choosing between *lotteries* or running a *bet*. It is thus all the more interesting to look for the mechanisms governing their choices. In this section we are going to focus our attention on this topic.

Economics typically explains decision making in terms of the so called "expected utility" model - EU model – (Bernoulli, 1738). According to this model, we choose between alternative courses of action by assessing the utility of all the possible outcomes, weighing it by the probability of their occurrence and then selecting the course of action that yields the greatest sum. In its classical form, the EU hypothesis claims that human beings take decisions by estimating these parameters as if they were engaged in an *explicit reasoning of declarative knowledge*. Yet, despite the great success of this model during the last century, this claim seems to be a limit of the EU theory. Over the last decades of the 20th century, research in neuroscience suggested that reasoning might be based on mechanisms that are – at least in part – different from those related to the production of propositional knowledge, so that such knowledge does not seem to be necessary. This insight has quickly become a cornerstone of neuroeconomics.

The evidence suggests a decision making model according to which we implicitly direct ourselves towards positive situations and far away from negative ones. Economists call this phenomenon "loss aversion" (Lowenstein, 2008). However, even if this model can account for several instances of our behaviour, there are some anomalies that still remain unexplained (see Starmer, 2000 for a review of such cases). In order to resolve this problem, the EU model was enriched by behavioural components (Thaler, 1980). The goal was to capture the fact that people tend to overvalue lower probabilities and undervalue higher ones. Once again, this modification was very useful, but it was not sufficiently strong. Indeed, even if this assumption contributed to solving many problems, there still remain many cases that cannot be accounted for in terms of a unitary mechanism, such as the one governing people's reactions to gamble games.

An alternative explanation consists in replacing the EU model - even in its behaviourally modified form – with a model which threats *Risk As Feelings* (RAF hypothesis)(Lowenstein et al., 2001). According to it, decision making is divided into two parts: one is concerned with purely *rational* processes, the other with *emotional* reactions to risky situations. Focusing on the former part, it is possible to build a general interpretative model adaptable to every particular circumstance. On the other way, focusing on the latter, it is possible to build a more specific decision making model: according to this view, people could react in different ways to the same situation.

Close but different to the RAF model, there is the so called somatic marker hypothesis (Damasio, 1994). According to this hypothesis the decision making process encodes possible outcomes of alternative courses of action affectively. In such a way, decision making is influenced by some "somatic markers" coming from the computation of bioregulator values like, for example, emotions (Bechara et al., 2004). It follows that one of the hot spot of neuroeconomics research - maybe the hottest one – concerns the weight of emotional states upon the decision making process.

In next section we will show how the conception of rationality underlying the game theoretical framework used in social sciences can be criticized by looking at some relevant neurobiological findings, while an overall view of the relevance of emotional states in decision making will be provided in section 5.

4. TRUST GAMES

Game theory is a part of applied mathematics that is largely used in economics to capture human behaviours in *strategic situations*, i.e. situations where the choice of an individual depends on the evaluation of the possible choices of others (von Neumann & Morgenstern, 1944). One of the main purposes of neuroeconomics is to provide a biologically based connection between functional and computational understanding of strategic choices.

We provide here a brief summary of some insightful findings concerning neurobiological studies on a paradigmatic kind of games called *trust games*. A typical case of trust game is the case where a player (the investor) has to decide how much from an initial endowment he should invest with a partner (the trustee) and where the partner has to decide whether to return money to the investor or not. If the trustee honours trust and returns money back to the investor, both players end up with a higher monetary payoff. However, if the trustee abuses trust and keeps the entire amount, the investor suffers a loss. The standard theoretic approach to this game predicts that a rational and selfish trustee should never honour the trust given by the investor and that a selfish investor who anticipates this behaviour will never transfer anything in the first place. This experiment mimics a vast number of real-life situations

characterised by monetary exchanges that take place in the absence of contracts enforced by a third party. Despite the predictions of the model, evidence from actual human behaviours shows that investors usually send amount of money to trustees, and that trust is frequently reciprocated (Sanfey, 2007).

Neurobiological evidence supports the hypothesis that trusting other individuals, by making *risky* investments characterized by the absence of any guarantee, is not a rational decision and it doesn't involve *only* the evaluation of monetary risks. An experiment performed by Kosfeld et al. (2005) shows that trust games also engage a complex social evaluation instead of only a monetary computation. Neurochemical results show that infusing the *synthetic neuropeptide oxytocin* in players of a trust game makes them more trusting than players in a control group and less fearful to be betreyed during social interactions. This effect is consistent with the hypothesis that oxytocine influeces amygdala activity (Kirsch, et al., 2005) a brain region involved in *trustworthiness of faces* and processing of ambiguous events with *social implications* (Adolph et al., 2005; Hsu, et al., 2005).

In an fMRI experiment, King-Casas, et al. (2005) investigated correlations between area-specific hemodynamic variations within players in a *multiround* trust game in which the same two individuals (one defined the "investor", and the other the "trustee") played ten consecutive rounds. By adopting the multiround format in the game trust becomes *bidirectional*, so that both the investor and the trustee assume the risk that money might not be returned by the partner. A multiround version makes it possible to evaluate how reputation evolves during the time, as players develop models of one another through iterated exchanges.

The authors were interested in studying how cross-brain correlations might change as the game develops and the subjects built more accurate models of each other. The experiment was conducted under the hypothesis that specific parts of the trustee's brain should be correlated with the neural consequences of the investor's decision to trust, so that the investor's ''intention to trust" would show strong cross-brain correlations. Results indicate that brain activity in the middle cingulated cortex (MCC) of the investor's brain and the anterior cingulated cortex (ACC) activity of the trustee's brain were strongly correlated, so that MCC of the investor was strongly active when the investor lodged a decision, and the ACC of the trustee was strongly activated as a reactive signal when the investor's decision was revealed.

Correlations measured during successive rounds showed a remarkable temporal transfer from the time immediately after the revelation of the investor's decision to the time immediately before the very same revelation, passing from a reactive correlation to an anticipatory one. The authors advanced the hypothesis that this shift is associated with the development of a model of the investor's intention to trust in the trustee's brain, figuring a typical *reinforcement learning process* in the context of a social exchange. Experiments such as this attempt to show how, in a multi-round game, trust becomes bidirectional and that social decision-making critically depends on internally represented models about the partners. Instead of only utility computation, experimental evidence indicates that trust-behaviours are dependent on repeated *social interactions* and *reputation-formation* (Fehr, et al., 2003).

Recently, Krueger et al. (2007) used an fMRI to measure brain hemodynamic variations while people interact with one another in a sequential trust game. In this experiment subjects were asked to make sequential decisions for monetary pay-offs such that the investor could either quit the game by not trusting the trustee, resulting in a small equal pay-off for both, or continue the game by trusting the other subject, hoping to receive a better pay-off. On the other side, the trustee was asked to decide whether to reciprocate the trust, determining a higher pay-off to both subjects, or defect the trust and keep all the money for himself. Results show that the brain regions involved in mentalizing processes such as the *rostro medial*

prefrontal cortex and the *anterior insula* are commonly recruited for both decisions to trust and decision to reciprocate, supporting the hypothesis that trust games dynamics involves a *mentalizing process*, i.e. the social ability to represent another person's psychological perspective and intentions (Krueger et al., 2008).

Several other studies linking trust games with mind reading abilities were performed in the past ten years. An example is the experiment made by McCabe, et al. (2001) which shows that prefrontal cortex region, generally considered as the region subserving mind reading processes (Fletcher, et al., 2004) are more active when subjects are playing with a human than when they are playing against a machine following a fixed and known probabilistic strategy. This result supports the idea that decision making within trust games involves the adoption of a different subjective perspective and that it's not reducible to a mere utility computation.

Moreover, another fMRI experiment performed by Singer et al. (2004) shows that displaying the partner's face in a trust and cooperative game triggers brain areas usually associated with emotions such as left amygdale and insula**.** This result supports the hypothesis that seeing another person's face emotively influences the subject's representation for a future trust evaluation. This experiment confirms the study made by Winston et al. (2002) where subjects' hemodynamic variations in the brain were measured while they viewed faces and assessed each other's trustworthiness. Results showed that trustworthiness ratings were correlated with increasing activity in bilateral amygdala and right insula in response to faces judged untrustworthy. The authors argue that these results support the hypothesis that trust judgments require *automatic* and *emotive* evaluations, rather than *conscious* and *intentional* judgments.

Neurobiological findings thus suggest that trust games involve shared social systems and that human subjects use both *mentalizing* and *emotive* processes to understand the other players in the game. Even if some traditional studies do recognize a role for mind reading in economics (see Harsanyi, 1977 on the role of imaginative empathy for interpersonal utility comparisons) a perfectly *rational model*, based on the assumption of an *ideal utilitarian* subject, as well on purely introspective methods, seems to be inadequate to capture the real cognitive processes underlying decision making, otherwise involving emotive and mentalizing processes. Today, neuroscience makes it possible to cast a new light on old problem, by simultaneously vindicating the naive idea that "decisions dictated by reason are not always good, while decisions dictated by emotion are not always bad" (Frith & Singer, 2008).

5. THE INFLUENCE OF FEARFUL STATES IN DECISION MAKING

In this section we intend to analyse the way in which emotional states can interfere with decision making. One interesting aspect of this issue is whether the outlook of negative ending can elicit an emotional state and to what extent. For example Kahn (2002) monitored the brain activity by means of an fMRI, while subjects were playing a game against the experimenter. In order to win, players had to occasionally bluff their opponent, hazard getting caught and suffer a loss. As a result, experimenters found that the activation of the amygdala was higher in cases of "bluff" than in cases of "honest" games. These data become very clear if we consider that the amygdala - a subcortical area involved in elaborating the negative value of stimuli (Philips et al., 1997) as well as those areas linked to rewards (Murray 2007)- play a pivotal role in processing information related to fear (Le Doux, 1995, Le Doux, 1996, Le Doux, 2003) an emotion with a high defensive value (Phelps et al. 2000). Giving this value, fear is traditionally linked to a conservative behaviour, like monitoring the balance between outlooked state of

things and real consequences of our actions, as in the case of "bluff" conditions of the game proposed by Kahn (Kahn et al., 2002).

At the very beginning of the 21st century Knutson et al. (2001) had shown an activation trend similar to the one of the amygdala for the nucleus accumbens (NAcc) - an area of the ventral striatum. In a test in which the experimenters make the entity of the gain be variable, they noticed a direct correlation between the NAcc activation and the probability to earn higher gains: the higher is the value of this probability, the greater is the NAcc activation. Finally, it is well known that anticipatory activation of NAcc correlates with self-reported positive arousal, in line with the idea that NAcc has a key role in processing positive emotions related to rewards.

In the light of the RAF model, which gives a great relevance to the emotional dimension in the explanation of choice, it is easy to hypothesise that decision making in neurologically sane people is very different from the one of people with damages in cerebral sites involved in processing emotional information. In *Descartes' Error* (1994) Damasio claims that one of these sites is the ventromedial prefrontal cortex (VMPFC). Some years later, Bechara and co-workers (Bechara et al., 1997) subjected two groups - the former composed by people with a damage in VMPFC and the latter composed by neurologically normal persons - to a simple game test in which subjects simply had to draw a card from one of the decks in front of them. The card could be either positive or negative, i.e. it could cause either a gain or a loss. In this study, the ratio between gains and losses was such to make the subjects believe in a negative overlook. The result of this test was that, giving a negative extraction in a certain deck, people with damages in VMPFC restart to draw from this deck before neurologically sane subjects. In other words, people with lesions in VMPFC forget the negative event faster than those of the control group.

More recently, in Shiv and co-workers' work (Shiv et al., 2005) this experimental paradigm was modified into a set involving an advantageous game instead of a non advantageous one. This test consisted in a "heads or tails" game, but with one additional condition: there was a strong positive asymmetry between the value of gains and those of losses. Participants were recruited according to the presence of cerebral lesions in proximity of sites involved in processing fearful states. Assuming that fearful states have inhibitory effects upon our behaviour, it is reasonable to claim that people with damages in the fear-system will obtain more gains than people normally impressionable by fear. The study of Shiv provided a good empirical corroboration of this theoretical prediction.

Although there is a broad consensus upon the idea that loss aversion is functionally linked to brain regions involved in processing fear (Camerer, 2005; Trepel, 2005; Hariri, 2002) we should mention that this is not the only possible interpretation. To illustrate this point, consider the so called "endowment effect", i.e. the phenomenon according to which people give a higher value to their objects than to objects owned by others (Thaler, 1980, Kahneman et al., 1990). Weber (Weber et al., 2007) combined an fMRI with the following test. The experimenters gave to the participants some mp3 songs, and the subjects had to decide at what price they were willing to sell these songs. Additionally, they were required to say how much they were willing to pay for other people's songs. The results were both behavioural and neuroscientific. Behaviourally, this study strengthened the idea of a ruling endowment effect. Neurologically, this study showed that the amygdala is more activated in the "sell" condition than in the "buy" one. Such result seems to be in line with the abovementioned ideas concerning the biological basis of loss aversion. Unfortunately, such an interpretation is not totally unambiguous. Indeed the words "buy" and "sell" are not synonymous of – respectively - "gain" and "loss": a "buy" condition implies a loss of a things and a gain in money, as well as a "loss" condition implies a loss of money and a gain of things. Thus, the amygdala activation

cannot be explained only as a cause of loss aversion, simply because it is impossible to speak, properly, about a loss. Further investigations are required in order to solve this ambiguity.

Stronger evidence comes from the study of the VMPFC complex. For example Tom (Tom et al., 2007) combined a gamble test with an fMRI study in which the experimenters monitored the activity of VMPFC. In this game, the probability of a gain was equal to the one of a loss, while the entity of the gains was made run. The experimenters noted that the activity of VMPFC increased when the gain value increased and not after its decrease. On the other hand, Xue and co-workers (Xue et al., 2007, Xue et al., 2009) have focused their attention on the dorsal division of the prefrontal cortex (DMPFC) in order to assess whether there is a correlation between this brain region and decision making in unpredictable situations, i.e. those in which cues about behavioural endings are lacking. They found a positive correlation. In the light of their data, it is reasonable to conclude that the DMPFC processes information related to risk, whereas VMPFC seems to be involved in the elaboration of reward signals.

6. CONCLUSION

Despite encouraging findings, some of which introduced in this review, at present neuroeconomics should be still considered at its early stage. Today neurobiology is in fact characterised by different levels of knowledge. Some topics are largely explored - e.g. the abovementioned processes underlying emotions – while others are still at the beginning.

As this review shows, neuroeconomics does not imply the adoption of a reductive stance. That is, neuroeconomics doesn't aspire to reduce all form of economic behaviour to mere neural processes. Instead it aims at suggesting new empirical data for the development of future studies in economics. Knowledge about the underlying neural mechanisms can generate or substantiate hypotheses about what modulating factors influence social activities (e.g. the role of emotions in decision making) which in turn might be generalised in new theoretical behavioural descriptions. This makes it possible to reject the basic charge that neuroscience cannot influence economic modelling *in principle*, but shows how seemingly disparate neural and social sciences can collaborate.

Finally, we suggest that both further methodological analyses concerning the possibility of defining experimental sets mirroring effective economical dynamics, as well as a better discussion of procedures adopted in brain imaging studies, are required in order to increase the interdisciplinary prospective of neuroeconomics.

BIBLIOGRAPHY

Adolphs R., Gosselin F., Buchanan T., Tranel D., Schyns P. & Damasio, A. (2005) A mechanism for impaired fear recognition after amygdala damage. *Nature*, 433, 68-72.

Bechara A., Damasio A.R. (2005) The somatic marker hypothesis: a neural theory of economic decision making, *Games Economic Behavior*, 25, 336372.

Bechara A., Damasio A.R., Damasio H., Anderson S.W. (2004) Insensivity to future consequences following damage to human prefrontal cortex, *Cognition*, 1994, 50, 7-15.

Bechara A., Damasio H., Tranel D., Damasio A.R. (1997) Deciding advantageously before knowing the advantageous strategy, *Science*, 275, 1293-95.

Bernheim B.D., Rangel A. (2004) Addiction and cue-triggered decision precesses, *American Economical Review*, 94, 1558-90.

Bernheim D. (2008) On the Potential of Neuroeconomics: a Critical (but Hopeful) Appraisal. (Working Paper 13954).

Bernoulli D. (1738) *Exposition of a new theory on the measurement of risk*. Transl. Sommer L., *Econometrica*, 1954, 22, 23-36.

Berridge K. (1999) Food reward: Brain substrates of wanting and liking. Neuroscience & Biobehavioral Reviews, 20 (1) 1-25.

Breiter H., Aharon I., KaHneman D., Dale A., & Shizgal P. (2002) Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains and Losses. *Neuron* , 619–639.

Camerer C.F. (2003) *Behavioral Game theory Experiments in strategic Interaction*, Princeton University Press.

Camerer C.F. (2008) Neuroeconomics: Opening the Gray Box, *Neuron*, 60, 416-419.

Camerer C., & Thaler R. (1995) Anomalies Ultimatums, Dictators and Manners. *The Journal of Economic Perspectives*, 9, 209-219.

Camerer C., Lowenstein G., & Prelec, D. (2005) Neuroeconomics: How Neuroscience can inform Economics. *Journal of Economic Literature*, XLIII, 9–64.

Chorvat T., & McCabe,K. (2004) The brain and the law. *Phil. Trans. R. Soc. Lond. B.*, 359, 1727-1736.

Damasio A.R. (1994) *Decartes' Error: Emotion, Reason and the Human Brain*, New York, Putnam.

Deppe M., Schwindt W., Kramer J., Kugel H., Plassmann H., Kenning P., Ringelestein E.B. (2005) Evidence for a neural correlate of a framing effect: Bias-specific activity in the ventromedial prefrontal cortex during credibility judgments, *Brain Research Bullettin*, 67, pp. 413-421.

Fehr E. & Fischbache U. (2003) The nature of human altruism. *Nature*, 425, 785-791.

Fletcher P. C., Happé, F., Frith U. Baker S. C., Dolan, R. J., & Frackowiak, R. S. (2004) Understanding Intentions in Social Interaction: The Role of the Anterior Paracingulate Cortex. *Journal of Cognitive Neuroscience*, 16, 1854 -1863.

Frith C. & Singer T. (2008) The role of social cognition in decision making. *Phil. Trans. R. Soc.*, 363, 3875-3886.

Gheslin M., & Landa, J. (1999) The emerging discipline of Bioeconomics aims and scope of the journal of bioeconomics. *Journal of Bioeconomics* , 5–12.

Glimcher P. & Rustichini, F. (2004) Neuroeconomics: The Consilience of Brain and Decision. Science.

Glimcher P., Camerer C. & Poldrack R. F. (2008) Neuroeconomics: Decision Making and the Brain. Academic Press.

Gul, F. & Pesendorfer W. (2008) The case for mindless economics. In A. Caplin, & A. Schotter, The Foundations of Positive and Normative Economics: *A Handbook* (pp. 3-42). Oxford University Press.

Hariri A.R. Mattay V.S., Tessitore A., Kolachana B., Fera F., Goldman D., Egan M.F., Weinberger D.R. (2002) Serotonin transporter genetic variation and the response of the human amygdala, *Science*, 297, 400.

Harmon-Jones E., & Winkielman, P. (2007) *Social Neuroscience. Integrating biological and psychological explanations of social behaviors.* The Guilford Press.

Harrison, G. W. (2008). Neuroeconomics: a Critical Reconsideration. *Economics and Philosophy*, 24, 303–344.

Harsanyi J. (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge, Cambridge University Press.

Hsu M., Bhatt, M. Adolphs R., Tranel, D., & Camerer, C. (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science*, 310, 1680–1683.

Kahn I., Yeshurun Y., Rothshein P., Fried I., Ben-Bashat D.et al. (2002) The role of the amygdale in signaling prospective outcome of choice, *Neuron*, 33, 983-984.

Kenning P., & Plassmann H. (2005) NeuroEconomics: An overview from an economic perspective. Brain Research Bulletin, 67, 343–354.

King-Casas B., Tomlin D., Anen C., Camerer C., Quartz, R., & Montague, R. (2005) Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science*, 308, 78 - 83.

Kirsch, P., & al., e. (2005) Oxytocin modulates neural circuitry for social cognition and fear in humans. *Journal of Neuroscience*, 25, 11489–11493.

Knutson B., Fong G.W., Adams C.M., Hommer D. (2001) Dissociation of reward anticipation and outcome with event-related fMRI, *NeuroReport*, 12, pp. 3683-87.

Kosfeld M., Heinrichs M., Zak P., Fischbacher, U., & Fehr, E. (2005) Oxytocin increases trust in humans. *Nature*, 435, 673–676.

Krueger F., Grafman J., & McCabe K. (2008) Neural correlates of economic game playing. *Phil. Trans. R. Soc.*, 363, 3859-3874.

Krueger, McCabe K., Mol J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007) Neural correlates of trust. *PNAS*, 20084–20089.

Le Doux J.E. (1995) Emotion: clues from the brain, *Annual Review of Psychology*, 46, 209-235.

Le Doux J.E. (1996) *The Emotional Brain*, New York, Simon & Schuster.

Le Doux J.E. (2003) The emotional brain, fear and the amygdala, *Cell. Mol. Neurobiol*, 23, 727-738.

Lee, N., Broderick, A., Chamberlain, L. (2007) What is neuromarketing? A discussion and agenda for future research. *International Journal of Psychophysiology* 63, 199–204.

Lowenstein G., Weber E.U., Hsee C.K., Welch N. (2001) Risk as feelings, *Psychological Bullettin*, 127, 267-286.

Lowenstein, G.; Rick, S.; Cohen, J. D. (2008) Neuroeconomics. *Ann. Rev. Psychol.*, 59, 647-672.

McCabe K., Houser D., Ryan L., Vernon Smith V., & Trouard S. (2001) A functional imaging study of cooperation in two-person reciprocal exchange. *PNAS*, 98, 11832–11835.

McClure S., Li J., Tomlin D., Cypert T., Montague, L., Montague, R. (2004) Neural Correlates of Behavioral Preference for Culturally Familiar Drinks. *Neuron*,. 44, 379–387.

Miller E.K., Cohen, J.D. (2001) An integrative theory of prefrontal cortex function, *Annual Review of Neuroscience*, 24, 167-202.

Montague R., & Berns G. (2002) Neural Economics and the Biological Substrate of Evaluation. *Neuron*, 265–284.

Murray E.A. (2007) The amygdala, reward and emotion, *TRENDS in Cognitive Sciences*, 11, 489-497.

Phelps E.A. O'Connor K.J., Cunningham W.A., Funayama E.S., Gatenby J.C., et al. (2000) Performance on indirect measures of race evaluation predicts amygdale activation, *Journal of Cognitive Neuroscience*, 12, 729-738.

Philips M.L. Young A.W., Senior C., Brammer M., Andrew C., Calder A.J., Bullmore E.T., Perret D.I., Rowland D., Williams S.C., Gray J.A., David A.S. (1997) A specific neural substrate for perceiving facial expression of disgust, *Nature*, 389, 495-498.

Platt & Glimcher. (1999) Neural correlates of decision variables in parietal cortex. *Nature*, 400, 233-238.

Rabin, M.A. (2002) Perspective on psychology and economics, *European Economical Review*, 46, 657-686.

Ross D. (2008) Two Styles of Neuroeconomics, *Economics and Philosophy*, 24**,** 473–483.

Sanfey A. (2007) Social Decision-Making: Insights from Game Theory and Neuroscience. *Science*, 318, 598-602.

Schultz W. (2002) Getting Formal with Dopamine and Reward. *Neuron*, 241–263.

Schultz W. (2008) Neuroeconomics: the promise and the profit. *Phil. Trans. R. Soc. B*., 3767-3769.

Shiv B., Lowenstein G., Bechara A., Damasio H, Damasio A.R. (2005) Investment behavior and the negative side of emotion, *Psychological Science*, 16, 435-439.

Singer T., Kiebel, S., Winston, J. S., Dolan, R., & Frith, C. (2004) Brain Responses to the Acquired Moral Status of Faces. Neuron, 41, 653–662.

Smith K, Dickhaut J., McCabe K., Pardo J.V. (2002) Neuronal substrates for choice under ambiguity, risk, gains and losses, *Management Sciences*, 48, 711-718.

Thaler R.H. (1980) Toward a positive theory of consumer choice, *Journal of Economical Behavior*, 1, 39-60.

Thaler R.H. (2000) From homo economicus to homo sapiens, *Journal of Economic Perspective*, 14, 133-142.

Tom S.M., Fox C.R., Trepel C., Poldrack R.A. (2007) The neural basis of loss aversion in decision-making under risk, *Science*, 315, 515-518.

Turgwell R.G. (1922) Human nature in economic theory, *Journal of Political Economy*, 30, 317-345.

Von Neumann, J., & Morgenstern, O. (1944) *Theory of Games and Economic Behavior*. Princeton University Press.

Vromen J. (2007) Neuroeconomics as a Natural Extension of Bioeconomics: The Shifting Scope of Standard Economic Theory. *Journal of Bioeconomics*, 9, 145-167.

Weber B., Aholt A., Neuhaus C., Trautner P., Elger C.E., Teichert T. (2007) Neural evidence for reference dependence in real-market-transactions, *NeuroImage*, 35, 441-447.

Winston J., Strange, A., B., O'Doherty, J., & Dolan, R. J. (2002) Automatic and intentional brain responses during evaluation of trustworthines of faces. Nature neuroscience, 5 (4) 277-284.

Xue G., Lu Z., Levin I.P., Weller J.A., Li X., Bechara A. (2009) Functional Dissociations of Risk and Reward Processing in the Medial Prefrontal Cortex, *Cerebral Cortex*, 19, 1019-1027.

Zak P. (2004) Neuroeconomics. *Phil. Trans. R. Soc. Lond. B*., 359, 1737–1748.

Intervista

Riccardo Dugini

Responsabile di Banca Etica Firenze

A cura di Laura Beritelli e Scilla Bellucci

LE BANCHE ALTERNATIVE IN ITALIA

1. Per iniziare vorremmo che ci parlasse di come nasce Banca Etica, di quali sono le tradizioni a cui si ispira e quali sono i meccanismi d'attuazione dell'etica in finanza.

Evitando di raccontare una storia troppo lunga, anche se accennerò ad alcuni movimenti rilevanti che ci sono stati d'ispirazione, direi che, nell'accezione moderna, si inizia a parlare di finanza etica all'inizio del '900, nel mondo anglosassone, con movimenti di ispirazione religiosa, protestante, il cui scopo era quello di dirottare risorse finanziarie verso attività che fossero scevre da implicazioni in campo bellico. Per quanto riguarda più da vicino il nostro territorio, le prime esperienze sono senz'altro quelle delle Società di Mutuo Soccorso, esistenti in Toscana da secoli e credo che rientrino a pieno titolo in questo tipo di attività.
È evidente che visibilità si comincia ad acquisirla globalmente a partire dalla costituzione della Grameen Bank, in Bangladesh, per iniziativa di Yunus, agli inizi degli anni '70. Tra il '73 e il '74 Yunus viaggiò nelle zone rurali del Bangladesh, anche per rendersi conto di quali fossero le potenzialità, da un punto di vista microeconomico, di alcune comunità rurali e scoprendo che proprio la possibilità di accesso al credito, totalmente preclusa a queste comunità, poteva diventare il motore per affrancarsi da meccanismi di schiavitù e per costruire realtà di microimprenditoria. A me piace ricordare il fatto che il primo finanziamento della Grameen Bank fosse qualcosa come un credito di 20 dollari, finalizzato a dotare un gruppo di donne di un villaggio delle strutture minime per proseguire l'attività che loro già facevano per un padrone, attività di produzione di cesti. La Grameen Bank ha avuto uno sviluppo che più o meno è noto, uno sviluppo straordinario che conta oggi parecchie migliaia di operatori in Bangladesh e un numero esorbitante di beneficiari. La sua attività si è internazionalizzata, non tocca solo sfere di quello che noi definiamo il Sud del mondo, ma anche aree metropolitane di città occidentali, Chicago fra le altre, ma anche nell'est europeo, in Kosovo. È evidente che, a livello europeo e a livello italiano, l'esperienza della Grameen Bank - per quanto peculiare di quel paese e non replicabile negli stessi termini - è diventato un modello a cui ispirarsi. In Italia, dagli anni '80, cominciano a nascere le Mutue Auto Gestioni, le MAG, coll'intento, almeno iniziale, di supportare l'attività del commercio equo e solidale. Credo sia riduttivo dire che le MAG poi abbiano fatto solo questo, hanno fatto molto di più, stanno facendo tuttora delle cose assolutamente pregevoli a sostegno dell'economia solidale e dei progetti di microcredito a livello locale. L'idea della Banca Etica, che risale agli inizi degli anni '90, discende dalla matrice del commercio equo e solidale e dall'esperienza delle MAG e nasce proprio nel momento in cui si ritiene necessario creare una struttura che sia complementare all'attività delle MAG in termini di servizi, per esempio per le potenzialità di raccolta, e servizi alla clientela come quelli di una banca qualsiasi. Da un punto di vista ideale, filosofico appunto, l'idea di Banca Etica viene dalla promozione del concetto di 'risparmiatore consapevole': consapevole in quanto trova una sua espressione solo nel momento in cui gli vengono dati gli strumenti per conoscere l'utilizzo che viene fatto del denaro che deposita in banca. Per questo la banca si è dotata da subito di uno strumento essenziale, quello dell'assoluta trasparenza

sulle iniziative da finanziare - che è cosa diversa dal principio della trasparenza sui prodotti, che è dovuta per legge. Chiunque, fatta eccezione naturalmente per i privati cittadini (per chi a noi chiede prestiti, microcredito o il mutuo prima casa), chieda credito in Banca Etica sottoscrive nella domanda l'autorizzazione a rendere pubblico questo tipo di dato: è la pubblicizzazione delle attività finanziate (il sito ne contiene un elenco molto consistente, aggiornato mensilmente).

2. Riguardo al passaggio dalle MAG a Banca Etica, si tratta di un'esperienza territoriale o è un processo verificatosi anche altrove? Quali sono stati i primi passi per arrivare a costituire la Banca?

L'esperienza della Banca Etica comincia a concretizzarsi nel '94-'95 a Padova, sede originaria e attuale della banca, con la costituzione di una 'associazione verso la Banca Etica'. Dopo un anno, l'associazione si converte in 'cooperativa verso la Banca Etica' ed inizia un percorso di raccolta di denaro con un appello alla cittadinanza, nell'intento di creare una banca di credito cooperativo. In seguito, avvertito il limite della banca di credito cooperativo, proprio perché essenzialmente territoriale, avviene in corsa il cambiamento per andare in direzione di una banca popolare; ossia di una banca che continui ad avere caratteristiche di cooperativa, ma che potenzialmente operi su tutto il territorio nazionale. L'obiettivo originario per riuscire ad avere le autorizzazioni dalla Banca d'Italia era di raggiungere 12 miliardi di lire di capitale sociale per costituire banca popolare, obiettivo che, a forza di sottoscrizioni - da più o meno 100.000 lire dell'epoca -, viene conseguito alla fine del '98. L'apertura della banca risale all'8 di marzo del '99. Tra i partner e sostenitori storici, che per noi significa, brutalmente, chi ci ha creduto per primo, oltre ad un buon numero di persone fisiche, ci sono molti soggetti del terzo settore, alcuni molto noti: si tratta di alcune sigle di ispirazione cattolica, religiosa come Acli Caritas, Emmaus, Manitese; alcune di ispirazione laica come ad esempio Arci, Amnesty International, alcune sigle sindacali e una rappresentanza piuttosto significativa di organizzazioni che operano in campo ambientale, come Greenpeace. La banca nasce dunque da una eterogeneità di soggetti: questa eterogeneità nel corso degli anni si è ampliata ed oggi tra i soggetti clienti e soci della banca ci sono rappresentanze delle esperienze più diverse, come quelle vicine all'antagonismo sociale. Noi riteniamo che ciò sia una ricchezza.

3. Data la genesi che ci ha descritto, che evidenzia un percorso di sviluppo a partire dall'esperienza delle cooperative e dell'associazionismo, viene spontaneo chiederle quali siano stati i soggetti che hanno messo a disposizione le competenze di scienza bancaria.

Nei primissimi mesi di attività, la banca contava più o meno su tredici risorse, 13 persone. La composizione, da un punto di vista di organico, e la provenienza di queste persone era la più disparata: c'erano soggetti che venivano dall'associazionismo del terzo settore, dalla cooperazione internazionale e c'erano dei bancari. Per ovvie ragioni, aumentando negli anni la complessità del lavoro ed essendo in presenza di uno sviluppo estremamente veloce – le vulnerabilità di questa organizzazione, dove può averne, sono proprio connesse ad uno sviluppo estremamente rapido - è stato necessario acquisire professionalità bancarie superiori, o perlomeno fare sì che la composizione all'interno della sfera socio-culturale della banca fosse ben contemperata dalla presenza di soggetti che venissero da un'esperienza bancaria forte. Siamo cresciuti molto rapidamente, in termini di mole di lavoro sicuramente, in termini di capitale sociale anche: è evidente che l'aspetto patrimoniale procede con una

velocità di accrescimento che non va di pari passo con altri tipi di sollecitazioni. Ancora oggi il capitale sociale è abbastanza contenuto. Questo perché chi concorre al patrimonio della banca sono soggetti finanziati; chi accede al credito di Banca Etica deve essere socio. Questo è un aspetto della nostra logica operativa: essendo una banca cooperativa, noi assistiamo i soci. Con l'eccezione del microcredito: ovviamente, ad una persona che si rivolge a noi con necessità di tipo socio assistenziale non chiediamo l'adesione. Possiamo contare, ogni tanto, sull'apporto, anche significativo, degli enti locali. Tanto per dire qualcosa sulla situazione locale: il Comune di Firenze è socio della banca, Provincia e Regione sono socie della banca e un buon numero di comuni della Toscana sono soci della banca. Sottoscrizioni molto significative è difficile averne, anche perché esiste un limite stabilito per legge oltre il quale non è possibile conferire capitale sociale.

4. Perché, dal momento che questo principio viene comunque rispettato al di là dell'apporto al capitale sociale di un singolo soggetto, esiste un limite?

Perché, se un soggetto che entrasse con un capitale molto importante, dovesse andarsene, questo creerebbe non pochi problemi. La banca conta oggi quasi trentamila soci sparsi per tutta Italia, che sono il patrimonio più importante della banca. Considerate che questa banca ha circa 170 addetti – in strutture di piccole-medie dimensioni - e 12 sportelli in tutta Italia. Questo di Firenze è l'unico sportello in Toscana e ci sono regioni in cui non ci sono ancora sportelli. C'è una rete di promotori piuttosto consistente, una trentina, perché lo scopo è quello di coprire tutte le regioni d'Italia. La rete dei soci ha un ruolo attivo - naturalmente è impensabile che trentamila soci siano tutti attivi, anche se le assemblee contano una larga partecipazione, sulle 1200-2000 presenze votanti - soprattutto sul territorio ed è organizzata in circoscrizioni locali con un proprio ordinamento, che è espressione 'politica' della rete. A Firenze naturalmente c'è un coordinamento di cinque, sette persone a fronte di 700 soci. All'interno del coordinamento si fanno attività di carattere divulgativo, culturale, ma anche attività di carattere molto pratico; ad esempio, a Firenze, ci sono due valutatori sociali. La banca, da sempre, all'istruttoria di tipo tecnico che fanno i professionisti della banca, affianca un'istruttoria di carattere socio-ambientale che valuti la sostenibilità di una richiesta di credito per un progetto ed indaghi sui meccanismi di partecipazione interna, sulle pari opportunità, sulle attenzioni di carattere ambientale e così via.

5. Quindi c'è un processo di selezione?

Sì, c'è un processo di selezione che funziona per iniziativa della rete dei soci. Questa istruttoria parallela avviene a cura di volontari, cioè di valutatori sociali che sono soci iscritti a un albo, formati adeguatamente. Questo processo in verità è perfettibile: la disponibilità di valutatori sociali non è sufficiente; però sta andando avanti in maniera sempre più puntuale.

6. Com'è organizzata la formazione dei valutatori sociali? Passa da voi o viene operata da altri?

Passa dall'area socio-culturale e dall'area progetti della banca, con dei corsi – quest' anno sono stati fatti per area territoriale perché l'attività è divisa per macroaree. La rete di conoscenze e il radicamento territoriale della banca, che per noi è una specie di fissazione, non è assolutamente conseguibile con gli operativi della banca, stante il loro numero esiguo: lo si ottiene con la presenza di soci attivi. Io credo che sarebbe stato impensabile riuscire a fare credito in alcune situazioni, in alcune zone geografiche ad alta penetrazione criminale,

mafiosa, tipo la Locride, se non ci fosse stata la serenità di poterlo fare perché qualcuno ci indirizzava alle situazioni e ai progetti giusti, soprattutto in momenti in cui la banca registrava ancora l'assenza totale di filiali nel Sud d'Italia. L'apertura della filiale di Napoli, se non erro, risale a tre, quattro anni fa: a Palermo nel 2007, a Bari nel 2008.

7. Può parlarci dei meccanismi con cui Banca Etica riesce a fare microcredito?

Noi riteniamo di giocare i nostri valori di caratterizzazione su concetti come la *partecipazione*, e sul concetto di *mutualità*: facciamo condizioni omogenee, uguali per tutti. Con una piccolissima forbice di differenziazione per i soggetti convenzionati o meno. Una delle anomalie rilevate, in ambito bancario - e lo dico con serenità -, avendo fatto il bancario per 14 anni in una banca ordinaria, è il livello omogeneo da Nord a Sud. Solitamente il costo del denaro nel Sud Italia è molto più elevato.

8. Di preciso quali sono queste caratterizzazioni cui vi siete orientati?

La banca si è imposta, a un certo punto, di caratterizzarsi su tre settori specifici: tutela ambientale – settore sul quale la banca ha sempre avuto l'ambizione di muoversi, ma che, per anni, ha sostenuto fondamentalmente con il sostegno all'agricoltura biologica e a qualche impianto significativo di produzione di rinnovabili. Però è innegabile che negli ultimi anni ha impresso un'accelerazione con un pacchetto di prodotti rivolti alla tutela ambientale a beneficio di persone giuridiche, associazioni, cooperative e quant'altro, ma anche a livello domestico, per esempio, c'è un prodotto a favore del fotovoltaico; vengono premiate le ristrutturazioni della prima casa qualora abbiano criteri di efficienza energetica, quindi con condizioni inferiori alle standard. Il secondo segmento di attività doveva essere quello dell'*housing* sociale e su questo siamo abbastanza indietro ancora, dobbiamo migliorare, però stiamo finanziando alcuni progetti di autocostruzione. Poi c'è il settore del microcredito: per ragioni ovvie, oggi come oggi, non aspiriamo a diventare banca di microcredito a livello nazionale, perché non avremmo i numeri. L'aspirazione è quella di creare sui territori piccoli progetti, di creare soprattutto buone prassi che siano replicabili. C'è chi, anche tra i soci, sostiene che l'obiettivo della banca sarebbe raggiunto quando della banca non ci fosse più bisogno, che essa agisca come una provocazione culturale. Per un operativo non è bellissimo sentirselo dire! Possiamo infatti rivendicare numeri anche piccoli, ma significativi.
Tornando al microcredito: abbiamo a macchia di leopardo in tutta Italia dei progetti di microcredito. Noi crediamo che nessuna banca, anche ben strutturata, sia in grado di portare avanti da sola un progetto di microcredito e che sia fondamentale avere una triangolazione di soggetti, dove la banca faccia la banca; un soggetto di rete - un'associazione o un network di associazioni - con forte radicamento territoriale faccia l'attività di accompagnamento dei soggetti richiedenti, sia che si lavori nell'ambito socio-assistenziale che nell'ambito microimprenditoriale; un terzo soggetto faccia da garante, cioè metta a disposizione un fondo di garanzia. Spesso il garante è lo stesso soggetto che fa l'attività di accompagnamento. Per cui, per esempio in Toscana, al momento, abbiamo quattro convenzioni già attive e una quinta in corso di attivazione: una su Pisa, una su Lucca, due su Calenzano e una sui comuni della fascia Nord intorno a Firenze. A Pisa il progetto *Le città sottili*, che è finalizzato al superamento del concetto di campo nomadi come situazione abitativa, a beneficio di soggetti la maggior parte macedoni arrivati alla fine degli anni '90, spesso disertori. Qui noi ci siamo convenzionati con Arci, titolare del fondo di garanzia – con risorse arrivate dalla Regione Toscana -, con la Società della Salute locale e con Caritas. Nel progetto operano anche un paio di cooperative

sociali con i propri operatori, non sono soggetti scritti in convenzione, ma Caritas si avvale del loro lavoro. Su Lucca è stato messo su un cartello di associazioni, capofila Caritas, Arci, S.Vincenzo dei Paoli, Comunità di S. Egidio e altri. A Calenzano c'è un progetto del Comune a beneficio degli immigrati per l'accesso alla casa e un altro, più generico, rivolto a soggetti in stato di necessità temporanea (si parla sempre di necessità temporanea proprio perché si fa microcredito a soggetti che saranno poi in grado di restituire). Nel caso dei comuni intorno a Firenze, capofila è Scandicci. Anche lì c'è Arci e la cooperativa sociale Arpat, oltre alla Società della Salute. Il microcredito finalizzato alla microimprenditoria è indubbiamente quello più difficile da realizzare, anche perché qui vengono richieste competenze da parte del soggetto accompagnatore sullo *start-up* d'impresa. Si parla di piccole imprese, magari il piccolo artigiano con problemi di accesso al credito, che deve provvedere alla dotazione minima del proprio laboratorio. È difficile in questo caso trovare partner attrezzati.

9. Ci sono buoni esempi?

Qui in Toscana, un progetto interessante l'aveva messo su la Regione qualche anno fa, con un prodotto che si chiama SMOAT, gestito da FidiToscana. L'attività di accompagnamento, come noi la intendiamo, è oggettivamente assente; la cosa interessante è come, nel corso degli anni, siano cresciuti i percorsi di formazione e quindi a livello di competenza, il tutto ad opera della rete di volontari della Banca del Tempo. La banca ha sperimentato processi di microcredito attraverso Etimos che è un consorzio finanziario che opera nel Sud del mondo; ha sperimentato il microcredito soprattutto in aree dell'America Latina e dell'Africa e con successo. Però noi riteniamo che, tutto sommato, sia paradossalmente più facile fare microcredito nel Sud del mondo che qua. La ragione è legata al fatto che, spesso, quando si va a fare microcredito in America Latina si parla ancora di società rurali, la comunità è forte, la responsabilità nei confronti di qualcuno che ti ha prestato denaro è, prima di tutto, responsabilità nei confronti della comunità. Da noi, per ovvie ragioni, per caratteristiche di tessuto sociale, che è quello che è, purtroppo, soprattutto nei contesti urbani, è molto difficile; tant'è che io, personalmente, credo che le esperienze cittadine siano di un'utilità straordinaria. L'esperienza delle Piagge - cito quella, ma ce ne sono altre due sul territorio -, è particolarmente interessante nella definizione di un ambito di intervento che è limitato ad un territorio circoscritto; l'assoluta certezza che la cosa può funzionare se si opera in un contesto in cui la gente si conosce. Chi accede al credito alle Piagge si presenta a un'assemblea, diventa socio del fondo del microcredito delle Piagge, denuncia i propri problemi, necessità ecc…
Se il progetto ha buone *chanches* di funzionare, io credo che sia proprio per questo.

10. Senza queste reti d'appoggio quindi non potreste concedere il credito?

No, anzi, credo che sarebbe una presunzione e un errore da parte della banca.

11. Vuol dire che sarebbe uno strumento finanziario qualsiasi?

Certo. La banca, o le banche di per sé, salvo che non abbiano l'opportunità o possano permettersi di avere degli operatori sociali sul territorio, finirebbero col ripetere un'esperienza già fatta in passato, con esiti molto discutibili, che si chiamava 'prestito d'onore'. Sostenuta dalla Regione Toscana, anche Banca Etica ha sperimentato il prestito d'onore, con il risultato che, anche se non ci ha rimesso perché c'erano i fondi di garanzia, non essendoci qualcuno a fare l'attività di accompagnamento per la richiesta, il fondo di garanzia, dopo un anno e mezzo, era bruciato ed il progetto finito. Riguardo a quelli che sono elementi tecnici delle convenzioni,

la banca, specialmente quando entra in contatto con soggetti, a livello territoriale, che non conosce, o comunque deve testare anche il livello dell'attività di accompagnamento, scrive nella convenzione che il primo anno, a fronte di 10 di garanzia, finanzia 10, quindi non rischia niente. Poi, previa verifica di due parametri, uno legato allo stato delle sofferenze - non superiori al 20 % dell'erogato, che è un limite comunque molto alto – e l'altro legato all'utilizzo del fondo di garanzia, si comincia ad applicare quello che si chiama 'moltiplicatore', per cui il secondo anno, a fronte di dieci depositato, la banca può erogare fino a venti, il terzo anno trenta e così via.

12. Banca Etica si muove secondo delle teorie finanziarie specifiche, socialmente fondate, o agisce piuttosto in base ad una prassi e ai risultati ottenuti attraverso di essa?

Banca Etica s'ispira per statuto ai principi della finanza etica. La nostra è una prassi, muove in maniera abbastanza spontanea dalle esigenze del territorio. Non c'è un orientamento precostituito. Senza voler tacere i limiti e i difetti che abbiamo, bisogna dire che sotto l'aspetto della trasparenza siamo stati innovatori. Secondo me questo è un capitolo su cui tutte le banche potrebbero guardarsi dentro e pubblicizzare quello che finanziano - e nessuno lo fa. L'aspetto della mutualità è un aspetto assolutamente dirompente, perché, come dicevo prima, ipotizzare di far credito in Emilia Romagna o in Veneto alle stesse condizioni con cui lo si fa in Sicilia è un gesto folle. Tant'è vero che la banca ha un livello di gradimento altissimo verso i prodotti nel Sud Italia e naturalmente qualche problema in più a rendersi concorrenziale al Nord.

13. La 'provocazione culturale' di cui parlavamo prima produce un esempio comportamentale per il cliente, ma anche per le altre banche. Perché queste non raccolgono l'esempio?

Me lo chiedo tante volte anch'io, è una domanda difficilissima. Soprattutto le banche piccole non dovrebbero avere problemi. Io credo che sia legato tutto, ancora oggi, a un problema di *appeal*: le banche, laddove non verifichino u'attrattiva verso certi temi da parte del mercato, vanno per la loro strada - senza voler aprire parentesi sulle criticità che fanno certe banche o i grandi istituti di credito; ma chi fa un'attività tutto sommato ordinaria, potrebbe andare in questa direzione. Non c'è una cultura indirizzata ad un comportamento trasparente sull'utilizzo del denaro. Tuttavia, un riflesso c'è stato: dagli anni '80 ad oggi, per esempio, il linguaggio utilizzato delle banche è stato modificato. Mi ricordo che negli anni '80 girava in televisione uno spot di una grande banca dove si invitava la gente a fare investimenti in borsa coll'intento di affrancarsi dal lavoro! Di questi tempi, invece, quasi tutti parlano di etica. Quello che è certo è che oggi le banche tendono a 'fare pulizia' rispetto a delle criticità, con prodotti orientati.

14. Ci sono altre banche che applicano principi simili a Banca Popolare Etica, In Italia e all'Estero?

La banca stessa aderisce ad una federazione di banche europee che si chiama FEBEA e che riunisce alcune delle esperienze più significative di 'banche alternative'. In Europa esiste una tradizione: le prime banche alternative nascono in Germania ed all'inizio erano ispirate dai movimenti ambientalisti degli anni '70 e '80; poi si sono lentamente convertite anche in altro, hanno cominciato ad affrontare temi di carattere più sociale. Esiste anche un progetto di realizzazione di una banca etica europea che dovrebbe portare, in un futuro non troppo

lontano – almeno nelle aspettative - alla costituzione di un soggetto-cooperativa europea alla quale concorrerebbero Banca Etica, La Nef francese - che è una finanziaria -, la fondazione basca Fiare. In Italia esiste Banca Prossima, nata un paio d'anni fa, all'interno del gruppo Intesa San Paolo ed è una banca che nasce con un patrimonio molto più consistente del nostro e con l'intento iniziale, credo, di finanziare soprattutto enti ecclesiastici; poi è entrata a pieno titolo nel terzo settore. Oggi come oggi dovrebbe essere un nostro concorrente, anche se noi stentiamo a riconoscerla come tale, giacché non opera con gli stessi criteri .

15. Voi come Banca Popolare Etica siete un ente indipendente o fate parte di un gruppo più ampio?

Siamo Banca Etica e Etica Sgr che è la società di gestione di risparmio di Banca Etica, ovvero la società che promuove fondi comuni d'investimento orientati. Questo progetto si è realizzato nel 2003 in partneriato con Banca Popolare di Milano, Banca Popolare di Sondrio e alcune banche di credito cooperativo. Anche su questa campagna c'è stata, tra i soci, una scissione tra chi sosteneva l'opportunità che la banca in qualche maniera entrasse nei mercati finanziari e chi no. Ovviamente qui si tratta di gestire risorse dei risparmiatori per acquisire quote azionarie di aziende, anche multinazionali, che rispondano ai nostri criteri ed interessi. Ci basiamo su 60 parametri di valutazione. Fra i titoli rappresentati di Etica Sgr, ad esempio, non ci sono titoli dello Stato Italiano in quanto l'Italia risulta tra gli ultimi posti per fondi destinati alla cooperazione internazionale; non ci sono titoli di stato di Stati che abbiano la pena di morte o che ledano le libertà individuali.

16. Quindi questa società gestisce risparmi anche tramite l'appoggio di altre banche, diciamo, 'normali'?

Sì e, a onor del vero, abbiamo avuto anche un problema non di poco conto. È vero che ci sono altre banche ma, per statuto, la presidenza di Etica Sgr l'abbiamo noi e quindi è la banca stessa che attrezza un comitato etico per la valutazione dei soggetti, avvalendosi di un *database* di Fondazione Eiris, una fondazione londinese. Il problema di cui parlavo è sorto perché Banca Popolare di Milano, che è il gestore tecnico, tre anni fa si è ritrovata nell'elenco delle banche armate, con tutte le conseguenze del caso. C'è stato un confronto teso fra le dirigenze. La scelta di quella banca non era stata tra l'altro casuale, dato che ha dei meccanismi di partecipazione interna, anche dei lavoratori, molto interessanti. Il confronto si è protratto fino a che Banca Popolare di Milano non ha preso la decisione solenne di affrancarsi da quelle operazioni.
Anche in questo ambito, agiamo quindi tramite un partneriato attivo, che significa partecipazione assembleare e proposizione (e destinazione) dello 0,50% della sottoscrizione all'attività di microcredito. Questo fondo, che ammonta a 4-500.000€, è stato impiegato, quando non esistevano garanti ma si voleva comunque finanziare un progetto, come fondo di garanzia, come una copertura a rischio. Infine, voglio ricordare che abbiamo dato vita alla nostra Fondazione Culturale, con sede a Firenze, che opera con la pubblicazione della rivista "Valori" e l'allestimento della manifestazione "Terra Futura".

17. Si può dire, allora, che in Italia non ci sono altre realtà simili alla vostra, dal momento che non riconoscete Banca Prossima come appartenente alla categoria?

Banca Prossima nasce da uno dei gruppi più armati – il quarto - e fa discrezione nell'applicazione delle condizioni di credito e *rating*. Per questo non la riconosciamo come

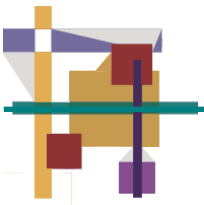
concorrente. Il Gruppo Intesa può anche fare una scelta d'immagine e rischiare di perdere su Banca Prossima senza troppi danni. Poi, tutto è relativo, ad esempio, al cospetto di tante realtà che si muovono nella finanza etica, noi siamo già un soggetto troppo istituzionale, è un appunto legittimo. Io penso che, quando si parla di finanza etica, nessuno vanti primati. Per parlare di esperienze più vicine a noi, qui a Firenze è nata l'associazione 'Verso Mag Firenze', che finora si appoggiava alla Mag6 dell'Emilia Romagna. Ho assistito alla conferenza di presentazione tenuta allo Stensen, ho letto lo statuto e per me è stato un momento di grande emozione. Credo che siano esperienze che vale la pena di rendere complementari, non necessariamente da un punto di vista tecnico ma, sicuramente, dal punto di vista culturale.

18. Quali sono i settori in cui investite maggiormente?

Gli investimenti sono rivolti ai soci, al microcredito e su quattro settori: servizi socio-sanitari educativi (tutto quello che attiene all'assistenza alla persona); cooperazione internazionale, microcredito, commercio equo e solidale, botteghe, Ong, tutela ambientale. Devo dire che esistono deroghe, perché, tendenzialmente, facciamo credito a soggetti *no profit*, ma oggi ci sono società, anche di capitale, che si occupano del sociale in maniera assolutamente ben fatta - com'è vero che ci sono cooperative sociali che mantengono l'insegna e basta. Quindi, nella tutela ambientale, si finanziano attività profit: aziende biologiche, installatori, progettisti, ricercatori in campo ambientale. Sugli impianti fotovoltaici, la banca ha deciso di finanziare alcune strutture ricettive dopo averne valutata la coerenza, ma anche un laboratorio di pelletteria. Infine il settore promozione qualità della vita, in cui rientra di tutto: attività ricreative, culturali e sportive.

19. Rispetto alla valutazione dei progetti da finanziare, come vi muovete? E qual è il rapporto che si instaura tra i vostri valutatori e il soggetto che riceve il loro controllo?

Il controllo viene preannunciato: per esperienza posso dire che il soggetto che lavora nel sociale è contento di poter riscontrare l'interesse della banca al rispetto di certi criteri. Di fronte a richieste anche importanti ma totalmente infondate da un punto di vista sociale, l'operatore si rivolge al comitato etico della banca. Posso citare un esempio clamoroso: molti anni fa è venuta qui un'organizzazione, chiedendoci un anticipo su contributi ministeriali, ci hanno detto, per la gestione di campi profughi. Siccome i campi profughi non esistevano più, alla nostra precisa domanda se si trattasse di Cpt, ci è stato risposto che così effettivamente era. Allora la richiesta di questi signori è passata al comitato etico. La filiale ha dato un parere negativo, il comitato l'ha amplificato per 10. È stata bocciata, con conseguenze anche abbastanza dirompenti nel rapporto con questo soggetto.



www.humana-mente.it

JOURNAL OF
PHILOSOPHICAL STUDIES
-
ISSN: 1972-1293