



Mind/Brain and Economic Behaviour: For a Naturalised Economics

Mario Graziano¹

Received: 3 August 2018 / Accepted: 12 September 2018 / Published online: 19 September 2018
© Springer Nature B.V. 2018

Abstract

Neuroeconomics is a science pledged to tracing the neurobiological correlates involved in decision-making, especially in the case of economic decisions. Despite representing a recent research field that is still identifying its research objects, tools and methods, its epistemological scope and scientific relevance have already been openly questioned by several authors. Among these critics, the most influential names in the debate have been those of Faruk Gul and Wolfgang Pesendorfer, who claim that the data on neural activity cannot find place in economic models, which should on the contrary be solely based on the data produced by choices. This paper aims at countering the gloomy and unsubstantiated claims of these two authors and those who believe that neuroscience cannot provide new and useful insights to the established knowledge of standard economics. The main point stressed here is that this perception is the product of a general misunderstanding of the advances made by neuroscience, which are incidentally of crucial importance.

Keywords Neuroeconomics · Therapeutic approach · Naturalism · Empathy

1 Introduction

Neuroeconomics, i.e. the science aimed at identifying the neural foundations on which decision-making is based through an integrated study of economics and neuroscience, has been the object of a growing interest by those economic researchers fascinated by the idea of going beyond the logical and formal models provided by standard economics.

Despite being initially perceived as a simple spin-off from behavioural economics—behavioural economics under a scanner (Ross 2008)—neuroeconomics has gradually freed itself from the aegis of behavioural economics and has become a full-fledged discipline towards the end of the first decade of the new millennium.

✉ Mario Graziano
mgraziano@unime.it

¹ Department of Cognitive Sciences, University of Messina, Messina, Italy

The main purpose of neuroeconomics is opening the “black box” represented by the human brain in order to expose how decisions are ultimately the product of the processing taking place in specific areas of the brain. To pursue this goal, neuroeconomists use advanced brain-imaging equipment to identify the cerebral mechanism involved in tackling specific decision-making tasks.

By doing so, neuroeconomists have introduced a series of “causal” factors that had always been overseen or simply ignored by standard economists in the analysis of decision-making. Therefore, while the main task of standard economics is to provide the tools to anticipate a certain behaviour starting from the acceptance of some rationality criteria at the basis of choices, the main task of neuroeconomics is to account for a certain choice behaviour starting from its underlying cerebral mechanisms.

In order to better understand this new approach, it is useful to observe how it tackles the main concept of standard economics, i.e. the utility function. From a neoclassical perspective, the utility function is a formal concept that defines a “preferable” relation between baskets of goods.¹ From a neuroeconomic perspective, conversely, the choice of one basket of goods is the result of a neural processing based on a mechanism composed of elements, activities, or cerebral structures. Neuroeconomics hence sees the utility of a choice as something that is not determined by formal preference relationships, but rather as the result of a complex cerebral machinery that entails specialised mechanisms in the triggering of pleasure, motivation, learning, attention, cognitive control, etc. According to neuroeconomists, the discoveries related to the functioning of the brain can allow for a reality-based and cause-oriented understanding of the basic intuitions related to the concept of utility function.²

The neuroeconomic approach does not limit its scope to the analysis of the brain of single decision-taking subjects, but it has the merit of opening up to the Theory of Games, focussing on the interactions between different agents. From this point of view, it is useful to mention the work of James Rilling and colleagues on the

¹ In neoclassical economics, stating that basket A is more useful than basket B means stating that the relationship existing between the two baskets will push a rational agent (whose utility function follows some formal constraints) to chose basket A rather than basket B.

² The experiment carried out by Platt and Glimcher (1999) on monkeys had precisely this target. The experiment was based on the idea of teaching monkeys to pick one point between two light points showed on displays placed on the right or on the left of the monkeys. The monkeys picked one point by moving their heads and when they picked the right point, they received a food reward. To maximize utility, monkeys had to remember the previous odds associated to their choices and the value of the reward. What was surprising was that the conclusions reached by Platt and Glimcher in their experiments were quite similar to those of later studies performed on humans with fMRI (cfr. Paulus et al. 2001; Knutson and Peterson 2005).

Prisoner's Dilemma (Rilling et al. 2002)³ and Alan Sanfey and colleagues on the Ultimatum Game (Sanfey et al. 2003).⁴

Neurosciences are undeniably seductive. However, several economists have raised doubts about the relevance of neuroeconomics and, broadly speaking, of neuroscientific data in economic inquiries. The most radical criticism has come from Faruk Gul and Wolfgang Pesendorfer, who have claimed that the data on neural activities cannot find place in economic models, which on the contrary should be exclusively based on the utility function of an agent in relationship to the concept of choice. From this perspective, stating that an option has a higher utility than the other, implies identifying it as the choice that a rational agent would take (Gul and Pesendorfer 2008). Economists do not necessarily need to know that choosing x over y "is motivated by the pursuit of happiness, a sense of duty, an obligation or a religious impulse" (Gul and Pesendorfer 2008, p. 24).

This defence of standard approach does not imply that economics is completely outside the inquiry scope of psychology. Nevertheless, it should only take into consideration relevant data, i.e. the ones related to choices. The data provided by psychology and experimental economics can be used to evaluate a model and anticipate choices or future equilibrium. On the contrary, the data or the variables that are not related to choices, such as the data on neural activities, cannot find place in theoretical models. Paraphrasing the words of Hilary Putnam, in Gul and Pesendorfer's theory it does not matter whether the brain is made of grey matter or Swiss cheese.

The economic perspective defended by Gul and Pesendorfer has been recently criticised by Dietrich and List (2016), who had the merit of highlighting to what extent such an attitude is wrong, both from an economic perspective and, most importantly, from a philosophical point of view. Nevertheless, Dietrich and List agree with Gul and Pesendorfer on the claim that the data provided by neuroscientific studies are useless in economic inquiries.

On the contrary, this paper will emphasize the assets and benefits that might be provided by the studies where neuroscience (as well as other disciplines such as neuropsychology, philosophy and obviously economics) is proactively involved in creating an inter-theoretical model that takes into account human mental processes. In

³ In this experiment, the Prisoner's Dilemma is used to study the cerebral activity of 19 participants (11 girls and 8 boys), both when they played against humans and when they played against a computer. The final results showed that participants tended to cooperate more (81% of times) when they played against humans. The researchers later tried to identify the activated cerebral regions during cooperation and defection, in order to analyse their dopaminergic impact, by recording the BOLD activation (Blood Oxygen Level Dependent Signal) of a part of the striatum and the ventromedial prefrontal cortex. It was then noticed that this activation was not significant when player A played against a computer: the nature of interactions hence plays an important role in the case of BOLD activation.

⁴ Sanfey used the fMRI to study the cerebral activation of 19 subjects playing the Ultimatum Game, focusing in particular to the cerebral activities of the participants that received lower offers (20% of the sum at stake). Starting from the assumption that this offer might cause a conflict between the emotional desire of rejecting the offer and the will to gain as much money as possible, Sanfey and colleagues identified the regions of the brain that could potentially process these tasks. The authors noticed in particular a higher level of activation in three cerebral areas: the anterior cingulate cortex, the anterior bilateral region (right and left insula), and the dorsolateral prefrontal cortex.

order to do so, the paper will be structured as follows. I will start by providing a brief overview of the ideas put forward by Gul and Pesendorfer, highlighting—in line with the claims of Dietrich and List—the limits and problems that such a theory might face when trying to explain some economic constructs. Nevertheless, I will distance myself from Dietrich and List as regards the assessment of neuroeconomics, since it is my opinion that neuroeconomics can account for the possible combination of structural and functional aspects, overcoming an understanding of cognitive processes that completely ignores the physical world and the influence that anatomic structures play in determining mental functions. Clinical studies have undoubtedly opened a field of research that has become a driver to look into the couple composed by brain and mind in a completely new way (and, generally speaking, their function as a structure), leaving behind the worst expression of the localisation practice, thus acknowledging the several ways in which an organism “fails” and the even more numerous ways in which it reacts by adapting and making up for a failure, recovering or replacing its functions. As a matter of fact, the input provided by the study of different pathologies has enabled the researchers to move away from an explanation scheme based on radical monist positions that required a simple association of neural sites and observable behaviours, pushing them towards integrated causal explanation schemes aimed at describing functions through the structures that allow for the mapping of the relationships between cerebral facts and mental facts, leading to their naturalisation. Nevertheless, contrary to an extreme conception of naturalisation—which claims that in order to understand the nature of the mind it is sufficient to understand the nature of the brain—the naturalism put forward in this paper refers to a liberalised (or pluralist) form of naturalism that is not eager to reunite the sciences that compose it in a unique brain science. Despite representing an ambitious challenge, this naturalisation scheme can provide an opportunity to explain a series of phenomena—including some that are dear to the hearts of economists—that up to now were considered unmanageable by the scientific method, hence placing them in a more realistic framework. From this point of view, it is therefore necessary that philosophy and economics entertain a strong relation with sciences, including neuroscience. In order to emphasize the advantages of this approach, I will conclude by providing the example of empathy as a promising case study in the field of pluralist economic naturalism.

2 Mindless and Brainless Economics

In *The Case for Mindless Economics*, Gul and Pesendorfer strenuously support the idea of “brainless economics”, providing arguments that support a “standard” economic approach. In particular, Gul and Pesendorfer contend that the science of economics does not normally focus on the same issues tackled by neuroscience and psychology, but it rather pursues different goals and must make use of abstractions to reach its purposes. The authors differentiate between the concepts of true utility and choice utility. According to the authors, economics should solely focus on the latter, defining the agent’s utility function only in relation to the concept of choice. From this perspective, saying that an option is more useful than the other simply means

that a rational agent would tend to pick it. Gul and Pesendorfer question instead the attitude of several neuroeconomists to be inclined to go back to the idea of looking for a “true utility” related to physical pleasure or other motivational factors.

The authors strongly defend the autonomy of economics in relation to other disciplines. They follow at any cost a nomological-deductive approach, hence regarding all non-economic factors as irrelevant for the purposes of economics.

In economics, researchers usually follow two approaches, both stemming from Pareto. The first one assumes that choice criteria are determined according to a given preference system and, hence, subjects are associated to a preference system in line with their actions and it is assumed that the choice criteria used by subjects are inferred from this system. In other words, in a set of possible actions, the set of possible choices is a subset of preferred actions. The second approach takes the choice as its starting point, on the condition that it is coherent with the preference system that makes it rational. This is the approach followed by Gul and Pesendorfer.

It is well known that, in this approach, an important role among coherence conditions is played by the principle of “revealed preference” (Samuelson 1938), which postulates that only the observable behaviour of agents provides reliable data. From this perspective, the actual preferences of subjects can be inferred from the behaviour of agents, since they act in order to maximise their utilities by selecting the most rational behaviour among those possible according to a decision-taking mechanism that can be logically deduced. Preferences are hence revealed by agents through their behaviour, giving them the same importance as those that cannot be observed.

In a recent article entitled “Mentalism versus behaviourism in economics: a philosophy-of-science perspective”, the priority given to the approach described above led the authors Franz Dietrich and Christian List to include Gul and Pesendorfer in the vast field of researchers belonging to behaviourism, where preferences, beliefs, and other mental statuses that play a role in social theories are simple constructions that provide a new description of the behavioural inclinations of subjects. In the words of Dietrich and List:

Behaviourism used to be the dominant view across the behavioural sciences, including not only economics, where it was pioneered by scholars such as Vilfredo Pareto (1848–1923), Paul Samuelson (1915–2009), and Milton Friedman (1912–2006), but also psychology and linguistics, where it was prominently expressed, e.g., by Ivan Pavlov (1849–1936), Leonard Bloomfield (1887–1949), and B. F. Skinner (1904–1990)... In psychology and linguistics, especially since Noam Chomsky’s influential critique (1959) of Skinner, behaviourism has long been replaced by some versions of mentalism (e.g., Katz 1964; Fodor 1975), though often under different names, such as ‘cognitivism’ or ‘rationalism’ [...]. In economics, by contrast, behaviourism continues to be very influential and, in some parts of the discipline, even the dominant orthodoxy. The ‘revealed preference’ paradigm, in many of its forms, is behaviouristic, though there are more and less radical versions of it. Recently, Faruk Gul and Wolfgang Pesendorfer (2008) have offered a passionate defence of what they call a ‘mindless economics’, a particularly radical form of behaviourism.

The declared target of their paper is therefore to reject the form of behaviourism heralded by Gul and Pesendorfer, defending instead a mentalist approach in line with the best scientific praxis.

To prove their point, Dietrich and List use the famous apple-choice problem postulated by Sen (1977),⁵ claiming that the situation requires to go beyond the simple evidence provided by choice behaviour, because otherwise an underdetermination issue would arise. The main objective of the authors is indeed emphasizing that—contrary to what Gul and Pesendorfer contended—it is necessary to go beyond a narrow interpretation of behaviourism and look deeper into the different agents' inner cognitive mechanisms that could lead to an enhanced ability to predict the different relationships that they could forge as well as anticipating those that they could establish under different circumstances that haven't yet been directly observed.

Despite the serious criticism made by Dietrich and List against Gul and Pesendorfer as regards their overestimation of choice behaviours, and their support for a mentalist approach over a behavioural one, the authors surprisingly seem to agree on the considerations about the potential intrusion represented by neuroscience in economic explanations. Dietrich and List are also very sceptical, as they believe that the explanation of economic phenomena should not rest on the analysis of the brain's neurophysiological processes. This fear has been surely fuelled by the reductionist mindset that is so popular among neuroeconomists (and broadly speaking among neuroscientists, see Churchland 1998). This approach aims at reducing the explanation of “high level” mechanisms to the explanations of “low level” physiological mechanisms (for example, the biochemical mechanisms of cell functioning, the cerebral areas devoted to the processing of a specific task or skill).

A reply to Gul and Pesendorfer's critical remarks against the reductionism heralded by some neuroeconomists was elaborated by Ross (2008), who emphasized that neuroeconomics is not limited to the approach criticised by Gul and Pesendorfer. Ross claimed indeed that there are at least two different approaches to neuroeconomics. The first one—and most renowned—is the one criticised by Gul and Pesendorfer, which relates to the “behavioural economics in the scanner” and has the purpose of identifying the neural mechanisms that lead to decision-making. The second and less known one is “neurocellular economics”. This approach aims at

⁵ According to the theories of revealed preference, the preference relation is inferred from the choices on the alternatives, according to the definition: $[x = S(\{x, y\}) \leftrightarrow x > y]$. On the basis of this behavioural assumption, the theory does not deny that agents may have diverging preferences, but it merely highlights that the “reason” of the preference is not important to the ends of the analysis. To emphasize how this does not account for the commitment action, Sen makes use of a story starring two agents, two objects [a big apple (m1) and a small apple (m2)], an action S (grasp one of the objects), and a preference relationship between the objects (>). The first agent asks the second agent to take the apple that he prefers and he actually picks the bigger apple. According to the theory outlined above therefore, $m1 > sm2$. While picking the smaller apple, the first agent complains for the shameless behaviour of the second agent, who replies by asking the first agent what he would have picked if he were in his shoes. The first agent answers that, out of courtesy, he would have picked the smaller one. “Why are you complaining then?” asks the second agent, “It is the one you have got”. The story shows that the dissatisfaction displayed by the first agent and the choice of the object that he does not want cannot be interpreted within the theory without incurring in contradiction.

applying economic methods to neural networks. In this case, Don Ross stresses that neoclassical economic constructs seem to adapt more easily to neural networks than they do in the case of real subjects. As a matter of fact, neural networks are less prone to violate the axioms of rational choice.⁶ Undoubtedly, the application of neoclassical economic theories to neural wires is a promising field of study and as such questions the idea that economics and neuroscience should be kept apart because they investigate different objects, with different methods and goals⁷ (Bourgeois-Gironde and Schoonover 2008).

The approach put forward by Don Ross, though, solely emphasizes the conceptual enrichment opportunities provided by neuroscience through the use of economic constructs, and not vice versa. Therefore, the critical remarks moved by Gul and Pesendorfer about the enrichment possibilities of economics through neuroscientific constructs remain unaddressed and unanswered.

A possible solution to the issues touched by Gul and Pesendorfer is provided by Colin Camerer, who contends that neuroscience has the merit of activating a “reformist” vision of economic theory in the economic constructs, since it supplies psychological data and variables that would otherwise be completely missing from formal economic theories. Therefore, according to Camerer, the reformist approach offered by neuroeconomics allows, for example, the widening of the theory of rational choice, so as to formally embrace the limits of rationality, free will, and personal interest (Camerer 2008). Furthermore, neuroeconomics would enable economic theories to “improve the capacity to understand and predict choice, while maintaining a mathematical discipline and use of behavioural (choice) data” (Camerer 2008).

It is therefore clear that Colin Camerer is strongly optimistic as regards the input of neuroscience on economic theories and the potential outcomes that could be obtained. Nevertheless, even if it is true that experimental data and the study of psychological variables may improve economic constructs, particularly as regards their ability to predict the real agents’ behaviour, going beyond their directly observable behavioural choices, a second question remains unanswered and even more elusive: if a phenomenon has already been studied at a psychological and behavioural level, why should we bother understanding also its neural correlates?

Glenn Harrison touches upon this very issue, claiming that it would be illogical for economic inquiries to make use of knowledge obtained through the study

⁶ For example, this research paradigm assumes that a specific cortical area (or dopaminergic wire) works as a market and that its behaviour might be modelled through the use of the general equilibrium theory. In this case, neurons are considered agents whose preferences are revealed by their activation (Ross 2008).

⁷ The possible existence of this convergence is indeed suggested by a paper published by Isabelle Brocas and Juan D. Carrillo, who postulate a hierarchical structure of the brain because of the existence of three types of conflicts (first, a conflict between the information available in different areas of the brain, which we refer to as an asymmetric information conflict; second, a conflict between the importance attached to temporally close versus temporally distant events, which we refer to as a temporal horizon conflict; and third, a conflict between the relative weight in utility attached to tempting versus non-tempting goods, which we refer to as an incentive salience conflict). On this basis, they put forward a model for discounting and an explanation for several behavioral anomalies (Brocas and Carrillo 2008).

of neural correlates. The author contends indeed that some psychological variables, such as aversion to risk or ambiguity, are so well known at a psychological level that several experimental methods allowing for the estimation of the right value that they have for economic agents have already been developed (Harrison 2008). Therefore, neuroeconomics could not provide anything new to the established knowledge existing at a psychological level.

It is true that, theoretically speaking, neuroeconomics might provide useful inputs to psychological research in order to identify the components of the mechanisms involved in the production of these variables.⁸ Nevertheless, the question remains unanswered: what advantages do these discoveries provide to economists? Ultimately, this is the question that must be answered.

In order to better explain what economists can gain from cerebral data, Camerer (2008) puts forward the idea that the study of the brain may help in empirically distinguishing between competing theories that have highly similar results in psychological experiments.

Camerer's expectations regarding the possibilities opened by neuroeconomics remain quite modest and limited to whether psychological theories are moving in the right direction, checking for example if different variables involve different components or processes.

Neuroeconomics might yet aspire to a bigger role and stop trailing behind psychology and behavioural economics, if it seriously started considering the importance of some options that, up to now, have largely been overlooked. An example of these options is the importance of the much decried "therapeutic role".

3 The Therapeutic Approach

As highlighted in the previous paragraphs, Gul and Pesendorfer believe that neuroeconomics studies should be disregarded because, basically, they reverse the object of economic research. The aim of this research program is not the prediction of the agents' behaviour, but rather to take the real economic agents' behaviour as a starting point to expose cognitive bias or describe more precisely their utility function (taking into account, for example, the dynamic aspect of preferences or the existence of social preferences).⁹ Due to this reversed perspective, according to Gul and Pesendorfer, the economic analysis turns into a "therapeutic" approach to individual behaviour (Gul and Pesendorfer 2005). On the basis of this approach, what is

⁸ Hsu et al. (2005) have for example established that there is a positive correlation between an ambiguous choice and the activation of the cerebral wire that connects the amygdala to the orbitofrontal cortex. These studies show therefore that two phenomena identified as different by behavioural economists (risk aversion and ambiguity aversion) are the products of different mechanisms.

⁹ In fact, as Gul and Pesendorfer (2005) emphasized, this inversion is similar to the one implemented by the preference analysis performed by Samuelson and used to interpret consumers' behaviour (Samuelson 1938). His analysis did not aim at anticipating the agents' behaviour by taking a hypothetical utility function as a starting point, but it rather tried to investigate the agents' utility function by observing their actual choices.

important is not to assess how individual choices interact in the framework of an institutional group once the individuals' objectives have been established, but rather to improve the individuals' objectives by differentiating between what is normal or pathologic. The cooperation between neuroscience and economics will eventually lead to the establishment of a "pathologic paradigm" that is quite far from the goals of economics, since in this framework the agents' choices cannot be used to analyse their well-being, because of the agents' inability to correctly maximise their utility function (Vallois 2012).

According to Hands (2009), the concept of pathology helps neuroeconomists to recover the normative role of the economic theory that is designed to develop a "behavioural engineering" of rationality. Neuroeconomics studies do indeed aim at correcting the mistakes that agents make in their economic choices, thus presenting itself as a kind of "economic psychiatric treatment". Furthermore, by doing so, neuroeconomists would betray the core values of the original research agenda of experimental economics launched by Kahneman and Tversky, who believed that normative and descriptive realms should always be strictly demarcated (see Kahneman 2003).

Conversely, through the concept of pathology, neuroeconomics questions the clear separation between what is normative and what is descriptive: in fact, while behavioural economics and psychology pursue the goal of describing human behaviour and show that the rationality norm defined by neoclassical economic theory is violated by real agents, neuroeconomics labels some behaviours as irrational and emphasizes their pathologic nature, hence formulating recommendations and prescriptions that should be followed when regulating or designing public policies aimed at improving the agents' well-being (see Bernheim and Rangel 2009). From this point of view, neuroeconomics would not strongly differ from the "libertarian paternalism" formulated, among others, by Thaler and Sunstein (2008):¹⁰ both approaches consider that agents are not rational actors, which leads to their making often irrational and illogical choices. The agents should therefore be helped in choosing wisely by pushing them towards the correct choice.

Nevertheless, despite the fact that some neuroscientists and philosophers have indeed theorised more or less explicitly that the discoveries in neuroscience might and should drive social choices and that they are about to provide the ultimate and comprehensive answers to all the most influential issues (see Churchland 2002, 2008; Gazzaniga 2008; Iacoboni 2008), it is also true that—historically speaking—this has been all but the goal of the therapeutic approach. This approach (and broadly speaking all clinical studies) has been praised because it has always offered, in line with experimental studies, a new way of looking at the relationship between brain and mind and, more generally, between structure and function. Clinical studies have indeed shown over the centuries their higher heuristic ability in the area of

¹⁰ Thaler and Sunstein formulated a form of *soft* paternalism called "libertarian paternalism" that, according to the authors, can be called upon when a general agreement on several important questions is lacking—climate issues, terrorism, earthquakes, etc., in other words on all sensitive questions in the eyes of the public opinion, freeing therefore the public debate from the mere personal remarks and beliefs of the citizens (often wrong), while giving more importance to the objective data collected through technical considerations.

research of behaviour and mental functions. As emphasized by Kandel (2005), the development of this field of research took place through two approaches in the study of the nervous system and mental functions. In his words:

Historically, neural scientists have taken one of two approaches to these complex problems: reductionist or holistic. Reductionist, or bottom–up, approaches attempt to analyse the nervous system in terms of its elementary components, by examining one molecule, one cell, or one circuit at a time. (...) Holistic, or top–down approaches, focus on mental functions in alert behaving human beings and in intact experimentally accessible animals and attempt to relate these behaviours to the higher-order features of large systems of neurons. (...) The holistic approach had its first success in the middle of the nineteenth century with the analysis of the behavioural consequences following selective lesions of the brain. Using this approach, clinical neurologists, led by the pioneering efforts of Paul Pierre Broca, discovered that different regions of the cerebral cortex of the human brain are not functionally equivalent (...) In the largest sense, these studies revealed that all mental processes, no matter how complex, derive from the brain and that the key to understanding any given mental process resides in understanding how coordinated signalling in interconnected brain regions gives rise to behaviour. Thus, one consequence of this top-down analysis has been initial demystification of aspects of mental function: of language, perception, action, learning and memory (Kandel 2005, p. 204).

In the ‘50s and ‘60s the clinical dimension became the main driver for the development of hypotheses and models on the relationship between mind-brain and brain-behaviour, breaking the stalemate suffered by localisation studies and laying the foundations for the connection between cerebral functioning and psychic activity. The clinical practice replaced the previous static and abstract conception of a rigid spatial localisation of cognitive functions with a new dynamic and temporal understanding of localisation, a concept that succeeded in filling a gap that had seemed unbridgeable during the first decades of the 20th century, the one between a strongly localising and associative clinical anatomic approach and a more psychological one, focused on the contrary on the integrated and consolidated character of psychic life.

The works of Alexander Lurija are undoubtedly an example of this altered perspective. Lurija reviews some concepts that, as explained above, were at the core of the traditional conception of the cerebral foundations of psychic processes. Lurija (1973) is the initiator of the concept of “functional system”, which postulates that the brain works dynamically, developing processes that do not link cognitive functions to strictly locatable cerebral areas but, on the contrary, to complex functional systems spread over different areas. In this way, the roots of higher psychic functions are not to be found in the activity of a single structure or specific area, but rather in complex functional systems (which incidentally are not genetically determined) that are shaped, according to the author, by social and cultural factors.

In addition to the concept of function, Lurija redefines even the concept of localisation by coining the term “dynamic localisation” of higher mental functions. In his words:

That is why mental functions, as complex functional systems, cannot be localized in narrow zones of the cortex or in isolated cell groups, but must be organized in systems of concertedly working zones, each of which performs its role in complex functional system, and which may be located in completely different and often far distant areas of the brain (Lurija 1973, p. 31).

Thanks to Lurija's work, neuropsychology acquires a new dimension over the second half of the 20th century, becoming the pathfinder that could lift it out of the stagnant swamp where this cognitive science found itself in the middle of the century, when investigating the mind and the general behavioural mechanisms without taking into account the brain, the body and the environment was considered indisputable.¹¹

As a matter of fact, today's state-of-the-art of neuroscience is characterised by the current consensus on the existence of multiple maps and on their dynamic and plastic nature, as proven by the studies carried out with powerful and advanced neuroimaging technology. These research activities clearly showed that all hypotheses or theories that tend to favour the mainstreaming of structural and functional considerations should be taken as main fact in the normal praxis of neuroscience.¹²

Therefore, if criticism must be levelled at all against the studies carried out by the theorists of neuroeconomics up to now, it is necessarily addressed to the opposite theories than those put forward by Gul and Pesendorfer. Neuroeconomics should not be criticised for addressing pathological or therapeutic issues, but rather for not sufficiently dwelling upon them and for subsequently failing at developing a model that could combine the structural and functional aspects involved in economic decision-making.

In fact, due to the communicative charm of the scans obtained through neuroimaging, the soundness of the research that embraced the idea that decisions were determined by genetic and neurological factors that could lead to behavioural answers at an emotional level—often working unconsciously—has often been overestimated, undermining the rational discussion of problems.

Nevertheless, there is no reason to believe that this conduct is based on sound theoretical assumptions emerging from neuroscientific or neuropsychological studies that, if generalised, could lead to a deterioration of our existence as such, strengthening the idea that only subliminal and emotional factors do play a role.

On the contrary, the most important lesson that can be drawn from contemporary neuroscience is not related to the localisation of cerebral areas that singly process one specific function, but rather to the fact that studying the brain of a species

¹¹ The concepts of *embodied* and *embedded* biological cognition, contrary to the solipsistic segregation of the mind supposed and simultaneously supported by classical cognitive science, are two of the main drivers that pushed the research and studies on motor cognition and the cognitive relevance of emotions, against the intellectualism of Cartesian origin.

¹² One of the most successful models in promoting the integration of neuroscience and neuropsychology concerns the study of linguistic functions and it was developed by Grodzinsky and Amunts (2006), who believed that the localisation of sub and neo-cortical areas in neuroscientific topography would have been impossible without the identification of actual behaviours, linguistic (or not), and their "intangible" interpretation—i.e. purely deductive—in terms of an explanatory theory of the interconnected functioning of systems of competencies (inference, mind reading, perception, syntax, semantics, etc.).

means studying also its relationship with the hands and feet, the musculoskeletal system, the circulatory and respiratory system, etc. In conclusion, it means studying the whole set of systems and structures that have determined the physiological characteristics of a species over the course of its evolutionary history. The same applies to mental functions. An animal who is able to develop language and culture—such as the human race—does not merely communicate in a different way, but perceives the world differently, reasons differently, remembers differently, yearns differently, gets excited and reacts to emotions differently, engages with others differently. This was indeed the breakthrough that pushed this cognitive science beyond its origins as a computational science, which saw cognitive processes as elements that could be fully simulated with the help of algorithmic processes, paving the way to the neuroscientific studies where the mind is not “artificial” but rather “natural”. Thanks to this change of paradigm, it was possible to leave behind the computational mind and the computer-like metaphor and launch a study of the brain where the latter is not considered a metaphor of the mind, but rather the mind, stemming from the basic assumption that mental phenomena should be tackled as natural phenomena.

Nevertheless, the term “naturalisation” implies a series of issues when it is used to identify the material sub layers of ideas and concepts in a specific field of research such as brain science, since research has the vocation of being managed with theories, interpretations and hypotheses. This is the issue that we will address in the next section.

4 What Kind of Naturalism?

As emphasized in the previous paragraphs, the term “mind” must not necessarily refer to a directly observable behaviour: in fact, “mind” refers to a wider set of cognitive, attentive and internal affective determinants, both private and subjective, which do not always translate into a perceivable and describable behaviour.

Furthermore, human beings react to the stimuli of the external environment in a subjective way, on the basis of their feelings and personal memories. In conclusion, individuals react to environmental stimuli according to the way in which their “mind” processes them, following the representation that they have created on the input received and the representation that they have of themselves.

In the past, these features led many to the conclusion that there are sophisticated levels of the mental life that go beyond and transcend the mere biological life, causing a division between mind and brain that was due to the apparent need of studying the mind by using categories, concepts, and tools different to those used to study other natural phenomena.

Even today, there are philosophers who are strongly convinced that all mental states are caused by neurobiological processes that take place in the brain while being sceptical towards the idea that the same phenomena described at a fundamental level (i.e. phenomena identified through the neurobiological language) are indeed sufficient to trigger higher mental phenomena (see Searle 2007; Dennett 1996). This is the idea at the basis of the normative models of ethical or social choice theory (at least in their utility-based version), according to which moral judgments do not

convey “moral intuitions” or value assessments of individual behaviours (“natural feeling”), but rather convey “reasons” that humans take into account to decide whether some moral “intuitions” are correct, reviewable, or contestable. From this point of view, moral norms are picked on the basis of rational criteria (which are at the basis of choices) that could vary, leading therefore to different choices.

In the other field we find the cognitive scientists contending that moral choices can, at the end of the day, also be taken down to cerebral physical statuses. An example is Michael Gazzaniga, who writes:

A series of studies suggesting that there is a brain-based account of moral reasoning have burst onto the scientific scene. It has been found that regions of the brain normally active in emotional processing are activated with one kind of moral judgment but not another. Arguments that have raged for centuries about the nature of moral decision and their sameness or difference are now quickly and distinctly resolved with modern brain imaging. The short form of the new results suggests that when someone is willing to act on a moral belief, it is because the emotional part of his or her brain has become active when considering the moral question at hand. Similarly, when a morally equivalent problem is presented but no action is taken, it is because the emotional part of the brain does not become active (Gazzaniga 2005, p. 167).

The studies mentioned by Gazzaniga have a long tradition and belong to the school of thought that postulates that human nature can only be fully and fairly explained by biology. In the words of Gazzaniga, all other possible explanations concerning the origins of human kind are only “stories that comfort, cajole, and even motivate—but stories nonetheless” (Gazzaniga 2005, p. 165).

Several authors have openly attacked these “stories” and the most prominent researcher in doing so is without any doubt Patricia Churchland, who in an article published in (1994) criticised the arguments that deny the neurobiological interpretation of the mind, categorising the doubts put forward by philosophers into five main groups (Churchland 1994) and strongly supporting the idea that the abilities of the human mind overlap with those of the human brain. The aim of Churchland’s strategy is to explain macrolevels (neuropsychological and cognitive abilities) in terms of microlevels (neurons’ properties), an approach that could be criticised, particularly considering its basic reductionist aspects (Legrenzi and Umiltà 2011). From this perspective, the criticism of Churchland by several authors is completely justified, especially from an epistemological point of view. Let’s take the example provided by Dietrich and List:

Consider, for example, how you would explain a cat’s appearance in the kitchen when the owner is preparing some food. You could either try (and in reality fail) to understand the cat’s neurophysiological processes which begin with (i) some sensory stimuli, then (ii) trigger some complicated neural responses, and finally (iii) activate the cat’s muscles so as to put it on a trajectory towards the kitchen. Or you could ascribe to the cat (i) the belief that there is food available in the kitchen, and (ii) the desire to eat, so that (iii) it is rational for the cat to go to the kitchen. It should be evident that the sec-

ond explanation is both simpler and more illuminating, offering much greater predictive power. The belief–desire explanation can easily be adjusted, for example, if conditions change. If you give the cat some visible or smellable evidence that food will be available in the living room rather than the kitchen, you can predict that it will update its beliefs and go to the living room instead. By contrast, one cannot even begin to imagine the informational overload that would be involved in adjusting the neurophysiological explanation to accommodate this change (Dietrich e List 2016, p. 27).

As Searle highlights, a reductionist approach may at best find the “correlations” between subjective statuses and cerebral ones, which might have a cause–effect relation but not an “identity” relation (Searle 2007). The Italian philosopher of the mind Michele Di Francesco goes in the same direction when he writes that:

Even recognising the causal role played by basic level phenomena cannot justify neurobiological fundamentalism. This is due to the fact that a specific event might be necessary for a specific effect without being its full and only cause; it can be a precondition, a secondary cause, one of the many conditions required for the establishment of a cause-effect link, and so on (Di Francesco 2007, pp. 136–137).

Therefore, it seems that the dilemma that must be solved boils down, on the one side, to the protection of a subjective and non-deterministic approach to human behaviour, while on the other there is the rejection of all forms of dualism, which represent pre-established walls dividing the study of the mind and the study of the brain, turning the mind into a natural phenomenon without falling for the charm of reductionism and biological determinism.

How is it hence possible to naturalise the mind without falling into the trap of the neurobiological fundamentalism described by Churchland or Gazzaniga?

Replying to this question is not easy, because the answer basically depends on the meaning of “naturalisation” or “naturalism”. In fact, these are two controversial terms and, in the history of philosophy, have acquired divergent and mutable meanings, becoming widely used in several philosophical fields (Ionian philosophers, Aristotelianism, Hume’s and Spinoza’s philosophy, the Positivism of the 19th century, logical empiricism and pragmatism, just to name but a few), all sharing just a vague reference to the natural world.

Luckily, over the last few years, contemporary naturalism acquired less vague connotations, to such an extent that nowadays the only two remaining big schools of thought are those of “scientific naturalism” and “liberalised or pluralist naturalism”. Both schools share the “constitutive thesis of naturalism”, i.e. the desire of using in their theories only laws, methods, and entities existing in nature, hence rejecting all those with supernatural features.¹³ In addition to this, both modern conceptions

¹³ As a matter of fact, several authors criticised the anti-scientific positions that referred to religious beliefs, mystical entities, demiurges, gods, etc. Furthermore, they strongly criticised the dualism related to the mind–body issue or the hypothesis of the Intelligent Design.

of naturalism agree in considering natural sciences the ideal model to which all forms of science should conform in order to be legitimised as knowledge-creating activities.

Yet, while the supporters of scientific naturalism are convinced that it is possible to reduce the scope of the phenomenon that one wishes to naturalise to the language of experimental science,¹⁴ the supporters of liberalised naturalism believe that it is possible to go beyond the clear division proposed by scientific naturalism between the phenomena belonging to the physical world and those that relate to the different fields of human existence. This allows the recovering of concepts (such as normativity, intentionality, free will, etc.) that cannot be sensed in the physical world, hence rightfully qualifying them as fully belonging to the natural world, without any form of metaphysical contamination (De Caro and Macarthur 2004).

The keystone of this new vision is indeed the compatibility (rather than the continuity) of philosophy and science, where a strong anti-reductionist character is salient, particularly as regards normativity.

But how is it possible to reconcile the normative level with the causal level typical of natural sciences?

John McDowell tries to answer this question with a proposal that epitomises the conciliating position taken by liberalised naturalists. McDowell contends that human beings have special characteristics that allow them to have a “second nature” (McDowell 2004). He claims that the best way to explain some human behavioural traits should not refer exclusively to the “causes” that govern body movements but, more importantly, to the “reasons” of acting. These “reasons” should not be considered to be completely disconnected and independent from human experience but, on the contrary, as an integral part of human nature (“second nature”). In this case, the liberalised naturalism proposed by McDowell complies with the first requirement (which could be defined as an ontological requirement), thanks to the inclusion among the explanations of all types of entities that are necessary to an explanatory end, without pre-established limits. In this way, the existence of elements of moral, modal, or intentional nature (and the truthfulness or falsehood of their judgment) becomes acceptable, because these elements are crucial to account for important aspects of our thought without including supernatural entities that go against the laws of nature in the explanation.

Therefore, McDowell’s liberalised naturalism¹⁵ does not encounter any difficulty in accepting the conceptual analysis (the methodological requirement) as a legitimate research method when it supplies productive inputs to explain specific phenomena, provided that it is not incompatible with natural sciences or, for example, neuroscientific studies. If this is true, then normativity is not incompatible with

¹⁴ The most radical form of scientific naturalism and of such a way of thinking can be traced back to Quine (1969), who contended that it was necessary to give up once and for all the “dream” of a *Philosophia Prima* (i.e. a form of philosophy that has the priority over natural science).

¹⁵ Nevertheless the same applies to all other authors who draw inspiration from this form of naturalism (see De Caro and Macarthur 2010).

descriptive and causal inquiries, which logically is tantamount to saying that it is compatible with them.

In this way, McDowell's liberalised naturalism represents a highly promising alternative to the reductionism promoted by scientific naturalism and, therefore, could provide a fruitful explanatory tool. Let us consider a couple of examples in the field we are currently analysing, i.e. economic theory.

As it is well known, according to the theory of rational choice, agents comply with a basic norm, which is the maximisation of the anticipated utility. The normative theory of rational choice postulates this norm and the choice principles it entails. The descriptive part of the theory has the task of determining whether the agents actually follow the norm. As several experimental psychologists have highlighted, under certain circumstances, agents do not follow the guidelines of normative theory, which leads to a series of paradoxes (the most famous of which is the paradox of Allais), where agents do not systematically maximise their utility. Several authors, particularly among psychologists, claimed that the misalignment between rational choice theory and reality is due to some reasoning mistakes or causal factors that influenced the agents' answer. In this case, as emphasized by Engel (2001), the interpretative approach seems not to work, because the causal factor is, so to speak, *obscure*, in that it does not refer to a norm but merely to a psychological process that is responsible of the mistake.

What Engel emphasized, i.e. the fact that cognitive modalities incompatible with the best epistemic practices available should be barred, is fully in line with the analysis of liberalised naturalism that, contrary to behavioural economics, can provide a rational answer to the abovementioned issue. In fact, the manifest paradoxes and mistakes incurred by agents can easily be explained as a product of "reasons" different than those established by the traditional neoclassical theory of utility maximisation. For example, the agent might decide to give up to maximisation today in order to get it tomorrow, or because by giving it up the agent believes that he/she will gain social prestige, etc. Therefore, according to the theories put forward by experimental psychologists or Engel himself, we do not always deal with real "mistakes".

The objection justifying this kind of argument basically relates to the fact that when it is time to evaluate the performance of subjects involved in reasoning tasks in a laboratory setting, it is easy to completely ignore the practical ends that, on the contrary, are typical of any spontaneous reasoning process in a natural context (Guala 2005). This kind of criticism against experimental economics implicitly entails that, outside of the laboratory, human inferences are always aimed at finalising and reaching a specific goal. Besides, as several studies have highlighted (Sperber et al. 1995; Bagassi and Macchi 2006; Sher and McKenzie 2006), some of the main phenomena that are traditionally considered as examples of the limits of human reasoning are instead the result of advanced inferential processes, largely influenced by considerations relating to semantics and meaning which strictly depend on the situation where the arguments are presented (Mercier and Sperber 2011).

As it happens in the case of experimental economists, not even the most radical form of naturalism can formulate an ontology that is able to include these explanations. On the contrary, liberalised naturalism takes on the challenge of keeping the notion that the facts taking place within the society are genuine facts and that the

mind should go beyond an intangible place enclosed inside a skull and include not only the concrete functioning structure of the brain and the rest of the body, but also the forms that the mind acquires in the environmental relations that it forges and from which it is influenced.

A field of study that patently clarifies the scope of the moderate naturalist approach is Game Theory.¹⁶

5 Game Theory and Scientific Naturalism

Over the last decades of study and research, Game theory and especially the Prisoner's Dilemma¹⁷ have been widely used by those neuroeconomists that strongly supported the partnership of Game theory and neurosciences to provide new and useful outcomes both for economists and the experts of the study of the brain. By applying the theories and methodologies developed by neurosciences and cognitive psychology to the Prisoner's Dilemma, researchers have tried to define a new and profoundly different framework compared to the one offered by the standard approach, in order to interpret differently the decisions taken by players during a game.

¹⁶ Game theory was initially developed as a part of applied mathematical sciences and it aims at studying rational individual choices in situations where different subjects interact, making the choices of one subject relevant for those of the others (often an adversary), according to a retroaction mechanism. The salient feature of game theory is that it aims at accounting for the strategic interactions in a situation where an agent does not face a passive environment, but rather an environment composed of—at least partially—other agents.

¹⁷ The Prisoner's Dilemma was developed in the 1950s by Merrill Flood and Melvin Dresher. It has been widely applied in economics and social philosophy, since it was believed that it could properly represent a wide range of social situations where the individual interest goes against the collective interest. The most famous formulation of the "Prisoner's dilemma" is as follows: Two suspects, A and B, are arrested by the police. The police have insufficient evidence for a conviction and, after having separated the two prisoners in two different cells, interrogates them and offer them the same deal: if one of them confesses (C) and the other does not confess (NC), then the one who did not confess will serve a 10-year sentence while the other will go free; if both of them do not confess, then they will both receive a 1-year sentence; while if both of them confess, then they will both be sentenced to 5 years in prison. Each prisoner can decide whether to confess or not. However, neither prisoner will know the other prisoner's choice. In this situation, all combinations of strategies are "Pareto optimal" with the exception of the mutual accusation strategy. If they were able to communicate with each other, then it is clear that the best strategy would be not confessing because both prisoners would be sentenced to only 1 year in prison. Nevertheless, as they cannot communicate, for the player who is questioned second, the best solution will always be to confess, without considering what the first prisoner might decide to do. In detail, if A confesses, then for B, it will be more convenient to confess, because in this way, both prisoners will serve a 5-year sentence; in any case, if A does not confess, then for B, it will be more convenient to confess, because in this way, he or she will go free, while the first prisoner will receive a 10-year sentence. This is the reason why the best solution for both players will always be to confess because independently of the other player's choice, the payoff of confessing is always higher than not confessing. A situation like this is called "Pareto inefficient," because even the most rational solution does not represent the best of all possible solutions. As a matter of fact, even if it is more convenient for both prisoners not to confess (because in this way, they will be sentenced to only 1 year in prison), then this strategy is less used because it is very risky: if the other prisoner were to confess (as it is rational for him to do), then the prisoner "taking the risk" would serve a 10-year sentence, letting the other free to go.

Among the earliest studies on the subject, it is worth highlighting the research by James K. Rilling and colleagues, who focussed on the relationship existing between the triggering of a specific cerebral area (the striatum, one of the regions of the brain associated with the feeling of satisfaction) and the cooperative behaviour of subjects involved in a Prisoner's Dilemma game (Rilling et al. 2002). To show that the subjects involved in the game got some satisfaction from the cooperation with other subjects (without considering monetary returns), the authors compared the triggering level of the striatum when subjects cooperated with other human partners with the triggering level of the same cerebral region when they cooperated with a computer. The authors noticed that even when the game parameters and monetary returns due to cooperation were exactly the same in both cases, the triggering levels of the striatum were markedly higher when the game partner was a human being. The authors came to the conclusion that subjects got a complementary dose of satisfaction (non-monetary) because of the mere fact of establishing a cooperation bond with another human being.¹⁸

Unfortunately, Rilling and colleagues do not provide an answer to the most important question highlighted by their experiments: why do we trust others and believe that they will treat us well? It is indeed logical to think that human beings have more than one reason (and many more than a computer) to behave egoistically and hence maximise their profits.

A group of Swiss researchers performed a much debated study that tried to answer this difficult question. Their research focussed on the role played by a very specific hormone called oxytocin (Kosfeld et al. 2005), which biologists have found in mammals when they are on the verge of establishing some kinds of social relationships, for example when they are in heat or when they develop maternal behaviours. According to these biological studies, oxytocin allows animals to approach other members of their species. The Swiss research team started from this assumption and put forward the idea that oxytocin allows human beings to establish social relationships and, in particular, develop feelings of trust and loyalty, just as it happens with other animals. To prove their point, they designed an experiment with 58 subjects divided in two equally distributed groups engaged in the trust game.¹⁹ Ahead of the game, the subjects of the first group were administered a massive dose of oxytocin through a nose spray, while the remaining 29 players received a placebo. The results showed that the average level of trust was significantly higher among the

¹⁸ The study performed by Rilling and colleagues was confirmed by the works of another neuroscience research team led by McCabe, who studied with fMRI the triggering levels of the prefrontal cortex of subjects playing a game similar to the Prisoner's Dilemma, i.e. the trust game (McCabe et al. 2001).

¹⁹ The participants to the trust game cannot communicate with each other. At the beginning of the game, they all receive a certain amount of money. The subjects belonging to the first group (investors) can give all their money, a part of their money or nothing to the subjects belonging to the second group (trustee). If they decided to give them no money, the game stops. If they give them money, the amount of money given is multiplied for a factor (e.g. three) and assigned to the trustees, who can then decide to give back to the investors a part of their money, all their money, or no money. The investor's payoff is equal to the initial amount of money received, minus the money given, plus the money received from the trustee; the payoff of the trustee is equal to the initial amount of money received, plus the money received from the investor, minus the money given to the investor.

subjects receiving the oxytocin compared to those receiving the placebo. In a nutshell: the oxytocin-exposed subjects invested much more money in the trust game compared to the others, proving that the hormone affected the subject's will to invest more in social relationships.

Another important result achieved by Kosfeld is related to the effect of oxytocin on the loyalty of trustees. Oxytocin-exposed players showed similar levels of loyalty to those receiving the placebo. In other words, these results proved that oxytocin affects the trust that one subject has towards another, but it does not affect the reliability of the latter and it does not make one more inclined to reply positively to the offers of the partner.

This result was partly confirmed by a subsequent study performed by Paul Zak and colleagues, who focussed on the class of players receiving the money (Zak et al. 2005). While performing their experiment, the authors noticed that when subjects saw the intention of the partner to invest trust in them, they released higher levels of oxytocin.

By combining the outcomes of both studies, it is clear that oxytocin is related to both the situation in which an investor trusts a trustee, and when a trustee recognises in the investor a genuine desire to trust him.

Nevertheless, the data on the role of oxytocin collected by these researchers have been abused over the years, leading to wrong (pseudo)scientific conclusions. These experimental data have been turned into some kind of proof that there is a cause-effect bond between the release of this specific hormone and the creation of trust bonds.²⁰ In fact, as outlined above, the release of oxytocin must be seen as one of the many correlations existing between a cerebral activity and a specific behaviour, not as the “cause” of trust. In other words, as far as oxytocin is concerned, the authors identify the coincidence of two events, nothing more, nothing less.

A recent study performed by Coren Apicella and colleagues corroborates this idea. Starting from a much more numerous statistical sample than the one used by Zak and Kosfeld in their experiments, they proved that between oxytocin and trust behaviours (even simple generosity) there is not even a simple correlation, while different behaviours are observed according to the population involved. This proves once again that cultural factors are much more relevant than genetic ones, at least as far as economic choices are involved (Apicella et al. 2010).

Another proof of the abuse of data related to oxytocin, or—even worse—their use for scientific marketing, comes from clinical studies.

It is well known that people suffering from Williams' syndrome have a well-developed empathic ability (Dykens and Rosner 1999) and, in particular, feel particularly attracted by strangers (Doyle et al. 2004). As shown by a study carried out at the prestigious Cedars-Sinai Medical Center in Los Angeles and published in 2012 on PLoS one, when subjects affected by Williams' syndrome are exposed to specific emotional inputs, they release quantities of oxytocin that are three times higher than those released by normal subjects. Nevertheless, the conclusions of this

²⁰ Even Paul Zak during several conferences referred to oxytocin as “trust molecule” or “moral molecule”.

study go in the opposite direction of those postulated by Zak. The authors do not refer to trust or prosocial behaviours by the subjects, but they rather believe that the excessive quantity of oxytocin is the cause of the behavioural anomalies caused by this syndrome. This study corroborates the literature that portrays a much less charming description of the relational skills of the subjects affected by the syndrome (see Sullivan and Tager-Flusberg 1999; Happé 1993; Gagliardi et al. 2003), highlighting the fact that the emotional response triggered in WS subjects by the presence of other people is often excessive and disproportionate. In brief, according to these studies, WS subjects have strong problems in understanding what other people are thinking or feeling.

Despite the fact that the extremely reductionist hypotheses put forward by researchers such as Zak fuel the interest of scientific naturalists (such as Gazzaniga) as much as they attract the criticism put forward by standard economists (such as Gul and Pesendorfer), these two different approaches do share a lot of similarities.

The most vocal supporters of reductionist naturalism “directly” associate some partial neural mechanisms to some economic behaviours, such as “trust”, “risk inclination”, “cooperation”, etc., without explaining what other connections turn those simple neural mechanisms into the concepts of “trust”, “risk”, and “cooperation”. Similarly, standard economists assume that individual choices are the manifestation of a maximising behaviour, without postulating how this happens (besides the rationality norm), completely disconnecting the process from reality.

These similarities are not something new. Alan Kirman and Miriam Teschl highlighted something similar in the “inequity aversion” model developed by Fehr and Schmidt (1999, 2010):

On the other hand, there are two behavioural and neuroeconomists, Fehr & Schmidt (1999, 2010), who have presented a model of an economic agent with particular other-regarding preferences, namely those exhibiting ‘inequity aversion’. In this model, the agent is supposed to care about the difference between others’ monetary outcomes and his own to the extent that any difference between the agent and someone else affects the agent negatively. We take this point to show the importance that some economists attribute to the stability of a given distribution of other-regarding preference types. Once again this allows one to fall back on the standard idea of fixed and immutable preferences (Kirman and Teschl 2010, p. 306).

To correct the wrong perspective used in the analysis of economic questions, Kirman and Teschl suggest to take a closer look to the relevance that empathy and mind reading have in decision-making processes.

6 Empathy and Pluralist Naturalism

Authors such as Rawls and Harsanyi have often highlighted the importance of empathy in the case of economic issues. When referring to utility interpersonal exchanges, Harsanyi contended that it was quite easy to understand their psychological basis: they were based on imaginative empathy, the agents’ ability to imagine being in

someone else's shoes (Harsanyi 1979). More recently, authors such as Ken Binmore highlighted the economic relevance of empathy by describing an imaginary society modelled and functioning on the basis of the strategic interactions of its subjects, hence placing empathy at its centre. According to Binmore, empathy should not be considered as “some auxiliary phenomenon to be mentioned only in passing”, but rather as a fundamental component that allows all subjects to understand the nature of economic interactions. Subsequently, “Homo economicus must be empathetic to some degree” (Binmore 1994, p. 28).

From the point of view of the researchers mentioned above, empathy is the particular ability of agents to mimic the internal feelings of other agents and, beyond its fundamental role in the establishment of positive interpersonal relationships and in the promotion of prosocial behaviours, it is extremely useful to explain decision-making situations where the positive outcomes for a specific individual depend on the individual's ability to strategically anticipate the behaviour of others.

When it comes to the specific issue of describing the processes at the basis of this empathic anticipation that allows individuals to understand and predict the behaviours of others, two different theories have been developed: the theory theory and the simulation theory.

According to the theory theory, human beings develop through their experience a form of naive theory (folk psychology) about the mental representations, motivations, aims, and emotions that drive different behaviours. The theory theory assumes that all individuals have at their disposal a quite advanced inferential system that provides them with the ability to anticipate what will happen by taking the experiential data that are the axioms of folk psychology as a starting point. The distinctive feature of theory theory is therefore the fact that it takes into account the skills expressed by common-sense psychology as part of a theory (Fodor 1983). Obviously, it would be a theory that—in the majority of individuals—would work implicitly. In fact, the theory theory does not assume that all individuals be able to ponder the single aspects of the theory or refer to it in order to make the right inferences. Yet, the systematic aspect of the anticipation mechanism would exist, even latently, in all human beings.

The limits of theory theory are striking. First of all, it takes for granted the existence of a naive theory of behaviour that consequently assumes that it is possible to consciously access complex inferential processes and skills such as the abstract representation of rules (Carruthers and Smith 1996).

Conversely, the simulation theory (Gordon 1995) rejects the use of a logical inferential chain, since it contends that inferences are too rich, burdensome and ant economical, rather focusing on a simpler explanation model of intersubjectivity based on emulation and imagination, as well as on the belief that beyond individual differences all human beings do share a mind that works similarly under similar circumstances.

Strong support to the simulation theory was provided at the beginning of the '90 s by the discovery of the features of a specific population of visuomotor neurons in the cerebral cortex devoted to the processing of information related to the behaviour

of others, the so-called “mirror neurons”,²¹ forcing anyone interested in intersubjectivity to face a new paradigm, hinged on the hypothesis of a neural correlate of empathy and, therefore, on the idea of a biological basis of sociability. In fact, from a neural perspective, there is no difference whatsoever between a subject performing an action and a subject observing another individual performing it (Rizzolatti and Sinigaglia 2006).

The discovery of mirror neurons confirmed the theory of social psychology on the automaticity and pervasiveness of emulation and empathy. As a matter of fact, mirror neurons are not simply triggered by looking at the motor schemes of others, but they are triggered also by looking at others’ emotional reactions: observing an emotion in someone else may trigger the activation of the same cortex regions in the observer that are normally activated when the observer feels that specific emotion (Rizzolatti and Voza 2008; Gallese and Sinigaglia 2012).

Nevertheless, besides the neural identity that allows us to explain the behavioural affinity among different agents (neural triggering that would be enough for a neuroeconomist supporting a naturalist approach à la Gazzaniga), there is still another problem to solve, i.e. the main object of our research: without taking into consideration the actions performed by single agents, which are the intentions that drive them and make them understandable to others? If the “common” character of this knowledge among agents can be explained by the mirror system, its intentional content seems nonetheless to be quite elusive. For example, the mirror system allows for the recognition of the rage in another person or the fact that another agent wishes to grasp an object. Yet, it is also necessary to explain the subsequent behaviour of an individual in complex social situations (such as those presumed by Game Theory). For the observer, it is not enough to recognise the rage in the other person, but rather to try to account for his rage, in order to trace and contextualise the reasons of his mood and understand the context that caused it.

Here is where the simulation theory based on the triggering of mirror neurons show its limits. If the main limitation of the theory theory is the fact that it is completely centred on the other subject, whose mind’s conceptual functioning must be understood in terms of logical-inferential argument chains, the simulation theory is instead too self-centred, since it entirely hinges on the abilities of the agents to put themselves in the others’ shoes. Yet, by taking a closer look, it is possible to understand that this egocentrism represents also a decentralisation of the Id, because it requires to put aside personal beliefs in order to take on those of others. In this way, the simulation theory—which is based on the special affinity existing between the observer and the observed agent whose behaviour can be understood and predicted—cannot explain on its own how it is possible to predict the behaviour of agents so different from the observer. Furthermore, it seems quite plausible that the

²¹ By recording what happens in single neurons, it was observed that in an area located in the premotor cortex of monkeys (area F5) there are neurons that get activated both when the monkey actively performed an action such as grasping or manipulating an object, and when the animal observed similar actions performed by another agent (another monkey or even a human being). These neurons have been therefore called mirror neurons, in order to emphasize the double activation involving them when performing or observing an action.

obvious goal of a simulation is not to understand what one would do if one were in someone else's shoes, but rather to imagine how the other will behave. In other words, when one places himself in the other's shoes, he should not take with him his mind, but he should try to see the world with the eyes of the one he is imagining to be.

The current state of the art of the debate is not entrenched on the positions of theory theory and simulation theory supporters: in fact, several attempts to combine the two positions and make concessions on specific points of the theoretical models are currently underway (Decety and Lamm 2006, 2007). These hybrid models provide an explanation of the fundamental characteristic of empathy, i.e. the recognition of the other as a fellow human being through a clear distinction between oneself and the other. Neuroscientific research has showed that this role is played by the lower right parietal cortex, in connection with the prefrontal areas and the front insular cortex (Keenan et al. 2003; Blakemore and Frith 2003). This cognitive mechanism represents a fundamental substratum for empathic experience, making a difference between simple emotional contagion (strictly based on the automatic bond existing between perception and action) and, as stated above, the empathy needed for “cooler” relationships, with a clear distinction between oneself and the other. The latter aspect requires therefore additional computational mechanisms that include the monitoring and managing of the inner information created by the activation of shared representations between oneself and the other.²²

Over the years, the neurobiological data and conceptual progress made on empathy led to useful experimental (and conceptual) developments in neuroeconomics.

To mention just one example, Singer et al. (2006) monitored the cerebral activity of individuals playing the Prisoner's Dilemma, with the clear intention of verifying whether the subjects could see the game not only from their perspective, but also from that of the adversaries. In order to do so, the experiment was divided into two stages. In the first stage, every player had the possibility to observe two types of behaviour by the other players: cooperation or systematic defection. The regularity of behaviours followed by the subjects had the aim of triggering opposed feelings of sympathy and dislike in the observers. In the second stage of the experiment, Singer and colleagues carried out a functional neuroimaging exam on the subjects, in order to understand whether the feelings of sympathy and dislike triggered during the first stage of the experiment affected or not the neural reactions of subjects when they saw the players exposed to painful stimuli. The authors recorded lower levels of empathy for those who followed a non-cooperative and defection-based behaviour in the first stage of the experiment.

The monitoring of cerebral activity of subjects seeing other individuals exposed to painful stimuli has been at the centre of several experimental studies on the neurobiological foundations of empathy. Generally speaking, the neural activation of

²² Once again, further studies supporting this hypothesis were carried out in clinical settings. They relate to some psychopathological disorders such as autism (McIntosh et al. 2006; Iacoboni and Dapretto 2006) or some forms of schizophrenia. In the latter case, for example, what is missing according to Gallagher (2004, 2007) is the sense of agency of the action, while the sense of ownership would remain unaltered.

the anterior insula (a region supposedly activated in physical or psychological situations felt with dislike or as unfair) and the anterior cingulate cortex (the region considered responsible of the processing of automatic warning reactions in situations of danger or social discomfort) both in the subjects feeling pain and in those watching them, led scientists to the conclusion that the observers seemed able to “feel” the pain of the others because they “shared” their pain, albeit in a different manner. As Singer and Fehr put it “[t]hese findings suggest that we use representations reflecting our own emotional responses to pain to understand how the pain of others feels” (Singer and Fehr 2005, pp. 5–6).

Therefore, the possible existence of a form of empathy can be of paramount importance for cooperation choices: by promoting the identification of others’ feelings and thoughts, it is possible to facilitate the self-interest opportunistic choices postulated by neoclassical economists. Empathy may yet push us away from this kind of choices and lead us towards less “selfish” choices: sharing emotions “renders” our emotions other-regarding, which provides the motivational basis for other-regarding behaviours (Singer and Fehr 2005, p. 2).

The results of the experiments carried out by Singer and colleagues show that all subjects seeing other cooperative subjects suffering record an empathic activation of pain mechanisms. Nevertheless, there is a clear difference between men and women in their empathic responses towards subjects who did not cooperate in previous experiences. In this case, men were less inclined to feel bad about the pain suffered by their opponents and, sometimes, it was even recorded the activation of the areas normally related to feelings of satisfaction. This behavioural difference has relevant impacts, because it flags the existence of different levels of empathy. The neuroscientific data collected by the authors prove that, even when seeing someone else experience a particular feeling (in the case of these experiments, the pain caused by an electric shock on one hand) automatically and unconsciously triggers a representation of that feeling in the observer’s brain (because of mirror neurons), this process cannot be inhibited or controlled. The management of empathy is an important factor of human lives, but it stops being possible when it becomes “too costly”: there is therefore a relative limit to the level of empathy that subjects can “bear” (see Decety and Lamm 2006).

This conclusion is supported by an experiment performed by Hichri and Kirman (2007) that analyses the behaviour of the subjects involved in a repeated public goods game.²³ The authors showed that in the early stages of the game, a majority of players generously contributed to the fund (in a higher proportion than what foreseen by Nash equilibrium), while later on the individual contributions decreased, even though the total contribution was unchanged. This means that the players that at the beginning contributed with more resources reduced their contributions during the game, and the other way around.

²³ In this game, each component of a group of n persons receives an x amount of money. Then, each participant must decide how much money to keep and how much money share in a common fund (the public good). When it is shared in the common fund, it is multiplied by a factor m by the researchers, dividing the final amount among the members of the group.

In brief, despite the fact that it is true that empathy plays a fundamental role in establishing the level of solidarity of individuals, the level of empathy is not constant, but it is rather calibrated by the players according to the context and the interactions with others. In other words, empathy is not a simple automatic bottom-up process (biologically based on mirror neurons), but individuals can manage their level of empathy by using top-down processes.

These remarks on empathy are not very much welcomed by economists or reductionist neuroeconomists. The perfect situation for economists would be for the studies on empathy to reveal a distribution of empathy in the population according to a series of stable preferences that could be used to develop economic predictions. This is due to the fact that the fundamental assumption of economics is that behaviour in general and choices in particular can be compared to stable and unchangeable preferences. As Samuel Bowles clearly describes, economists usually believe that ““for preferences to have explanatory power they must be sufficiently persistent to explain behaviours over time and across situations” (Bowles 1998, p. 79).

Similarly, reductionist economists would prefer dealing with individuals with a specific pre-established level of empathy, unaffected by the context or the people that they face in their game. Should this happen, then, they could design neurophysiological experiments aimed at finding for each individual his/her empathic predisposition to cooperation.

On the contrary, the naturalist pluralist model described above and suggested by some neuroeconomists for empathy postulates an integrated approach based on the joint role of cognition and affections. When we combine this with the neuroscientific studies on the correlations between behaviour and its underlying neurophysiological mechanisms, both in its normal and pathological versions, we reach the last frontier of neuroeconomic studies, where the choices of individuals are explained not only by using cerebral physiology, but also by properly referring to the human nature.

7 Conclusions

As we have seen, trying to capture the essence of social behaviour and intelligence without falling into the temptation of exploiting reductionist ideas obviously implies the need of taking into account the various description levels of the question tackled: from neural dynamics to social dynamics, from the adaptive logics of its development in the phylogenetic tradition up to its emergence during the ontogenetic development of different species. Our analysis emphasized that not all normative categories that account for human agency may be translated and explained in neuroscientific terms. Following this thesis, though, does not imply the rejection of a naturalistic interpretation of the mind. Naturalism per se does not imply that all discoveries of partial aspects of brain functioning should be considered a part of a consistent theoretical system that will reduce all cognitive and behavioural functions to neuropsychological or neurobiological functions. This is the first and main reason why it is preferable to opt for the liberalised and pluralist version of naturalism that tends to include in its research the authenticity of some significant cultural phenomena, their normative aspects (for example laws or social conventions)

and the qualitative aspects of mental states that imply awareness, without opening the door to the possibility of questioning the legitimacy of neuroscientific research on higher human functions. Liberalised naturalists agree on the idea that our decisions are the expression of thoughts that have some kind of neurophysiological basis that, obviously, would be highly interesting to identify. Yet, at the same time, they believe that their neurophysiological basis does not completely explain them and, on the contrary, an inter-theoretical model developed by different disciplines (neuroscience, psychology, economics, and philosophy) that could combine different inquiry techniques in a multidisciplinary framework might provide a better naturalistic explanation of mental phenomena. By using these models, phenomena such as decision-making, empathy or the establishment of social groups and the exchanges that take place inside them might be finally brought in the research framework of other natural phenomena that were successfully investigated over the last few centuries, without having to follow an unbearable form of scientism.

References

- Apicella CL, Cesarini D, Johannesson M, Dawes CT, Lichtenstein P, Wallace B, Beauchamp J, Westberg L (2010) No association between oxytocin receptor (OXTR) gene polymorphisms and experimentally elicited social preferences. *PLoS ONE* 5:e11153
- Bagassi M, Macchi L (2006) Pragmatic approach to decision making under uncertainty: the case of the disjunction effect. *Think Reason* 12(3):329–350
- Bernheim BD, Rangel A (2009) Beyond revealed preference: choice theoretic foundations for behavioral welfare economics. *Q J Econ* 124(1):51–104
- Binmore K (1994) *Playing fair*. Vol. 1 of game theory and the social contract. MIT Press, Cambridge
- Blakemore SJ, Frith C (2003) Self-awareness and action. *Curr Opin Neurobiol* 2:219–224
- Bourgeois-Gironde S, Schoonover C (2008) Une hybridation de l'économie et des neuro sciences a-t-elle un sens. *Revue d'économie politique* 118:35–50
- Bowles S (1998) Endogenous preferences: the cultural consequences of markets and other economic institutions. *J Econ Lit* 36:75–111
- Brocas I, Carrillo JD (2008) The brain as a hierarchical organization. *Am Econ Rev* 98(4):1312–1346
- Camerer CF (2008) The potential of neuroeconomics. *Econ Philos* 24:369–379
- Carruthers P, Smith PK (1996) *Theories of theories of mind*. Cambridge University Press, Cambridge
- Chomsky N (1959) A review of B. F. Skinner's verbal behavior. *Language* 35(1):26–58
- Churchland PS (1994) Can neurobiology teach us anything about consciousness? *Proc Addresses Am Philos Assoc* 67(4):23–40
- Churchland PS (1998) What should we expect from a theory of consciousness? *Adv Neurol* 77:19–30
- Churchland PS (2002) *Brain-wise*. Studies in neurophilosophy. MIT Press, Cambridge
- Churchland PS (2008) The impact of neuroscience on philosophy. *Neuron* 60(6):409–411
- De Caro M, Macarthur D (2004) *Naturalism in question*. Harvard University Press, Cambridge
- De Caro M, Macarthur D (2010) *Naturalism and normativity*. Columbia University Press, New York
- Decety J, Lamm C (2006) Human empathy through the lens of social neuroscience. *Sci World J* 6:1146–1163
- Decety J, Lamm C (2007) The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13:580–593
- Dennett DC (1996) *Kinds of minds*. Basic Books, New York
- Di Francesco M (2007) Neurofilosofia, naturalismo e statuto dei giudizi morali. *Etica Politca* 2:126–143
- Dietrich F, List C (2016) Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Econ Philos* 2:175–229
- Doyle TF, Bellugi U, Korenberg JR, Graham J (2004) "Everybody in the world is my friend" hypersociability in young children with Williams syndrome. *Am J Med Genet* 124:263–273

- Dykens EM, Rosner BA (1999) Refining behavioral phenotypes: personality-motivation in Williams and Prader–Willi syndromes. *Am J Ment Retard* 104:158–169
- Engel P (2001) *L' espace des raisons est sans limites?*. Gallimard Folio, Paris
- Fehr E, Schmidt K (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114:817–868
- Fehr E, Schmidt KM (2010) On inequity aversion: a reply to Binmore and Shaked. *J Econ Behav Organ* 73:101–108
- Fodor JA (1975) *The language of thought*. Harvard University Press, Cambridge
- Fodor JA (1983) *The modularity of mind: an essay in faculty psychology*. MIT Press, Cambridge
- Gagliardi C, Frigerio E, Burt DM, Cazzaniga I, Perrett D, Borgatti R (2003) Facial expression recognition in Williams syndrome. *Neuropsychologia* 41:733–738
- Gallagher S (2004) Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology* 37:8–19
- Gallagher S (2007) Sense of agency and higher-order cognition: levels of explanation for schizophrenia. *Cogn Semiot* 32–48
- Gallese V, Sinigaglia C (2012) Response to de Bruin and Gallagher: embodied simulation as reuse is a productive explanation of a basic form of mind-reading. *Trends Cogn Sci* 16(2):99–100
- Gazzaniga MS (2005) *The ethical brain*. Dana Press, New York
- Gazzaniga MS (2008) *Human. The science behind what makes us unique*. Ecco Books, New York
- Gordon RM (1995) The simulation theory: objections and misconceptions. In: Davies M, Stone T (eds) *Folk psychology*. Blackwells, Oxford, pp 100–122
- Grodzinsky Y, Amunts K (2006) *Broca's region*. Oxford University Press, Oxford
- Guala F (2005) *The methodology of experimental economics*. Cambridge University Press, Cambridge
- Gul F, Pesendorfer W (2005) *The case for mindless economics*. Working paper
- Gul F, Pesendorfer W (2008) *The Case for Mindless Economics*. In: Caplin A, Schotter A (eds) *Foundations of positive and normative economics*. Oxford University Press, New York, pp 3–40
- Hands DW (2009) *The positive-normative dichotomy and economics*. In: Gabbay Dov M, Thagard Paul, Woods John (eds) 2011 *Handbook of the philosophy of science*. Elsevier, Amsterdam
- Happé FGE (1993) Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48:101–119
- Harrison GW (2008) Neuroeconomics: a critical reconsideration. *Econ Philos* 24(3):303–344
- Harsanyi J (1979) Bayesian decision theory, rule utilitarianism, and Arrow's impossibility theorem. *Theory Decis* 3:289–317
- Hichri W, Kirman A (2007) The emergence of coordination in public good games. *Eur Phys J B Condens Matter Complex Syst* 55:149–159
- Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310:1680–1683
- Iacoboni M (2008) *Mirroring people: the new science of how we connect with others*. Farrar, Straus and Giroux, New York
- Iacoboni M, Dapretto M (2006) The mirror neuron system and the consequences of its dysfunction. *Nat Rev Neurosci* 7:942–951
- Kahneman D (2003) *Maps of bounded rationality: psychology for behavioral economics*. *Am Econ Rev* 93:1449–1475
- Kandel E (2005) *Psychiatry, psychoanalysis, and the new biology of mind*. American Psychiatric Publishing, Washington, DC
- Katz JJ (1964) Mentalism in linguistics. *Language* 40(2):124–137
- Keenan JP, Gallup GG, Falk D (2003) *The face in the mirror: the search for the origins of consciousness*. HarperCollins, New York
- Kirman A, Teschl M (2010) Selfish or selfless? The role of empathy in economics. *Philos Trans R Soc Lond B Biol Sci* 365:303–317
- Knutson B, Peterson R (2005) Neurally reconstructing expected utility. *Games Econ Behav* 52:305–315
- Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. *Nature* 435:673–676
- Legrenzi P, Umiltà C (2011) *Neuromania: on the limits of brain science*. Oxford University Press, Oxford
- Lurija AR (1973) *The Working Brain. An Introduction to Neuropsychology*, Harmond- sworth
- Mc Intosh DN, Reichman-Decker A, Winkielman P, Wilbarger J (2006) When the social mirror breaks: deficits in automatic, but not voluntary mimicry of emotional facial expressions in autism. *Dev Sci* 9:295–302

- McCabe K, Rigdon M, Smith V (2001) Cooperation in single play, two-person extensive form games between anonymously matched players. In: Zwick R, Rapoport A (eds) *Experimental business research*. Kluwer, Boston, pp 49–68
- McDowell J (2004) Naturalism in the philosophy of mind. In: De Caro Mario, MacArthur David (eds) *Naturalism in question*. Harvard University Press, Cambridge
- Mercier H, Sperber D (2011) Why do humans reason? Arguments for an argumentative theory. *Behav Brain Sci* 34:57–111
- Paulus MP, Hozack N, Zauscher B, McDowell JE, Frank L, Brown GG (2001) Prefrontal, parietal and temporal cortex networks underlie decision-making in the presence of uncertainty. *Neuroimage* 13:91–100
- Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233–238
- Quine WVO (1969) *Epistemology naturalized, ontological relativity and other essays*. Columbia University Press, New York
- Rilling JK, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002) A neural basis for social cooperation. *Neuron* 35:395–405
- Rizzolatti G, Sinigaglia C (2006) *So quello che fai. Il cervello che agisce e i neuroni specchio*. Raffaello Cortina, Milano
- Rizzolatti G, Vozza L (2008) *Nella mente degli altri. Neuroni specchio e comportamento sociale*. Zanichelli, Bologna
- Ross D (2008) Two styles of neuroeconomics. *Econ Philos* 24(3):473–483
- Samuelson PA (1938) A note on the pure theory of consumer's behaviour. *Economica* 5(17):61–71
- Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. *Science* 300:1755–1758
- Searle J (2007) Biological naturalism. In: Velmans Max, Schneider Susan (eds) *The Blackwell companion to consciousness*. Blackwell, London
- Sen AK (1977) Rational fools: a critique of the behavioral foundations of economic theory. *Philos Public Aff* 6(4):317–344
- Sher S, McKenzie CRM (2006) Information leakage from logically equivalent frames. *Cognition* 101:467–494
- Singer T, Fehr E (2005) The neuroeconomics of mind reading and empathy. Discussion Paper No. 1647. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=747904
- Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan R, Frith C (2006) Empathetic neural responses are modulated by perceived fairness of others. *Nature* 439:466–469
- Sperber D, Cara F, Girotto V (1995) Relevance theory explains the selection task. *Cognition* 57:31–95
- Sullivan K, Tager-Flusberg H (1999) Second-order belief attribution in Williams syndrome: intact or impaired? *Am J Ment Retard* 104:523–532
- Thaler RH, Sunstein CR (2008) *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press, New Haven
- Vallois N (2012) The pathological paradigm of neuroeconomics. *Oeconomia* 1(4):525–556
- Zak PJ, Kurzban R, Matzner WT (2005) Oxytocin is associated with human trustworthiness. *Horm Behav* 48:522–527