Epistemological Foundations for Neuroeconomics

Elise Payzan, Sacha Bourgeois-Gironde

First draft

Epistemological foundations for Neuroeconomics

August 2005

1 Introduction: Neuroeconomics, an inchoative field

1.1 What's Neuroeconomics?

«*Neuroeconomics*», a hitherto blurred concept, mirroring an inchoative field The concept of «*Neuroeconomics*» appeared recently in the field of economics. Colin Camerer and some other pioneer peers (Camerer et al., 2003) wrote a survey, «*Neuroeconomics*», which connotes the desire to institutionalize the new symbiosis between imaging neuroscience and experimental economics. We can assign this scientific project to the development of modern techniques – like the *Functional Magnetic Resonance Imaging* and the *Positron Emission Topography* – used in the lab by neuroscience to visualize the cortical regions activated while performing controlled tasks. As pointed out by the authors, these recent technical devices allow to go beyond behaviorism by «*opening the black box of the brain*»[1] and thereby match the realistic project to understand the human being.

Yet the epistemological nature of such an attempted integration is not addressed by the authors of this handout, aimed at presenting together some nice neural results liable to arouse the economist's interest. Such a «captatio benevolentiae» is not designed to pin down the paradigm of neuroeconomics, nor its institutional borders and functioning. To that extent, the extensive use of the label «Neuroeconomics» is more of a practical nature than of a reflexive one. That's what we call the «*indigenous*» concept of «Neuroeconomics», that we distinguish from the «*epistemological*» concept of neuroeconomics. It is not our goal to minimize this first step, which is having a great impact in making a bunch of recent researches converge toward the idea of more realism in economics and more unity in behavioural sciences[2]. Our objective is rather to outline basic aspects of

[1] An analogue opening was done in the past regarding the firm, first trivialized by the neoclassic economists before March and Simon's insights suggested new scientific examinations.

[2] It definitely helps to define new lines and new positions in the moving field of experimental and behavioural economics. Though it lets the epistemological foundations open up for the future.

a possible epistemological coherence of the research that sprouts today under the label of neuroeconomics. We think it is time for neuroeconomics to position itself within the field of economics, to construe its aims and scope[3].

Epistemological definition of the concept of neuroeconomics A preliminary definition of the epistemological concept of neuroeconomics is the very minimal condition for this to be carried out. We define neuroeconomics as the *joint experimental production* between neural sciences and experimental economics. It is a natural outgrowth of laboratory economics, aimed at studying behaviour in a controlled environment.

Two remarks are to be made at this stage:

- First it is necessary to justify our definition of neuroeconomics as *joint* production between different sciences. Indeed it presupposes a real interdisciplinary collaboration between economists and neuroscientists, one that would fulfil the promise of incorporation or integration. It appeals some common principle, since the cross-breeding of several fields – psychology, neurobiology, normative economics – into some, to be sure, quite elaborate experimental settings, is not by itself a unification of these fields, of their methods, nor of the principles that found them[4].

  One way to foresee epistemological soundness is to follow Popper in asserting that a common principle of rationality provides the ultimate unity of behavioural and social sciences (Popper, 1967). This common principle constitutes the cement of neuroeconomics. Indeed some epistemological coherence can be fostered by the consideration that the ways of interpreting behaviour that each of these fields applies rely on a common principle of rationality. Namely the epistemological unity of the different aspects – psychological, neurobiological and economical – of experimental data provided by neuroeconomics studies, are implicitly unified because it is presumed that in one way or another the agent systematically follows some reasons that can explain her apparent behaviour. This principle needs not be qualified in general, but it can be qualified in the various situations that the agent meets; and what specific experimental data may contribute to provide is precisely those

[3]This epistemological enterprise is, of course, quite compatible with a more practical or strategic one, as its outcome will be to get a better appraisal of the situation of neuroeconomics in the field: in particular it is not insignificant to know whether it stands closer to experimental economics than to psychology-oriented behavioural economics.

[4]Indeed, why would the cross-breeding between the two different fields not reduce to a mere collaboration between neuroscience and economics? The latter would «pick and choose», among the neural solidified data, the ones that match their interests. They would use them as «inputs» in the production of economic knowledge. Neurobiological data would provide then an external source of information for the understanding of human behaviour pursued in usual economic terms, but there would not be, as such, a more profound epistemological reflection as to know why those neurobiological data can constitute data that import for the understanding of economic behaviour. Once cognitive functions have been attributed to certain brain areas, it would be the purely external job of the so-called neuroeconomist to relate them to cognitive tasks or states that have an interest for the economist tout court: choice, decision, anticipation, regret, etc. Then neuroeconomics would not be a joint production, but an «epistemic outsourcing».

local justifications that are needed in view of the whole coherent account of human behaviour.

Therefore we are entitled to think of neuroeconomics as a real incorporation leading to the previously alleged «joint production», because of the shared rationality principle.

- Second notice that our provisional definition purportedly restricts neuroeconomics to the *experimental* domain of economics. We will spend a few time in this article to raise some major issues related to the experimental specificities of neuroeconomics[5]. Though we will mainly focus on the relation between neuroeconomics and the «a priori domain»[6] of economics: papers in economics can be now illustrated with pictures of the brain, but how these pictures can be an element of economic theorizing? The purpose of this article is to show and illustrate the following thesis. There are three possible main views of how neuroeconomics may relate to economics: neural data can be used to test, to hypothesize, or to induct. According to which of these three views – respectively *Deduction*, *Abduction*, and *Induction* – is privileged, neuroeconomics may be said to be established as epistemologically autonomous or not. The first view is an hypothetical-deductive understanding of economics. Neural data obtained in specific experimental settings are confirmatory data, relevant only to the extent that they constitute a way of testing existing theories. They lend the latter a veneer of realism in the sense that descriptions of brain-mechanisms can be seen as specific tests of behavioural predictions made by the theory. We show that this hypothetical-deductive approach leads to an «epistemological circle» and should definitely be considered as a dead-end.

  Now the connections between neural functioning and economics can be encompassed in quite a different epistemological framework, one which we recommend and illustrate in the following. An abductive understanding of those connections amounts to consider that neural data – now endowed with scientific primacy – may help build hypotheses or refine extant ones in order to create maximal compatibility between data and theoretical premisses. Abduction, here, means that neural data are not, as such, predicted or deduced by theoretical premises of economics. Differently, they serve to update the models by suggesting new hypothesis and urging new insights.

  The abductive conception of economics urges behavioral economists toward a greater heuristic (adaptation of one's hypotheses to account for new data) and hermeneutic (emphasis put on the purpose of explaining human behavior in all its relevant manifestations for economists) understanding of their task. Still the primacy of the existing theories remains, to the extent that the new hypothesis are a way of

[5] In particular we suggest important methodological differences between neuroeconomics experiments and laboratory economics.

[6] The theoretical domain of economics, namely the domain of the economic models, is referred to by Machlup as the «*a priori domain*» (Machlup, 1980). In the following development we will use the two expressions indifferently.

rationalizing – by updating them – the latter. The essence of neuroeconomics is not abduction according to us. We point to a somewhat ideal epistemological relation between neural data and economics that would constitute neuroeconomics as a fully epistemologically autonomous scientific discipline. This relation is inductive. It means an extension of the behavioral topics of interest. Empirical exploration comes first, then the models intervene in a second – still essential – step[7]. Once the researcher is involved in an inductive logic, each new empirical fact helping at understanding the way human-being actually functions will be meaningful, even though no economic model has ever dropped an anchor in this new direction. This inductive pathway is possible only to the extent that neuroeconomics rests on the common rationality principle mentioned before. The latter both suppresses the institutional delimitation between the traditional objects of interest proper to each field, and the primacy of preexisting theories.

1.2 Aims of this article

Which epistemological pathway for neuroeconomics The necessity of such an effort at epistemological foundations has been foreseen by some authors. A recent instance is an introduction to a volume of *Games and Economic Behavior* dedicated to the new field of neuroeconomics (Rustichini, 2005). The author, Aldo Rustichini, encompasses the use of tools, concepts and methods of neuroscience in economic analyses under the general idea of understanding the processes that underlie behaviour which are predicted by axiomatic Decision Theory and Game Theory. It seems to us that Rustichini programmatically addresses important questions for anyone involved in the deductive approach we previously mentioned: what can theoretical economics learn from neuroeconomics? Can neuroeconomics pretend to contradict the axiomatic models? Though it also seems that he simply suggests a very broad outline of the epistemology that could provide satisfactory answers to them.

Rustichini's broad epistemological perspective is to re-unite current economics with the research program that was once the one of early classics like Hume and Smith, who, according to Rustichini, sought to provide a unified theory of human behaviour. Isolated experimental data and observations won't make sufficient sense, then, and can always be the prey of sceptical rebuttal, if one does not endeavour to say why these new types of data about human behaviour enter into a unified view of human sciences whose economics is the backbone. Now special justification is provided in the cases of a phenomenon like cooperation in competitive games, once thought an irrational deviation from utility maximization. Rustichini, at this juncture, directly echoes Adam Smith's views: what holds society together and makes possible trade, competition and the whole realm of

[7] The models would be heuristic devices to illuminate the empirical facts. Our use of the term «heuristic» must be understood in a pragmatic sense, as an alternative criterion to the ontological one to judge the economic models. We deny that the ontological ingredient – is this model true or false? – is necessary to have a warranted epistemological belief that some models are better than others: a given model is good if it helps to understand some targeted aspects of the real order built by the experimental domain. Echoing Ian Hacking's terminology, we will call our epistemological viewpoint «*anti-realist about theories and realist about entities*» (Hacking, 1983).

economics, is a feeling such as sympathy which counterbalances selfishness. There are some social emotions which provide the cement of society. Interestingly enough, as we will see in more details below, this is what neuroeconomics, in its current stage, looks for in priority: it explains, for example, deviations from pure rationality in the ultimatum game, by the presence in the subject, well attested by neural data, of emotions such as trust, disgust, or revenge. This is, indeed, an indication of where neuroeconomics might head to, but this is far from offering, or even pointing to, a unified theory of human behaviour as such, and we expect more epistemological discussion, apart from prestigious ancestry, in favour of this view. Our guess is that Rustichini implicitly equates economics with the traditional objects of interest of social science, in particular the individual underpinnings of social exchange. In this logic, the move toward more realism, allowed by new information provided by neuroeconomics, would lead to a new look on these preexisting objects.

As it should clearly appear, Rustichini's Smithian viewpoint epitomizes our abductive view of economic science. Our Popperian answer to the raison d'etre of neuroeconomics is quite different: first we do not restrict a priori the scope of economics, and postulate that neuroeconomics will encompass new objects of interest, to the extent that the latter will be integrated in the rationalization process. Second the common rationality principle implies a change in the relation between theory and experience within economics, shifting it from deduction towards induction.

The first part of this article is descriptive: our motivation is to shed light on the emergence of neuroeconomics as a scientific project. We point to the conditions of the emergence of neuroeconomics, by emphasizing intellectual factors, in addition to strategic and institutional ones. We show that the neuroeconomics project can be equated with the final step of an ongoing evolution of economics toward more realism. While the axiomatic approach is still flourishing, an alternative humanistic ethos leads to experimental economics and its theoretical counterpart, behavioural economics. Both aim at understanding the real world. Neuroeconomics, an «outgrowth» of experimental economics, epitomizes this modern ethos.

The remaining of the article is normative: the objective is to determine the optimal relation between neuroeconomics and the theoretical domain. First we explain why the hypothetical-deductive approach is not relevant, by pointing to some specific aspects of neuroeconomics experiments, that prevent the confirmatory use of neural data. In particular we show that a qualitative (interpretative) stage is intrinsic to the construction of the experimental fact in neuroeconomics. Then we define the abductive pathway and illustrate its relevance through examples from the recent neuroeconomics literature. Last we try to convince the reader that the essence of the neuroeconomics project is inductive, and suggest some potentially fruitful research plans within this epistemological scheme.

Scientific stakes Before taking the reader any further in the epistemological assessment of the field of neuroeconomics, a conceptual precision is required. We think of the

production of scientific knowledge[8] as a dynamic Polya process. Before justifying such a claim, let us recall briefly what a Polya process is. Polya studied the evolution of the number of balls in an urn with infinite capacity. At the beginning of the process, the urn contains one red ball and one white ball. A new ball is added at each period, and its color – red or white – depends on the proportion of the balls having the same color, already present in the urn[9]. Hence the transition probabilities are endogenous. The characterization of the equilibrium of the process is twofold. While the proportion of red balls tends toward a limit, the solution is undetermined[10]. To put it differently, the arbitrary «starting point» is crucial, since the final situation (lock-in) depends on it, through the endogenous transition probability. That's what B.W.Arthur and al. call «*path dependency*» (Arthur, 1994) and (Arthur et al., 1987), intrinsic to non ergodicity[11]. The analogy with the epistemic process is as follows: at each period of time, the scientific community is to choose a particular object of interest from a set of potential scientific objects. The reasonable assumption we make is that the probability to choose a particular object of interest crucially depends on the previous choices[12]. As this process is characterized by non ergodicity, we should be particularly cautious regarding the initial conditions, since they largely determine the «borderline», the scientific horizon of the discipline. Therefore the *ex ante* epistemological moment is fundamental. The initial choices must be objectified and construed before neuroeconomics becomes shaped into a structured «*field*» (Bourdieu, 1978). Let us say metaphorically that it is necessary to pin down the nature of the different «balls» in our epistemic urn. This article is a tentative move in favor of this enterprise. More precisely, we claim that good starting conditions for neuroeconomics privilege hitherto unexplored objects over the ancestral topics directly derived from social sciences. Induction is a necessary condition for such an extension of the set of relevant topics.

2 How did «*Neuroeconomics*» appear?

2.1 Intellectual and institutional factors: axiology of Economics as a structured scientific field

Motion towards more realism It is necessary to outline some important institutional transformations of the economics field in order to fully understand the current develop-

[8] Henceforth we will use the term «epistemic» to refer to this production of knowledge and its scientific outputs. Therefore in our terminology «epistemic» and «epistemological» are not synonymous. The latter encapsulates the notion of reflexivity: epistemology takes some epistemic phenomena as its objects of reflection.

[9] For instance, assume that the new ball will have the same color as the ball it hits first.

[10] The value susceptible to emerge is uniformly distributed within the interval $[0, 1]$.

[11] Ergodicity is an attribute of stochastic systems that tend in probability to a limiting form that is independent of the initial conditions.

[12] The microfoundations leading to endogenous transition probabilities can be resumed as follows: each researcher is to some extent a beckerian actor aimed at her reputation within the field and thereby does not have an individual incentive to deviate from the mainstream topics. To put it differently it is one's best interest to imitate the ongoing choices. Hence a deviation from the initial trend is, at the aggregate level, unlikely.

ment of neuroeconomics.

Indeed the personality of such «empiricist» leaders like Colin Camerer and Ernst Fehr is not a sufficient reason for the emergence of the collaboration between neurologists and economists. Nor is sufficient the recent development of the new imaging techniques, that makes possible the examination of the neural underpinnings of observed behavioral patterns. Actually this scientific project is the manifestation of some deeper structural factor, some «habitus» [13]. At this stage it is noteworthy that the realistic ethos suggesting to «*open the black box of the brain*» is not the mainstream in the economics of decision theory. It mirrors a modern trend towards more realism, born in the seventies thanks to Kahneman and Tversky's insights[14]. While it is clearly opposed to the formalistic approach (the Chicago School one), which crowds out any problem that cannot be presented in mathematical form, it rests on the humanistic desire to understand the human-being and its manifold behavior. Kahneman and Tversky showed in an elegant and understandable fashion the weakness of the predictions delivered by the classical theories of individual decision-making[15]. Therefore the testability criterion, put forth by the dominant economists to legitimate their mathematical constructions, could no longer deceive the profession.

In addition to this intellectual milestone, the legitimacy of these two polyvalent figures – Amos Tversky was psychologist, philosopher, and excellent mathematician – deeply embedded in the American academic field, their scientific proximity with the prominent economist Richard H.Thaler, were conducive to a real «intellectual revolution». Thus, according to David Laibson and Richard Zeckhauser (Laibson and Zeckhauser, 1998), they initiated «*the modern behavioral economics*»[16] of the late nineties – the label comes from Richard H.Thaler (Thaler, 1997). Indeed Richard H.Thaler, George Loewenstein, David Laibson, Richard Zeckhauser, all prominent leaders in the field, refuse the escapism consisting of neglecting the real world to preserve the tractability of the models. Their common claim is «*to start updating our equations to include these behavioral factors*» (Thaler, 1997).

To some extent, this move towards more realism is a return to Marshall[17] – and more broadly to the pre-thirties epistemic ethos – after the enduring formalism period which started with Samuelson's «*Foundations*»(Samuelson, 1947). Rustichini's recent article – the one we previously mentioned – is a manifest illustration of this new mood. So is the actual proximity between economics and psychology. While some researchers can't be subsumed under one peculiar field[18], others like Jonathan Cohen, an eminent neurosci-

[13] See Bourdieu (1990) and Bourdieu (1977).

[14] See in particular Kahneman and Tversky (1972).

[15] The deviations from the predictions, for being systematic, could not be assigned to individual noises any more (Kahneman and Tversky, 1996).

[16] To many aspects the microeconomic analogue of the neokeynesianism vis-a-vis the classical macroeconomics.

[17] See Pigou's apology of Marshall (Pigou, 1925). Pigou stresses Marshall's cautiousness regarding mathematical formalism in economics: the mathematical constructions should not become «*intellectual toys*» freeing us from questioning the real world.

[18] Daniel Kahneman and Amos Tversky illustrate evidently this syncretism. Other researchers do, for example Eldar Shafir and Robyn A.Leboeuf, researchers in psychology whose interests and hitherto

entist and doctor, have a clear and asserted open-mindedness towards other disciplines. Experimental designs used by neuroeconomics also signals hybridization borrowing as much to experimental psychology as to experimental economics[19].

We view neuroeconomics as the latest manifestation of this realistic frame of mind. As we explained in introduction, Karl Popper's «*rationality principle*» envelops this realist ethos, common to the cognitive sciences: neuroeconomics is a piece in the whole rationalization enterprise. The unified behavioral science focuses on the *reasons* underlying our behaviors, and each piece of evidence that helps at illuminating these *reasons* shall be entitled within the new discipline.

Thus we exhumed the condition of possibility of the neuroeconomics project, namely the new realistic ethos and the technical possibility to go beyond behaviorism, to investigate the observed behaviors, through imaging neurology.

Neuroeconomics and Laboratory Economics are coterminous Actually neuroeconomics is the natural outgrowth of the behavioral experiments. More precisely, the realistic ethos is conducive to two intertwined programs, a quantitative one, and a qualitative one. The first one refers to the experimental designs (collection of the empirical data)[20], while the second refers to the interpretations of the neural data (construction of the empirical facts). Let us describe the quantitative practice first. The project to encapsulate the real behaviors through rigorous scrutiny leads to the cautious construction of behavioral facts through experimental designs in a controlled environment (the lab). Then, the desire to flesh out puzzling phenomena is conducive to the brain imaging phase. Actually brain imaging analysis is an inductive experimental phase to answer the question: what do these experimentally established behavioural facts mean, how do they hint to basic information about human nature?

At this stage some epistemological lucidity should be fully exerted: the visualization of specific neural regions which are activated during the controlled experiments does not allow per se to correlate the activation of a zone to a concomitant behaviour. Only the replication of the same co-occurrence may lead to the association between a given neural region and a behavioral pattern[21]. To put it differently, a correlation does exist if it has

scientific contribution perfectly match the questions of those economists who are sensitive to psychology (Shafir and LeBoeuf, 2002). Dan Ariely's joint production with eminent economists – in particular Drazen Prelec and George Loewenstein – is another manifest example of the actual cross-breeding between psychology and economics. See for instance Ariely et al. (2003) and Ariely et al. (forthcoming).

[19] For example the «*non deception rule*» is a «sacred principle» for experimental economics, whereas psychologists commonly use deception within their experiments. However, some experiments in neuroeconomics introduce deception in their design. See for example (Rilling et al., 2002). In one of the sessions, the subject thinks she is playing with a human, while she actually plays with the computer.

[20] This quantitative stage is itself twofold: core behavioral experiments are complemented, if necessary, by sequences of brain imaging; the data are therefore sequentially behavioral and neural.

[21] Let us illustrate this point with a specific example. Resorting to brain-imaging, neurologists designate the striatum as a «*zone of reward*» encoding satisfaction feelings. This brain area has been found to be systematically activated during pleasurable experiences. The label «zone of reward» is not a reification, but a handy practical concept used to refer to this established correlation.

been observed a sufficiently large number of times. Therefore neuroeconomics data are the product of a «solidification» process, and if not «solidified», they cannot pretend to be established results.

Once we get real neuroeconomics data, the interpretative phase intervenes. This heuristic stage is not empirical in the strict sense, it lies in the construction of the empirical fact by qualitative reasoning: what does this observed correlation *mean*? Here a new caveat emerges, because such a qualitative discourse does not rest on strong inference (Platt, 1964)[22]. More precisely, the correlation between typical patterns of behaviour and brain areas is not necessarily one to one. For example, a region on which neuroeconomists have especially focussed their attention, the insula, has been implicated in the recognition and experience of disgust, sadness and fear. As J.Morris emphasizes: «*The diverse nature of these activations suggests that insula has a generalized role that is not specific to any particular emotion of behavioural context*» (Morris, 2002). Moreover insula is highly anatomically connected to other regions in the brain such as amygdala, equally well known for their role in processing emotions. What we find, then, is a set of neurobiological data connected to a manifold of patterns of behaviour and manifest emotional reactions.

Hence the fallacious association between one given neural region and one given psychological state can be equated with what we shall call «a bijection fallacy». The latter is a real epistemological issue, because it makes the inference process far more tricky and thereby vitiates the neuroeconomics project if not seriously withstood. Barring this concern it remains true that the controlled behavioural experiments, by triggering typical manifest reactions, such as arousal of a variety of emotions, might help charting the functional map of specific areas in the brain. And reciprocally, it might be that neural data will help to appraise the emotional and cognitive underpinnings of our behaviors. This possible scientific «symbiosis»[23] between the disciplines is naturally enticing. Still we should keep in mind what we call the «bijection fallacy», and we advocate an extreme cautiousness in the interpretation of the neural data.

This point leads to the following recommendation: neuroeconomics should not serve as a way of testing preexisting theories, to the extent that the risk of bias – towards the confirmation of the tested theories – in the interpretative phase of the experiments is high, due to the weak inference property inherent to this qualitative stage. To put it in another way, using neuroeconomics to test some existing theories would lead to a pernicious «epistemological circle»: the achieved experimental result is actually derived from the theory, and thereby is tautological, at least a very weak signal in favor of the

[22] Assume that one has a (testable) theoretical proposition A in mind and makes an experimental design so as to test this proposition. The experimental design produces two possible outcomes, a and $\bar{a}$. If $\bar{a}$ is observed, then the probability that A is true is 0. This is strong inference. For example, Kahneman and Tversky chose peculiar problems to get clear results: if the subject follows an intuitive approach, she will give a wrong answer. The error is an unambiguous signal that the subject uses intuition when facing the cognitive task. But for lots of similar analytical problems, the intuitive answer and the rational one match (Lopes, 1991).

[23] Experimental economics may contribute to the elaboration of data and hypotheses in the investigation of brain functions, and behavioral economics would use the functional map of specific regions to «read» behind observed behaviors.

theory. Therefore a deductive use of neuroeconomics is the wrong track.

We indicate two possible epistemological understandings of the relation between neuroeconomics and theoretical economics, an abductive vision, and an inductive one. The first one bypasses the epistemological circle issue, and leads instead to extremely fruitful insights. still it is not inherent to the neuroeconomics project either; only a real inductive frame of mind is such, according to us.

3 Neuroeconomics as abduction

3.1 New look on old objects

Modern behavioral economics seeks to make *classical* economics models sensitive to recalcitrant empirical facts. A canonical example is Fehr and Schmidt's model of fairness and cooperation (Fehr and Schmidt, 1999), which updated the hitherto oversimplified economics models. Indeed experimental economics established in the lab the existence of both positive and negative reciprocity. While playing the ultimatum game, subjects who are facing an unfair offer have a propensity to reject it, even though from a purely rational viewpoint, they should accept the deal[24]. This behavioural pattern is puzzling for the standard economist, since the rejection of the unfair offer is – economically speaking – costly. No less puzzling is the high degree of mutual cooperation when subjects play prisoner dilemma games in the lab. Standard economists predict the opposite, namely the disastrous suboptimal outcome of a mutual defection[25]. Actually the empirical negation of the standard economics benchmark turned out to be extremely fruitful, to the extent that it was conducive to new hypotheses. Not only did they serve as core assumptions in the improved behavioural models, but they also allowed new investigations, carried out by neuroeconomists[26]. The original models are not thrown away for being empirically refutated, but «*updated*». This suggests that in the rationalization process, new hypotheses[27] are going to be construed, to accommodate the models (henceforward we shall use the term premises to design the latter) with the empirical data.

We characterize this dialectical process of serial hypothesis construction as an abductive one. We call it abductive[28] in the sense that what is properly targeted in abductive reasoning is not properly a conclusion, but the intermediary theoretical premises that would, given general scientific laws, lead to an already observed conclusion. There is an interpretative leeway that is opened by abduction, to the extent that what is primarily

[24] By rejecting the unfair offer the player gets nothing (and the proposer also gets nothing), while by accepting the player gets something (and the proposer takes the remainder). Under the commonly accepted non satiation assumption, the player should accept even the mere nothing – still better than nothing at all.

[25] When the maximization of one's own payoff is the objective of each player – and this is a common knowledge – the Nash equilibrium arising is the inefficient mutual defection.

[26] For example, once we take for granted that pro-social behavior is a common pattern in a specific context, it is the role of neuroeconomics to flesh out the mechanisms involved.

[27] To wit new *reasons*, new causes explaining what is observed.

[28] We collapse here scientific and individual cognition as joint instances of idealized processes of reasoning and we follow Niiniluoto (Niiniluoto, 1999) in our definition of abduction.

sought, given a sufficiently articulated set of theoretical major premises in a problem and a sufficiently well observed phenomenon that is supposed to derive from a particular application of the major premises to a case, is a partition of possible answers to a problem rather than a single one. So likeliness, relevance, and further confirmatory data will help determine which set of hypotheses is liable to derive the observed results: the partition becomes less and less coarse. Neuroeconomics, by exploring the neural underpinnings of controlled behaviors, largely participates to the refinement of the partition.

Notice that the use of an abductive methodology does not present itself as the rival of the hypothetical-deductive methodology of normative economics. By essence it does not impart any substantial modification over its basic epistemological structure which may well lay beyond the reach of local adaptation of the theoretical sources of predictive power of economics and freshly acquired relevant data. This spirit of local modification and adaptation of economics is what seems to have been adopted by behavioural economics. This could be considered a major reason why neuroeconomics, when understood and fostered in priority by behavioural economists, does not, at institutional and epistemological levels, as we have already hinted, lead to a deep epistemological reshaping of economics[29]. Behavioural economics, extended to the use of neuroscience techniques, data and interpretation does not entail an epistemological overturn of economics. The research program of «economics and psychology», in the hand of researchers like Camerer or others, is sensitive to all kinds of data that reveal systematic deviations from the predictions of the rational theory of choice, but once we have accommodated these deviations in the utility functions by incorporating relevant parameters, we have in no way changed the deep epistemological functioning of normative economics. We have, however, disclosed its conceptual openness[30].

Various recent experiments in neuroeconomics are instances of the abductive process we have just described. More precisely neuroeconomics intervene at the level of the interpretative leeway mentioned before, aimed at reconciliate the theoretical premises with experience. For instance the rejection of an unfair offer in the ultimatum game is puzzling insofar as the only heuristic tools which are available to us are the standard economics models. The partition of hypothesis aimed at rationalizing this empirical fact within the previous – properly updated – theoretical frame can be refined thanks to neuroeconomics experiments. To illustrate the way neuroeconomics constructs plausible hypotheses from its experimental designs, we develop three examples issued from the recent neuroeconomics literature.

[29] This adaptative motion leaves the possibility of an incredulous or unconvinced stare in the eyes either of those who wish to preserve the actual and parsimonious normative models of rational behaviour or of those who wish to encompass neurobiological data in a more comprehensive view of rationality and human behaviour.

[30] Notice that in this abductive process, the epistemological status of the theoretical models, in weberian terminology, is tantamount to an «*ideal-type*»: far from being considered as empirically valid, the model is a heuristic «tool» to appraise reality (Weber, 1949).

3.2 Illustrative Examples

Reciprocity is a manifest «Smithian» object of interest. Behavioral economics showed that reciprocity is not an imaginary notion, but has an experimental robust anchoring. As standard economics cannot rationalize properly reciprocity, reciprocity plays therefore the role of compelling empirical fact in our abductive scheme.

Understand negative reciprocity How to rationalize the rejection of the unfair offer in the ultimatum game? The partition of hypothesis offered by neuroeconomics reintroduces the hitherto neglected limbic (emotional) dimension to explain how we really behave, and makes negative reciprocity more understandable.

First Jonathan Cohen and al. construed an experimental design under *functional magnetic resonance imaging* (*fMRI*) in order to pin down the neural underpinnings of the decision to reject an unfair offer in the ultimatum game (Cohen et al., 2003). The high level of sensitivity of the right anterior insula to the degree of unfairness, in addition to the negative correlation between the right anterior insula activation and the acceptance rate, suggests that an emotional factor might be conducive to the refusal. More precisely the rejection might be the manifestation of moral disgust[31] while facing an unfair offer. The decision to reject would mean that the cognitive goal of accumulating money – mirrored by the steady activation of the prefrontal cortex – is overwhelmed by the emotional goal of resisting unfairness. Therefore standard economics, by focusing exclusively on the rational motives, neglected the limbic dimension and trivialized the decision-making process, reduced to only one dimension.

In the same vein Dominique de Quervain and al. (de Quervain et al., 2004) provided a complementary insight on negative reciprocity, through a *Positron Emission Topography* (*PET*) experiment. The objective is to understand the willingness to punish an untrustworthy person in the trust game. In this dynamic exchange game, the *first movers* start by giving their endowment to the *second movers*, who then decide either to share, or to keep all. The *first movers* are under imaging scrutiny: in an additional step, they are allowed to punish the unfair partner, and they choose the amount of the punishment during a one minute deliberation period. The *PET* measure is restricted to this period. To appraise the hypothesized satisfaction derived from the punishment, the experimental design contrasts a «symbolic punishment» condition with an «effective punishment» one[32]. The anterior striatum and thalamus specific activations suggest that there is a satisfaction derived from the punishment.

The underpinnings of social exchange and trust On the other hand, positive reciprocity has been studied by James Rilling and al. using an *fMRI* procedure during an

[31] The correspondence between insula activation and disgust (physical or moral) is consensual among the imaging neuroscientists. As previously explained, it is the result of an inductive assignment, likened to the replication of experimental conditions involving disgust.

[32] The subject knows that the (economic) punishment is purely symbolic in the first case (it won't hurt the punished partner), and in the alternative condition she knows that the punishment will have an effective impact on the partner.

iterated prisoner dilemma game (Rilling et al., 2002). The neural substrates of cooperative behavior and the specific impact of *human* exchange on brain are under scrutiny. For this purpose, they contrasted the sessions of the game played against a computer, with the sessions played against a human being. They observed that the frequency of the mutual cooperation outcome is significantly lower in the computer session. Moreover, an emotional process is associated with reciprocity, to wit the observation of symmetric outcomes – mutual defection and mutual cooperation – vis-a-vis asymmetric ones. Actually this emotional reaction seems to be more salient in the case of positive reciprocity. This echoes the results achieved recently in an experiment led by Tania Singer and al., in which the vision of faces of intentional *cooperators* activates the insula region, in addition to the amygdala, the striatum, and regions commonly associated to social cognition and its corresponding inference process. This activation is found to be more important than the one while facing faces of intentional *deviators* (Singer et al., 2004a).

Besides, a masking procedure[33] reveals that this specific activation is not observed in the computer session[34]. This result is consistent with an analogue observation made by Cohen and al. in the ultimatum game experiment previously evoked: for the same amount of unfairness the emotion response is stronger when the second partner is human than when he has been replaced by a computer[35].

What's empathy? The previous experiment suggests that human interactions are multidimensional and cannot be encapsulated without integrating the limbic component and fleshing out its interactions with cognitive mechanisms. To that extent, a last example of the recent contribution of neuroeconomics to social science is an fMRI experiment run by Tania Singer and al. to investigate the nature of empathy – Cf Singer et al. (2004b) and Singer and Fehr. When the subjects of the experiment learn that their lover is going to suffer a painful treatment, a specific activation of the anterior insula and the anterior cingulate cortex is observed. As these two areas are known to be the affective components of the «pain matrix», this observation leads to equate empathy with self-pain. This finding is consistent with existing results in the same research field, and definitely constitutes a major step to understand empathy feelings.

Therefore it seems that social exchange rests on a *social-emotional* experience. Standard economics, by equating mutual cooperation in an iterated game with a «*grim-trigger strategy*», might be relevant to understand a collusive relation between firms, but definitely not to shed light over social interactions. More generally the utilitarian *trade-off* model might be irrelevant to understand lots of our actions, not only because it involves a cognitive process that is not necessarily possible (bounded rationality), but also because emotion plays a major role in the decision-making process, especially under some circumstances. For example Jonathan Cohen, Joshua Greene and al. have shown the

[33] The idea is to look at an eventual activation of the regions which have been found to be significantly activated in the first stage.

[34] Conversely the orbitofrontal cortex appears to be an overlapping region.

[35] A specific activation of the anterior striatum is also observed during the decision-making period, when the choice to cooperate with the human partner is made. This is another cue suggesting positive feelings towards the *human* partner.

importance of emotional mechanisms in moral dilemmas involving a conflict between the cognitive (rational-utilitarian) goal and the gut reaction related to the emotionally unacceptable choice (Greene et al., 2004).

Hence neuroeconomics, by incorporating the emotional dimension – more precisely the interactions between cognitive motives and emotional ones, makes the observed phenomena much more graspable.

4 Neuroeconomics as real syncretism

4.1 Widen the scope of economics

True, it may be extremely fruitful to flesh out compelling questions arising from the gap between some theoretical premises and experience. Nevertheless the abductive path brings about both an ontological critic and an epistemological one. First, in the abductive perspective the way economists use neuroeconomics data has the drawback to rationalize theoretical models – by sharpening them up – instead of real behaviors. Second the abductive process leaves aside important questions[36] and is only a portion of an enveloping set of questions: economics should not limit to its – historically – designated «Smithian» objects[37]. We advocate rather an expansion of the scope of economics, and point to Karl's Popper «*rationality principle*» as cementing the relation between neuroeconomics and the theoretical domain. This paradigm, backed by the realistic ethos (defined as the motive to provide *reasons* for observed puzzling phenomena), declines itself through two main aspects: one is methodological, the other is institutional. The methodological manifestation of the Popperian rationality principle is the shift from deduction to induction. The relation between the a priori domain and the experimental domain is reviewed: henceforward neuroeconomics is exploratory, while economics models are no longer inquiry engines, but rather ex post heuristic devices serving to understand the empirical observations[38]. Theoretical contributions to established psychological facts still exists. For example, Ronit Bodner and Drazen Prelec (Bodner and Prelec) have fleshed out the *self-signaling motive* put forth by George Quattrone and Amos Tversky in the article mentioned before. They construct a «*diagnostic utility*» directly echoing the «*deceptive diagnosis*» observed empirically. In the same vein, Roland Benabou and Jean Tirole's models related to the self construction process, also focus on *personal rules* and *self-deception*[39]. The insights they provide constitute important heuristic devices, evidently helpful in the opening of the «*black box*» of human psychology. Might neuroeconomics findings entail similar theoretical contributions in the future.

The institutional aspect is the unification of behavioral sciences. This motion – char-

[36] Because the interpretative leeway depends necessarily on the theoretical premises.

[37] These inherited objects are basically those of social sciences, for instance social exchange, cooperation and trust...

[38] The latter are not a priori constrained, to the extent that specific experimental data may contribute to provide those local justifications that are needed in view of a full account of human behavior. Thus the expansion of the scope of economics is intrinsic to the inductive perspective.

[39] See in particular Benabou and Tirole (2002) and Benabou and Tirole (2004).

acterized by a syncretism with neurologists, psychologists, philosophers – has already started. Below are some examples to convince the reader it is the case.

4.2 Examples of new investigations

Cognitive control and the construction of feelings Rilling and al.'s experiment suggests that social exchange rests on an emotional experience. On the other hand it also puts forth the important contribution of the orbitofrontal cortex in the cooperative choice. The delayed reward of cooperation certainly involves some «forward-looking disposition», that would inhibit the selfish impulse to defect – to get the immediate reward. Actually this interpretation is consistent with studies tending to show the essential role played by the ventromedial frontal lobe in the control of emotion. The emotional stimulus is not a feeling – implying some degree of consciousness – before the intervention of the the ventromedial frontal lobe. This means that an impairment between this region and the limbic system damages learning from the emotional experience. This is salient in the famous «gambling game» (Damasio et al., 1994). In this experiment the subjects are asked to choose between four different packs of cards. Two of the packs are associated with a nice payoff of 100 dollars, but occasionally one must repay 1250 dollars, whereas the other two packs provide a comparatively modest payoff of 50 dollars, but the repayment involves only minor sums. The subjects ignore the cards composition of the packs. They play 100 trials. In practice the choice of the two first packs provides systematic «bad news», hence normal subjects basically try the risky packs during the first trials, then they prefer the other packs. They develop a hunch, learn from experience. Conversely the subjects suffering a damage in the ventromedial frontal lobe do not learn that there is a danger associated with the choice of the risky packs and they continue to choose the «bad packs» throughout the trials even though they lose a lot. This is due to the fact that they have the *emotion* following «bad news» – when making the risky choice – but not the *feeling*; hence they do not have the motivation to modify their behavior[40].

Hence cognitive control is a manifest object of interest for behavioral economists, not only to flesh out typical economics themes such as compulsive gambling and inefficient risk taking, in addition to the study of social exchange and trust, but also to understand other psychological puzzles. For instance, the distinction between feeling and emotion allows some new insights related to the control of pain. In a famous article related to «*deceptive diagnosis*», George A.Quattrone and Amos Tversky show that people select actions that are diagnostic of favorable outcomes even though those actions do not cause those outcomes (Quatronne and Tversky, 1984). This is clear in the following experiment. In a control condition subjects are asked to immerse their arm into cold water, and their tolerance to coldness is measured. Then the experimentalists divide the subjects into two random groups – let's say A and B. At an intermediary stage, group A is informed that a higher tolerance to cold water is a sign of a long, healthy life, whereas group B receives the opposite piece of information – to wit the less tolerant one is, the

[40]For a very clear presentation of this experiment, see Kandel et al. (2000), p 994.

more healthy he is likely to be. Follows the experimental condition: subjects are asked to do the tolerance to coldness test again. The striking result is that the tolerance of group A subjects (those who have been told that a higher tolerance to cold water was a sign of a healthy life) typically increases; and symmetrically, the dominant pattern for group B is a decrease in the tolerance to cold water, as compared with the initial condition (which serves as a benchmark).

The interpretation made by the authors to explain the changes in the physiological tolerance is that subjects have a «calvinist motivation» that leads them to choose behaviors signalling good outcomes, even though they are aware that these actions do not facilitate the occurrence of these future outcomes. For this comforting process to be at work, the subjects must ignore the fact that they are selecting the action for being a good signal of future destiny. They must be pure behaviorists who neglect the inner mental states originating the action. A denial is necessary, hence the expression «*deceptive* diagnosis».

Now we stress the following point: while a self-deceptive «cheating» might well explain group B's decreasing tolerance to cold water[41] group A's increasing resistance to coldness is a real puzzle, since during the benchmark condition, subjects did their best[42]. Therefore, for those who resisted more the second time, some objective change must have occurred. Neuroeconomics should definitely explore this phenomenon. The distinction emotion (automatic)/feeling (involving neocortex) might help to pin down the mechanisms that are involved.

Quattrone and Tversky's paper had a big impact on behavioral economics. Hopefully it shall have a similar influence on neuroeconomics.

Rationality Neuroeconomics seen, partly, as a continuation of the *economics and psychology* programme and as a specific experimental development of behavioural economics, bears a special interest in the investigation of the neural bases of cognitive biases and limitations. One can try to look at the neural underpinnings of the cognitive defects while cognitive tasks are being performed in order to flesh out the hypothesis that the automatic (intuitive) response associated to directly accessible mental representations (the intuitive contents of mind) leads to the error, when the mind does not detect and correct the automatic response (Cohen et al., 2004). This hypothesis, defended by Daniel Kahneman who sees many of the psychological biases he has attracted the attention to as falling under this characterisation (Kahneman, 2003), is also referred to by some as the «*dual process theory*» (Stanovich and West, 2000) or again the System 1/ System2 view of the mind. An alternative possible explanation for the error is a deficiency of our cognitive systems, whose limited capability restricts the ability to make right calculation, for example. In that prospect a fMRI experiment using Shane Frederick's *Bat and Ball problem* is currently being run by the authors. Its main goal is to to explore the nature of the cognitive defect in that particular case. Subjects are asked to answer the following question:

[41] Unconsciously, one convinces oneself that coldness cannot be suffered any longer, while physiologically one could resist more.

[42] They had no reason to falsify their tolerance to cold water.

A bat and a ball cost together 1 $ 10, the bat costs 1 $ more than the ball. How much does the bat cost?

About 70% of subjects give the wrong intuitive answer 1$ even when the correct answer, 1 $ 05 is made salient and available in different ways. It is predictable, as other studies of neural underpinnings of cases of contrasts between System 1/ System 2 types of answers have shown, that we observe, when the intuitive answer 1 $ is given, significant activity in the parietal or temporal lobe (Goel (2003) and Goel et al. (2000)). In the case of the non-intuitive, effortful, condition, the right prefrontal cortex should be significantly activated.

Experiments on such cognitive biases, failures or limitations, make sense in a widened view of what neuroeconomics can ascribe to itself as a conceptual and epistemological goal. In other words one important aspect of the endeavour of bringing forward a new hybridization of sciences is to acquire a better sense of the natural basis of our various normative definitions of rationality. It is important to know whether a subject is in principle capable to accomplish a cognitive task or whether she is neuronally wired in such a fashion that the chances that the task be standardly fulfilled are quite low. In the case of the *Bat-and-Ball problem* the apparent paradox is that the standard accomplishment of the task is not so cognitively costly, as any numerate subject, in particular when put in front of the solution, is capable to recognise it. Yet most subjects fail to solve the problem. The presence of two separate available cognitive approaches to the problem at hand, linked to two distinct neural circuits triggered according to the different elicited answers, points in fact to a certain level of indeterminacy in the application of the normative notion of rationality to actual cognitive performances. It is of epistemological importance to determine whether we face, in the case of a systematic biased response corresponding to the use of some available cognitive and neural resources, a serious cognitive limitation or rather the use of a heuristics that leads to wrong or sub-optimal results but that can be bypassed when the good conditions are met or the right signals sent.

One aspect of the debate, as one can foresee, is the nature of bounded rationality. In a certain vision of Simon's Bounded Rationality Paradigm heuristics are fast and frugal procedures that lead generally to correct outputs and are related to neural mechanisms and processes. In particular emotions can be viewed as a type of neural mechanisms that short-circuit the more cognitively costly information processing that would lead to 18 optimal decision making (Muramatsu and Hanoch, 2005). The presence of serious cognitive deficiencies would not flatly point to the coexistence of two types of mind or mental systems, but it would raise the question as to know whether we continue to call rational, even «boundedly rational», any contribution of the mind that, in spite of its intrinsic cognitive low-cost, can lead to far from correct and satisfying outputs. One goal of neuroeconomics, through the systematic study of psychological biases that has long interested behavioural economists, is then to help draw the boundary between what can reasonably be called rational behaviour or cognition – and be, as such, introduced in standard economic models – and what cannot.

References

Dan Ariely, George Loewenstein, and Drazen Prelec. Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics*, 118 (1):pp. 73–105, February 2003.

Dan Ariely, George Loewenstein, and Drazen Prelec. Tom sawyer and the construction of value. *Journal of Economic Behavior and Organization*, forthcoming.

W.B. Arthur. *Increasing Returns and Path Dependence in the Economy*. The university of michigan press edition, 1994.

W.B. Arthur, Y.M. Ermoliev, and Y.M. Kaniovsky. Path dependent processes and the emergence of macro-structure. *European Journal of Operation Research*, 1:pp. 294–303, 1987.

Roland Benabou and Jean Tirole. Self-confidence and personal motivation. *The Quaterly Journal of Economics*, August 2002.

Roland Benabou and Jean Tirole. Willpower and personal rules. *Journal of Political Economy*, 112(4), 2004.

Ronit Bodner and Drazen Prelec. Self-signaling and diagnostic utility in everyday decision-making. October 2001.

Pierre Bourdieu. *Questions de Sociologie*. 1978.

Pierre Bourdieu. *In Other Words: Essays Towards a Reflexive Sociology*. Stanford University Press, 1990.

Pierre Bourdieu. *Outline of a Theory of Practice,*. Cambridge University Press, 1977.

Colin Camerer, George Loewenstein, and Drazen Prelec. Neuroeconomics. July 2003.

Jonathan D. Cohen, Alan G. Sanfey, James K. Rilling, Jessica A. Aronson, and Leigh E. Nystrom. The neural basis of economic decision-making in the ultimatum game. *Science*, 300:pp. 1755–1758, June 2003.

Jonathan D. Cohen, Nick Yeung, and Matthew M. Botvinick. The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, 111(4):pp. 931–959, 2004.

Antonio R. Damasio, Antoine Bechara, Hanna Damasio, and Steven W. Anderson. Insensitivity to future consequences following damages in human prefrontal cortex. *Cognition*, 50:pp. 7–15, 1994.

Dominique J-F de Quervain, Ernst Fehr, Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, and Alfred Buck. The neural basis of altruistic punishment. *Science*, 305:pp. 1254–1258, August 2004.

Ernst Fehr and Karl Schmidt. A theory of fairness, competition and cooperation. *The Quaterly Journal of Economics*, 114:pp. 817–868, 1999.

Vinod Goel. Evidence for dual neural pathways for syllogistic reasoning. *Psychologica*, 32:pp. 301–309, 2003.

Vinod Goel, V.Buchel, C.Frith, and R.Dolan. Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage*, 12(5):pp. 504–514, 2000.

Joshua D. Greene, Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44:pp. 389–400, October 2004.

Ian Hacking. *Representing and intervening*. Cambridge university press edition, 1983.

Daniel Kahneman. A perspective on judgment and choice – mapping bounded rationality. *American Psychologist*, 58(9):pp. 697–720, September 2003.

Daniel Kahneman and Amos Tversky. Subjective probability: a judgment of representativeness. *Cognitive Psychology*, 3:pp. 430–454, 1972.

Daniel Kahneman and Amos Tversky. On the reality of cognitive illusions. *Psychological Review*, 103:pp. 582–591, 1996.

Eric R. Kandel, James H. Schwartz, and Thomas M. Jessel. *Principles of Neural Science*. McGraw-Hill Medical, 4 edition, January 2000.

David Laibson and Richard Zeckhauser. Amos tversky and the ascent of behavioral economics. *Journal of Risk and Uncertainty*, 16, 1998.

Lola L. Lopes. The rhetoric of irrationality. *Theory and psychology*, 1(1):pp. 65–82, 1991.

Fritz Machlup. An interview with fritz machlup. *The Austrian Economics Newsletter*, 3(1), Summer 1980.

J. Morris. How do you feel. *Trends in Cognitive Sciences*, 6:pp. 317–320, August 2002.

Roberta Muramatsu and Yaniv Hanoch. Emotions as a mechanism for boundedly rational agents: the fast and frugal way. *Journal of Economic Psychology*, 26(2):pp. 201–221, 2005.

I. Niiniluoto. Defending abduction. *Philosophy of Science*, 66(supp.):S436–S451, 1999.

Alfred C. Pigou. *Memorials of Alfred Marshall*. New York: Augustus M. Kelley, 1925.

John R. Platt. Strong inference. *Science, New Series*, 146(3642), 1964.

Karl Popper. *The rationality principle*. New Brunswick: Popper Selections, Princeton, NJ, princeton university press edition, 1967.

George A. Quatronne and Amos Tversky. Causal versus diagnostic contengencies: on self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, 46(2):pp. 237–248, 1984.

James K. Rilling, David A. Gutman, Thorsten R. Zeh, Giuseppe Pagnoni, Gregory S. Berns, and Clinton D. Kilts. A neural basis for social cooperation. *Neuron*, 35:pp. 395–405, July 2002.

Aldo Rustichini. Neuroeconomics: Present and future. *Games and Economic Behavior, Editorial*, 52(2):pp. 201–212, August 2005.

P.A. Samuelson. *The Foundations of Economic Analysis*. Harvard University Press, 1947.

Eldar Shafir and Robyn LeBoeuf. Rationality. *Annual Review of Psychology*, 53:pp. 491–517, 2002.

Tania Singer and Ernst Fehr. The neuroeconomics of mind reading and empathy.

Tania Singer, Stefan J. Kiebel, Joel S. Winston, Raymond J. Dolan, and Chris D. Frith. Brain responses to the acquired moral status of faces. *Neuron*, 41:pp. 653–662, February 2004a.

Tania Singer, Ben Seymour, John O'Doherty, Holger Kaube, Raymond J. Dolan, and Chris D. Frith. Empathy for pain involves the affective but not sensory components of pain. *Science*, 303:pp. 1157–1161, February 2004b.

Keith E. Stanovich and Richard F. West. Individual differences in reasoning: implications for the rationality debate? *Behavioral and Brain Siences*, 23:pp. 645–726, 2000.

Richard H. Thaler. Irving fisher: Modern behavioral economist. *The American Economic Review*, 87(2):pp. 439–441, May 1997.

Max Weber. *The methodology of the social sciences*. New York, 1949.