Neuroeconomics and modern neuroscience

Daniel Serra

▶ **To cite this version:**

Daniel Serra. Neuroeconomics and modern neuroscience. 2019. ⟨halshs-02160907⟩

HAL Id: halshs-02160907

https://shs.hal.science/halshs-02160907v1

Preprint submitted on 20 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

CEE-M
CENTER FOR ENVIRONMENTAL ECONOMICS – MONTPELLIER

Neuroeconomics and Modern Neuroscience

Daniel Serra

CEE-M Working Paper 2019-12

CEE-M

Neuroeconomics and Modern Neuroscience

Daniel Serra*[†]

CEE-M, Univ. Montpellier, CNRS, INRA, Montpellier SupAgro, Montpellier, France

May 30, 2019

Abstract.

The paper is an overview of the main significant advances in the knowledge of brain functioning by modern neuroscience that have contributed to the emergence of neuroeconomics and its rise over the past two decades. These advances are grouped over three non-independent topics referred to as the "emo-rational" brain, "social" brain, and "computational" brain. For each topic, it emphasizes findings considered as critical to the birth and development of neuroeconomics while highlighting some of prominent questions about which knowledge should be improved by future research. In parallel, it shows that the boundaries between neuroeconomics and several recent subfields of cognitive neuroscience, such as affective, social, and more generally, decision neuroscience, are rather porous. It suggests that a greater autonomy of neuroeconomics should perhaps come from the development of studies about more economic policy-oriented concerns. In order to make the paper accessible to a large audience the various neuroscientific notions used are defined and briefly explained. In the same way, for economists not specialized in experimental and behavioral economics, the definition of the main economic models referred to in the text is recalled.

Keywords: *Neuroeconomics, Neuroscience, Behavioral economics, Experimental economics*

Classifications (JEL): *D87, A12, B41, C7, C9, D81*

---

* Corresponding Author: *Daniel Serra, CEE-M, Univ. Montpellier, CNRS, INRA, Montpellier SupAgro, Montpellier, France ; Faculté d'économie, Site Richter, Avenue de la mer – Raymond Dugrand, 34960, Montpellier, CS : 79606 ; e-mail: daniel.serra@univ-montp.fr.*

[†] The author thanks Sacha Bourgeois-Gironde, Philippe Mongin, Pierre Livet, Christian Schmidt, Nicolas Vallois, Marc Willinger for useful comments and conversations over the years related to the topics of this paper. He also thanks Thierry Blayac and Guillaume Cheikbossian for extremely useful comments and suggestions on both earlier and current versions of the paper.

By the late 1990s, several converging trends in economics, psychology, and neuroscience had set the stage for the birth of a new scientific field known as “neuroeconomics.” The take-off of experimental economics in the 1980s undoubtedly favored the emergence of the first studies in neuroeconomics by offering a set of well codified experimental designs. Nevertheless, it is generally agreed that this new field is mainly based on the neuroscientific revolution of the 1990s, with the provision of sophisticated investigating tools, primarily functional magnetic resonance imaging (fMRI), in which metabolic correlates of neural activity can be measured non-invasively. In a more fundamental way, neuroeconomics has largely built on the fundamental knowledge developed by several branches of modern neuroscience.

Neuroscience has always been a multi-disciplinary field, covering different explanatory goals, concepts and vocabularies, and different techniques and methods. One explicit aim of the Society for Neuroscience, which came into existence in 1970, was to integrate all these fields with the common goal of understanding the nervous system (Craver, 2007). Overall, neuroscience is usually divided into two vast fields: *molecular and cellular neuroscience* and *cognitive and behavioral neuroscience*. The former studies neurons at a cellular level and examines the biology of the nervous system, while the latter is devoted to the study of neural mechanisms of mental and behavioral activities, or more generally, the relationships among the brain, mind, and action (Gazzaniga and Mangun, eds., 2014). Neuroeconomics is closely, but not exclusively, associated with cognitive and behavioral neuroscience.

More specifically, within the mosaic of neuroscience, several branches support the field of neuroeconomics, including neurobiology, neuroimagery, neuroanatomy, neuropsychology, neurophysiology, neuroendocrinology, and computational and theoretical neuroscience. *Neurobiology* focuses on the overall biological functioning of the nervous system in relation to genetics and is typically regarded as the flagship branch of neuroscience. *Neuroimagery* enables the identification of brain regions active (i.e. “firing”) when a subject is performing specific tasks or is receiving information (brain imaging studies). *Neuroanatomy* increasingly improves the knowledge of the structure and organization of the nervous system. *Neuropsychology* is devoted to the relationships between brain structures and functions by comparing healthy subjects and patients with brain damage or neurological or psychiatric disorders (lesions studies). *Neurophysiology*—by inserting fine electrodes into neural tissue immediately adjacent to the neurons of interest—allows the recording of electrical activity when, for example, a non-human primate is performing a specific task (single neuron recording or electrophysiological studies). *Neuroendocrinology* deals with the relationship between the nervous and hormonal systems, revealing, in particular, the impact of some hormones or neuropeptides on behaviors and mental states (pharmacological interventions).

Computational and theoretical neuroscience studies computational abilities of the nervous system and develops formal models reporting on algorithms for implementing well defined functions often in relation to artificial intelligence.

On the other hand, cognitive and behavioral neuroscience include several other sub-disciplines, in particular *decision neuroscience*, *affective neuroscience*, and *social neuroscience*, three cutting-edge fields whose boundaries with neuroeconomics are occasionally blurred due to the shared focus on decision-making, emotions, and behaviors. Decision neuroscience is broadly defined as a wide converging field between cognitive neuroscience and decision sciences (such as psychology and economics), while affective and social neurosciences pursue neighboring but more limited ends: the former studies neural mechanisms involved especially in emotion and affects, whereas the latter is devoted to understanding how biological systems implement social processes and behavior.

The goal of the present paper is to provide an overview of the main significant advances in the knowledge of brain functioning by modern neuroscience that have contributed to the emergence of neuroeconomics and its rise over the past two decades, while highlighting some parallels between this new scientific field and, in particular, decision, affective, and social neurosciences (for a first brief history of neuroeconomics, refer to Glimcher and Fehr, 2014). These key neuroscientific advances are grouped over three non-independent topics.

The first topic is devoted to emotion processing by the brain and the findings regarding the interconnection of emotions and higher cognitive processes. The key role of emotion in economic decisions is now unanimously recognized as a teaching of neuroeconomics, in the wake of behavioral economics (Sanfey et al., 2006; Engelman and Hare, 2018). Yet, neuroanatomy, affective neuroscience, and neuropsychology are at the heart of identifying neural structures and mechanisms involved in both cognitive processes and emotional responses. The second topic refers to the human brain's considerable flexibility and ability to undertake complex patterns of social cognition; social neuroeconomics focuses on decisions made in a social context and seeks to explain prosocial behaviors, such as trust (Fehr and Camerer, 2007; Sanfey, 2007; Rilling and Sanfey, 2014; Fehr and Krajbich, 2014; Engelman and Fehr, 2017). This neuroeconomic subfield relies partly on findings of social neuroscience about the neural networks that are responsible for interpreting other people's thoughts and feelings, sympathizing with their states of mind, and acting in a moral manner, namely "mentalizing", mirror neurons, and empathy systems. The third topic deals with reward learning as a new theoretical framework for neuroscience and the identification of brain mechanisms deployed for learning and valuing the many stimuli the brain is continuously subjected to. Neuroeconomics can now draw the contours of a computational model of how the brain makes "simple" economic choices, and recent studies have explored how this structural model may extend to more complex decisions, such as risky decisions,

intertemporal choices, and social decisions (Fehr and Rangel, 2011; Rangel and Clithero, 2014; Berridge and O'Doherty, 2014; Glimcher, 2011a, 2014a; Ruff and Fehr, 2014; Hutcherson, Bushong, and Rangel, 2015; Krajbich et al., 2015). Here, these studies are mainly rooted in neurobiology and computational and theoretical neuroscience while taking into account findings from studies relating to the other two topics.

The paper is organized as follows. Once briefly mentioned the aim of neuroeconomics (section 1), I will deal with these three topics by talking respectively about the "emo-rational brain" (section 2), "social brain" (section 3), and "computational brain" (section 4).

1. The Aim of Neuroeconomics

The pioneers of this young scientific field have different definitions for neuroeconomics. Originally, several definitions stand out from two separate communities: one predominantly (although not exclusively) behavioral economic (what will be called the "behavioral economics in the scanner" program) and the other predominantly (although not exclusively) neuroscientific (called the "neural or neurocellular economics" program). However, it can be argued that these two trends today are converging on a largely shared research agenda.

Two early distinct programs: "behavioral economics in the scanner" and "neural economics"

According to the most frequently used definitions given by economists, the purpose of neuroeconomics is to study influence of the brain and of the nervous system on economic behaviors. The now available neural and physiological measures should be used for a better understanding of the nature of deliberative and affective processes underlying decision making by economic agents (McCabe, 2003; Camerer, Loewenstein, Prelec, 2005; Camerer, 2007, 2008a). Neuroeconomics is recognized as a transdisciplinary domain using neuroscientific tools of measure in order to identify the neural bases of economic decisions (Zac, 2004). At the crossroads of economics and neuroscience, this new approach seeks to a better knowledge of choice models by taking advantage of each ones (Sanfey et al., 2006). In other words, neuroeconomics simply seeks to complete the behavioral economics approach by inquiring upon the neurobiological origin of psychological traits that these new economic models assign to individuals. A large part of economics is now soaked with psychology; with the help of neuroeconomics, this discipline should turn into a biological science.

The shared characteristic of these first definitions of neuroeconomics is emphasizing the benefits of using neuroscientific tools for studying economic behaviors. The whole range of tools used today in cognitive neuroscience can be divided into two main categories: (1) measurement techniques, that measure changes in brain function while an experimental

subject (human or animal) engages in some cognitive activity, and (2) manipulation techniques, which examine how perturbations of the brain's function (either by transiently changing neural firing rates or neurotransmitter levels or by permanently damaging tissue) change cognitive functions or behavior. Single-unit recording, electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), and fMRI are measurement techniques commonly used in neuroscience; however, fMRI is by far the most frequently used tool in neuroeconomics[1]. Manipulation techniques can be grouped into three classes: 1) brain stimulation techniques, including transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (TDCS); 2) neuropharmacological interventions (classically, manipulation of neuromodulators, including neurotransmitters such as dopamine, serotonin, and noradrenaline/norepinephrine, as well as hormones such as oxytocin and testosterone); and 3) approaches that study the consequences of brain lesions and neurological or psychiatric disorders. In neuroeconomics, only a few experiments have used brain stimulation techniques (Knoch et al., 2006a; Knoch et al., 2006b; Knoch et al., 2008; Karton and Bachmann, 2011; Baumgartner et al., 2011; Ruff, Ugazio and Fehr, 2013), while a lot of experiments now involve neuropharmacological interventions, mainly concerning hormones (see Crockett and Fehr, 2014)[2]. Ruff and Huettel (2014) (see also Genon et al., 2018) have evaluated the strengths and limitations of these various cognitive neuroscience approaches, which are added to older psychophysiology tools and methods for experimentally studying emotional and affective responses. Whether it concerns direct observation of body physiological activation (heart rate, blood pressure, galvanic skin response, eye-tracking, response delays, and recording of subjects' activity) or measuring emotions via facial expressions (Darwin, 1859; Ekmann, 1982; Fernandez-Dols and J. Russell, 2017), these methods are commonly used today in neuroeconomic experiments.

A different neuroeconomics definition was initially adopted by other researchers. Don Ross used the name "behavioral economics in the scanner" for the works corresponding to previous definitions, and identified another trend in the literature, which he named "neurocellular economics" (Ross, 2005, 2008) (Vromen, 2007, also identified these two distinct programs). Here, the goal is to borrow from microeconomics concepts and mathematics of equilibrium and optimum for modeling neural cells activity and testing hypotheses about neural learning. According to this viewpoint, standard economic theory may be suitable for modeling neuron networks than current individuals as economic "agents". The objective functions optimized by "agents," their utility functions, as we suppose in microeconomics, may in principle apply to all kinds of "agents," including sub-personal

[1] Near-infrared spectography is another method very recently introduced in experiments (see Kopton and Kenning, 2014).
[2] Additional invasive stimulation methods (as microstimulation and optogenetics) have recently been introduced in nonhuman primate experiments (for optogenetics, see Zalocusky and Deisseroth, 2014, and Galvan et al., 2017).

agents such as the neuron, neurotransmitter system, or quasi-modular circuit (Ross, 2008, p. 479). In other words, this author rejects the idea of anthropomorphism as an initial necessary condition for neuroeconomics. Thus, neurocellular economics, which I prefer to call "neural economics" after Montague and Berns (2002), borrows the conceptual arsenal of economic theory for understanding neural mechanisms through which the brain is valuing and comparing the multiple stimuli it is subjected to in decision-making. In a way, the project is in contrast to the goal of behavioral economics in the scanner; rather than seeking to improve economic theory by borrowing tools from neuroscience, the purpose is to use standard economic theory for analyzing brain functioning (Glimcher, 2003; Montague, 2007).

It is well known that neuroeconomics can be divided into these two distinct topics based on theoretical roots and project programs. In the 1st edition of the *Handbook of Neuroeconomics* (Glimcher et al., 2009), the editors have agreed to this internal division among neuroeconomists. This distinction is useful because many critics toward neuroeconomics expressed by some economists were aimed at the "behavioral economics in the scanner" program[3]. Nevertheless, in the light of recent research evolution, its relevance is going to fade. Currently, these two historical neuroeconomics programs are in the process of converging on a largely shared research agenda.

"Value-based" decisions: the main research topic assigned to neuroeconomics

For many scholars, now, the transfer of knowledge from economic theory to neuroscience (i.e. the first goal of "neural economics") would have constituted only a first step during information exchange between the two scientific disciplines. A transfer in the reverse direction, from neuroscience to economic theory, has to be followed. Indeed, the final goal of neuroeconomics is first and foremost improving the predictive power of economic theory[4]. This can be achieved by inserting the biological constraints that brain functioning poses to behaviors into economic theory. In this respect, the viewpoint of Paul Glimcher, a pioneer in neural economics, is meaningful when he writes in the introduction of his last influential book, "we know that there are things human brains can and cannot do. If economic theories about how people make choices should be constrained by what human brains actually can do, then

[3] In short, the argument is that if a phenomenon is already well known in psychological and behavioral terms, knowledge of neural correlates and mechanisms would be useless for economists (e.g. Smith, 2008, Chap. 14; Harrison, 2008a, 2008b; Rubinstein, 2008). In addition to this issue of utility for economists, controversial debates about neuroeconomics bear on reliability of findings, in relation particularly with the so-called "reverse inference fallacy" and the non-trivial statistical analysis of fMRI data (Poldrack, 2006; Harrison, 2008b; Harrison and Ross, 2010; Bourgeois-Gironde, 2010; Poldrack et al., 2017) but also on legitimacy of the approach from an epistemological standpoint, in particular the "mindless economics" argument (Gul and Pesendorfer, 2005/2008) (see e.g. Serra, 2016, Chap. 3, for a recent statement of the discussions).

[4] In the same time, neuroeconomics results are useful in psychiatry for analyzing a constellation of mental and neurological disorders including frontotemporal dementia, obsessive-compulsive disorder, and drug addiction (see Zald and Rauch, eds., 2006, Part 3; Maia and Frank, 2011; several Chapters in Schutt, Seidman, and Keshaven, eds., 2015; Dreher and Tremblay, 2017, Part IV; Allos-Ferrer, 2018).

quite a bit of contemporary neuroscience can be of use in the construction of economic theory" (Glimcher, 2011a, p. xvii).

Vromen (2011) appears to be the first scholar who noted this convergence between the two main historical neuroeconomics programs by focusing just on Glimcher's thinking evolution from early to late 2000s (Glimcher, 2003, 2011a). However, this evolution is strongly detected in recent works of many neuroeconomists, including, but not limited to, B. Douglas Bernheim, Colin Camerer, Ernst Fehr, Joseph Kable, and Antonio Rangel (see Serra, 2016, Chap. 1).

Perceptual decisions and value-based decisions

In decision-making, how does the brain choose among options? To answer this question, we have to first precisely make the decision nature relevant. Cognitive psychologists and decision neuroscientists now distinguish between two great decision families: (1) "perceptual" decisions, which refer to processes by which a subject is reacting to a sensorial input (e.g., at the airport, for the employee who is scanning the personal luggage of passengers, to decide instantly whether it is a gun or a hair-drier, or for a woman who is about to cross a street, to decide how quickly is a particular car moving toward her)*;* (2) "value-based" decisions (VBDs), which correspond to the subject idiosyncratic preferences (for instance, to choose between eggs or cereals for the breakfast, or between different financial investments in a retirement plan). Contrary to perceptual decisions, VBDs are subjective by nature. In this regard, they correspond to behaviors that economists typically are studying in their models. Glimcher (2014a) and Wang (2014) proposed an overview of these "twin approaches" of decision-making in neuroscience: "The theoretical background of these approaches is clearly distinct: perceptual decisions are based on the standard neurobiological theory of perceptual categorization, the "signal detection theory," whereas VBD refers to the standard microeconomic theory, the expected utility theory.

The modern idea that perceptual experience could be studied as a mental phenomenon distinct from external physical measurements has its origins in the work of the 19th century German physiologist Ernst Weber who showed that human perceptual judgments are imperfect and, although probabilistic, quite lawful (Weber, 1834). His more important finding that this confusability of perceptual judgments scales as a constant fraction of stimulus intensity became codified as a psychological law. Signal detection theory was largely developed by Green and Sweets (1966) who put a set of assumptions that are meant to correlate stimulus properties directly with perceptual judgments about confusability. The key assumptions are that real-world stimuli give rise to percepts through a random process like drawing from a Gaussian distribution (Macmillan and Creelmann, 2004). This notion that subjects have variable internal experiences from the same stimulus is far removed from

expected utility theory; standard economic theory and perceptual psychology are quite different (Gold and Heekeren, 2014). Although randomness is a concept known to economists, the theory of revealed preference assumes that there is a stable determinist choice correspondence from choices to observable behaviors. Yet, taking up an idea set forth by Luce (1959/2005), McFadden (1974) proposed that, like the percept curves of psychophysics, the utility curves of economics should be considered variable, or, said differently, that the very same tools used to study confusability in perceptual judgments should be brought to bear on "errors" in choice observed under economic conditions. This class of theory is now called "random utility models" in economics (see McFadden, 2005; Gul and Pesendorfer, 2005; Karni and Safra, 2016). The computational model arising from the neuroeconomics literature (see section 4) should be regarded as providing a neurobiological foundation for these random utility models (Fehr and Rangel, 2011; Glimcher, 2011a).

The subjective value of decision and choice

Economic theories of choice behavior have a cornerstone in the concept of value. While choosing, individuals are supposed to assign values to available options and a decision is then made by comparing these values. Thus, value represents a common unit of measure for making comparisons. Neuroeconomists seized this concept from economists, and in the past 10 years, considerable research has focused on neural representations of value and selection mechanisms of a decision by comparing the values. Like for economists, the notion of "subjective value" is a core concept for neuroeconomists. However, there exists a crucial difference. In economics, the concept of value is behavioral and analytical, nor psychological. The "as if" stance captures a fundamental limit: based on behavior alone, values cannot be measured independently of choice. The assertion that choices maximize values is intrinsically circular. Neuroeconomics breaks this circularity by establishing that values are computed in the brain. By showing correspondence between a neural signal and a behavioral measure of value, that signal in the brain provides an independent measure of value, in principle dissociable from choices. So, the assertion that choices maximize values becomes potentially falsifiable and thus truly scientific (Padoa-Schioppa, 2011).

In brief, today, for most neuroeconomics researchers, the ultimate goal assigned to this new scientific field is to understand more profoundly these VBDs by studying the neurobiological processes and cognitive mechanisms that implement human decisions. Neuroeconomics intends to discover the neurobiological and computational underpinning these kinds of decisions that we can identify with "economic behaviors" (Montague, 2007; Schultz, 2008; Rangel, Camerer, and Montague, 2008; Balleine, Daw, and Doherty, 2009; Kable and Glimcher, 2009; Fehr and Rangel, 2011; Padoa-Schioppa, 2011; Rangel and Clithero, 2014; Glimcher, 2014a; Padoa-Schioppa and Conen, 2017).

2. The Emo-Rational Brain

The dual-process framework: automatic versus controlled systems

When studying decision making, psychologists often use a valuable distinction between automatic processes (fast, specialized, rigid, intuitive, unconscious, and heuristics-based) and controlled processes (slightly slow, generic, flexible, deliberate, conscious, rule-based, and using high cognitive faculties, such as reasoning). It was indeed an old dichotomy. William James is one of the first psychologists who defended this view by the end of the 19th century (James, 1890). This general "dual-process" framework is a simplified and reductionist way for analyzing cerebral activities. It is always debated in cognitive psychology (see Melnikoff and Bargh, 2018 for critics, and Pennycook et al., 2018, for several arguments in favor of it). Its main merit is facilitating the understanding of many decision biases.

We find this dichotomy today in many theoretical analyses of modern behavioral economics that oppose two systems: one would depict a quasi-automatic or short-sight behavior while the other would reflect optimization. Different labels have been used to refer to this dichotomy, including "emotional" versus "deliberative" systems (Loewenstein and O'Donoghue, 2004), "automatic" versus "controlled" systems (Benhabib and Bisin, 2005), "short-term" versus "long-term (and controlled)" systems (Fudenberg and Lenine, 2006), and "warm (and automatic)" versus "cold" systems (Bernheim and Rangel, 2005). In the formal analysis built by Kahneman (2003)—maybe the most general one—these two systems are called "system 1" (intuitive system) and "system 2" (deliberative system), respectively. The main features of system 1 are its automatic operation and minimal demands on working memory, acting mostly through components of associative memory, while the main features of system 2 are the active engagement of working memory and analytic thinking (see also Stanovich and West, 2000; Sloman, 2002; Evans and Frankish, eds., 2009; Evans, 2010; Kahneman, 2011; Evans and Stanovich, 2013). Within the cognitive architecture, system 1 occupies a central position midway between the merely automatic functioning of perception and the merely deliberative functioning of system 2. The latter is similar to processes implicitly involved in standard microeconomic theory, whose many results were disproved by an increasing number of economic experiments since 1960s (for outlines of history of experimental economics, see Guala, 2008, Serra, 2012, Svorencik, 2015, Cot and Ferey, 2016). In addition, the primary goal of behavioral economics is to build new empirically more relevant models by integrating, in a formal way, some features of system 1 (for a recent overview of these new models of behavioral economics, see e.g. Serra, 2017, Chap 4).

From a neurobiological point of view, the rough distinction between *emotional* and *cognitive* systems is largely akin to the duality between *automatic* and *controlled* systems from a psychological point of view (Sanfey et al., 2006). The overview that Camerer,

Loewenstein, and Prelec (2005) proposed is expected to be more comprehensive; by supposing mutual independence between both systems (i.e. automatic versus controlled systems and emotional versus cognitive systems), these authors define by crossing four kinds of systems of which the only one corresponding to “controlled and cognitive system” may be identified to the standard economic model. But what modern neuroscience teaches is that in reality there exists a set of interactions among the four kinds of systems.

To deal with these interactions at the anatomical-functional level, the cognitive-emotional distinction is adopted for convenience. The study of neural learning mechanisms offers a more relevance dynamic framework for analyzing the links between automatic and controlled systems; this topic will be addressed later in section 4.

Cognitive systems

Brain, cortex, frontal cortex, and prefrontal cortex

Broadly speaking, the primate (and hence human) brain can be divided into four main divisions whose boundaries are based on converging evidence from developmental, genetic, physiological and anatomical sources: (1) the telencephalon (or forebrain), including mainly the cerebral cortex, basal ganglia, amygdala, and hippocampus; (2) the diencephalon, that corresponds essentially to the thalamus (around three-quarters of this part of the brain); (3) the mesencephalon (or midbrain), including the superior and inferior colliculus, substantia nigra, and tegmental area; (4) the brainstem (or hindbrain), including the pons and medulla. In addition, from a macroscopic viewpoint, cerebral hemispheres are divided into lobes: in the main, frontal, parietal, occipital, temporal lobes. These are not functional subdivisions but rather names of convenience. Until recently, the insula (or insular cortex) was considered an independent fifth lobe but it is now often included in the frontal lobe (Glimcher, 2014b). Other subdivisions are usually used: lobules (internal divisions in some lobes), sulcus (troughs in the cortex, the deepest ones are called fissure), gyrus (convolutions in lobes bounded by some sulcus). Furthermore, it should be noted that the terms used for specifying the position of the brain and its various regions are the same that those used for describing the macroscopic anatomy of the rest of the body: the terms “anterior” and “posterior” denote the front and back, “dorsal” and “ventral” denote upper and lower body, “medial” and “lateral” denotes the center and the sides, “rostral” and “caudal” denoting the head or tail’s direction (these qualifiers will be frequently used in the paper).

Among the many brain divisions of the cerebral cortex based on histological criteria that were evolved, the map including 52 areas built by the German neuroanatomist Korbinian Brodmann in the early 20th century is the most well-known (Brodmann, 1909). Yet, this cytoarchitectonic organization (each area being characterized by a specific cellular organization) faces a problem related to heterogeneity of data from different organisms,

which makes the comparison of their neural activities difficult. The standard proportional stereotaxic space, a method conceived by the French neurosurgeon Jean Talairach in the sixties but published in English about twenty years later (Talairac and Tournoux, 1988), allows responding to this problem. Frequently used in brain imaging experiments, the method presents three advantages: first, it allows a more thine computing identification of all brain points in a tridimensional geometric system (x, y, z), where x denotes the axis left-right, y the axis anterior-posterior, and z the axis superior-inferior; second, it offers a spatial transformation in order for drawing up a correspondence among different brains; third; it suggests an atlas describing a "standard" brain anatomically and cytoarchitectonically[5]. Most brain localization studies in neuroscience are based on this idea of a standard brain. Yet we need to be aware that the discovery of "neuroplasticity" (i.e. the fact that experience directly changes the brain throughout an individual's life) weakens the reliability of this static view (for a short overview see Schutt, Seidman, and Keshavan, eds., 2015, Chap. 2).

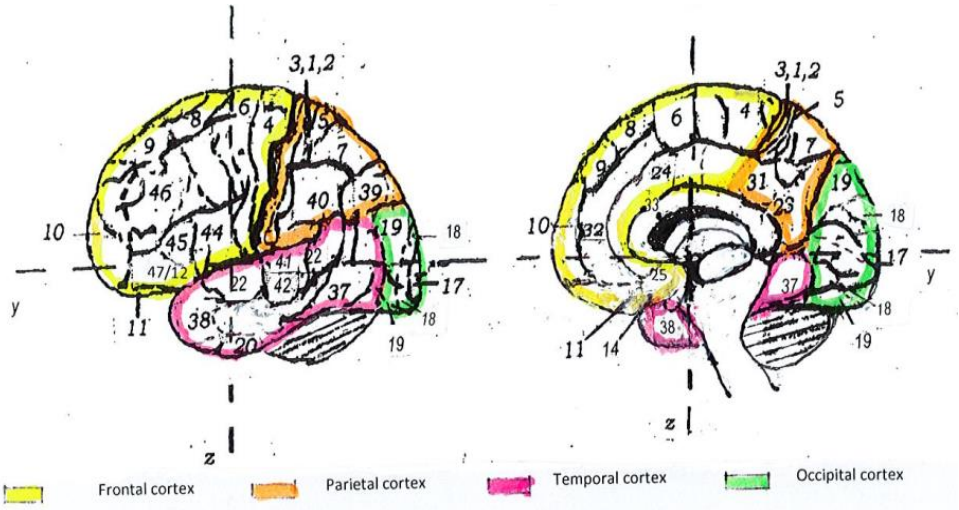
Traditionally, "association areas" in the cortex are considered as the cognitive abilities center. Association areas fill the greatest part of human brain surface (maybe about four-fifths), the other part corresponds to sensorial and motor cortices, which encode sensory information (mechanical or thermal sensitivity, vision, audition, sense of smell, taste) as well as movements control (see Fuster, 2008 or Purves et al., 2011, Chap. 25). For example, some Brodmann areas (BA thereafter) in the occipital cortex intervene for the vision (BA 17, 18, 19), some areas in the parietal cortex (in the back of the central sulcus) correspond to the primary somatosensory cortex (BA 3, 1, 2), some areas in the frontal cortex (in the front of the central sulcus) are specialized with motor activity (primary and secondary motor areas: BA 4, 6) whereas others areas in the temporal cortex deal with audition (BA 41, 42)[6] (see Figure 1 below for the map of the main cortical areas).

All mammalians have a frontal cortex; however, its relative size in the brain varies widely with species. While in non-human primates, such as monkeys or lemurs, its size is enormous compared with other species (e.g. rats or hedgehogs), in humans, the frontal cortex is simply gigantic (a third of the cerebral volume). Moreover, the metabolic cost of this additional cerebral tissue is very high; more than 20 % of what we eat is spent in feeding our brain, although the brain accounts for only 4 % of our body weight, and much of the fuel is utilized to drive the frontal cortex. Despite this human singularity, our information about frontal cortex connectivity in humans has been essentially derived from studies in monkeys (see the detailed studies of cytoarchitectonic comparisons between the frontal cortex of humans and monkeys published in Petrides and Pandya, 1994, Ongür and Price, 2000, Ongür, Ferry, and

[5] Today, neuro-imagery studies use more frequently the Montreal Neurological Institute (MNI) space, which slightly differs from Talairach-Tournoux normalization by relying on a highly number of fMRI images (see Poldrack, Mumfort, and Nichols, 2011, Chap. 4).
[6] Notice that the different neural regions referred to in the text often include only a part of the BAs mentioned in bracket.

Price, 2003). Although some studies suggest a strong similarity in this respect between humans and monkeys, caution must be executed when directly correlating the results obtained from non-human primates, among other mammals, with humans (Dehaene et al., 2005); accordingly, monkeys have brain structures that rodents lack, and humans have brain structures that both monkeys and rodents lack. This restricts the conclusions we could draw about the human brain from studies of animals with much smaller cortices. For instance, what is known as the orbitofrontal cortex—the part of the frontal cortex just above the orbits—differs in humans and monkeys, suggesting that this frontal region plays a specific role in humans (Glimcher, 2011a, p. 310 and 358). The localization of neural areas in humans similar to those in monkeys remains to be studied in the neuroanatomy research agenda (see in particular Mackey and Petrides, 2010, 2014; Wallis, 2012).

Figure 1. **Brodmann's cytoarchitectonic map of the human cerebral cortex**[7]



Lateral view *Medial view*

The human frontal cortex plays a fundamental function. It collects complex perceptive information from the sensory and motor cortices as well as the parietal and temporal associate cortices. This results in an evaluation of individual relationships with the world that allows oneself to experience properly planned and fulfilled behaviors. The frontal cortex is view as the "executive" region of the brain (Pribram, 1973)[8]. In front of the motor and premotor areas in the frontal cortex stands the prefrontal cortex—a wide cortical region involved in cognitive processes.

[7] BA 13 and 14 were added by Beck (1949). BA 47/12, added by Petrides and Pandya (1994) for reconciling human and monkey brain architectonic organization, corresponds to BA 12 in monkey. All these areas (like BA 10 and 11) are the subject of finer subdivisions (see e.g. Ongür, Ferry, and Price, 2003).

[8] While some cognitive processes are automatic (such as visual perception or language), most of them are controlled: they need some attention and so imply subject consciousness.

Cortical regions involved in cognitive processes

The prime importance of the prefrontal cortex (PFC) in highly intellectual functions has always been recognized. Yet, knowing if these various functions are performed by well-defined specific zones has been a matter of debate for a long time. Today, neuroscientists agree that a certain specialization exists (a partial one, at any rate), even if there is a better understanding of the neural mechanisms sending back to loops or networks (I will return below to this important point). In this respect, it is generally agreed that highly complex cognitive processes (i.e. reasoning, planning, problem-solving, decision-making) strongly involve some anterior and lateral zones in PFC, namely the anterior cortex (BA 10)[9] and dorsolateral PFC (BA 8, 9, 46)[10].

However, nowadays it is recognized that other neural areas that are engaged in high-level cognitive processes are in the more posterior cortical zones, including the anterior cingulate cortex (ACC) (BA 24, 32, 25)[11], posterior cingulate cortex (PCC) (BA 23, 31)[12], temporo-parietal junction (TPJ)[13], and posterior parietal cortex (PPC) (BA 7, 40) (Miller, 2000; Miller and Cohen, 2001; Semendeferi et al., 2001; Ramnani and Owen, 2004; Zald and Rauch, eds., 2006; Fuster, 2008; Passingham and Wise, 2012).

Using a comprehensive battery of neuropsychological tasks on a large set of individuals with damage to the frontal lobes, Gläscher et al. (2012) suggest that it make sense to speak of a "cognitive control network" in the brain, including primarily the dorsolateral PFC and the ACC, that is thought to draw on multiple processes, such as task switching, response inhibition, error detection and conflict monitoring, and working memory. Figure 2 indicates the approximate locations of several regions involved in cognition or/and emotion processing and provides a qualitative depiction of the main brain regions of interest to the neuroscientists and mentioned in the text).

[9] The anterior cortex (or frontopolar cortex) is the most rostral zone of the frontal lobe. It performs a function of cognitive control in the most complex situations; it is involved to monitor completely unknown situations or forcing the subject to think about one's own thoughts (i.e. metacognition).

[10] The dorsolateral PFC corresponds to the superior part of the frontal lobe exterior. It is seen as the most "rational" part of the brain.

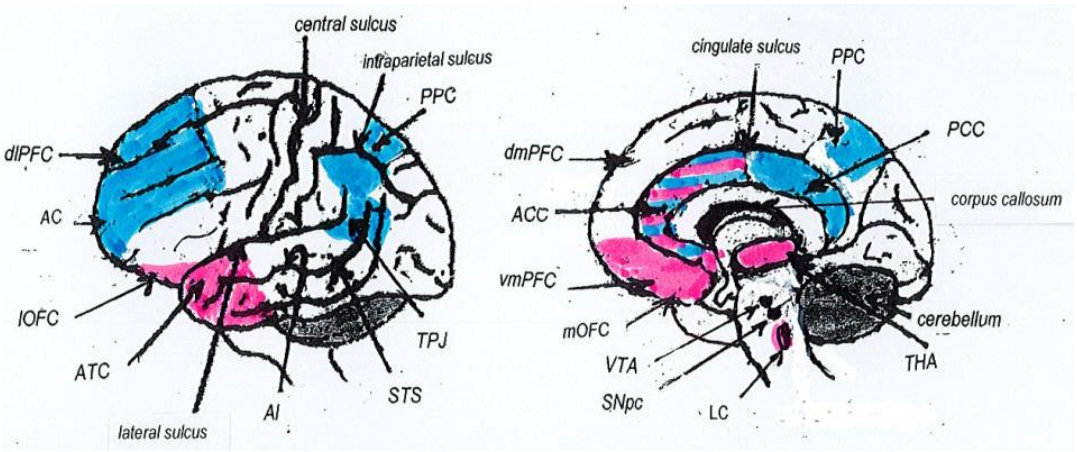
[11] The cingulate cortex is an internal zone located along the interhemispheric fissure above the corpus callosum. It is divided into an anterior (ACC) and a posterior (PCC) parts. The ACC has long been known to play a role in decision making, especially when subjects made errors in simple decision-making tasks and detected those errors. It is traditionally known as mainly implicated in the monitoring of internal conflicts, namely when conflicting signals are sent by several neural areas and that selection of an action may be tricky.

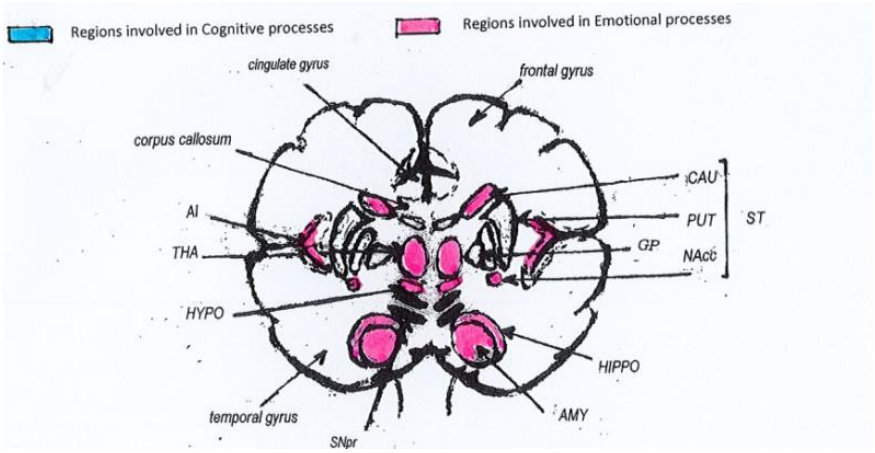
[12] The PCC is typically known as devoted to several high-level cognitive functions, including attention, working memory, and more broadly, "external consciousness", but its ventral part seems to show a functional integration with the whole areas belonging to the cerebral "default mode" (i.e. the brain's intrinsic activity when it is undertaking no task whatsoever); this network is supposed to accommodate what some authors called "internal subjective consciousness" (Buckner et al., 2008; Demertzi et al., 2013; Raichle, 2015).

[13] The TPJ is the part of the temporal cortex at the edge of the parietal cortex.

Figure 2. **Main neural regions involved in cognition and/or emotion processing**

Lateral view *Medial view*





Internal structures (axial plan)

AC: anterior cortex, dlPFC : dorsolateral PFC, dmPFC :dorsomedial PFC, vmPFC : ventromedial PFC, lOFC : lateral orbitofrontal cortex, ACC: anterior cingulate cortex, PCC: posterior cingulate cortex, AI: anterior insula, ATC: anterior temporal cortex (temporal pole), PPC: posterior parietal cortex, STS: superior temporal sulcus, TPJ: temporoparietal junction, THA: thalamus, HYPO: hypothalamus, ST: striatum, NAcc: nucleus accumbens, CAU: caudate nucleus, PUT: putamen, GP: globus pallidus, AMY: amygdala, HIPPO: hippocampus, SNpc: substantia nigra pars compacta, SNpr: substantia nigra pars reticula, VTA: ventral tegmental area, LP: locus coerulus

Emotional systems

From a neurobiological viewpoint, emotions and affect expression are closely associated with the autonomous nervous system. The centers coordinating emotional responses have been historically pooled as the “limbic system”.

The limbic system

Analysis of emotional behavior control systems has a long history marked by Papez and his identification of an emotional circuit that later became known as the “Papez circuit” (Papez, 1937). Paul MacLean contributed to the well-known three-part brain architecture (MacLean, 1970): (1) the reptilian brain, the oldest cerebral structure in terms of evolution

(i.e. the basal ganglia[14]), is seen as the seat of primitive emotions (as fear or aggressiveness); (2) the "old" mammalian brain (originally called the "visceral brain") broadens the set of emotional responses by including social emotion (as guilt, shame or envy), and corresponds to many of the components of Papez circuit (the thalamus, hypothalamus, hippocampus, and cingulate cortex) and additional important structures, such as the amygdala[15]; (3) the "new" mammalian brain (i.e. the neocortex) interfaces emotion with cognition and exerts top-down control over the emotional responses driven by other systems. The term "limbic system", introduced by MacLean (1952) for the "visceral brain", survives today as the dominant conceptualization of the "emotional brain", even though over the years its configuration has evolved by including some cortical areas (see Dalgleish, 2004).

It is commonly recognized today in "affective" neuroscience[16] that in addition to the many subcortical structures involved in emotional responses (the amygdala, hippocampus, thalamus, hypothalamus, and some structures in the basal ganglia, such as the caudate and nucleus accumbens in the ventral striatum), several other cortical zones are viewed as engaged in emotion processing: the orbitofrontal cortex (OFC) (BA 11, 14, 13, 47/12), ventromedial PFC (10, 11, 14, 32)[17], and anterior parts of insula[18], cingulate cortex (ACC) (BA 24, 32, 25), and temporal cortex (BA 38). Both these subcortical structures (i.e. the "classic" limbic system) and cortical structures are now thought to be involved in emotion and mood expression and processing[19] (Dalgleish, 2004; LeDoux, 1996, 2007; Phelps and LeDoux, 2005; Whalen and Phelps, eds., 2009; Pessoa, 2010; Purves et al., 2011, Chap. 28; Lempert and Phelps, 2014; Engelman and Hare, 2018; Fox et al., eds., 2018).

---

[14] All vertebrates (fish, amphibians, reptiles, birds, and mammals) possess such a neural structure, of one form or another. It consists of a set of functionally diversified nuclei embedded in cerebral hemispheres depth, behind the frontal lobes and encircling the thalamus, including the striatum. The striatum includes itself three structures connected to different neural regions: the caudate nucleus, the putamen, and the nucleus accumbens (NAcc). They receive extensive inputs from the frontal cortex and send almost all of their outputs to two other nuclei in the basal ganglia, the globus pallidus and the substantia nigra pars reticula (see Figure 2). Today, many researchers simply divide the striatum into ventral versus dorsal sections, with the ventral striatum encompassing the NAcc and inferior part of the caudate and putamen (interacting with regions engaged mainly in emotion and motivation), while the dorsal striatum encompasses the superior parts of the caudate and putamen (interacting with regions implicated in movement and memory).

[15] The amygdala corresponds to a group of nuclei in the medial temporal lobe in front of the hippocampus. This structure plays a central place in emotion and motivational processing and is implied both in the emotional component of sensorial stimuli and emotional stimuli memorization. The hippocampus, with near structures with whom it is closely connected, is related to memory in general and spatial memory and is crucial for complex spatial representations; it is part of a "human navigation network" (Maguire et al., 1998).

[16] As already said, affective neuroscience is the field that studies the neural mechanisms of emotion. The term was coined by Jaak Panksepp in the late 1990s (Panksepp, 1998).

[17] In the wide orbitomedial region of the PFC (the region encompassing all internal and orbital neural areas), several specific zones are identified but not all researchers agree on their boundaries. By moving up from the zone located just above the orbits to the top of the skull, are typically defined the orbitofrontal cortex (OFC) (whose medial, caudal, and lateral parts are differentiated), ventromedial PFC and dorsomedial PFC (sometimes named globally medial PFC). The ventromedial PFC often is defined as including the medial OFC (see Zald and Rauch, eds., 2006; Wallis, 2012).

[18] The insula (or insular cortex) is a part of the cortex moved in depth of the lateral sulcus, at the junction between the frontal and temporal lobes. The insula is sometimes called the "paralimbic structure" (Moll et al., 2006). Its anterior part is strongly involved in emotion expressing: it is acting as a monitoring system that informs the brain about high-risk or unpleasant situations that may be a source of danger, harm, or pain. Some authors call this structure the "interoceptive" cortex because it is implicated in the processing of internal representations signals of body states (Craig, 2002; Singer and Tusche, 2014).

[19] It is known today that motor and sensory areas also are engaged during practically all emotions.

The cerebral "geography" of emotions

Whether we may identify a specialization of some neural zones in perception, expression, or experience of certain emotions, or whether all emotions are function of the same basic brain circuit has long been an open question (Dalgleish, 2004). Today, many argue in favor of specialization based on a litany of clinical studies dealing with patients suffering from brain damage or pathologies and, more recently, on brain imaging works (Damasio, 1994, 2003, 2010; Rolls 1999, 2007, 2014; LeDoux, 1996; Saarimäki et al., 2018). The idea that a link would exist between a body function and a brain zone dates from the 19th century to Franz Joseph Gall. Phrenology enjoyed popular success but was soon called into question before the great biologist Paul Broca proved some truth in Gall's work; by performing patient autopsies, he showed the existence of correlations between brain damage and neurological deficiencies. Nevertheless, even though some neural zones are devoted to specific functions[20], networks, circuits, and loops are engaged most of the time. Although the central principle for understanding representation in the brain remains as the notion of "modularity" (Fodor, 1983; Glimcher, 2014b), interpretation of network connectivity outweighs interpretation of brain localization, particularly for high-order complex functions (Fuster, 2009; Fuster and Bressler, 2012; Lindquist and Barrett, 2012; Fehr, 2013; Pessoa, 2017b).

Alternatively, there is a longstanding debate about whether measures of emotion organize themselves into categories or deploy in a more continuous way along affective dimensions. For some authors, emotion differentiation may arise according to some distinct and specific categories—fear, anger, disgust, happiness, sadness, surprise (i.e. the canonical "basic" or "primary" emotions). Many authors have enlarged the list of emotions including, but not limited to, shame, compassion, guilt, envy, contempt, discomfort, amusement, irony, satisfaction, excitation, and pride (Darwin, 1872; Plutchik, 1980; Ekmann, 1982, 2003; Panksepp, 1998). For others, subjective experience of emotions formulated by individuals may be described according to a main underlying dimension—the pleasant/unpleasant dimension (or positive/negative dimension). Another dimension is often involved—the awareness/depression dimension (or low/high activation dimension). So, an individual describes their emotional personal experience by saying they feel good or unwell (positive or negative emotion) but also aroused or quiet (high or low activation). It is critical to distinguish these conscious subjective experiences of emotions, often called "feelings", from emotions

[20] For instance, the Wernick area (a part of BA 22) and the Broca area (a part of BA 44 and 45) are involved in language. The first one is the region of language structuration and emission whereas the second one is the place where happen the syntactical coordination and implementation of linkage with the adjacent primary supplementary motor area.

as internal functional states (Scherer, 2005; Damasio, 2003, 2017; Adolphs, 2017; de Waal, 2018; Adolphs and Anderson, 2018)[21].

By crossing both approaches one gets a certain image of some neural zones specialization (Lindquist et al., 2016; Cowen and Keltner, 2017; Clark-Polner, Johnson, and Barrett, 2017; Barrett, 2017). Clearly, several distinguishable neural structures are associated with some negative emotions: the amygdala (fear, anxiety, aggressiveness, stress, sadness), hypothalamus (anger, fear, aggressiveness), anterior insula (disgust, distress, anger), locus coerulus[22] (fear, anxiety), and ACC (sadness, pain, anxiety). But the amygdala and the ACC are also activated in perception of some positive emotions. Several areas in the ventral striatum including the NAcc are strongly associated with positive emotions. And several neural areas in the wide orbitomedial region, such as the OFC and ventromedial PFC, are thought to be generally related to the psychological component of emotions.

Thus, a multitude of various emotions are represented in the brain in a distinguishable manner, yet in partly overlapping regions: the same region possibly plays a different role[23] (see Kragel and Labar, 2016, and Saarimäki et al., 2018, for recent reviews). Some scientists go farther by categorically refusing the modular ideas that underlie this typology and argue that human emotions result from interaction of broadly distributed functional networks (Lindquist and Barrett, 2012; Pessoa, 2013, 2017a). There would not be macroscopic brain structures dedicated specifically to emotions; instead, there would be specificity at the level of circuits and cell populations (Adolphs and Anderson, 2018).

Emotion and rationality interplay

Anatomic and physiological observations dealing with the modern system limbic conception teach that understanding of emotion neural supports also requires considering the role of the cerebral cortex. The amygdala and the PFC—with their connections in the striatum and the thalamus—are involved both in emotion processing and in complex cognitive processing that underlies "rational" decision. So, a main tenant of contemporary neuroscience is that, contrary to what was always generally assumed, emotion and rationality interplay. The current image of opposing "cold" regions in the frontoparietal cortex to "hot" regions in the classic limbic system regarding emotional state processing is no longer

[21] Psychologists distinguish another notion, "mood", considered as an affective state more diffuse, less intense but more durable than emotion. The term "affect" often is used as a generic term that involves both emotion and mood (Frijda, 1986; Scherer, 2005).

[22] The locus coerulus, located in the cerebral pons, is in close contact with the amygdala. It is associated with noradrenaline/norepinephrine, a chemical substance related to adrenaline considered as neurotransmitter; it is seen as active in waking, sleeping and feeding behavior but it also interplays with cortical regions for modulating attention (Aston-Jones and Cohen, 2005).

[23] However, several meta-analyses showed that often there are differences in response intensity of a same structure depending of the emotion: for instance, both fear and happiness active the amygdala but the activation level is significantly stronger with fear than with happiness, or both disgust and anger actives insula but the activation level is significantly stronger with disgust than with anger. Hemispheric lateral effects also were observed: for instance, the right amygdala is more involved in negative emotions and the left in positive.

really correct (Kelso and Engstrom, 2006; Pessoa, 2008, 2013; Koziol and Budding, 2009; Phelps, 2009; Lempert and Phelps, 2014; Richter, Shackman, and Okon-Singer, 2017; Okon-Singer et al. 2018)[24]. Several neuropsychology studies from the 1990s go even farther by pretending that emotion processing often would be *necessary* for making rational decisions.

Antonio Damasio was the first neurologist who established the essential role for emotions in rational decision making by taking as subjects several patients with brain damage in the ventromedial PFC (the cortical region known as mainly specialized in emotional signals processing)[25]. When they face a task consisting of getting rid of risky lotteries that seem attractive but harmful in the long run (the famous "Iowa Gambling task" experiment), his patients were incapable of adapting their behavior to choosing lotteries seemingly less attractive but profitable in the long term (Bechara et al., 1994). Initially, the game was intended for assessing the decision abilities of schizophrenics compared with healthy subjects. In their experiment, the authors recorded that the healthy subjects were drawing a card originally among the 4 decks proposed, and after 40 or 50 trials they were drawing only in the "advantageous" decks; conversely, schizophrenics or patients with damage to the ventromedial PFC continued to draw from all decks by focusing on immediate gains, seemingly indifferent to the whole game's result. Furthermore, measure of subjects' electrodermal responses showed that the observed behavioral deficiencies were accompanied by no specific reaction from the vegetative nervous system before making decisions (Bechara et al., 1996; Bechara et al., 1997). Studies on subjects with damage to the amygdala reached the same result (Bechara et al., 1999).

These experiments show that emotions are not always harmful to those which are feeling them[26]. Without emotional substrate, a rational choice is hard if not impossible. Higher cognitive abilities do not universally govern behavior. The "somatic marker" hypothesis (Damasio, 1994, 1996; Bechara and Damasio, 2005; Reiman and Bechara, 2010), where

[24] Indeed, a more comprehensive understanding of the cerebral functioning should require including motor activity too. Traditionally, coordination of movements was supposed filled out by subcortical structures, including the cerebellum. Today it is recognized that these structures also play a role in cognition and emotion (Koziol and Budding, 2009).

[25] It was back in 1994 that Damasio depicts for the first time the now famous history of this young American railway worker named Phineas Gage who, in 1948, was suffering a serious injury in the brain (a crowbar of 6 kg was going through his brain), an accident whose consequences, against all odds, were not physical but behavioral (for further details see Macmillan, 2000). Interested in pathological consequences of patients with frontal lobe lesions, Damasio had the opportunity to observe subjects like Gage: Elliot history, a patient suffering from a benign brain tumor, is now as famous as Gage history (Damasio, 1994).

[26] Of course, this is not to say that emotions are only beneficial effects for subjects. Damasio himself acknowledges that the participation of emotion to reasoning process may be advantageous or detrimental according to both the decision circumstances and the decision-maker's past history (see Engelman and Hare, 2018, for a recent overview of various emotional bias promoting specific harmful behavioral tendencies).

emotions should be physiological states before being psychological states for making good decisions easier, is widely confirmed in the literature[27].

Good decision making is often dependent on good emotion expression through neural circuits combining classic limbic system areas that are always seen as taking place in emotional and visceral behaviors, such as the amygdala and ventral striatum, with some areas in PFC seen as taking place in rational thinking. Interestingly, the results of this famous experiment can be interpreted in terms of regret. The most common formulation of regret is a result of counterfactual thinking where a realized outcome is compared with, "what might have been". Damasio remarked that his patients, even though they could not adopt an adaptive behavior, had a certain consciousness of better choices. Lacking any emotional feeling, they did not expect the regret of not choosing the good decision.

There are numerous neuroeconomic experiments dealing with the role of regret in decision that have gone further into this topic. For example, some fMRI studies have investigated brain activity involved in regret and disappointment—two negative emotions—by manipulating the feedback participants saw after deciding to choose a certain risk: full-feedback (regret: participant sees the outcomes from both the chosen and unchosen gamble) versus partial-feedback (disappointment: participant only sees the outcome from chosen gamble). These studies showed that the feeling of regret is associated with the dorsal ACC (BA 24), medial OFC (BA 11, 14), and anterior hippocampus, while disappointment activated middle temporal gyrus and dorsal brainstem. They also showed that both regret and disappointment activate the anterior insula, part of dorsomedial PFC (BA 8), and lateral OFC (BA 47/12), with activity stronger for regret. Thus, regret exerts a more substantial influence on choice than disappointment, and the emotional impact of regret is stronger than that of disappointment. The cortical differentiation between regret and disappointment helps to better understand their role in decision-making and how they differ in biasing choice (Camille et al*.,* 2004, Coricelli et al*.,* 2005; Chua et al., 2009; Giorgetta et al*.,* 2013)[28].

Now let us think about the famous metaphor imagined by Plato. The mind is seen as a chariot pulled by two horses. The rational brain is the charioteer; it holds the reins and chooses where the horses run. If the horses get out of control, the charioteer just needs to take out his whip and reassert his authority. One of the horses is well bred and well behaved, but even the best charioteer has difficulty controlling the other horse. According to Plato, this obstinate horse represents negative, destructive emotions. The job of the charioteer is to

[27] Over the years several studies have questioned the somatic marker hypothesis (Dunn, Dalgleish, and Lawrence, 2006). Nevertheless, this hypothesis has played a central role in affective neuroscience in that it was one of the first which links emotional responses and brain systems to behavioral decision patterns.

[28] The emotion of regret not only has an important impact in "normal" decision-making but has also been implicated in several clinical disorders such as schizophrenia, depression, obsessive-compulsive disorder, and "chasing" behavior in pathological gambling.

keep the dark horse from running wild and to keep both horses moving forward. With that simple metaphor, the mind was seen as conflicted, torn between reason and emotion. This dual division of the mind is one of the more enshrined ideas in Western culture. Lehrer (2009, Chap. 1) paints a vast fresco of Western thought, from René Descartes to Sigmund Freud, and including Francis Bacon, Auguste Comte, and Emmanuel Kant, a large set of influential philosophers who all stand for various forms of this duality—until the modern metaphor of the brain as a computer proposed by the cognitive psychology—for which feelings are seen as antagonists of rationality. Aristotle in *The Nicomachean Ethics* is viewed of as an exception by claiming that rationality was not always in conflict with emotion. One of the critical functions of rational thinking is checking that emotions are intelligently applied to the real world; the key to "cultivating virtue" was learning how to manage one's own passions. Another widely known exception is Spinoza, a contemporary of Descartes, who by pooling body and mind saw emotions and feelings as a central aspect of humanity[29].

Therefore, when the brain's "black box" is opened, we find that the horses (one symbolizing negative and destructive emotions) and charioteer (symbolizing rational brain) are dependent on each other. Where there is no emotion, reason will not exist. We can use the term "emo-rationality" for translating this complementarity between emotion and reason (Oullier, 2010). From an economic viewpoint, this is an essential finding. We cannot suppose that economic agents' rational decisions are free from any emotional interference. This is just what was confirmed by a deal of early neuroeconomic experiments within the "behavioral economics in the scanner" program, in the wake of behavioral economics.

A multitude of neuroeconomic experiments prove that, in cerebral regulation of behaviors, coordination between emotional and cognitive systems is necessary. A great number of arguments, in terms of psychology in behavioral economics, for explaining so-called "anomalies" or "biases" in decision making generated a meaningful explanation in more objective terms. If we distinguish "external" and "internal" assessment of rationality (i.e. evaluation of the optimality in achieving certain goals versus evaluation of the coherence of intentions, actions, and plans), it can be argued that neuroeconomic studies show how "externally" irrational choices can be "internally" rational. This teaching follows mainly from brain imaging experiments in very various contexts, such as purchasing consumer goods, financial decisions (i.e. risky and ambiguous decisions), intertemporal choices (decisions involving trade-offs among payoffs available at different points in time) or social decisions (interactive decisions between several individuals). Considering a more complete explanation of social decisions from a neuroeconomic perspective requires also taking account social

[29] Opposition between Descartes and Spinoza is highlighted by Antonio Damasio in the title he chose for two of his books: *Descartes' Error: Emotion, Reason, and the Human Brain* (1994) and *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain* (2003).

cognition processes such as “mentalizing” and empathy (see section 3) (see Serra, 2016, Chap. 4, for a recent comprehensive presentation of these various experiments).

Consider, as an example, two contexts where neuroeconomics provides more precise explanations than behavioral economics in individual decision-making. The first one is concerned with what happens inside the brain when a person makes typical consumer choices, such as buying an item in a retail store or choosing a cereal. In an fMRI experiment, Knutson et al. (2007) discovered that when subjects were first exposed to an object their nucleus accumbens (NAcc) was turned on; it is known that this neural area is strongly required for positive emotion and is taking part in the reward pathway (see section 4), so the intensity of its activation reflects desire for the item. Then, when the subject was exposed to the cost of the product, the anterior insula and the PFC were activated. The anterior insula produces “aversive” feelings, so people tend to try to avoid anything that makes their insula excited and this includes spending money. The PFC was activated because this rational area was computing the risk and reward. The PFC was most active when the item cost on display was significantly lower than normal. Importantly, by measuring the relative amount of activity in each brain zone, the authors could accurately predict the subject’s shopping decisions: if the negativity of the anterior insula exceeded the positive feelings generated by the NAcc, then the subject always chose not to buy the item, while if the NAcc was more active than the anterior insula or if the PFC was convinced that it had found a good deal, the object proved irresistible. It turns out that during many decisions, the rational PFC is largely a spectator, standing silently while the NAcc and anterior insula interact and argue with each other. So, whichever emotion we feel most intensively tends to dictate one’s shopping decisions[30].

Neuroeconomics also provides more precise explanations than behavioral economics for loss-aversion and ambiguity-aversion, two behaviors frequently observed in financial decisions. Instead of postulating abstract cognitive heuristics, neuroeconomics explains loss-aversion as the interaction of neural structures involved in the anticipation, encoding, and computation of the hedonic effect of a risky decision. The amygdala, shown to be involved in fear and emotional learning, encodes the emotional impact of the loss; the ventromedial PFC (BA 10, 11, 14, 32) predicts that a loss will result in each affective impact; the reward pathway computes the probability and magnitude of the loss (Naqvi, Shiv, and Bachara, 2006; Tom et al., 2007). Thus, subjects are loss-averse because they tend to have or already had a negative response to losses. Then, they expect a loss to occur, they anticipate their affective reaction. Similar studies also can shed light on ambiguity-aversion. In many experimental settings, subjects have a strong preference for risky prospects (i.e. those for

[30] This “shopping brain” model is implicitly at the root of several marketing practices in retail stores: put in the most prominent places the most covered items (exciting the NAcc), repeatedly assured that low prices are “guaranteed”, or that a certain item is on sale, or that it is getting the “wholesale price” (inhibiting the anterior insula).

which the probabilities are known) over ambiguous ones (i.e. those for which the probabilities are unknown). Neuroeconomic studies show that, although decision theory treats ambiguity as a special case of risky decision-making, ambiguous and risky decision-making are supported by two distinct mechanisms. In an fMRI experiment, Huettel et al (2006) showed that activation of a part of the dorsolateral PFC (BA 8, 9, 46) is correlated with ambiguity preference while activation of the posterior parietal cortex (PPC) (BA 40, 7) is correlated with risk preference; so, the dorsolateral PFC would play a specific role in the ambiguity resolving mechanism (i.e. when ambiguous situations are progressively transforming in risky situations). Activation of this neuronal zone involved in highest cognitive functions would reflect the operation of information searching for decreasing ambiguity, a situation that is negatively felt in emotional terms. Thus, ambiguity-aversion appears to happen because people have a stronger negative affective reaction to ambiguity than risk (see Platt and Huettel, 2008, and Taya, 2012, for a broader review of neuroeconomic experiments on ambiguous decisions).

3. The Social Brain

So far, we have ignored one striking characteristic of human societies from an evolutionary point of view—societies are based on work division and a large-scale cooperation between genetically heterogeneous people compared to most animal species where cooperation is restricted to smaller groups (Tomasello, 2000; Bowles and Gintis, 2011). Charles Darwin was the first scientist in the 19th century to reinforce the importance of these abilities to explain human dominance over other species (Darwin, 1859). Human brains have great flexibility and the ability to interpret complex forms of social interactions based on sophisticated beliefs. So, we may imagine that there are neural networks devoted to belief formation and updating. The "social brain" refers to this function and can be roughly identified with "social cognition and emotion" (Alos-Ferrer, 2018). This issue has been largely studied over the past few decades by researchers in social neuroscience—a new interdisciplinary field that has emerged from the union of classical cognitive neuroscience and social psychology[31].

At the source of these studies, there is a theory initially built in social psychology known as "theory of mind" (Premack and Woodruff, 1978) or "mentalizing" (Frith and Frith, 2003);

[31] In the beginning, much research in social neuroscience (the term was coined by Cacioppo and Berntson, 1992) has been driven by mental illnesses, because many of them often involve a breakdown of the "social brain" (in particular, schizophrenia). Remember that, likewise, the study of brain lesions has been a starting point for much of the early progress in neuroscience. Yet, in the last fifteen years, research in social neuroscience has increasingly focused on the social behavior of mentally healthy decision makers, encompassing many social phenomena as social interactions.

some authors also refer to this as "mind-reading" or "cognitive perspective-taking"[32]. Nevertheless, today it is known that, in addition to the mentalizing system, the mechanism for understanding other people's minds (i.e. social cognition) also involves recently identified nervous cells called "mirror neurons" and a sensory system linked to emotional character, the empathic system (some authors also refer to "emotional perspective taking") (for comprehensive reviews, see Cacioppo, Visser, and Pickett, eds., 2006; Decety and Cacioppo, 2011; Singer, 2012; Lieberman, 2012; Schutt, Seidman, and Keshavan, eds., 2015; Alos-Ferrer, 2018).

Even though the fields of social neuroscience and neuroeconomics are still perceived as two distinct fields, we have to agree that the topics they are concerned with overlap substantially both in content and methodology. Researchers in both fields are interested in understanding the nature of human social interaction and human decision making and aim to determine the neural mechanisms underlying these complex skills. Economic decision making, for example, frequently takes place in the context of social interactions and game theory—developed in economics and has come to provide an effective quantitative framework for studying how information, incentives, and social knowledge influence optimal strategies for social interaction (Singer and Tusche, 2014).

Theory of mind in social psychology

Theory of mind refers to the human cognitive capacity to distinguish mental states of self (beliefs, intentions, desires) from others' mental states and, thus, the ability to use this distinction for anticipating others' behavior and, in return, for their one's own behavior. More precisely, theory of mind addresses the process by which healthy human adults attribute unobservable mental states to other people (the "target") and integrate these attributed states into a single coherent model that can be used for explaining and predicting the target's behavior. This ability amounts to perceive the mind as a sort of "mental representations designer". Of course, these representations are not always necessarily correct. It should be emphasized that mentalizing amounts to a metacognitive ability—it enables an individual to represent one's own psychological representations and that of others, and implies the ability to simultaneously hold several world representations in the mind (Frith, 2012)[33].

[32] See Schlinger (2009) for a vast review about mentalizing and an appeal in favor of a behaviorist perspective in terms of Skinner's old works (1945, 1953).

[33] It was recognized that ability to mentalize is severely delayed in autism. That could explain observed failure in communication and social interaction by most autistic children. Today, the autistic brain is at the heart of social neuroscience because it helps to clarify the missing links between brain and social behavior (Baron-Cohen et al., 2000; Frith, 2001). Temple Grandin (an American professor in animal science) was one of the first high-functioning autistic woman (people with Asperger syndrome) whose brain was scanned by fMRI toward the end of the 1980s. Like Gage and Elliot cases, mentioned by Damasio (1994), Grandin case is become paradigmatic in cognitive neuroscience (Sacks, 1995).

The most stringent test for the presence of mentalizing would be to see whether someone is able to predict someone else's actions on the basis of that person's false belief ("false belief" test). Children age four starts to correctly attribute false beliefs to others and give verbal explanations when asked. At age five, over 90 % of children understand this task, and all do by at age seven (Frith and Frith, 2003, 2007; Hyde et al., 2018)[34].

Several theories were proposed to explain mentalizing. The "theory of simulation" seems to account for most results in social neuroscience, proposing that access to other people's mental states is carried out via a mental simulation process, by which "observer" individuals are taking the perspective of the "observed" individuals by seeking to imagine oneself in the situation or circumstances of another person or "put themselves in their shoes" (see Goldman, 2006; Decety and Grèzes, 2006). In social psychology, social cognition refers to two sorts of inference processes: (1) inferences based on transitory states (goals, intentions) and (2) inferences based on individuals' stable psychological traits (e.g. personality traits). Inferences based on transitory states relate to observed behaviors is of a perceptive nature, while inferences based on individuals' stable psychological traits driven by individuals' idiosyncratic traits would be more abstract and implies further developed mentalizing abilities (Van Overwalle, 2009). Most models studied in social psychology agree with the concept of these two processes. Among them, the "identification/attribution" model is undoubtedly one of the most popular (Trope and Gaunt, 2000); here, the observer individual identifies and categorizes the observed behavior to then attribute a psychological state to the observed individual. Therefore, the two-involved socio-cognitive processes are not mutually independent.

The mirror neuron system and mentalizing

Recent studies in social neuroscience do not infirm this type of model proposed in social psychology. The central nervous system would involve two sorts of complex neurocognitive systems implied in understanding of other people: (1) the "mirror neuron" system ("mirror system")—the term "mirror neuron" was first used by Gallese et al., 1996, and Rizzolatti et al., 1996a—and (2) the intrinsic mentalizing system. They correspond to two different levels of "mentalizing". A first level of mentalizing—an automatic, pre-conceptual and unconscious phase—would allow the fast identification of mental states (the mirror system), and a second level—a voluntary and conceptual phase—would provide the individual the ability to simulate others' mental states via one's own decision making system (the intrinsic mentalizing system) (Coricelli, 2005).

[34] The ability to mentalize is absent in monkeys, but is not an exclusively human trait. It is likely to be present, in varying degrees, in all species of apes (Povinelli and Bering, 2002; Call and Tomasello, 2008; Singer, 2009). Besides, it was recently shown that the great apes pass the classical false-belief test (Krupenye et al., 2016).

This literature on the “mirror system” draws on the conceptualization of the motor system established in non-human primates in the second half of 1980s. Instead of only being organized into three areas of the frontal cortex (i.e. the primary motor area, premotor cortex, and supplementary motor area), the motor cortex actually is formed by a constellation of different regions. It turns out that some areas in the parietal cortex not only receive outputs from sensorial zones but also have motor properties similar to that of the frontal cortex (for comprehensive reviews see Rizzolatti and Craighero, 2004; Rizzolatti and Sinigaglia, 2006; Rizzollatti, Fugassi, and Gallese, 2009).

Mirror neurons, which have visual-motor properties, were originally discovered in the motor frontoparietal cortex in macaque using electro-physiological studies (Di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996a) and then were identified in corresponding human neural structures using non-invasive tools (e.g. fMRI, positron emission tomography) (Fadiga et al., 1995; Rizzolatti et al., 1996b; Kilner et al., 2009; Mukamel et al., 2010). These neurons possess a singular property: they are activated when we are taking a particular goal-directed action as well we observe someone else taking the same action directed to a similar goal. Accordingly, one of the primary functions of mirror neurons is to understand the goals of the actions of the person being observed; here, a representation of this action is generated by the observing individual’s premotor cortex to resonate with the observed individual via an internal simulation process (Rizzollatti et al., 2001). Thus, the mirror system appears to quickly, almost anticipatory, identify other peoples’ intentions and affective states enabling an intuitive understanding of what the other is doing (Iacoboni et al., 2005). Being widely automatic and pre-conceptual, these phenomena of sensorimotor resonance do not imply the use of costly cognitive resources (Rizzolatti and Sinigaglia, 2006).

Studies of human mirror neurons have consistently shown the involvement of a dorsal frontoparietal network that links two large zones: (1) a zone located in front of the brain including the ventral precentral gyrus (BA 6) and the posterior inferior frontal gyrus (BA 44, the posterior part of Broca area) and (2) a zone located at the back of the brain formed by the inferior parietal lobule (BA 40). A third region, although it does not possess neurons with strictly speaking mirror property, in the posterior STS is often included in this network; this zone is involved in input for mirror system by providing high-level visual representations that spread up to the parietal cortex before being passed to premotor zones in the PFC (Rizzolatti and Sinigaglia, 2006; Molenberghs et al., 2012).

Recent meta-analyses about theory of mind are also largely based on many fMRI studies that have identified a set of neural regions involved when subjects are engaged in a mentalizing task. The network mainly consists of two zones: (1) a zone located in front of the brain including the dorsomedial PFC and extending up to the rostral ACC (known as the

paracingulate cortex) (BA 9, 46, 24) and (2) a set of neural areas in the temporal cortex including the temporal pole (ATP) and the posterior superior temporal sulcus (STS), extending up to the temporoparietal junction (TPJ) (BA 38, 42, 22) (Farrer and Frith, 2002; Gallagher and Frith, 2003; Amodio and Frith, 2006; Frith and Frith, 2006; Young, Dodell-Feder, and Saxe, 2010; Bzdok et al., 2012; Carter et al., 2012; Olson et al. 2013; Mahy, Moses, and Pfeiffer, 2014; Molenberghs et al., 2016). Farrer and Frith (2002) suggested that activation of the dorsomedial PFC would show that the neurocognitive mechanism devoted to mentalizing is limited to a well-defined part of the cortex. Correspondingly, activations observed in other zones would reflect interventions of non-specific mechanisms—activation in the STS could match the physiological counterpart for processes classically implied in biological movement perception, activation in the temporal poles could be interpreted in the larger context of retrieval in autobiographical memory, while implication of the TPJ could be conceived as a system that allows to distinguish one's own mental states from the others' mental states. Thus, the current consensus is that the mentalizing structure is built around frontotemporal pathways connecting frontal executive regions in the PFC to phylogenetically older regions in and near the temporal lobes.

Several additional results have to be mentioned. First, even though meta-analyses have identified central systems for mentalizing, great variability in the activation localization at the individual level has been observed (Singer and Tusche, 2014). Second, when during childhood mentalizing arises in the brain is still unknown. Yet, in a recent study using the emerging technique of near-infrared spectroscopy (Hyde et al., 2018), the TPJ, but no other temporal or prefrontal regions, was shown to have functional organization that is relevant to high-level social cognition by around seven months of age. Finally, it must be stressed that the "core network" of mentalizing is perhaps less broad than generally thought. In a meta-analysis examining fMRI data for several mentalizing task classes, Schurz et al. (2014) showed that there is a minimal active network common for all tasks including only the dorsomedial PFC and the bilateral TPJ but that general activation profiles are significantly different from each other according to the task class with supplementary activation of the inferior frontal gyrus and temporal poles. Comparable results were obtained by Spunt and Adolphs (2014) with a new "why/how" task designed to single out the basic idea of mentalizing. Indeed, what these recent studies show is that, so far, the literature on mentalizing has suffered from the lack of a standardized task.

For several years, it was generally thought that the two systems broadly involved in mentalizing—mirror system and intrinsic mentalizing system—were both anatomically and functionally completely independent (Van Overwalle and Baetens, 2009). Yet, it turns out that several studies deeply call this hypothesis into question because the mirror system and

intrinsic mentalizing system keep cooperative interactions during the accomplishment of social cognition tasks (Zaki and Oschner, 2012).

These findings have led to the development of new integrative models in social neuroscience that are similar to models for social psychology. In these models, the mirror system and the mentalizing system are supposed to interact when subjects are carrying out a social task while playing distinct roles. The mirror system would be involved in the mentalizing "identification" component (i.e. decoding of intentional and affective signals) whereas the intrinsic mentalizing system would be active in mentalizing "attribution" component (Sperdutti et al., 2014). The respective weight of the two systems in social situations understanding would depend both on the context and the task (Lombardo et al., 2010). More generally, "social flexibility"—our ability to understand a social situation and to fit it in the most effective way—would be fulfilled by the creation of transitory meta-systems expressing a real link between different networks (Cochi and al., 2013).

Of course, these social neuroscience findings about the human capacity to anticipate other people's intentions by imagining oneself in the situation or circumstances of these persons ("putting oneself in their shoes") provide an obvious advantage for economists. Besides, we immediately see the closeness between this ability and the one granted to players in game theory by economists assuming that players can think about the game from their own perspective just like the one adopted by their protagonists for predicting their actions. In game theory, it is necessary to develop a theory regarding other players' actions, and mentalizing could fit the bill (see Singer and Tusche, 2014). However, mentalizing is broader than what is usually assumed in game theory in which we only refer to a specific probability vector describing actions likelihood into a set of strategies. In theory of mind, this is referred to by individuals' general ability for mentally simulating other people's mode of reasoning (Hsu and Zhu, 2012).

Today, "social neuroeconomics" (the term was coined by Fehr and Camerer, 2007) parallels these studies by looking for neural foundations of other-regarding behaviors in studies employing well-established paradigms of experimental games (for an overview of these paradigms see e.g. Houser and McCabe, 2009, 2014). Many fMRI experiments have shown that when subjects are interacting with human partners rather than a computer partner the brain functions differently by activating regions involved in the mentalizing network. For instance, one of the first neuroeconomic studies (McCabe et al., 2001), in which economic Nobel Prize winner Vernon Smith participated, confirmed that trusting in the *trust game*[35] implies anticipating others' behavior and activating neural areas of the mentalizing

[35] Two participants are randomly and anonymously matched, one as *investor* (player I) and one as *trustee* (player T), and play a one-shot game. Both participants are endowed with an amount of money. Player I can send some, all or none of her endowment to player T. Every

network; the authors showed that a stronger activation of the medial PFC is observed in the brains of cooperative players when interacting with players localized in the lab rather than against a computer (see also King-Casas et al., 2005; Deldago, Franck, and Phelps, 2005). On their side, Rilling et al. (2004) examined subjects playing two other well-defined paradigms—the *ultimatum*[36] and the *prisoner's dilemma games*[37]—with both human and computer partners and observed stronger activation for human partners in the typical mentalizing areas. A similar finding was found in the *P-beauty contest game* (or *guessing game*) (Coricelli and Nagel, 2009)[38], which examines the subject's "strategic reasoning depth" in a framework that does not refer to any social or moral motivation (like previously mentioned games) but in which a psychological problem of coordination arises; this study showed that activation of the paracingulate cortex and superior temporal sulcus (STS) is strengthened when subjects are playing against human partners rather than against a computer. Not surprisingly, thus, the regions of interest in neuroeconomic studies using classical game-theoretic paradigms coincide partly with the mentalizing network[39]. Clearly, the mentalizing is a key function of the social brain for economics.

The system of empathy

Empathy is usually defined as the ability to share another person's emotions and feelings, following the definition given by the German philosopher Theodor Lipps at the very beginning of the 20th century (Lipps, 1903). This ability obviously is not without a relationship with the mentalizing system; the mirror system was proposed to be used as a neural scaffold for empathy and, from an evolutionary perspective, as a biological substrate for human evolution of sophisticated sociability and morality. As Lipps anticipated, in an fMRI study in which subjects observed or imitated emotions in facial expressions, Carr et al. (2003) showed that empathy is drawn from a special form of "resonance" implemented by the mirror system or

---

amount sent by player I is tripled. Player T observed the tripled amount send, and can send some, all or none of the tripled amount back to player I. The amount send by the investor is view as a measure of trust; the amount returned by the trustee is view as a measure of trustworthiness.

[36] Two participants are randomly and anonymously matched, one as *proposer* (player P) and one as r*esponder* (player R), and told they will play a one-shot game. Player P is endowed with an amount of money, and suggests a division of that amount between herself and player R. Player R observes the suggestion and then decides whether to accept or reject. If the division is accepted then both earn the amount implied by the player P's suggestion. If rejected, then both players earn nothing for the experiment. It is a simple take-it-or-leave-it bargaining environment.

[37] As is well known, Prisoner's Dilemma (PD) games are used to study "social dilemmas" that arise when the welfare of a group conflicts with the narrow self-interest of each individual group member. In a typical two-player PD, each player can choose either to *cooperate* or *defect*. Payoffs are symmetric, and chosen so that the sum of the payoffs is greatest when both choose to defect. However, each player earns the most if she chooses to defect when the other cooperate.

[38] In the simplest variant of the game, each player simultaneously chooses a number between 0 and 100. The person whose number is closest to 2/3 times the average of all chosen numbers wins a fixed amount of money; others receive noting; ties are broken randomly.

[39] Other games with very different logical structures are also concerned by this specificity of subjects' behavior when they know (or believe to know) that they are interacting with humans and not with computers. For instance, in one of the first experiments conducted with positron emission tomography (PET), Gallagher et al. (2002) showed that in the well-known *rock-paper-scissors game*, the paracingulate cortex was strongly more activated when subjects thought they were playing against another human player rather than against a computer (in reality they always were faced with random choices). For a review of recent neuroeconomic works dealing with strategic thinking, see Camerer and Hare (2014).

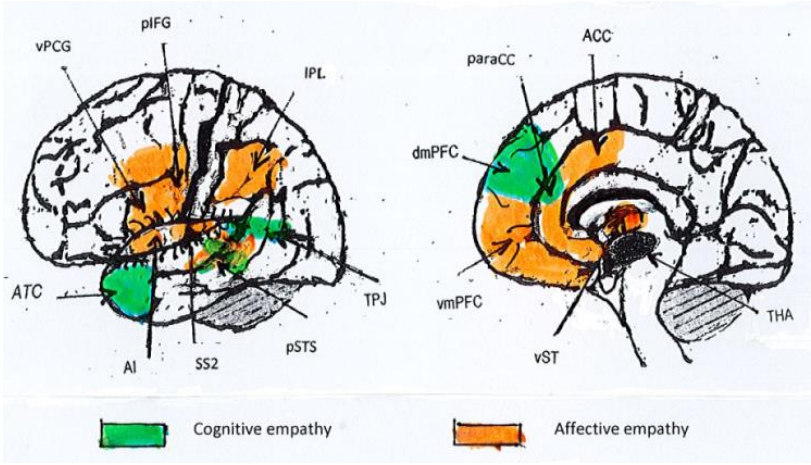
specifically on the *non-motor* side of another mirror system. This experiment proves that, besides the dorsal frontoparietal network that characterizes the mirror system, some regions into the limbic system also are activated; in parallel to the neural region devoted to tactile perception (i.e. the parietal cortex), regions linked to emotional perception are involved. For example, in experiments where subjects felt pleasure or received a reward while viewing images of people expressing disgust or suffering indicated activation of neural structures involved in disgust or suffering perception (the anterior insula), pleasure perception (the medial PFC), or the brain's "reward network" (the ventral striatum) (see section 4 for further detail). In parallel, somatic symptoms typically associated with these emotions (sweating, feeling of pressure, increased heart rate, and so on) were observed. In empathic situations, several regions involved in behavioral and sensorial perception regulation—the secondary somatosensory cortex and the medial ACC—were also activated (Singer and Tusche, 2014)

Since empathy engages the motivational and emotional brain, empathic motivation is a better predictor for engagement in other-regarding behavior than mentalizing. The psychopath exemplifies this; they may lack empathy but not cognitive perspective-taking, explaining why they can engage in antisocial behavior while being very good at manipulating and fooling other people—an ability that requires an understanding of other people's beliefs and intentions (Blair, 2005, 2008; Singer, 2009). Notice, however, that empathy is not always a direct avenue to moral behavior. Morality includes ideas such as justice, fairness, and rights, and comprises norms regarding how humans should treat one another. At times, empathy can interfere with morality by introducing partiality, for instance by favoring in-group members (Decety and Cowell, 2014).

In summary, in a somewhat reductive manner, we can acknowledge that the human brain has three partially distinct neural systems which provide the following abilities: (1) to understand others' motor intentions and actions (what other people *do*) (the mirror system), (2) to understand others' beliefs and thoughts (what other people *think*) (the mentalizing system), and (3) to understand and share others' feelings (what other people *feel*) (the system of empathy) (Frith and Singer, 2008; Singer, 2009; Singer and Tusche, 2014). Some authors, founding their analysis on empirical evidence, are making more explicit relationships between empathy, mentalizing, and the mirror system, suggesting to separate empathy into two distinct dimensions (Cox et al., 2011; Zaki and Ochsner, 2012): (1) "cognitive" empathy (*understanding* of other people's feelings) specifically in relation to the mentalizing system (Shamay-Tsoory, 2011a, 2011b) and (2) "affective" empathy (*sharing* of other people's feelings) specifically related to the mirror system (Schnell et al., 2011). Yet, during an empathic process, generally, the two systems are likely to maintain cooperative interactions

(Schnell et al., 2011; Zaki and Ochsner, 2012; Oliver et al., 2018)[40] (see Figure 3 below for a schematic representation of the brain areas typically involved in cognitive and affective empathy).

Figure 3. **Main neural regions involved in social cognition**



Lateral view *Medial view*

dmPFC: dorsomedial PFC, vmPCF: ventromedial PFC, ACC: anterior cingulate cortex, paraCC: paracingulate cortex, vPCG: ventro precentral gyrus (BA 6), pIFG: posterior inferior frontal gyrus (BA 44), AI: anterior insula, SS2: secondary somatosensory cortex, ATC: anterior temporal cortex (temporal pole), IPL: inferior parietal lobule (BA 40), pSTS: posterior superior temporal sulcus, TPJ: temporoparietal junction

Typically, this system of empathy is regarded as providing human's ability to adopt cooperative behaviors on a larger scale and explains the evolutionary success of the human species. However, in economics, the strategic anticipation of others' intentions have been included in game theory's conceptual background since the very beginning, and only recently has the emotional nature of interactive decisions been introduced by behavioral economists. Building on a large body of evidence that many people exhibit "social" preference (say, roughly, non-selfish or other-regarding preference), they proposed several models to explain observed behaviors inconsistent with standard game theory by referring to psychological factors, such as trust, altruism, reciprocity or inequity aversion (for a review, see Fehr and Schmidt, 2006; Serra, 2017, Chap. 4).

Social neuroeconomics, in a lot of recent studies, provides insights into the different neural pathways that lead to these prosocial decisions and reveal explanatory mechanisms for why humans deviate from the canonical self-interest model. Behaviors, such as altruism in *charitable donations game* (Moll et al., 2006; Harbaugh et al., 2007; Mayr, Harbaugh, and

[40] Some authors introduce additional distinctions. For example, Blomm (2016) adds to cognitive and affective empathy two other senses of empathy: "emotional contagion", understood as sharing the feelings of those in your immediate vicinity while for affective empathy others does not have to be present or even exist, and "compassion", "kindness" or "sympathy", that would replace affective empathy as a moral motivation. When one empathizes with another person, there does not have to be a prosocial motivation attached to it; when one sympathizes or shows compassion for another person, there is. However, in general, empathy is viewed as a first necessary step in the process that begins with affect sharing, which motivates other-related concern and finally engagement in helping behavior. Empathy and prosocial behavior are closely linked (Singer, 2009).

Tankersley, 2009; Hare et al., 2010) and mutual cooperation in *prisoner dilemma* or *public good games* (McCabe et al., 2001; Rilling et al. 2002; Rilling et al., 2004; Rilling et al., 2008) as well as direct punishment of "deviants" in *public good with punishment*[41] or *trust game* (de Quervain et al., 2004) or indirect punishment in *ultimatum games*[42] (Sanfey et al., 2003; Knoch et al., 2006b; Knoch et al., 2008; Baumgartener et al., 2011), would all be viewed as rational when accounting for the "emotional" utility or disutility that these behaviors incite and elicit neural traces observed in brain regions involved in emotion processing including the medial PFC, thalamus, NAcc, and anterior insula, depending on the specificity of each experiment.

Social neuroeconomics is the domain in which are found the largest number of experiments with pharmacological intervention, sometimes in combination with fMRI. To date, oxytocin is by far the chemical substance subject to the most experimental inquiries. Evidence has indicated that this neuroactive hormone is essential for prosocial behavior, particularly for trust (see Riedl and Javor, 2012, for a survey). Yet, even though oxytocin treatment has been suggested to have implications for at least patients suffering from disturbed social behavior, such as autism or social anxiety disorder, the relationship between oxytocin and human trust behavior is still an open question. The evidence that oxytocin directly influences trust behavior remains unclear, and, if there is an effect, it is likely moderated by a variety of factors (Nave, Camerer, and McCullough, 2015; Koppel et al., 2017). Several experiments have also studied the effects of nasal inhalation of oxytocin on, including but not limited to, generosity, betrayal aversion, empathy, positive and negative reciprocity, and inequality aversion (for a recent overview of some results about the effects on social behaviors of various chemical substances including oxytocin, vasopressin, testosterone, serotonin, and benzodiazepine, see Serra, 2016, Chap. III)[43].

In summary, in view of the evidence accumulated in recent decades by social neuroeconomists, it is now clear that neural networks dedicated to social cognition—both

[41] Recall that the structure of Public Good (PG) games is similar to that of Prisoner's Dilemma games, but they are typically played in larger groups. In a typical PG game, each member of a group of four people is allocated an amount of money, say 10 dollars. Group members simultaneously decide how to allocate their endowment between two "accounts", one private and one public. The private account returns one dollar to the subject for each dollar allocated to that account. In contrast, every dollar invested in the public account doubles, but is then split equally among the four group members (0.50 dollar each). Thus, like the PD game, group earnings are maximized at 80 dollars if everybody cooperates and contributes everything to the public account, in which case each of the four participants will earn 20 dollars. However, if three subjects contribute 10 dollars each, and the fourth free-rides and contribute nothing, then the free-rider will earn 25 dollars. Like the PD game, each group member has the private incentive to contribute nothing (free-riding). PG games with punishment are sequential PG games where players have the option to punish non-contributors and to reward the highest contributors after a round of the game. We know that the funding of public goods is a balancing act, both voluntary and involuntary mechanisms. In general, modern societies rely much more on taxation than on voluntary giving to provide public goods. However, for specific goods (e.g. the arts or some kinds of medical research) voluntary giving can be quite important. Charitable donation games allow to experimentally study altruistic giving in a PG framework.

[42] In ultimatum games, the act of rejection of the Proposer's offer by the Responder represents an act of costly punishment because both players suffer a cost.

[43] A more complete panorama of this neuropharmacology literature, that also includes the effects of chemical substances on time and risk preference, can be found in Crockett and Fehr (2014).

affective and cognitive empathy—are consistently recruited when people face social dilemmas and economic exchanges in cooperative and bargaining environments[44]. These neural networks act together or in competition with those dedicated to cognitive control (see above section 2) and reward processing (for reviews on social neuroeconomics see Fehr, 2009; Sanfey and Rilling, 2011; Rilling and Sanfey, 2011; Fehr and Krajbich, 2014; Declerck and Boone, 2016; Serra, 2016, Chap. 4; Engelmann and Fehr, 2017; Dreher and Tremblay, 2017, Part III).

4. The Computational Brain

Since the first cognitive science studies in the early 1950s, the brain has popularly been compared to a computer. Warren McCullogh, one of the founders of computational neuroscience, was the first to propose this comparison (McCullogh and Pitts, 1943; see also McCullogh, 1965). This comparison, although appropriate in many ways, disregards one critical functional constraint of the brain—it is a computer specially designed for making survival and reproduction easier (Glimcher, 2003; Montague, 2007). In this respect, the brain must value information subjectively received in the context of that final goal. This is the function of the "reward cerebral system" in all mammals. From an evolutionary perspective, this explains interest in experiments with animals (e.g. non-human primates or rodents) for understanding the function of the human brain (Santos and Platt, 2014).

A new paradigm for neuroscience: from the "stimulus-reflex" framework to the "reward learning" framework

The cerebral reward system is crucial for survival, providing the motivation necessary for making adapted behaviors (e.g. search for food, reproduction, danger avoidance) to enable preservation of the individual and species. Contemporary neuroscience research showed that this system is involved in three specific psychological components: (1) an *emotional* component, pleasure or pain caused by positive or negative reinforcements ("appetitive" or "aversive" stimuli respectively); (2) a *motivational* component, the motivation for getting reward or avoiding punishment (perception of the reinforcing by the sensorial organs or mental representation of this reinforcing, which trigger the motivation for getting reward or avoiding punishment); (3) a *cognitive* component, learning that combines reinforcement with emotional or behavioral responses. In the animal kingdom, survival and reproduction determines most rewards. In modern humans, the physiological model of threat to the

[44] It should be noted that social neuroscience literature covers a much broader thematic domain than questions of social cognition. A lot of studies concern what is called "moral dilemmas", which differ from "social dilemmas" by the fact that *all* solutions of a given problem generate a not morally desirable outcome (for example, the famous "trolley problem") (Christensen and Gomila, 2012).

species is secondary; threat happens when humans face a subjective constraining situation, such as complex decision making (Rolls, 2014).

The understanding of the reward system demonstrates a radical change in a neuroscience paradigm: actions are no longer supposed to be only governed by reflexes but also and mainly by *motivations and intentions* (Glimcher, 2003). Reflexes are fixed, stereotyped behaviors automatically elicited by specific types of stimuli (Sherrington, 1906). Such stimuli do not require learning over the lifetime of the organism to come to elicit such responses but rather have innate activity tendencies. These reflexes are behaviors that have been shaped over the course of evolutionary history because they provide an adaptive solution to environmental challenges. According to the "theory of reflex", the function of the nervous system consists of directly connecting a muscular response to a sensory stimulus so that a complex behavior can be broken down into a set of simpler behaviors. Neurophysiology sets the challenge describing the way the nervous system logically decomposes sensory signals and encodes motor commands. The theory of reflex is still present in neuroscience as a reference framework; however, it is acknowledged today that this theory is not relevant for explaining *all* behaviors (Glimcher, 2003, Chap. 4).

To date, three distinct learning systems have been identified and are dissociable by psychologic, neural, and computational terms. There are three different routes that enable organisms to draw on previous experience to make predictions about the world and to select behaviors appropriate to those predictions. Since these different sorts of predictions ultimately concern events relevant to biological fitness, such as rewards or punishments, they can also be thought of as different forms of value. These systems are: (1) *Pavlovian systems,* which learn to predict biologically significant events to trigger appropriate responses; (2) *habitual* systems, which value a great number of actions by a simple trial-and-error process, repeating previously successful actions; (3) *goal-directed systems,* which value actions on the basis of their anticipated impacts (Daw and O'Doherty, 2014) (see below in the section).

Although this theoretical turn in neuroscience was clear by the 1990s with the birth of electrophysiology and experimental study of neural activity in non-human primates, experimental work by some "neo-behavioral" psychologists with rats and pigeons carried the seeds for this revolution since the 1960s (Herrnstein, 1961; see also Herrnstein and Prelec, 1991). These biologically and evolutionary inspired works—going further into classic research of Skinner on conditioning (Skinner, 1953)—tested the hypothesis whereby reward frequency or amount is quantitatively changing learning and behavior (Commons, 2001). Hence, we observe the early stages of the convergence between reward expectation and expected utility of standard microeconomic theory, which were explicitly suggested several decades later by neurophysiologist Paul Glimcher. These studies, sometimes named the

"quantitative approach of behavior" (Richard Herrnstein, then Howard Rachlin, Georges Ainslie, George Loewenstein, Drazen Prelec), shed light on impulsivity and intertemporal choice, a question initially studied by Strotz (1955), but subsequently dropped by economists. Authors such as Ainslie, Loewenstein, and Prelec moved closer to economics progressively. Today, Loewenstein and Prelec are leading figures in neuroeconomics. The theoretical debt of this new discipline to these neo-behavioral psychologists is fully recognized by several pioneers of neuroeconomics (Glimcher, 2003). Importantly, there are significant differences between this branch of psychology, which is heir to the "behavioral school" born in the early 20th century in opposition to introspective psychology (Watson, 1913), and modern behavioral economics born in the 1980s fueled by studies in cognitive psychology, primarily Daniel Kahneman and Amos Tversky's works (see Heukelom, 2014, for a history of behavioral economics).

To analogize between reward expectation in neurobiology and expected utility in economics, Glimcher coined the term "physiological" expected utility (Glimcher, Dorris, and Bayer, 2005). With the new experimental tool of microelectrodes, which allows recording of an individual neuron's activity in conscious animals, experiments are no longer drawing on behavioral observations only but also on neural data. This is considerable progress compared to experiments made by neo-behaviorists. Observation of neural activity, to the extent that it gets *before* decision, makes possible the understanding of the involved mechanism—we can understand how the monkey, conditioned to make eye movements, is constructing and assigning "subjective value" to each option offered.

In the quest for rewards and to avoid punishments, how is the nervous system is organized for representing and valuing stimuli, making predictions about when and where they will be felt and using them to guide behaviors? Can these different aspects of the system impute by component or does the system have to work together as a whole? Are different sections of the system preferably activated when a subject is expecting reinforcement or when feeling it? Do rewards and punishments activate distinct components of the system or is it a question of activation intensity in the same areas? These are crucial questions that the decision neuroscience literature has begun to address and neuroeconomists are taking advantage of the first answers for improving predictions from economic choice models. Before outlining the main components of that model, I focus on a major discovery of contemporary neurobiology, namely the leading role of "dopamine" in the reward system and regulation of neural activity.

Dopamine and the reinforcement learning system

For several years, most neuroscientists suspected that dopamine—one of the main neurotransmitters—as well as the neurons using it (i.e. "dopaminergic" neurons) were playing

a decisive role in the reward system. Since the 1920s, it has been widely known that vertebrate neurons communicate with one another chemically—that neurons employ neurotransmitters to send signals across the synapses. Prior to 1950s, however, it was widely assumed that all neurons employed a single neurochemical to achieve this communication, the neurotransmitter acetylcholine. In the late 1950s and early 1960s, several neurochemists showed that this was incorrect; their studies revealed that highly localized clusters of cell bodies synthetized compounds including dopamine and serotonin, that these cells sent those compounds down their axons, and that these cells released those compounds from their terminals in response to depolarization (Doya, 2008). Dopaminergic neurons (DNs) are sometimes considered as "the retina of reward system", alluding to the central role played by this internal nervous membrane in the visual system (Tobler and Weber, 2014, p. 159).

Reward prediction error (RPE)

In the 1980s, the most frequent hypothesis was that DNs were the "pleasure center of the brain." Dopamine was supposed to play a crucial role in behavior by mediating reward and acting in the service of primary motivational characteristics (Wise, 1982). In the 1990s, this hypothesis was challenged by a set of experiments revealing the role that expectations were playing in modulation of dopamine activity too. While DNs are specialized in information transmission within the reward system, they are active if the reward is *unexpected*—these neurons do not respond to reward predicted by anterior stimuli. Thus, dopaminergic function rests not only on reward but also on beliefs. DNs do not react to reward per se; they implement a learning based on the difference between *prediction* (expected reward in relation to anterior reward) and *valuation* (reward really perceived). Learning is driven by a "reward prediction error." That error, the difference between expected reward and perceived reward, is used to continuously update value that the brain is attaching to stimuli (Houk, Adams, and Barto, 1995; Montague, Dayan, and Sejnowski, 1996; Schultz, Dayan, and Montague, 1997)[45].

These pioneer experiments revealing the role of dopamine in reward system were carried out in non-human primates. However, a similar mechanism was shown to exist in honeybees, which employ a close chemical homologue of dopamine called "octopamine" (Real, 1991; Montague et al*.,* 1995). As Glimcher points out, "the fact that the same basic

[45] It was well known for a long time that the dopamine circuit is crucially involved in motor functions and cognitive processes behind organization of motor behaviors; it was shown that motor difficulties affecting Parkinson's disease patients are caused by DNs degeneration. It's accidently that dopamine role in reward system was discovered in the 1950s by Olds and Milner (1954) in an experiment with rats. But it's really in the mid-1990s that, in an experiment designed to study the dopamine role in some movements, Wolfram Schultz found that dopaminergic responses are more closely associated to some "important" event, such as reward in the form of food. However, he thought unlikely that dopamine is encoding reward per se. Peter Dayan, the first one, established the link between the Schultz's discovery and reinforcing learning models evolved in computer science (Montague, 2007).

system occurs in species separated by something like 500 million years of evolution suggests how strongly evolution has conserved this mechanism" (Glimcher, 2011a, p. 302).

A slightly more formal presentation of this reinforcement learning system can be advanced. Assume that an organism maintains a set of predictions of the reward associated with each stimulus, *s*, called $V(s)$ (*i.e*. its subjective value), and also assume that these predictions determine the organism's conditioned response to whichever stimulus is observed. Then, upon observing stimulus received s_k on trial k, and receiving a reward on that trial, R_k, the prediction error is: $RPE_k = R_k - V_k(s_k)$. Dopamine is actively taking part in this system of reinforcement learning that is supposed to guide behavior: either there is a correspondence between the expected event and that which occurs—and the belief is strengthened—or there is no correspondence and a change in the anterior mental representation and a new learning arises. In this last case, the brain updates the prediction in the sense of the prediction error for reducing it. In the next trial, $k+1$, the reward prediction associated to stimulus s_k is: $V_{k+1}(s_k) = V_k(s_k)+\alpha RPE_k$, in which α stands for the learning rate parameter (between 0 and 1) determining the size of the update step. Understanding of that expression is easier if we transform it as follows: $V_{k+1}(s_k) = (1-\alpha)V_k(s_k)+\alpha R_k$. We see, thus, that the error-driven update accomplishes a weight average between the observed reward (with weight α) and the previous reward prediction (with weight $1-\alpha$). Finally, a further algebraic manipulation of this expression may allow us to realize that the prediction updating amounts to compute a weight running average of all rewards received previously in the presence of the stimulus, with the most recent reward weighted most heavily and the weight for prior rewards declining exponentially in their lag (Daw and Tobler, 2014).

These pioneering studies on the role of dopamine in reward learning date from the 1990s before the paper written by Platt and Glimcher (1999) that generally is accepted as the first neuroeconomics study. Although, before Platt and Glimcher's experiment, several studies by psychologists or neuroscientists wished them with economic influences (see section 2), it was the first study that put forward the hypothesis for a neural signal dealing with the anticipation of a reward (in that case a food reward) that was neither of sensorial origin nor linked to a motor command (later, Glimcher will use the term of "physiological expected utility", Glimcher, Dorris and Bayer, 2005). So, we may consider them as a kind of "theoretical prehistory" of neuroeconomics. Further they own a strong specificity—results exclusively concern modelling of brain functioning in *non-human primates* undertaking an "*action*" (most often, an eye-tracking task) in electrophysiology experiments.

If for a neurobiologist such an experimental context may be relevant, for an economist it must be expanded to human experiment with choices in a "good-based framework" (i.e. abstract decisions independent of the sensorimotor contingencies of choice) so that the neurons do not just encode movements. Theoretical models of decision making, indeed,

establish the advantages of separating processes related to selecting the good to acquire (i.e. “economic choice”) from those responsible for selecting the action necessary to acquire that good (Padoa-Schioppa, 2011). This is an important point on which I will come back further.

Thanks to cerebral imaging, the RPE paradigm was extended to humans in the early 2000s. Several experiments with gustatory or financial rewards generalized the theory, giving rise to the “neural economics” program (Delgado et al., 2000; Knutson et al., 2000; Elliot et al., 2000; Knutson et al., 2001; Berns et al., 2001; Breiter et al., 2001). Even if some cytoarchitectonic differences between the monkey brain and the human brain are observed, it appears that the cerebral processes are similar.

“Stochasticity and sequentiality” are two prominent properties of the choice mechanism in neural terms (Serra, 2016, p. 146). Uncertainty is fundamental to understanding brain functioning. Neural activity is stochastic by its very nature and, thus, the neural computations necessary for making choices are stochastic. Neuronal action potential rates are typically described as near Poisson-like stochastic processes. Although the precise source of this stochasticity in firing remains to be fully understood, the latest available data suggest that this stochasticity results from a mixture of thermal noise and that membrane voltages are driven by very small numbers of atomic-scale events (Glimcher, 2014b). Faced with this uncertainty, the brain would be able to make hypothesis and expectations, reinforcing or changing beliefs after checking, and continuously updating via various learning mechanisms that scholars have begun to identify (Niv and Montague, 2009; Balleine, Daw, and O’Doherty, 2009; Balleine and O’Doherty, 2010; Daw and Tobler, 2014). This is one of the main findings of “neural economics” that will be briefly outlined later in the section.

On the other hand, in the standard approach, economic choice is fundamentally choice between “actions”. Intuitively, it is a less than satisfactory hypothesis; we have a strong intuition that, for example, when consumers choose durable household goods like refrigerators or cars they do not rely on action-based decision making. Some authors have proposed an alternative to this standard approach called the “goods-based model”, which suggests that economic choice occurs within the space of goods and is computationally removed from sensory and motor representations. The key feature of this model is that economic choice fully takes place in the space of goods (Padoa-Schioppa and Assad, 2006, 2008; Padoa-Schioppa 2011). In any case, both classes of models suppose that representations of value based on “actions” and “economic choices” are closely linked. A serial process is generally postulated: the organism first values the potential “goods” in the environment and then recalculates those values based on the “actions” necessary to acquire them (Kable and Glimcher, 2009; Rangel and Hare, 2010). Importantly, however, these models differ on one critical point—according to the goods-based model, choice should be

completely processed within an abstract representation of goods. Knowing how goods-based valuation, which unarguably occurs, and action-based valuation interact in decision making remains, in fact, a controversial issue (see Glimcher, 2014a; Platt and Plassmann, 2014; Padoa-Schioppa and Conen, 2017).

Another central finding of "neural economics" refers to the existence of a "common currency" within the neural system that can be used to compare the valuation of diverse behavioral acts or sensory stimuli. For these scholars, who draw a parallel between economic systems and biological systems, "a currency is an abstract way to represent the value of a good or service… it provides a common scale to value fundamentally incommensurable stimuli and behavioral acts. Without internal currencies in the nervous system, an organism would be unable to assess the relative value of different events like drinking water, smelling food, scanning for predators, sitting quietly in the sun, and so forth" (Montague and Berns, 2002, p. 276). For choosing an appropriate behavior, the nervous system must estimate the value of each potential actions, it must convert them in a common scale and use this scale to determine a course of action; that common scale may be requested to value reward predictions as well as rewards themselves. It follows that to value a predictor, a neural system must have a way to compute the predictor's value *before* the reward that it promises actually arrives. And within this neural mechanism able to resolve in a *centralized way* the comparability problem among *a priori* non-comparable stimuli, it turns out that the dopamine production might be one of the essential components (Montague and Berns, 2002; Landreh and Bickle, 2008; Schultz, 2009; Balleine, Daw, and O'Doherty, 2009; Niv and Montague, 2009; Rangel, 2009; Chib et al., 2009; Levy and Glimcher, 2012).

Today, the "common currency" hypothesis is widely accepted in the neuroscientific community; however, some authors do not fully agree with it. Rolls (2007, Chap.3; 2014, Chap.9) has argued that different specific rewards must be represented "on the same scale" but not necessarily converted into a "common currency". The key difference between the two concepts of common scaling and common currency lies in the specificity with which rewards are represented at the level of *single* neurons. While a common currency view implies convergence of different types of reward onto the same neurons, a common scaling view implies that different rewards are represented by different neurons with the activity of the different neurons scaled to be in the same value range. Due to the limited resolution of the tool, fMRI studies cannot answer whether the same or different neurons are encoding the value of different rewards; only single neuron recording studies may provide such evidence (Grabenhorst and Rolls, 2011).

The dopamine circuit and the centralized regulation of neural activity

Today, there is vast literature dealing with the role of dopamine in the reinforcement learning process related to RPE and neural activity regulation (Pessiglione et al., 2006; Niv and Montague, 2009; Schultz, 2010, 2013; Cools, 2011; Glimcher, 2011a, 2011b; Daw and Tobler, 2014; Daw, 2014; Doya and Kimura, 2014; Rutledge et al., 2015; Schultz, 2016). However, the role of this neurotransmitter does not stop there; in addition to its central function in motor functions (as already mentioned), dopamine also acts as a powerful regulator in other aspects of cognitive cerebral functions, such as attention and working memory (Nieoullon, 2002; Arnsten, Wang, and Paspalas, 2012)[46]. In fact, even if it appears that a consensus exists in neuroscience about the multifunctional nature of dopamine and its major contribution in terms of motor performance, some disagreements persist with its role in learning and motivation[47]. Particularly, Rolls (2014, Chap. 6) agrees that there is evidence for DNs action in encoding of RPE signals and that this could present a problem; according to Rolls, the alternative hypothesis that DNs reflect the effects of many stimuli *salience* (i.e. a property less dependent to reward) is more consistent with experimental data. This is also explicit in the survey written by Berridge and O'Doherty (2014), in which each co-author has a slightly different point of view: for O'Doherty, dopamine is a prediction-error mechanism of reward learning, while for Berridge, dopamine mediates incentive salience. Indeed, there has been considerable debate over the role of dopamine activity in processing non-rewarding events (i.e. signals related to salient, surprising, and novel events). A series of studies provide evidence that DNs are more diverse than previously thought. Rather than encoding a single homogeneous motivational signal, they come in multiple types that encode both reward *and* non-reward events in different manners. Thus, these results pose a problem for general theories that identify dopamine with a single neural signal or motivational mechanism.

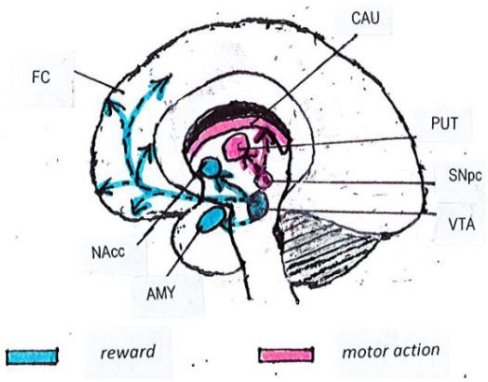
Whether with the human or monkey brain, neurobiologists identified the "dopamine circuit"—the brain zones in which neural information circulates mostly due to this neurotransmitter. Dopamine is synthesized in over a dozen locations in the brain, but only a few clusters of dopaminergic neurons sent axons along a long-distance trajectory that can influence brain activity in many areas. These long-distance projections have their origin in two zones lying within the midbrain and the high part of the brain stem—the ventral

[46] Attention allows for the voluntary processing of relevant over irrelevant inputs in line with the current behavioral goal of the organism. Working memory can be conceived as an active process whereby stimulus or internal representations are stored "on-line" to prevent temporal decay or intrusion from competing or distracting stimuli that are outside the current focus of attention. So, dissociating effects of attention from those of working memory is difficult, and in practice the two processes are interactive (Awh and Jonides, 2001). Dopamine is not the only neuromodulator implicated in attention; acetylcholine, noradrenaline and serotonin also play a role in top-down attentional control (for a recent review see Thiele and Bellgrove, 2018).

[47] The dopaminergic system also is a primary pharmacological target for psychiatric disorders such as attention deficit, hyperactivity disorder, schizophrenia, and Parkinson's disease, which are associated with attention deficits (Arnsten and Rubia, 2012).

tegmental area (VTA) and the substantia nigra pars compacta (SNpc) (Bjorklund and Dunnett, 2007). DNs irradiate in the brain by following two main networks: (1) from the VTA clusters up to the ventral striatum and the PFC, and (2) from the SNpc cluster up to the dorsal striatum (caudate nucleus and putamen). Functional experiments support the preferential link of VTA DNs to reward (the “reward network”) and SNpc DNs to motor action (Howe and Dombeck, 2016). In all, five regions from the PFC project to five zones in the basic ganglia, which in turn project via output neurons to the initial cortical regions. Some internal structures, such as the NAcc (in the ventral striatum), appear to be playing an important role in the reward network, but many other cortical and subcortical regions also may be involved (Glimcher, 2011a, Chap 13) (see a schematic representation of the dopamine circuit in Figure 4).

Figure 4. **The dopamine circuit**



SNpc: substantia nigra pars compacta, VTA: ventral tegmental area,
FC: frontal cortex, AMY: amygdale, NAcc: nucleus accumbens, CAU: caudate, PUT: putamen

In fact, all regions targeted by dopamine in the reward network do not verify the encoding conditions of RPE that might be stated (Caplin and Dean, 2008, 2009). Learning from the axiomatic method commonly employed by economists in modern theoretical economics, these authors wondered whether it is possible to identify a small number of axioms that might characterize all classes of RPE algorithms. They prove that three axioms are enough. In Caplin et al. (2010), these axioms were empirically tested with fMRI neural data from all zones targeted by dopamine. It turns out that only the NAcc verifies the three axioms. Then, some authors, such as Caplin and Glimcher (2014), suggest generalization of this approach in the field of neuroeconomics for building “axiomatic” neuroeconomics, which may have the advantage of solving some disagreements revealed by the literature.

It should also be noted that some areas belonging to the reward network are targeted by other “afferents”—specifically serotonin. Several scholars suggest that serotonin, producing feelings of satiety and inhibitions, is active in the field of *negative* RPE (Daw, Kakade, and

Dayan, 2002; Niv and Montague, 2009)[48]. As noted by Glimcher (2011a, p. 329), this hypothesis is rather attractive because it resolves the contradiction that exists between some experimental observations of the RPE model in the case of negative errors and many robust experimental economic results that prove existence of an asymmetry between losses and gains in terms of expectation. Schultz, Dayan, and Montague (1997), showed that negative prediction errors are less valued than positive ones by non-human primates, whereas a lot of experiments showed that both human and non-human primates are more sensitive to losses than to gains with respect to their expectations (Tversky and Kahneman, 1981, 1986, for humans; Chen, Lakshminarayanan, and Santos, 2006, for non-human primates). When dopamine is shown to no longer be the only transmitter that intervenes for negative errors, these results will no longer be conflicting. In any event, our understanding is still rudimentary and contested concerning encoding of negative RPE (Glimcher, 2011a, Chap. 13; Daw, 2014).

Learning and valuation: the three systems

To overcome uncertainty, the brain develops a learning mechanism. As mentioned above in this section, there is agreement that several learning processes coexist in the brain. More precisely, psychologists and neuroscientists nowadays agree that the brain employs three different systems (Bouton, 2007; Rangel, Camerer, and Montague, 2008; Balleine, Daw, and O'Doherty, 2009; Balleine and O'Doherty, 2010; Dolan and Dayan, 2013; Daw and O'Doherty, 2014; Rolls, 2014):

(1) *Pavlovian systems* are a category of learning mechanisms that only value a small number of behaviors soundly anchored in terms of evolution, acquired as responses to some positive or negative stimuli; they are independent of any actions. With this mechanism, an organism can learn to make predictions about when biologically significant events are likely to occur and to learn the stimuli tend to precede them (Pavlov, 1927). Pavlovian behaviors are more flexible than simple reflexes to emit behaviors shaped by predictive learning, but they also are inflexible since the responses themselves are stereotyped. Pavlovian learning is known to be present in vertebrates, including humans, as well as many invertebrates, including insects such as drosophila, and even in the sea-slug aplysia.

(2) *Habitual systems* can learn to value many actions with a simple trial-and-error process. Such a learning mechanism works simply by repeating actions that were previously successful. However, it has an odd and sometimes maladaptive inflexibility owing to its

[48] Broadly, serotonin is implicated in a variety of motor, cognitive and affective functions, such as locomotion, sleep-wake cycles and mood disorders. It was argued that this neurotransmitter would play a role in impulsive behaviors: reduced levels of serotonin would promote impulsive actions (i.e. the failure to suppress inappropriate actions) and choices (i.e. the choice of small immediate rewards over larger delayed rewards) (Dalley, Everitt, and Robbins, 2011).

foundation in the stimuli-response reflex. Habitual systems are relevant only for stable enough environments, without sudden change, and at slow rates.

(3) *Goal-directed systems* value actions on the base of their anticipated impact. This learning mechanism evaluates actions more prospectively, as with a cognitive map. It depends on a representation of the action-outcome contingency and on the outcome as a desired goal or incentive. For these reasons, goal-directed systems can calculate the value of an action in wholly new situations and in quickly changing environments, but they are generally associated with high informational cost.

Each system evaluates decisions and behaviors in a specific way—they correspond with *values* of specific nature. The main criterion for distinguishing between these various learning processes refers to the *intentional* nature of actions (Balleine, Daw, and O'Doherty, 2009). The test rests both on contingency and on control of actions by knowledge of their relationship to consequences. Goal-directed decisions pass both factors while habitual decisions pass none of them. In habitual systems, the only previously "reinforced" behaviors are valued. Concerning Pavlovian decisions, they pass no factor too. Learning is merely passive—behaviors are independent of their outcomes. However, the same is not true of behaviors acquired in other conditioning preparations, notably *instrumenta*l conditioning (Thorndike, 1911). In Pavlovian conditioning, the delivery of a biologically important event (e.g. food, water, and a predator) is predicted by, or made conditional upon, a relatively neutral stimulus, and then the reflexive unconditional response induced by exposure to the event is elicited by the stimulus. Let us think of Pavlov's popular experiment in which he studied salivary responses to food delivery in dogs and the conditioning of those reflexes produced by pairing a neutral stimulus (i.e. the sound of a bell) with food. For instrumental conditioning, access to a motivationally valuable commodity (e.g. food when hungry and safety when apprehensive) is made conditional on the performance of the animal's own actions, such as pressing a lever or pulling a chain. Here, learning is based on stimulus-response or response-outcome relationships rather than on direct stimulus-outcome relationships in which learning becomes active (e.g. how to teach pressing such or such lever to the animal). However, there is now considerable evidence confirming that goal values and instrumentally conditioned reflex values are mediated by distinct neural processes (Balleine, Daw, and O'Doherty, 2009).

Broadly, we may consider that Pavlovian and habitual systems correspond to two distinct classes of automatic (or quasi-automatic) behaviors for which learning is fast, whereas goal-directed systems are associated with the class of controlled behaviors for which learning is slower. For automatic behaviors, decisions are mainly of an instinctive nature, while with goal-directed systems, they would be made more consciously.

Nowadays, many questions persist. First, we do not know whether only three kinds of learning systems exist—Lengyel and Dayan, 2007, advance the hypothesis of a fourth "episodic" system managed by the hippocampus—or whether there exist multiple Pavlovian, habitual, and goal-directed systems, each one specialized in a class of actions or outcomes (Rangel, Camerer, and Montague, 2008). Second, the question of status assigned to each system in terms of state of consciousness is still open. Although Rolls (2014, Chap. 10) suggests linking only goal-directed systems to consciousness, which he calls "explicit" systems in contrast to the other systems called "implicit" systems, some cognitive neuroscientists do not exclude that habitual systems should be compatible with some aspects of consciousness (Dehaene et al., 2006; Dehaene and Changeux, 2011). Finally, knowing why the brain needs these different regulatory mechanisms and how the trade-off is conducted when the systems come into conflict has always been a matter of debate. Some elements of response have been proposed for explaining complementarity and not conflict between habitual and goal-directed systems depending on the circumstances each one is able to efficiently perform the same function—that is, to minimize error risks (Daw, Niv, and Dayan, 2005; Daw and O'Doherty, 2014; Rolls, 2014, Chap. 9). In this respect, we find, from the study of neural mechanisms involved in the treatment of uncertainty, a general qualitative conclusion that converges to the one I highlighted in the anatomical-functional identification of neural regions specialized in emotional and cognitive treatments (see section 2). Like emotional and cognitive systems that interfere in decision making, automatic and controlled processes cooperate to regulate behaviors "for the best" according to the circumstances[49].

From a neural viewpoint, Pavlovian, habitual, and goal-directed systems are implemented in dissociable substrates, each involving loops through different parts of the basal ganglia. Habitual actions are encoded in loops involving sensorimotor cortical inputs to the dorsolateral striatum. A parallel circuit linking the medial PFC and the dorsomedial striatum appears to support goal-directed behavior. Finally, Pavlovian responses appear to involve a ventral loop linking the OFC and the ventral striatum, with important contributions also from the central and lateral nuclei of the amygdala. All three loops are innervated by dopaminergic inputs from the VTA and the SNpc (Daw and O'Doherty, 2014)[50].

[49] In Pavlovian systems, the conditioned response is supposed to be proportional to the predicted reward. How these responses coordinate (competition or cooperation) with conditioned answers provided by model-free or model-based RL? Dayan et al. (2006) have proposed some hypothesis but we still do not have works that could rationalize principles of interaction similar to those explaining the efficient trade-off between habitual and goal-directed systems (Daw and O'Doherty, 2014).

[50] It was argued that degeneration of cortico-striatal circuits underlying these distinct learning systems result in different forms of pathology such as Huntington's disease and Tourette's syndrome for goal-directed systems and Parkinson's disease for habitual systems. Furthermore, interaction between these two learning systems, and particularly the situation when habits come to dominate behavior, has become a topic of great interest in neuropsychology of addiction and others psychiatric disorders involving compulsive behaviors, such as obsessive compulsive disorder (Maia and Frank, 2011; Daw and O'Doherty, 2014).

Reinforcement learning and computational algorithms

Reinforcement learning has been primarily developed within computer science. Two waste classes of algorithms were defined to identify optimal decisions in formal abstract tasks known as *model-free* and *model-based* reinforcement learning (RL) (Barto, 1995; Sutton and Barto, 1998; Dayan, 2008). Incidentally, the expression "reinforcement learning"—also named "routine" learning—may be confusing for economists in that it is not used in the same sense in computer science and neuroscience as in economics. In computer science and neuroscience, RL refers broadly to learning in the context of decision problems and comprises many sorts of learning including both model-free and model-based approaches. In contrast, in economics RL refers more specifically to only one approach, the model-free strategy (Daw, 2014, p. 302). See in Camerer (2003, Chap. 6) or Montet and Serra (2003, Chap. 7) for an overview of learning models in game theory in that it does not assume sophisticated behavior by players who are supposedly concerned only with their recent experiences (success or failure of their own followed strategy) and completely ignoring information they could infer from success or failure of other players' strategies.

Importantly, this implies correspondence between these two classes of algorithms and some aspects of real-word decisions by biological organisms. More precisely, there is a proposed link for habitual values to model-free RL and goal-directed values to model-based RL. Although model-free RL has received the majority of attention in neuroscience, there has been an increasing understanding for how the brain uses model-based methods (Daw, Niv, and Dayan, 2005; Balleine, Daw, and O'Doherty, 2009; Dolan and Dayan, 2013; Daw and O'Doherty, 2014).

The two algorithms rest upon a formal class of Markov decision processes (MDPs). The core elements of an MDP are a set of situations or *states*, *actions*, *transitions* (i.e. how states and actions lead to next states), and *rewards*. In this framework, the expected cumulative future reward, called the *state-action value function*, is given by the following recursive equation:

$$Q(s_t, a_t) = R(s_t) + \Sigma\, T(s_{t+1}/a_t, s_t)\, max_a\, [Q(s_{s+1}, a_{t+1})] \qquad (MP)$$

where s_t stands for the state at time t, a_t for the action at time t, Q for the value function state-action, R for the reward function, and T for the transition function determined by the conditional probability of s_{t+1} with respect to s_t and a_t. The most important simplifying assumption of the MPD is that the state transition probability depends only upon the *current* state and action (i.e. the new state is independent of all earlier states and actions).

It should be emphasized that the difference between model-free RL and model-based RL does not lie in the nature of conveyed information but rather in how this information is supposedly inferred from experience by a trial-and-error process (i.e. by trying various

actions and observing their results). A key insight into the early developments of RL was that learning problems could equally be attacked by focusing on estimating the quantities appearing on either side of the equation *MP*. Let us consider each one of these learning algorithms.

In model-based RL, the brain is supposed to implicitly solve equation *MP* by building up a picture of the dynamic "state-action-next state" for each action by requesting its experience with actions and their outcomes—meaning it estimates the right-hand side of *MP*. Learning is said to be "model-based" because the brain implicitly turns to an "internal model" of the task. In psychologic terms, since rewards value is supposedly dependent on expected outcomes, the control is viewed as goal-directed. The system is very flexible but the calculations are complex and costly in terms of working memory. The flip side of the simplicity in learning an internal model is computational complexity in using it.

In model-free RL, the brain is supposed to directly link the given system state, action, and resulting consequence and learn how to directly approximate the left-hand side of equation *MP* (Sutton, 1988). Here, reinforcement learning is said to be "model-free" because it lies on no representation of MPD for estimating the right-hand side of equation *MP*. In psychological terms, in this system, learning is habitual. This approach corresponds exactly to the RPE paradigm (see above in the section). *Q* can be updated using error-driven learning, increased or decreased depending whether the reward is larger or smaller than expected. This system has the benefit of directly representing reward value; however, the system is not very flexible (Doya et al., 2007). The "temporal difference" algorithm (TD) is an often-quoted example of model-free RL in computer science (Sutton, 1998; Suton and Barto, 1998) and may be interpreted as an extension of the Rescorla-Wagner algorithm (Rescorla-Wagner, 1972), which supposes that reward predictions are continuously improved by comparing them with current rewards (see e.g. Glimcher, 2011a, Chap. 13).

Among the several variants of TD algorithms present in the literature, the most known variant is the "actor-critic" model (Baro, 1995). It is imagined that the selection of the action is implemented by two fictitious components: (1) a component that learns to predict future rewards associated with specific environmental states (the "critic"), and (2) a component that chooses specific actions allowing to move the agent from one state to the next according to the selected policy (the "actor"). The "critic" encodes the states of the world value, characteristic of Pavlovian reward prediction signal, while the "actor" stores probabilities associated with each action in any state of the world and then chooses actions according to these probabilities. The aim of the model is to change the "policy" adopted by the "actor" so that, over time, actions associated with the highest rewards are chosen most frequently. Some analogies have been drawn between the anatomy and connections of the basal ganglia (in the classic limbic system) and the implementation mechanisms of RL models,

such as the actor-critic model—here, the ventral striatum plays the role of "critic" whereas the dorsal striatum is the "actor" (Montague, Dayan, and Sejnowski, 1996). However, the question arises of how to rectify this old model with the more recent distinction between habitual and goal-directed decisions.

Daw, Niv, and Dayan (2005) (see also Balleine, Daw, and O'Doherty, 2009) suggest that the actor-critic model would only apply to habitual decisions, and they propose another model that can integrate the goal-directed component of learning; the "forward" model includes the sensibility of actions to online changes into the consequences of value. Some experiments confirm the existence of these signals reflecting the goal-directed value computed by the "forward" model in the ventromedial PFC (Hampton, Bossaerts, and O'Doherty, 2006). Yet, although this model is likely to provide an alternative conceptual framework that may better integrate learning and motivation processes in the brain, a few of its implications are difficult to accept (Balleine and O'Doherty, 2010). One implication of this model is that the function of dopamine in 'weighing-up' value would only concern habitual decisions and not goal-directed decisions, which is inconsistent with a largely accepted conclusion today that dopamine plays its role for both decisions types as dopaminergic neurons project both to the PFC and the striatum (see above).

Finally, I would like to emphasize that at the very heart of the RPE paradigm is statistical reasoning. The brain is supposed to react to environmental information by continuously updating existing memory with new observations. According to this perspective, evolution would have endowed all living beings with a learning ability and the Bayesian form would be the most natural expression. Today the "Bayesian brain" hypothesis is commonly used in theoretical neuroscience (Dayan and Abbot, 2001; Knill and Pouget, 2004; Doya et al., 2007; Rolls and Deco, 2010; Friston, 2012). This hypothesis is linked to the "free-energy principle" built up by Karl Friston (Friston, Kilner, and Harrison, 2006; Friston, 2010) that began with an idea moved forward by the great physiologist and physicist Hermann von Helmholtz in the 19th century. Essentially, the free-energy principle states that any self-organizing system that is at equilibrium with its environment must minimize its free energy. The principle is essentially a mathematical formulation of how adaptive systems—that is, biological agents, like animals or brains—resist a natural tendency to disorder. However, even if this hypothesis seems widely accepted concerning perceptual decisions (Gold and Heekeren, 2014), this may be problematic for VBDs.

Towards a computational model of economic choice

Driven by the theory of revealed preference, economics traditionally has not been interested in the neural processes underlying human choice. As a result, standard economic models are "as if" as opposed to "as is" descriptions of decision making. In contrast,

neuroeconomists are interested in the actual computational and neurobiological processes behind human behavior. Neuroeconomics aims for “structural” models of decision making (Glimcher, 2011a; Fehr and Rangel, 2011). The contours of such a model emerge today in neuroeconomic literature. They arise from the theoretical and computational neuroscience teachings coupled with the various experimental findings of the past 20 years. This model applies to goal-directed decisions[51] and deals mainly with “simple” choices, namely choices between a small number of familiar goods, with no informational asymmetries, strategic consideration, self-control problems, and financial, temporal, or social dimensions (e.g. choosing between an apple or an orange for desert). Of course, researchers hope that insights learnt in this simple case will also be applied to more complicated and interesting problems. As suggested by the first experimental results dealing with risky decisions and intertemporal choices, this hypothesis appears to be reasonable (see Levy et al., 2010, for risky decisions, and Hare, Camerer, and Rangel, 2009, Fehr and Rangel, 2011, and Kable, 2014, for intertemporal decisions). Social decision making is more complicated, which relies on neural representation of oneself and others (Hare et al., 2010). Yet, recent findings suggest that there is likely a unified mechanism for motivational control of behavior that may incorporate both social and non-social factors, even though some aspects of these findings suggest that there are also differences between social and non-social neural valuation (Ruff and Fehr, 2014; Hutcherson, Bushong, and Rangel, 2015; Krajbich et al., 2015; Wake and Izuma, 2017).

Schematically, the key components of the model are based on two mechanisms that are supposed to proceed sequentially: (1) a *valuation* mechanism that learns, store and retrieves the values of options offered to the choice; (2) a *choice* mechanism that selects one of these options by comparing values. The process of choice implementation in the brain is better understood as functionally separate from the encoding of values. Moreover, both computation and comparison of decision values are modulated by attention—the brain’s capacity to vary the computational resources that are used in different circumstances insofar as these resources are scarce and costly in terms of consuming energy (Kable and Glimcher, 2009; Fehr and Rangel, 2011; Glimcher, 2014a).

The valuation mechanism

The concept of “subjective value” (SV) of a decision is at the heart of this device (see section 1) and is linked with the notion of “experienced utility”. The SV of a decision corresponds to a neural signal computed “online” at the time of choice by forecasting the eventual hedonic impact of taking the differing options, whether chosen or not chosen, while

[51] Glimcher’s model is more widely dealing with VBD (i.e. it is supposed to also include habitual decisions), but the switch among the two neural systems is not explicitly mentioned.

the experienced utility represents a neural signal computed at the time where the organism is effectively experiencing the hedonic impact of the chosen option (Kable and Glimcher, 2009; Rangel and Hare, 2010; Fehr and Rangel, 2011; Padoa-Schioppa, 2011; Berridge and O'Doherty, 2014; Padoa-Schioppa and Conen, 2017). Of course, it is difficult not to draw a parallel with the current distinction made in behavioral economics between "decision" utility, which corresponds to utility concept of standard economic theory, and "experienced" utility, which symbolizes subjective welfare (Kahneman et al., 1997; Kahneman, 2003, 2011)[52].

Some authors do not explicitly refer to this notion of "experienced utility". Rangel and Clithero (2014) talk about the "stimulus value" and the "outcome value". Glimcher (2011a, 2014a), presents the matter differently; he discriminates between neural regions where an "expected subjective value" is "represented—he no longer uses the confusing expression of expected "physiological utility" proposed earlier in Glimcher, Dorris, and Bayer, 2005)—and neural regions where this value is "learnt and stocked" (i.e. where it is "constructed")[53]. I will follow this way of presenting the valuation mechanism.

Regarding the "representing" phase, many fMRI experiments with human subjects suggest that the same neural region, the ventromedial PFC (including the medial OFC), is activated for valuing decisions at the time of choice. This general finding is corroborated by clinical studies examining patients with damage in this brain zone (see above section 1). Furthermore, experiments using manipulation techniques (i.e. experiments able to prove causality and not just correlation), such as brain stimulation, end up with similar results. Brain activity in the ventromedial PFC clearly reveals the idiosyncratic values people place on goods, actions, or rewards, whether one is talking about food rewards, fluid rewards, monetary rewards, gains, losses, social rewards, or abstract rewards (e.g. viewing beautiful faces). Finally, the evidence available in single-cell recordings show that this neural region is the final common path for valuation in the human and monkey brain (for a review of these studies see Fehr and Rangel, 2011, Rushworth et al., 2011, Wallis, 2012, Levy and Glimcher, 2012, Bartra, McGuire, and Kabel, 2013; Platt and Plassmann, 2014).

---

[52] Specifically, Kahneman makes a distinction between "predicted utility" and "experienced utility" and this last one is itself separated into "remembered utility" (*due to* consumption experience) and "real time utility" (*over* experience). Remark there is a strong similarity between remembered utility and the "somatic marker" mechanism proposed by Damasio and his colleagues (see section 1). Other distinctions are developed in the literature: for instance, Bossaerts, Preuschoff, and Hsu (2009) mentioned "true" preferences (what individuals want) and "revealed" preferences (what individuals do), Berridge and O'Doherty (2014) separated what is "wanting" and "liking" for an outcome, "it is possible to want what is not expected to be liked, not remembered to be liked, as well as what is not actually liked when obtained" (p. 242).

[53] Glimcher's variant of the model, with what he names "hard" expected utility, stands out somewhat from other variants in the literature. The author draws a parallel between the standard economic choice model and the neurobiological decision model. Briefly, the economic model links "utility' to "choice" by a maximization operation (argmax), while the neurobiological model would link "subjective value" to "action" by a neural mechanism of the type "winner-take-all" or "drift-diffusion". He emphasizes the fact that "because neurobiological notions of subjective value and action (unlike modern economic notions of utility and choice) are measurably distinct objects" … they "are linked by an observable mechanism" (p. 194). These models will be mentioned later in the section when talking about the choice mechanism.

Several authors add the ventral striatum (including the NAcc and the inferior parts of the caudate nucleus and putamen) to the ventromedial PFC as contributing to SV, at least for non-human primates (Knutson et al., 2009; Glimcher, 2011a; Levy and Glimcher, 2012). However, it has been argued that signals in the ventral striatum could be simply RPE signals and not subjective value signals *per se* (Hare et al., 2008). In response to this criticism, Glimcher (2011a) admits that it is very difficult to discriminate SV and RPEs in many experimental designs and that the dense dopaminergic projections to the ventral striatum from the ventral tegmental area (VTA) makes it nearly certain that activity in this area should track RPEs. Yet, Glimcher has argued that this view is not well aligned with either evolutionary data or single-unit monkey physiology data. He adds that we know that animals without frontal cortices (e.g. reptiles) can learn and represent SVs, a finding that clearly would suggest a role for the striatum in SV representation throughout the vertebrate line. In any event, this is still a controversial hypothesis, at least as how it pertains to good-based values in humans (Padoa-Schioppa, 2011; Rangel and Clithero, 2014).

In this encoding mechanism of an SV by the brain, some additional aspects occur. First, the position on a value scale is not absolute in that SV is *relative* to the decision-making context and other options offered to the choice (Elliot et al., 2008; Padoa-Schioppa, 2009). Second, in some circumstances, the brain also takes into account the "action cost" that measures effort or unpleasant character associated to decision making independently of its expected benefits. This is the case when the action cost associated with acquiring the option are not negligible relative to the benefits from consuming them or when the action costs of the options under consideration are very different. For Wallis and Rushworth (2014), in the supposed serial process between good-based choices and action choices, if one does not initially act costs into account, then the potential goods space is vast. Thus, actions costs can help constrain this abstract space from the outset. If action costs arise, then, net decision value is supposedly given by the decision value minus the cost value (Basten et al., 2010; Ranger and Hare, 2010; Ranger and Clithero, 2014)[54].

The "constructing" phase of decision SV refers to processes by which the brain is using experienced utility signals supposed to be computed at the time of reward or punishment "consumptions"—when the organism is experiencing the actual consequences of chosen options). How these signals are used to update future decision values is still an exploratory issue[55]. However, authors agree that the reward system and RL processes are likely taking a central place in decision value construction by favoring the building of a stimuli valuation

[54] The costs nature issue in encoding of decision is addressed somewhat differently by Grabenhorst and Rolls (2011) (see also Rolls, 2014, Chap. 9). These authors draw a distinction between "extrinsic" costs (such as action costs, time delay, risk in getting reward) and "intrinsic" costs (such as motivation state, impulsiveness, risk and ambiguity attitude of the subject).

[55] The more preliminary evidence with experienced utility arises in a large part from technical difficulties: difficulty to measure it reliably in animals, difficulty to induce controlled consumption experiences in humans when they are lying inside an fMRI scanner.

common scale and by feeding the memorization system that keeps track of relevant environmental information collected and continuously updated (see above in this section).

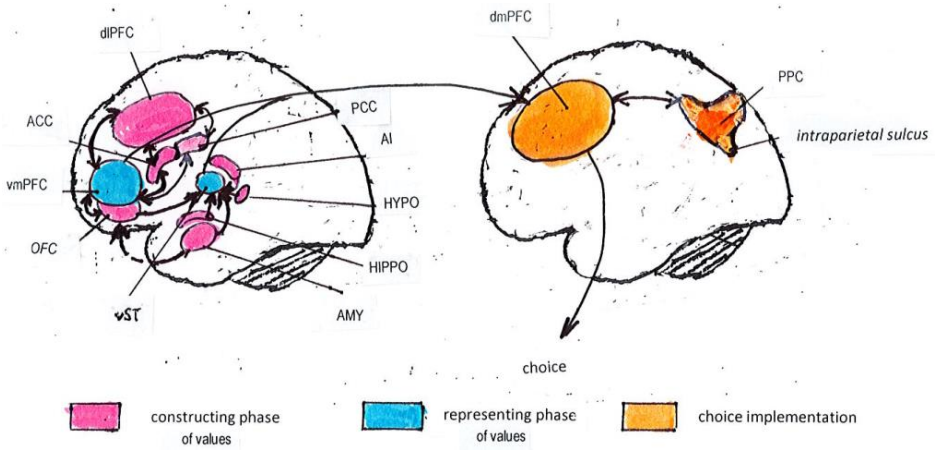
First, evidence suggests that a wide neural network is taking part in the computation of experienced utility signals, including: (1) the OFC for the valuation of many sorts of primary rewards (various consumer goods: food, drinks, music, works of art…) and secondary rewards (monetary gains) during rather short time courses; (2) the NAcc (in the ventral striatum) for the valuation of several sorts of primary and secondary rewards too; (3) the dorsolateral PFC, which provides a complementary function to the OFC in value building, for the valuation over generally much longer time courses and in more abstract terms and the valuation of social/symbolic rewards or consumer goods requiring or referring to self-control, particularly in humans; (4) the amygdala in emotional regulation of reward/punishment value (in particular, effects of fear and stress on value), such as risky or ambiguous decisions; (5) the anterior insula, which is traditionally associated to visceral sensation of disgust and to some classes of events generating punishments, such as ambiguous decisions; (6) the ACC, which usually plays a role in conflict resolution, in particular when subjects are nearly indifferent among largely disparate options or when the valuation subsystems are providing contradictory outputs; (7) the posterior cingulate cortex (PCC) for the valuation of preferred risky options in a choice task or of delayed options in an intertemporal choice task; (8) the hypothalamus for the valuation of risk aversion for primary rewards; (9) the hippocampus and near structures with which it is closely connected in the median temporal lobe for value storing and modulation as well as its critical role in working memory processes. Thus, what emerges is a fairly complex network of brain areas that construct, in the ventromedial PFC and probably in the striatum, a subjective value signal that guides choices (for a review of the studies proving these results see Glimcher, 2011a, 2014a, Wallis and Rushworth, 2014, Fehr and Krajbich, 2014, Platt and Plassmann, 2014, Rangel and Clithero, 2014, Lempert and Phelps, 2014, Jung et al., 2016, Padoa-Schioppa and Conen, 2017).

Some scholars (Glimcher, 2011a, Chap. 14; Rolls, 2014, Chap. 9) stress on the more *absolute* nature of this learned and stored value in these various neural structures, in comparison to the *relative* nature of value taking place in the representing phase. It was found that the responses of some OFC neurons encoding the value of a specific stimulus did not depend on which other stimuli were simultaneously available (Padoa-Schioppa and Assad, 2008); this has been referred to as *menu invariance*, a property that could provide a neurobiological foundation for transitivity—a fundamental trait of economic choice.

Thus, it is important to note that the OFC is playing a central role in the representing phase of decision SV, essentially in the medial OFC (often included in the ventromedial PFC), but this neural region also would be active in the constructing phase, if only for primary and secondary rewards (Padoa-Schioppa and Assad, 2006; Plassman, O'Doherty and

Rangel, 2007; Rushworth et al., 2011, Rolls, 2014, Chap. 9; Padoa-Schioppa and Conen, 2017) (see Figure 5 below for a schematic representation of the valuation network distinguishing the constructing and representing phases).

***Figure 5*. Schematic valuation and choice implementation networks**



ACC: anterior cingulate cortex, dlPFC: dorsolateral PFC, vmPFC: ventromedial PFC, OFC: orbitofrontal cortex, vST: ventral striatum, AMY: amygdale, HIPPO: hippocampus, HYPO: hypothalamus, AI: anterior insula, dmPFC: dorsomedial PFC, PPC: posterior parietal cortex

The choice mechanism

The last phase taking place in a VBD corresponds to the selection of one particular option among those offered a choice. Once SV has been computed for each option, the "best" one has to be selected by comparing them. This raises two questions: what kind of choice mechanism is likely to be implemented, and what are the neural structures taking part in this process of choice implementation?

The topic of which neural substrates are mobilized in the final phase of decision making remains unclear. Some neuroscientists suggest, however, that a frontoparietal network is involved. Their argument is essentially based on non-human primate experiments using behavioral paradigms that include decisions leading to *actions* (i.e. involving motor circuits). The most obvious evidence comes from a decision system familiar to neurophysiologists—the monkey visio-saccadic system, which, for widely technical reasons, was used as a model to understand sensorimotor control in general since the 1980s (Anderson and Buneo, 2002). The core of this frontoparietal network that is playing a critical role for oculomotor tasks involves areas known as the lateral intraparietal area (LIP) in the intraparietal sulcus, the frontal eye field (FEF) (in the prefrontal cortex), and the superior colliculus in the midbrain (Platt and Glimcher, 1999; Glimcher, 2003). These findings were later generalized to body movements; it has been shown that the primary motor cortex (M1) (BA 4) (analogue to the colliculus), premotor cortex (BA 6) (analogue to the FEF), some anterior areas of the parietal cortex (including parts of BA 5 and 7) (analogue to area LIP), and supplementary motor area (SMA) (homologue to the supplementary eye field of the eye movements system) play

equivalent roles (see Glimcher, 2011a, Chap. 11, and 2014a). Knowing if these results may be transposed from “action” to “economic choice” (i.e. a decision that a priori do not necessary imply a movement) is still an open question. Yet some cerebral imaging studies suggest that human’s neural zones involved in comparison of VBD correspond to a specific network including the dorsomedial PFC, the posterior parietal cortex (PPC), and the nearby intraparietal sulcus (Kable and Glimcher, 2009; Hare et al., 2011; Glimcher, 2011a, 2014a). These areas are suggested to implement a comparison process, and that the output of these so-called “comparator regions” modulates activity in the motor cortex to implement the choice (see Figure 5 above for a schematic representation of this presumed choice implementation network).

Concerning the mechanism of choice implementation, two basic models of brain functioning were commonly developed in theoretical neuroscience: (1) the “winner-take-all” (WTA) and (2) the “drift-diffusion” (DD) models (Glimcher, 2011a, Chap. 2; Deco et al., 2013; Rolls, 2014, Chap. 8). The DD model was initially developed by the psychologist Roger Ratcliff in a study on memory for explaining accuracy and response times in any task involving binary responses that can be elicited in a handful of seconds (Ratcliff, 1978) In the first model, the brain is supposed to compare all feasible options according to their respective SV assessed on the common scale of internal valuation, and then to choose the one with the highest value. Basically, the brain would behave like it is assumed in the standard economic model. In the second model, it is supposed there is a fixed *threshold* of expected reward that is considered satisfactory by the organism, options are assessed one after another, and as soon as an option exceeds this threshold, it is chosen. How is fixed this threshold? It would depend on decisions’ complexity. The brain would undertake a trade-off between speed and accuracy: if the threshold is low, the choice is highly simplified, at the risk possibly of not choosing the best option, while if it is high, the choice is more difficult and longer but the error risk is decreasing (Roitman and Shadlen, 2002; Palmer, Huk, and Shadlen, 2005).

We can think a priori that perceptual decisions should be rather ruled by the DD model whereas value-based decisions (VBD) should result from the mechanism described by the WTA model, i.e., the natural neural counterpart of the “arg max” operation in the standard economic model (Kable and Glimcher, 2009; Glimcher, 2014a; Wang, 2014). However, if the DD model is widely accepted today for perceptual decisions (Palmer et al., 2005; Gold and Shadlen, 2007; Glimcher, 2011a, Chap. 9), one can be skeptical about relevance of the WTA model for VBD. Actually, it appears that anatomically, the same frontoparietal network is involved in the choice mechanism, irrespective of the nature of decisions (Glimcher, 2014a). Thus, the question is whether we have to conclude that neural structures of this network fulfil these two different functions (perceptual decisions and VBD) or whether there is one

integrate decision-making model (Glimcher, 2014a). Soltani and Wang (2008) were the first to suggest that there would be a *unique neural system*, using the same computational properties, which would generate the two kinds of behaviors. In the recent literature, there is evidence that the integrate model would be not only be possible but perhaps necessary (see Wang, 2014, for a survey). It should be remarked that if this finding were to be confirmed, the specificity of neuroeconomics in relation to decision neuroscience would tend to lose part of its specificity.

What are the contours of this integrate decision-making model? An increasing consensus in the neuroeconomic literature favors a slightly more sophisticated variant of the DD model's basic version. I briefly mention here the model set out by Fehr and Rangel (2011). In case of a binary choice, for instance a choice between decisions x or y, the model provides a representation of the dynamic computing that the brain is supposedly making. The brain computes a relative SV signal, denoted by *R*, which measures the value difference of *x* versus *y*. This signal starts at zero and at every instant *t* evolves according to the following equation:

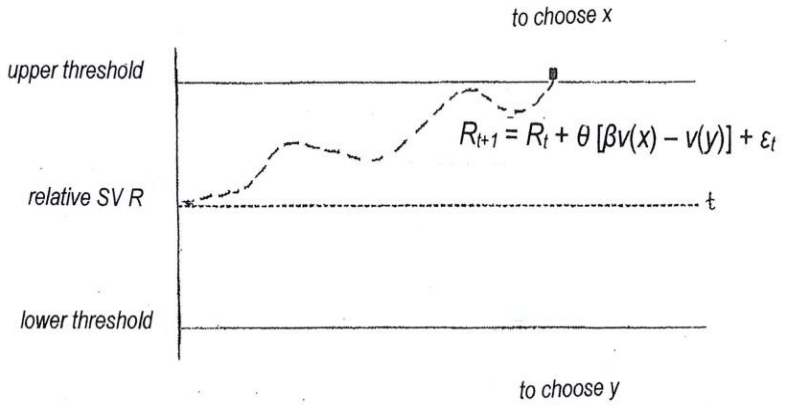
$$R_{t+1} = R_t + \theta\,[\beta v(x) - v(y)] + \varepsilon_t\,,$$

where R_t denotes the level of the signal at time t (measured from the start of the choice process), $v(x)$ and $v(y)$ denotes the SV assigned to the two options, θ being a parameter that affects the speed of the process (drift rate), β being a parameter that measures the "attentional bias" towards the attended option (β > or = 1), and ε_t an independent and identically distributed error term with variance s^2. The process continues until one of the predefined thresholds (upper or lower) is crossed: x is chosen if the upper threshold is crossed first, and y is chosen if the lower threshold is crossed first. Notably, Glimcher (2011a, Chap. 9) suggests a parallel between this notion of "threshold" in terms neuronal and the usual notion of reservation price in economics (i.e. the minimal price at which choosing an option is subjectively profitable) by referring to familiar idea of "bounded rationality" and "satisficing" introduced by Herbert Simon (1955, 1979). However, it was just an illustration and, by the way, misleading. Because this notion of a threshold in neuronal terms is fully compatible regarding an optimizing behavior once the information processing cost is integrated (i.e. the trade-off between speed and accuracy). Maybe one can interpret it as the required level of trust so that the corresponding option is chosen (Krajbich, Oud, and Fehr, 2014) (see below Figure 6).

This model has an important feature—since the relative SV signal evolves stochastically, choices are inherently noisy. Of course, the stochastic nature of *R* is a direct result of the inherent stochasticity of neural activity, as stated above. Furthermore, interestingly, this model makes quantitative predictions about the correlation between attention, choices, and

reaction times (when $\beta > 1$), and these correlations can be tested using eye-tracking testing methods. Another important prediction of this model is that exogenous increases in the amount of relative attention paid to an appetitive item (for instance, through experimental or marketing manipulations) should bias choices in its favor by increasing the probability that it is chosen. Several studies have confirmed this prediction through manipulations of visual attention (Fehr and Rangel, 2011).

Figure 6. **The drift-diffusion model (binary choice)**



The components of this DD model have empirically been tested using both behavioral and neural data with food choice, temporal discount decision, and social decision (Krajbich, Armel, and Rangel, 2010; Krajbich et al., 2012; Krajbich et al., 2015). This has been generalized to the case of three-way choice, and these findings suggest that the underlying processes might be robust for small numbers of items (Krajbich and Rangel, 2011). However, additional research is necessary for precisely delimiting the field in which the DD model can be useful (Krajbich, Oud, and Fehr (2014).

Final Remarks

Neuroeconomics is still a nascent scientific field, two decades old at the most. Although much remains to be done, a great deal of results has already been proven about how the human brain makes choices, and these findings provide insights into the understanding of economic behavior in many domains. Undoubtedly, without the availability of an extensive variety of experimental designs for dealing with individual and social decision-making provided by experimental economics, many neuroeconomics studies could not have been developed. Indeed, it is very likely that, for future historians of economics, lab experiments will be “one of the most stunning methodological revolutions in the history of science” (Guala, 2009, 152). At the same time, without the significant progress made in modern neuroscience for grasping and understanding brain functioning, neuroeconomics would have never seen the light of day.

In this paper, I proposed an overview of these advances in neuroscience by grouping them together over three non-independent topics referred to as the "emo-rational" brain, the "social" brain, and the "computational" brain. For each topic, I emphasized findings that I consider critical to the birth and development of neuroeconomics while highlighting some of prominent questions about which knowledge should be improved by future research.

In parallel, I show that the boundaries between neuroeconomics and several recent branches of cognitive and behavioral neuroscience, such as affective, social, and, more generally, decision neuroscience, are particularly porous. In this regard, I suggest that a greater autonomy of neuroeconomics should perhaps come from the development of studies about more economic policy-oriented concerns, such as for instance the approach undertaken by Bernheim and Rangel to find in neuroeconomics the foundations for a "behavioral public economics" (Bernheim and Rangel, 2005, 2007, 2009), in the wake of the debates around the "libertarian paternalism" (Sunstein and Thaler, 2003; Thaler and Sunstein, 2003, 2008; Loewenstein and Haisley, 2008; Sunstein, 2014).

References

Adolphs R. (2017), "How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences", *Social Cognitive and Affective Neuroscience*, 12: 24-31.

Adolphs R. and D. Anderson (2018), *The neuroscience of emotion: A new synthesis*, Princeton University Press.

Allos-Ferrer Carlos (2018), "A review essay on *Social neuroscience*: Can research on the social brain and economics inform each other?" *Journal of Economic Literature*, 56(1): 234-264.

Amodio D.M. and C.D. Frith (2006), "Meeting of minds: The medial frontal cortex and social cognition", *Nature Review Neuroscience*, 7(4): 268-277.

Anderson R.A. and C.A. Bueno (2002), "Intentional maps in posterior parietal cortex", *Annual Review of Neuroscience*, 25: 189-220.

Arnsten A.F. and K. Rubia (2012), "Neurobiological circuits regulating attention, cognitive control, motivation, and emotion: Disruptions in neurodevelopmental psychiatric disorders", *Journal of the American Academy of Child and Adolescent Psychiatry*, 51: 356-367.

Arnsten A.F., M.J. Wang, and C.D. Paspalas (2012), "Neuromulation of thought: Flexibilities and vulnerabilities in prefrontal cortical network synapses", *Neuron*, 76: 223-239.

Aston-Jones G. and J.D. Cohen (2005), "An integrative theory of locus coerulus-norepinephrine function: adaptive gain and optimal performance", *Annual Review of neuroscience*, 28: 403-450.

Awh E. and J. Jonides (2001), "Overlapping mechanisms of attention and spatial working memory", *Trends in Cognitive Sciences*, 5: 119-126.

Balleine B.W, N.D. Daw, and J.P O'Doherty (2009), "Multiple forms of value learning and the function of dopamine", in Glimcher P. *et al*. (eds.) (2009), Chap.: 24: 367-388.

Balleine B.W and J.P. O'Doherty (2010), "Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action", *Neuropsychopharmacology*, 35: 48-69.

Baron-Cohen S. (2000), "Theory of mind and autism: A review", *International Review of Research in Mental Retardation*, 13: 169-184.

Barrett L.F. (2017), *How emotions are made: The secret life of the brain*, Hougton-Mifflin-Harcourt.

Barto A.G. (1995), "Adaptive critics and the basal ganglia" *in* Houk J.C., J.L. Davis, and D.G. Beiser (eds.), *Models of information processing in the basal ganglia*, Cambridge, MA: MIT Press, 215-232.

Bartra O, J.T. McGuire, and J.W. Kable (2013), "The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value", *NeuroImage*, 1(76): 412-427.

Basten U., G. Bille, H.R. Heekeren, and C. Fiebach (2010), "How the brain integrates costs and bebefits during decision making", *Proceedings of the National Acacademy of Science USA*, 107: 21767-21772.

Baumgartner T., T. Knoch, D. Hotz, P. Eisenegger, and C. Fehr (2011), "Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice", *Nature Neuroscience*, 14(11): 1468-1474.

Bechara A. and A.R. Damasio (2005), The somatic marker hypothesis: a neural theory of economic decision, *Games and Economic Decision*, 52: 336-372.

Bechara A., A.R. Damasio, H. Damasio, and S.W. Anderson (1994), "Insensitivity to future consequences following damage to human prefrontal cortex", *Cognition*, 50: 7-15.

Bechara A., H. Damasio, A.R. Damasio, and G.P. Lee (1999), "Different contributions of the human amydalia and ventromedial prefrontal cortex to decision-making", *Journal of Neuroscience*, 19: 5473-5481.

Bechara A., H. Damasio, D. Tranel, and A.R. Damasio (1997), "Deciding advantageously before knowing the advantageous strategy", *Science*, 275 (February 28): 1293-1295.

Bechara A., D. Tranel, H. Damasio, and A.R. Damasio (1996), "Failure to respond autimatically to anticipated future outcomes following damage to prefrontal cortex", *Cerebral Cortex*, 6: 215-225.

Benhabib J and A. Bisin (2005), "Modeling internal commitment mechanisms and self-control: A neuroeconomic approach to consumption-saving decisions", *Games and Economic Behavior*, 52(2): 460-492.

Bernheim D.B. and A. Rangel (2005), "From neuroscience to public policy: A new economic view of addiction", *Swedish Economic Policy Review*, 12: 11-46.

Bernheim D.B. and A. Rangel (2007), "Toward choice-theoretic foundations for behavioral welfare economics", *The American Economic Review,* 97(2): 464-470.

Bernheim D.B. and A. Rangel (2009), "Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics", *Quarterly Journal of Economics*, 124(1): 51-104.

Berns G.S., S.M. McClure, G. Pagnoni, and P.R. Montague (2001), "Predictability modulates humain brain responses to reward", *Journal of Neuroscience*, 21: 2793-2789.

Berridge K.C. and J.P. O'Doherty (2014), "From experienced utility to decision utility", in Glimcher P. and E. Fehr (eds.) (2014), Chap.17: 335-351.

Bjorklund A. and S.B. Dunnett (2007), Dopamine neuron stems in the brain: An update", *Trend Neuroscience*, 30: 194-202.

Blair R.J.R. (2005), "Responding to the emotions of oters: Dissociating forms of empathy through the study of typical and psychiatric populations", *Consciousness and Cognitions*, 14(4): 698-718.

Blair R.J.R. (2008), "Fine cuts of empathy and the amygdala: Dissociable deficits in psychopathy and autism", *Quarterly Journall of Experimental Psychology,* 61(1): 157-170.

Bloom P. (2017), "Empathy and its discontents", *Trends in Cognitive Sciences*, 21(1): 24-31.

Bossaerts P., K. Preuschoff, and M. Hsu (2009), "The neurobiological foundations of valuation in human decision making under uncertainty", in Glimcher P., C.F. Camerer, E. Fehr, and R.A. Poldrack, (eds.) (2009), Chap. 23: 353-366.

Bourgeois-Gironde S. (2010), "Is neuroeconomics doomed by the reverse inference fallacy?" *Mind and Society*, 9(2): 229-249.

Bouton M.E. (2007), *Learning and behaviour: A contemporary synthesis*, Sunderland, Massachussets: Sinauer Associates Inc.

Bowles S. and H. Gintis (2011), *A cooperative species: Human reciprocity and its evolution*, Princeton: Princeton University Press.

Breiter H.C.,I. Aharon, D. Kahneman, A. Dale et P. Shizgal (2001), "Functional imaging of neural responses to expectancy and experience of monetary gains and losses", *Neuron*, 30: 619-639.

Brodmann K. (1909), *Vergleichhende lokalisationslehre der grosshimrinde*, Leipzig: Verlag von Johann Ambrosius Barth (English translation: Garey L.J., *Brodmann's "Localisation in the cerebral cortex"*, London: Smith-Gordon, 1994).

Buckner R.L., J.R. Andrews-Hanna, and D.L. Schacter (2008), "The brain's default network: Anatomy, function, and relevance to disease", *Annals of the New York Academy of Sciences*, 1124(1): 1-38.

Bzdok D., L. Schibach, K. Vognely, K. Schneider, A.R. Laird, R. Langner, and S.B. Eickhoff (2012), "Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy", *Brain Structure and Function*, 217: 783-796.

Cacioppo J.T. and G.C. Berntson (1992), "Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis", *American Psychologist*, 47(8): 1019-1028.

Cacioppo J.T., P.S. Visser, and C.L. Pickett (eds.) (2006), *Social neuroscience: People thinking about thinking people*, Cambridge and London: MIT Press.

Call J. and M. Tomasello (2008), "Does the chimpanzee have a theory of mind? 30 years later", *Trends in Cognitive Sciences*, 12(5): 187-192.

Camerer C.F. (2003), *Behavioral game theory. Experiments in strategic interaction*, Princeton: Princeton University Press.

Camerer C.F. (2006), "Behavioral economics", Working Paper Caltech, Div HSS 228-77 (prepared for the Worl Congress of the Econometric Society, august 2005, London).

Camerer C.F. (2007), "Neuroeconomics: using neuroscience to make economic predictions", *Economic Journal*, 117: C26-C42.

Camerer C.F. (2008a), "The case of mindful economics", in Caplin A. and A. Schotter (eds.), *Foundation of positive and normative economics*: *A handbook*, New York: Oxford University Press, 2008, Chap. 2: 43-69.

Camerer C.F. (2008b), "The potential of neuroeconomics", *Economics and Philosophy*, 24: 369-379.

Camerer C.F. and T.A Hare (2014), "The neural basis of strategic choice", in Glimcher and Fehr (2014), Chap. 25: 479- 511.

Camerer C.F., G. Loewenstein, and D. Prelec (2004), "Neuroeconomics: Why economics needs brains", *Scandinavian Journal of Economics*, 106(3): 555-579.

Camerer C.F., G. Loewenstein, and D. Prelec (2005), "Neuroeconomics: how neuroscience can inform economics", *Journal of Economic Literature*, 43: 9-64.

Camille N., G. Coricelli, J. Sallet, P. Pradat-Diehl, J-R. Duhamel, and A. Sirigu (2004), "The involvement of the orbitofrontal cortex in the experience of regret", *Science*, 304, 21 may : 1167-1170.

Caplin A. and M. Dean (2008), "Dopamine, reward prediction error, and economics", *Quarterly Journal of Economics,* 123(2): 663-701.

Caplin A. et M. Dean (2009), "Axiomatic neuroeconomics", in Glimcher P., C.F. Camerer, E. Fehr, and R.A. Poldrack (eds.) (2009), Chap. 3: 21-32.

Caplin A., M. Dean, P.G. Glimcher, and R. Rutledge (2010), "Testing the reward prediction error hypothesis with an axiomatic model", *Journal of Neuroscience*, 30: 13525-13536.

Caplin A. and P. Glimcher (2014), "Basic methods from neoclassical economics", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 1: 3-17.

Carr L., M. Iacoboni, M.-C. Dubeau, *et al*. (2003), "Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas", *Proceedings of the National Academia of Science USA*, 100: 5497-5502.

Carter R.M., D.L. Bowling, C. Reech, and S.A. Huettel (2012), "A distinct role of the temporo-parietal junction in predicting socially guided decisions", *Science*, 337: 109-111.

Chen M.K., V. Lakshminarayanan, and L.R. Santos (2006), "How basic are behavioral biases? Evidence from capuchin monkey trading behavior", *Journal of Political Economy*, 111(3): 517-537.

Chib V.S., A. Rangel, S. Shimojo, and J.P. O'Doherty (2009), "Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex", *Journal of Neuroscience*, 29: 12315-12320.

Christensen J.F. and A. Gomila (2012), "Moral dilemma in cognitive neuroscience of moral decision-making: A principle review", *Neuroscience and Biobehavioral Reviews*, 36(4): 1249-1264.

Chua H.F., R. Gonzalez, S.F. Taylor, R.C. Welsh, and I. Liberzon (2009), "Decision-related loss: Regret and disappointement", *NeuroImage*, 47: 2031-2040.

Clark-Polner, T.D. Johnson, and L.F. Barrett (2017), "Multivoxel patern analysis does not provide evidence to support the existence of basic emotions", *Cerebral cortex*, 27: 1844-1948.

Cochi L., A. Zalesky, A. Fornito, and J.B. Mattingley (2013), "Dynamic cooperation and competition between brain systems during cognitive control", *Trends in Cognitive Sciences*, 17(10): 493-501.

Commons M.L. (2001), “A short history of the Society for the quantitative analysis of behavior”, *Behavior Analyst Today*, 2 (3): 275-279.

Cools R. (2011), “Dopaminergic control of the striatum for high-level cognition”, *Current Opinion in Neurobioloy*, 21: 402-407.

Coricelli G. (2005), “Two-levels of mental states attribution: From automaticity to voluntariness”, *Neuropsychologia*, 43(2): 294-300.

Coricelli G., H.D. Critchley, M. Joffily, J.P. O’Doherty, A. Sirigu, and R. J. Dolan (2005), “Regret and its avoidance: A neuroimaging study of choice behaviour”, *Nature Neuroscience*, 8(9): 1255-1262.

Coricelli G. and R. Nagel (2009), “Neural correlates of depth of strategic reasoning in medial prefrontal cortex”, *Proceedings of the National Academia of Science USA*, 106: 9162-9168.

Cot A.L. and S. Ferey (2016), « La construction de ‘faits’ économiques d’un nouveau type : éléments pour une histoire de l’économie expérimentale », *L’Actualité économique. Revue d’analyse économique*, 92(1-2) : 11-47.

Cowen A.S. and D. Keltner (2017), “Self-report captures 27 distinct categories of emotion bridged by continuous gradients”, *Proceedings of the National Academy of Science USA*, 114: 7900-7909.

Cox C.L., L.Q. Uddin, F.X. Castellanos, M.P. Milham, and C. Kelly (2011), “The balance between feeling and knowing: Affective and cognitive empathy are reflected in the brain’s intrinsic functional dynamics”, *Social Cognitive and Affective Neuroscience*, 7(6): 727-737.

Craig A.D. (2002), “How do you feel? Interoception: The sense of the physiological condition of the body”, *Nature Reviews Neuroscience*, 3: 655-666.

Cravel C.F. (2007), *Explaining the brain: mechanisms and the mosaic unity of neuroscience*, Oxford: Oxford University Press.

Crockett M.J. and E. Fehr (2014), “Pharmacology of economic and social decision making”, in Glimcher P. and E. Fehr (eds.) (2014), Chap. 14: 259-279.

Dalgleich T. (2004), “The emotional brain”, *Nature Reviews Neuroscience*, 5: 583-589.

Dalley J.W., B.J. Everitt, and T.W. Robbins (2011), “Impulsivity, compulsivity, and top-down cognitive control”, *Neuron*, 69(4): 680-694.

Damasio A.R. (1994), *Descartes’error: Emotion, reason, and the human brain*, New York: A. Grosset/Putnam Books (French translation: *L’erreur de Descartes*, Paris : Odile Jacob, 1995)

Damasio A.R. (1996), “The somatic marker hypothesis and the possible functions of the prefrontal cortex”, *Philosophical Transactions of the Royal Society of London B*, 351: 1413-1420.

Damasio A.R. (2003), *Looking for Spinoza: Joy, sorrow, and the feeling brain*, New York: Harcourt, Inc. (French translation: *Spinoza avait raison. Joie, tristesse, le cerveau des émotions*, Paris : Odile Jacob, 2003).

Damasio A.R. (2010), *Self comes to mind. Constructing the conscious brain*, New York: Pantheon Books (French translation : *L’autre moi-même : les nouvelles cartes du cerveau, de la conscience et des émotions*, Paris : Odile Jacob, 2010).

Damasio A.R (2017), *The stranger order of things. Life, feeling, and the making of cultures*, New York: Pantheon Books (French translation: *L’ordre étrange des choses. La vie, les sentiments et la fabrique de la culture*, Paris : Odile Jacob, 2018).

Darwin C. (1859), *On the origin of species by means of natural selection*, New York: D. Appleton and Company (6th edition: 1872).

Darwin C. (1872), *The expression of the emotions in man and animals*, London: John Murray (Chicago University Press, 1965).

Daw N.D. (2014), “Advanced reinforcement learning”, in Glimcher P. and E. Fehr (2014) (eds.), Chap. 16: 299-320.

Daw N.D., S. Kakade, and P. Dayan (2002), “Opponent interactions between serotonin and dopamine”, *Neural Networks*, 15: 603-616.

Daw N.D., Y. Niv, and P. Dayan (2005), “Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control”, *Nature Neuroscience*, 8(12): 1704-1711.

Daw N.D. and J.P. O’Doherty (2014), “Multiple systems for value learning”, in Glimcher P. and E. Fehr (eds.) (2014), Chap. 21: 393-410.

Daw N.D. and P.N. Tobler (2014), "Value learning through reinforcement: The basics of dopamine and reinforcement learning", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 15: 283-298.

Dayan P. (2008), "The role of value systems in decision making", in Engel C. and W. Singer (eds.), *Better than conscious? Implications for performance and institutional analysis*, MIT Press.

Dayan P. and L.F. Abbot (2001), *Theoretical neuroscience*, Cambridge, MA: MIT Press.

Dayan E., Y. Niv, B. Seymour and N.D. Daw (2006), "The misbehavior of value and the discipline of will", *Neural Network*, 19: 1153-1160.

Decety J. and J.T. Cacioppo (eds.) (2011), *The Oxford handbook of social neuroscience*, New York: Oxford University Press.

Decety J. and J.M. Cowell (2014), "The complex relation between morality and empathy", *Trends in Cognitive Sciences*, 18(7): 337-339.

Decety J. and J. Grèzes (2006), "The power of simulation: imaging one's own and other's behaviour", *Brain Research*, 1079-4-14.

Declerck C. and C. Boone (2016), "The neuroanatomy of prosocial decision making", in Declerck C. and C. Boone (eds.), *Neuroeconomics of prosocial behaviour. The compassionate egoist*, London; Elsevier

Deco G., E.T. Rolls, L. Albantakis, and R. Romo (2013), "Brain mechanisms for perceptual and reward-related decision-making", *Progress in neurobiology*, 103: 194-213.

Dehaene S. and J-P. Changeux (2011), "Experimental and theoretical approaches to conscious processing", *Neuron*, 70(2): 200-227.

Dehaene S., J-P. Changeux, L. Naccache, J. Sackur , and C. Sergent (2006), "Conscious, preconscious, and subliminal processing: A testable taxonomy", *Trends in Cognitive Sciences*, 10(5): 204-211.

Dehaene S., J.-R. Duhamel, M.D. Hauser, and G. Rizzolatti (2005), *From monkey brain to human brain. A Fyssen Foundation symposium*, MIT Press.

Delgado M.R., R.H. Franck, and E.A. Phelps (2005), "Perceptions of moral character modulate the neural systems of reward during the trust game", *Nature Neuroscience*, 8: 1611-1618.

Delgado M.R., L.E. Nystrom, C. Fissel, D.C. Noll, and J.A. Fiez (2000), "Tracking the hemodynamic responses to reward and punishment in the striatum", *Journal of Neurophysiology*, 84: 3072-3077.

Demertzi A., A. Soddu, and S. Laureys (2013), "Consciousness supporting networks", *Current Opinion in Neurobiology*, 23: 1-6.

De Quervain, U. Fishbacher, V. Treyer, M. Schellhaller, U. Schnyder, A. Buck, and E. Fehr (2004), "The neural basis of altruistic punishment", *Science*, 305: 1254-1258.

De Waal F. (2018), *Mama's last hug*, Norton W.W. & Company (French translation: *La dernière etreinte*, Les liens qui libèrent, 2018).

Di Pellegrino G., L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti (1992), "Understandind motor events: A neurophysiological study", *Experimental Brain Research*, 91: 176-180.

Dolan R.J. and P. Dayan (2013), "Goals and habits", *Neuron*, 80(2); 312-325.

Doris M.C. and P.W. Glimcher (2004), "Activity in posterior parietal cortex is correlated with the relative subjective desirability of action", *Neuron*, 44(2): 365-378.

Doya K. (2008), "Modulators of decision making", *Nature Neuroscience*, 11(4): 410-416.

Doya K., S. Ishii, A. Pouget, and R. Rao (2007), *Bayesian brain: Probabilistic approach to neural coding and learning*, Cambridge, MA: MIT Press.

Doya K. and M. Kimura (2014), "The basal ganglia, reinforcement learning, and the encoding of value", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 17: 321: 334.

Dreher Jean-Claude and L. Tremblay (2017), *Decision neuroscience. An integrative perspective*, Academic Press.

Dunn B.D., T. Dalgleish, and A.D. Lawrence (2006), "The somatic marker hypothesis: A critical evaluation", *Neuroscience and Biobehavioral Review*, 30: 239-271.

Ekmann P. (1982), *Emotion in the human face. Studies in emotion and social interaction,* Cambridge: Cambridge University Press.

Ekmann P. (2003), *Emotions revealed: Understanding faces and feeling*, London: Weidenfeld and Nicolson.

Elliot R., Z. Agnew, and J.F. Deakin (2008), "Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in human", *European Journal of Neuroscience*, 27: 2213-2218.

Elliot R., K.J. Friston, and R.J. Dolan (2000), "Dissociable neural responses in human reward system", *Journal of Neuroscience*, 20: 6159-6165.

Engelmann J.B. and T.A. Hare (2018), "Emotions can bias decision-making processes by promoting specific behavior"", in Fox A.S., R.C. Lapate, A.J. Shackman, and R.J. Davidson (eds.), *The nature of emotion. Fundamental questions*, 2nd edition, Oxford University Press.

Engelmann J.B. and E. Fehr (2017), "The neurobiology of trust and social decision: the important role of emotions", in Van Lange P.A.M., B. Rockenbach, and T. Yamagishi (eds.), *Trust in social dilemmas,* New York: Oxford University Press.

Evans J.St.B.T. (2010), *Thinking twice: Two minds in one brain*, Oxford: Oxford University Press.

Evans J.St.B.T. and K. Frankish (eds.) (2009), *In two minds: Dual processes and beyond*, Oxford: Oxford University Press.

Evans J.St.B.T. and K.E. Stanovich (2013), "Dual-process theories of higher cognition: Advancing the debate", *Perspectives on Psychological Science*, 8(3): 223-241.

Fadiga L., L. Fogassi, G. Pavesi, and G. Rizzolatti (1995), "Motor facilitation during action observation: A magnetic stimulation study", *Journal of Neurophysiology*, 73: 2608-2611.

Farrer C. and C.D. Frith (2002), "Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency", *NeuroImage*, 15(3): 586-603.

Fehr E. (2009), "Social preferences and the brain", in Glimcher P. et al. (2009), Chap. 15: 215-232.

Fehr E. and C.F. Camerer (2007), "Social neuroeconomics: The neural circuitry of social preferences", *Trend in Cognitive Sciences*, 11(10): 419-427.

Fehr E. and I. Krajbich (2014), "Social prefrences and the brain", in Glimcher P. and E. Fehr (2014), Chap. 11: 193-218.

Fehr E. and A. Rangel (2011), "Neuroeconomics foundations of economic choices – Recent advances", *Journal of Economic Perspectives*, 25 (4): 3-30.

Fehr E. and K.M. Schmidt (2006), "The economics of fairness, reciprocity and altruism. Experimental evidence and new theories", in Kolm S.-C.and J. Mercier-Ythier (eds.), *Handbooks of the economics of giving, altruism and reciprocity*, vol. 1: Foundations, Amsterdam: Elsevier (Chap. 8).

Fehr T. (2013), "A hybrid model for neural representation of complex mental processing in the human brain", *Cognitive Neurodynamics*, 7(2): 89-103.

Fernandez-Dols J.-M. and J.A. Russel (eds.) (2017), *The science of facial expression*, Oxford University Press.

Fodor J.A. (1983), *The modularity of mind*, Cambridge and London: MIT Press.

Fox, A., R. Lapate, A. Shackman and R.J. Davidson (eds.) (2018), *The nature of emotion. Fundamental questions*, 2nd ed., Oxford: Oxford University Press.

Frijda N.H. (1986), *The emotions*, Cambridge: Cambridge University Press.

Friston K.J., J. Kilner, and L. Harrison (2006), "A free-energy principle for the brain", *Journal of Physiology-Paris*, 100: 70-87.

Friston K.J (2010), "The free-energy principle: A unified brain theory?" *Nature Review Neuroscience*, 11: 127-138.

Friston K.J. (2012), "The history of the future of the Bayesian brain", *Neuroimage*, 62: 1230-1233.

Frith U. (2001), "Mind blindness and the brain in autism", *Neuron*, 32(6): 969–979.

Frith U. (2012), "The role of metacognition in human social interactions", *Philosophical Transactions of the Royal Society of London, Series B, Biological sciences,* 367(1599): 2213-2223.

Frith U. and C.D. Frith (2003), "Development and neurophysiology of mentalizing", *Philosophical Transactions of the Royal Society B: Biological Sciences,* 358(1431): 459-473.

Frith U. and C.D. Frith (2006), "The neural basis of mentalizing", *Neuron*, 50(4): 531-534.

Frith U. and C.D. Frith (2007), "Social cognition in humans", *Current Biology*, 21: R724-R732.

Frith U and T. Singer (2008), "The role of social cognition in decision making", *Philosophical Transactions of the Royal Society B: Biological Sciences*", 263: 3875-3886.

Fudenberg D. and D.K. Lenine (2006), "A dual-self model of impulse control", *American Economic Review,* 96(5): 1449-1476.

Fuster J.M. (2008), *The prefrontal cortex: Anatomy, physiology and neuropsychology of the frontal lobe*, 4th ed., Philadelphia: Lippincott-Raven (1st ed.: 1980).

Fuster J.M. (2009), "Cortex and memory: Emergence of a new paradigm", *Journal of Cognitive Neuroscience*, 21(11): 2047-2072.

Fuster J.M. and S.L. Bressler (2012), "Cognitive activation: A mechanism enabling temporal integration in working memory", T*rends in Cognitive Sciences*, 16(4): 207-218.

Gallagher H.L. and C.D. Frith (2003), "Functional imaging of 'the theory of mind' ", *Trends in Cognitive Sciences*, 7(2): 77-83.

Gallagher H.L., A.I. Jack, A. Roepstorff, and C.D. Frith (2002), "Imaging the intentional stance in a competitive game", *NeuroImage*, 16: 814-821.

Gallese V., L. Fadiga, L. Fogassi, and G. Rizzolatti (1996), "Action regognition in the premotor cortex", *Brain*, 119: 593-609.

Galvan A., W.R. Stauffer, L. Acker, Y. El-Shamaykleh, K.-I. Inouei, S. Ohayon, and M.C. Schmid (2017), "Nonhuman primate optogenetics: Recent advance and future directions", *Journal of Neuroscience*, 37(45): 10894-10903.

Gazzaniga M.S. and G.R. Mangun (eds.) (2014), *The cognitive neurosciences*, 5th ed., London: MIT Press.

Genon S., R. Andrew, R. Langner, K. Amunts, and S.B. Eickhoff (2018), "How to characterize the functions of brain region", *Trends in Cognitive Sciences*, 22(4): 350-364.

Giorgetta C., A. Grecucci, N. Bonini, G. Coricelli, G. Demarchi, C. Braun, A.G. Sanfey (2013), "Waves of regret: A meg study of emotion and decision making", *Neuropsychologia*, 51: 38-51.

Gläscher J., R. Adolphs, H. Damasio, A. Bechara, D. Rudrauf, M. Calamia, L.K. Paul, and D. Tranel (2012), "Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex", *Proceedings of the National Academy of Sciences*, 109(36): 14681-14686.

Glimcher P.W. (2003), *Decisions, uncertainty, and the brain: The science of neuroeconomics*, Cambridge, MA: MIT Press.

Glimcher P.W. (2011a), *Foundations of neuroeconomic analysis*, New York: Oxford University Press.

Glimcher P.W. (2011b), "Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis", *Proceedings of the National Academy of Science USA*, 108(Suppl. 3): 15647-15654.

Glimcher P.W. (2014a), "Value-based decision making", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 20: 373-391.

Glimcher P.W. (2014b), "Introduction to neuroscience", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 5: 63-76.

Glimcher P.W., C.F. Camerer, E. Fehr, and R.A. Poldrack (eds.) (2009), *Neuroeconomics. Decision making and the brain,* Amsterdam: Elsevier.

Glimcher P.W., M.C. Dorris, and H.M. Bayer (2005), "Physiological utility theory and the neuroeconomics of choice", *Games and Economic Behavior*, 52(2): 213-256.

Glimcher P.W. and E. Fehr (eds.) (2014), *Neuroeconomics. Decision making and the brain*, 2nd ed., Amsterdam: Elsevier.

Glimcher P.W. and E. Fehr (2014), "Introduction: A brief history of neuroeconomics", in Glimcher and Fehr (eds.) (2014), xvii-xxviii.

Gold J.I. and H.R. Heekeren (2014), "Neural mechanism for perceptual decision making", in Glimcher and Fehr (eds.) (2014), Chap. 19: 355-372.

Gold J.I and M.N. Shadlen (2007), "The neural basis of decision making", *Annual Review of Neuroscience,* 30: 535-574.

Goldman A.I. (2006), *Simulating minds: the philosophy, psychology, and neuroscience of mindreading*, Oxford University Press.

Grabenhorst F. and E.T. Rolls (2011), "Value, pleasure, and choice systems in the ventral prefrontal cortex", *Trends in Cognitive Sciences*, 15: 56-67.

Green D.M. and J.A. Sweets (1966), *Signal detection theory and psychophysics*, New York: Wiley (reprinted: Los Altos, CA: Peninsula Publishing, 1988).

Guala F. (2008), "Experimental economics, History of", in Durlauf S. and L. Blume (eds.), *The new Palgrave dictionary of economics*, Palgrave-Macmillan, London, 152-156.

Gul F. and W. Pesendorfer (2005), "Random expected utility", *Econometrica*, 74: 121-146.

Gul F. and W. Pesendorf (2005/2008), "The case of mindless economics", in Caplin A. and A. Schotter (eds.), *The foundation of positive and normative economics: A handbook*, New York: Oxford University Press, 2008, Chap. 1: 3-39.

Hampton A.N., P. Bossaerts, and J.P. O'Doherty (2006), "The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans", *Journal of Neuroscience*, 26: 8360-8367.

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

Hare T.A., C.F. Camerer, D.T. Knoepfle, and A. Rangel (2010), "Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition", *Journal of Neuroscience*, 30: 583-590.

Hare T.A., C.F. Camerer, and A. Rangel (2009), "Self-control in decision-making involves modulation of the vmPFC valuation system", *Science*, 324: 646-648.

Hare T.A., O'Doherty, C.F. Camerer, W. Schultz, and A. Rangel (2008), "Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors", *Journal of Neuroscience*, 28(22): 5623-5630.

Hare T.A., W. Schultz, C.F. Camerer, J.P. O'Doherty, and A. Rangel (2011), "Transformation of stimulus value signals into motor commands during simple choice", *Proceedings of the National Academy of Science USA*, 107: 18120-18125.

Harrison G.W. (2008a), "Neuroeconomics: a critical reconsideration", *Economics and Philosophy*, 24: 303-344.

Harrison G.W. (2008b), "Neuroeconomics: rejoinder", *Economics and Philosophy*, 24: 533-544.

Harrison G.W and D. Ross (2010), "The methodology of neuroeconomics", *Journal of Economic Methodology*, 17(2): 185-196.

Herrnstein R.J. (1961), "Relative and absolute strength of response as a function of frequency of reinforcement", *Journal of the Experimental Analysis of Behavior*, 4(3): 267-272.

Herrnstein R.J.and D. Prelec (1991), "Melioration: A theory of distributed choice", *Journal of Economic Perspectives,* 5(3): 137-156.

Heukelom F. (2014), *Behavioral economics. A history*, New York: Cambridge University Press.

Hill M.R., E.D. Boorman, and I. Fried (2016), "Observational learning computations in neurons of the human anterior cingulate cortex", *Nature Communications*, 7: 12722.

Houk J.C., J.L. Adams, and A.G. Barto (1995), "A model of how the basal ganglia generate and use neural signals that predict reinforcemement", in Houk J.C., J.L. Davis, and D.G. Beiser (eds.), *Models of information processing in the basal ganglia*, Boston: MIT Press, 249-270.

Houser D. and K. McCabe (2009), "Experimental neuroeconomics and non-cooperative games", in Glimcher P. *et al.* (eds.) (2009), Chap. 5: 47-62.

Houser D. and K. McCabe (2014), "Experimental economics and experimental game theory", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 2: 19-34.

Howe M.W. and D.A. Dombeck (2016), "Rapid signaling in distinct dopaminergic axons during locomotion and reward", *Nature*, 535 (7613): 505-510.

Hsu M. and L. Zhu (2012), "Learning in games: Neural computations underlying strategic learning", *Recherches Economiques de Louvain*, 78(3): 47-72.

Huettel S.A. (2010), "Ten challenges for decision neuroscience", *Frontiers in Neuroscience*, 4: 171-185.

Huettel S.A., C.J. Stowe, E.M. Gordon, B.T. Warner, and M.L. Platt (2006), "Neural signatures of economic preferences for risk and ambiguity", *Neuron*, 49: 765-775.

Hutcherson C.A., B. Bushong, and A. Rangel (2015), "A neurocomputational model of altruistic choice and its implication", *Neuron*, 87: 451-462.

Hyde J.W., C.E. Simon, F. Ting, and J.I.Nikolaeva (2018), "Functional organization on the temporo-parietal junction for theory of mind in preverbal infants: A near-infrared spectroscopy study", *Journal of Neuroscience*, 38(18): 4264-4274.

Iacoboni M., I. Molnar-Szakacs, V. Gallese, G. Buccino, J.C. Mazziotta, and G. Rizzollatti (2005), "Grasping the intentiions of others with one's own mirror neuron system", *PLoS Biology*, 3: e79.

James W. (1890), *Principles of psychology*, New York: Holt, Rinehart and Winston.

Jung W.H., S. Lee, C. Lerman, and J. W. Kable (2016), "Amygdala functional and structural connectivity predicts individual risk tolerance", *Neuron*, 98: 394-404.

Kable J.W. (2011), "The cognitive neuroscience toolkit for the neuroeconomist: A functional review", *Journal of Neuroscience, Psychology, and Economics*, 4(2): 63-84.

Kable J.W. (2014), "Valuation, intertemporal choice, and self-control", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 10: 173-192.

Kable J.W. and P.W. Glimcher (2009), "The neurobiology of decision: Consensus and controversy", *Neuron*, 63: 733-745.

Kahneman D. (2003), "Maps of bounded rationality: Psychology for behavioral economics", *American Economic Review*, 93(5): 1449-1475.

Kahneman D. (2011), *Thinking, fast and slow*, London: FSG (French translation: *Système 1, Système 2. Les deux systèmes de la pensée*, Paris : Flammarion, 2012).

Kahneman D., P. Wakker, and R. Sarin (1997), "Back to Bentham? Explorations of experienced utility", *Quarterly Journal of Economics*, 112: 375-405.

Karton I. and T. Bachmann (2011), "Effect of prefrontal transcranial magnetic stimulation on spontaneous truth-telling", *Behavioral Brain Research*, 225(1): 209-214.

Kelso J.A.S and D.A. Engstrom (2006), *The complementary nature*, Cambridge, MIT Press.

Kilner J.M., A. Neal, N. Weiskopf, K.J. Friston, and C.D. Frith (2009), "Evidence of mirror neurons in human inferior frontal gyrus", *The Journal of Neuroscience*, 29: 10153-10159.

King-Casas B., D. Tomlin, C. Anen, C.F. Camerer, S.R. Quartz, and P.R. Montague (2005), "Getting to know you: reputation and trust in a two-person economic exchange", *Science*, 308: 78-83.

Knill D.C. and A. Pouget (2004), "The Bayesien brain: The role of uncertainty in neural coding and computation", *Trends in Neurosciences*, 27: 712-719.

Knoch D., L.R. Gianotti, A. Pascual-Leone, V. Treyer, M. Regard, M. Hoffman, and P. Brugger (2006a), "Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior", *Journal of Neuroscience*, 26: 6469-6472.

Knoch D., M.A. Nitsche, U. Fischbacher, C. Eisenegger, A. Pascual-Leone, and E. Fehr (2008), "Studying the neurobiology of social interaction with transcranial direct current stimulation – The example of punishing unfairness", *Cerebral Cortex*,18(9): 1987-1990.

Knoch D., A. Pascual-Leone, K. Meyer, V. Treyer, and E. Fehr (2006b), "Diminishing reciprocal fairness by disrupting the right prefrontal cortex", *Science*, 314: 829-832.

Knutson B.K., M.R. Delgado, and P.E.M. Phillips (2009), "Representation of subjective value in the striatum", in Glimcher P. et al. (eds.) (2009), Chap. 25: 389-406.

Knutson B.K., G.W. Fong, C.M. Adams, J.L. Varer, and D. Hommer (2001), "Dissociation of reward anticipation and outcome with event related fMRI", *Neuroreport*, 12: 3883-3687.

Knutson B.K. and S.A. Huettel (2015), "The risk matrix", *Current Opinion in Behavioral Sciences*, 5: 141-146.

Knutson B.K., S. Ricks, G.E. Wimmer, D. Prelec, and G. Loewenstein (2007), "Neural predictors of purchases", *Neuron*, 53: 147-156.

Knutson B.K., A. Westdorp, E. Kaiser, and D. Hommer (2000), "FMRI visualization of brain activity during a monetary incentive delay task", *NeuroImage*, 12: 20-27.

Kolling N., T.E.J. Behrens, M.K. Wittmann, and M.F.S. Rushworth (2016), "Multiple signals in anterior cingulate cortex", *Current Opinion in Neurobiology*", 37: 36-43.

Koppel L., D. Andersson, I. Morrison, D. Västfjäll, and G. Tinghög (2017) "The (null) effect of affective touch on betrayal aversion, altruism, and risk taking", *Frontiers in Behavorial Neurosciences*, 11 (December), Art 251: 1-11.

Kopton L.M. and P. Kenning (2014), "Near-infrared spectroscopy (NIRS) as a new tool for neuroeconomic research", *Frontiers in Human Neuroscience,* 8, 549.

Koziol L.F. and D.E. Budding (2009), *Subcortical structures and cognition: Implications for neuropsychological assessment*, New York: Springer.

Kragel P. and K.S. LaBar (2016), "Decoding the nature of emotion in the brain", *Trends in Cognitive Sciences*, 20(6): 444-455.

Krajbich I., K.C. Armel, and A. Rangel (2010), "Visual fixations and the computation and comparison of value in simple choice", *Nature Neurosciences*, 13: 1292-1298.

Krajbich I., T. Hare, B. Bartling, Y. Morishima, and E. Fehr (2015), "A common mechanism underlying food choice and social decision" *PLos Computational Biology*, 11, e1004371.

Krajbich I., D. Lu, C. Camerer and A. Rangel (2012), "The attentional drift-diffusion model extends to simple purchasing decisions", *Frontiers in Psychology*, 3:1-18

Krajbich I., B. Oud and E. Fehr (2014), "Benefits of neuroeconomic modeling: new policy interventions and predictors of preference", *American Economic Review*, 104(5): 501-506.

Krajbich I. and A. Rangel (2011), "Multialternative drift-diffusion model prdicts the relationship between visual fixations and choice in value-based decisions", *Proccedings of the National Academy of Science USA*, 108: 13852-13857.

Krupenye C., F. Kano, S. Hirata, J. Call, and M. Tomasello (2016), "Great apes anticipate that other individuals will act according to false belief", *Science*, 354(6308): 110-114.

Landreh A. and J. Bickle (2008), "Neuroeconomics, neurophysiology and the common currency hypothesis", *Economics and Philosophy*, 24: 419-429.

LeDoux J.E. (1996), *The emotional brain: Mysterious underpinnings of emotion life*, New York, NY: Simon and Schuster, 1996) (French translation: *Le cerveau des émotions*. *Les mystérieux fondements de notre vie émotionnelle*, Paris : Odile Jacob, 2005).

LeDoux J.E. (2007), "The amygdala", *Current Biology*, 17(20): R868-R874.

Lehrer J. (2009), *How we decide*, New York: Houghton Mifflin Harcourt Publising Company.

Lempert K.M. and E.A. Phelps (2014), "Neuroeconomics of emotion and decision making", in Glimcher P. and E. Fehr, eds. (2014).

Lengyel M. and P. Dayan (2007), "Hippocampal contributions to control: The third way", *Advances in Neural Information Processing Systems,* 20: 889-896.

Levy D.J. and P.W. Glimcher (2012), "The root of all value: A neural common currency for choice", *Current Opinion in Neurobiology*, 22: 1027-1038.

Levy I., J. Snell, A.J. Nelson, A. Rustichini, and P.W. Glimcher (2010), "Neural representation of subjective value under risk and ambiguity", *Journal of Neurophysiology*, 103: 1036-1047.

Lipps T. (1903), *Asthetic. Psychologie des Schönen und der Kunst*, 2 vol., Hambourg et Leipzig: Leopold Voss.

Lieberman M.D. (2012), "A geographycal history of social cognitive neuroscience", *NeuroImage*, 61: 432-436.

Lindquist K.A and L.F. Barrett (2012), "A functionbal architecture of the human brain: Emerging insights from the science of emotion", *Trends in Cognitive Sciences*, 16(11): 533-540.

Lindquist K.A., A.B. Satpute, T.D. Wagner, J. Weber, and L.F. Barrett (2016), "The brain basis of positive and negative affect. Evidence from a meta-analysis of the human neuroimaging literature", *Cerebral Cortex*, 26: 1910-1922.

Lombardo M.V. et al (2010), "Shared neural circuits for mentalizing about the self and others", *Journal of Cognitive Neuroscience*, 22(7): 1623-1639.

Loewenstein G and E. Haisley (2008), "The economist as therapist: methodological ramifications of 'light' paternalism", *in* A. Caplin et A. Schotter (eds.), *Perspectives on the future of economics: positive and normative foundations,* Oxford: Oxford University Press, 210-245.

Loewenstein G. and T. O'Donoghue (2004), "Animal spirits: Affective and deliberative processes in economic behavior", Working paper 04-14, Center for Analytic Economics, Cornell University.

Luce, R. D. (1959/2005) *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley (reprinted by Dover Publications).

Mackey S. and M. Petrides (2010), "Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains", *European Journal of Neuroscience,* 32: 1940-1950.

Mackey S. and M. Petrides (2014), "Architecture and morphology of the human ventromedial prefrontal cortex", *European Journal of Neuroscience*, 40(5): 2777-2796.

MacLean P.D. (1952), "Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). Electroencephalography: The basal and temporal region", *Yale Journal of Biology and Medicine*, 22: 407-418.

MacLean P.D. (1970), *The triune brain in evolution: Role in paleocerebral function plenum*, New York: Plenum Press (French translation: *Les trois cerveaux de l'homme*, Paris : Robert Laffont).

Macmillan M. (2000), *An odd kind of fame: Stories of Phineas Gage*, Cambridge, MA: MIT Press.

Macmillan N.A. and C.D. Creelmann (2004), *Detection theory: A user's guide*, Lawrence Erlbaum, Mahwah, NJ.

Maguire et al (1998) (human navigation network) A COMPLETER p. 8

Mahy C.E.V., L.J. Moses, and J.H. Pfeifer (2014), "How and where: Theory-of-mind in the brain", *Developmental Cognitive Neuroscience*, 9: 68-81.

Maia T.V. and M.J. Frank (2011), "From reinforcement learning models to psychiatric and neurological disorders", *Nature Neuroscience*, 14(2): 154.

Mars R.B., J. Sallet, U. Schüffelgen, S. Jbabdi, I. Toni, and M.F.S. Rushworth (2012), "Connectivity-based subdivisions of the human right temporoparietal junction area: Evidence for different areas participating in different cortical networks", *Cerebral cortex*, 22(8): 1894-1903.

Mayr U., W.T. Harbaugh, and D. Tankersley (2009), "Neuroeconomics of charitable giving and philanthropy", in Glimcher P. *et al*. (eds.) (2009), Chap. 20: 303-320.

McCabe K.A. (2003), "Neuroeconomics", *in* Nadel L. (ed.), *Encyclopedia of cognitive science*, Nature publishing Group, New York: Macmillan Pub., 294-298.

McCabe K.A., D. Houser, L. Ryan, V. Smith, and T. Trouard (2001), "A functional imaging study of cooperation in two-person reciprocal exchange", *Proceedings of the National Academy of Science USA*, 98: 11832-11835.

McCullogh W. (1965), *Embodiments of mind*, Cambridge, MA: MIT Press.

McCullogh W.S. and W. Pitts (1943), "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, 5(4): 115-133.

McFadden D. (1974), "Conditional logit analysis of qualitative choice behaviour" in Zarembka P. (ed.), *Frontier in econometrics*, New York: Academic Press, 105-142.

McFadden D. (2005), "Revealed stochastic preference: a synthesis", *Economic Theory*, 26: 245-264.

Melnikoff D.E. and J.A. Bargh (2018), The mythical number two, *Trends in Cognitive Sciences*, 22: 280-293.

Miller E.K. (2000), "The prefrontal cortex and cognitive control", *Nature Review neurioscience*, 1: 59-65.

Miller E.K. and Cohen J.D. (2001), "An integrative theory of prefrontal cortex function", *Annual Review of Neuroscience*, 24: 167-202.

Moll J., F. Krueger, R. Zahn, M. Pardini, R. Oliveira-Souza, and J. Grafman (2006), "Human fronto-mesolimbic networks guide decisions about charitable decisions", *Proceedings of the National Academy of Science USA,* 103: 15623-15628.

Mollenberghs P., R. Cunnington, and J.B. Mattingley (2012), "Brain region with mittot properties: A meta-analysis of 125 human fMRI studies", *Neuroscience and Biobehavioral Reviews*, 36: 341-349.

Mollenberghs P, H. Johnson, J.D. Henrey, and J.B. Mattingley (2016), "Understanding the mind of others: A neuroimagenry meta-analysis", *Neuroscience and Biobehavioral Reviews*, 65: 276-291.

Montague P.R. (2007), "Neuroeconomics: A view from neuroscience", *Functional Neurology*, 22(4): 219-234.

Montague P.R. and G.S. Berns (2002), "Neural economics and the biological substrates of valuation", *Neuron*, 36: 265-284.

Montague P.R., P. Dayan, C. Person, and T.J. Sejnowski (1995), "Bee foraging in uncertain environments using predictive hebbian learning", *Nature,* 377: 725-728.

Montague P.R., P. Dayan, and T.J. Sejnowski (1996), "A framework for mesencephalic dopamine systems based on predictive Hebbian learning", *Journal of Neuroscience*, 16: 1936-1947.

Montet C. and D. Serra (2003), *Game theory and economic*, New York and London: Palgrave Macmillan (Chinese translation, Economy & Management Publishing House, 2004).

Morrison, S.E. and C.D. Salzman (2010), "Revaluing the amygdala", *Current Opinion in Neurobiology*, 20: 221-230.

Mukamel R., A.D. Ekstrom, J. Kaplan, M. Iacoboni, and I. Fried (2010), "Single-neuron responses in humans during execution and observation of actions", *Current biology*, 20(8): 750-756.

Murray E.A. (2007), "The amygdala, reward and emotion", *Trend in Cognitive Sciences*, 11: 489-497.

Murray E.A., J.P. O'Doherty, and G. Scoeenbaum (2007), "What we know and do not know about functions of the orbitofrontal cortex after 20 years of cross-species studies", *Journal of Neuroscience*, 27: 8166-8169.

Naqvi N., B. Shiv and A. Bechara (2006), "The role of emotion in decision making. A cognitive neuroscience perspective", *Current directions in Psychological Science*, 15(5): 260-264.

Nave G., C. Camerer, and M. McCullugh (2015), "Does oxytocin increase trust in humans? A critical review of research", *Perspectives on Psychological Science,* 10: 772-789.

Nicolle A., D.R. Bach, J. Driver, R.J. Dolan (2011), "A role for striatum in regret-related choice repetition", *Journal of Cognitive Neuroscience,* 23(4): 845-856.

Nieoullon A. (2002), "Dopamine and the regulation of cognition and attention", *Progress in Neurobiology*, 67: 53-83.

Niv Y. and P.R. Montague (2009), "Theoretical and empirical studies of learning", in Glimcher P. *et al.* (eds.) (2009), Chap. 22: 331-351.

Ochsner K.N. and M.D. Liebermann (2001), "The emergence of social cognitive neuroscience", *American Psychologist*, 56: 717-734.

O'Doherty J.P., B. Shiv, and A. Rangel (2008), "Marketing actions can modulate neural representations of experienced pleasantness", *Proceedings of the National Academy of Sciences*, 105(3): 1050-1054.

Okon-Singer H., D.M. Stout, M.D. Stockbridge, M.D. Gamer, M. Fox, and A.J. Shackman (2018), "The interplay of emotion and cognition", in A. Fox, R. Lapate, A. Shackman and R.J. Davidson (eds.) (2018).

Olds J. and P. Milner (1954), "Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain", *Journal of Comparative and Physiological Psychology*, 47(6), 769-776.

Oliver L.D., J.B. Vieira, R.W.J. Neufeld, I. Dziobek, and D.G.V. Mitchell (2018), "Greated involvement of action simulation mechanisms in emotional vs cognitive empathy, *Social Cognitive and Affective Neuroscience*, 15 February, 367-380.

Olson I.R., D. McCoy, E. Klobusicky, and L.A. Ross (2013), "Social cognition and the anterior temporal lobes: A review and theoretical framework", *Social Cognitive and Affective Neuroscience*, 8: 123-133.

Ongür D., A.T. Ferry, and J.L. Price (2003), "Architectonic subdivision of the human orbital and medial prefrontal cortex", *Journal of Comparative Neurology*, 460: 425-449.

Ongür D.and J.L.Price (2000), "The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans", *Cerebral Cortex*, 10: 206-219.

Oullier O. (2010), "The useful brain: How neuroeconomics might change our views on rationality and a couple of other things", in Michel-Kerjan E. and P. Slovic (eds.), *Irrational economist: Making decisions in a dangerous world*, Public Affairs ed.

Padoa-Schioppa C. (2009), "Range-adapting representation of economic value in the orbitofrontal cortex", *Journal of Neuroscience*, 29: 14004-14014.

Padoa-Schioppa C. (2011), "Neurobiology of economic choice: A good-based model", *Annual Review of Neuroscience*, 34: 333-359.

Padoa-Schioppa C. and J.A. Assad (2006), "Neurons in the orbitofrontal cortex encode economic value", *Nature,* 441: 223-226.

Padoa-Schioppa C. and J.A. Assad (2008), "The representation of economic value in the orbitofrontal cortex is invariant for changes in menu", *Nature Neuroscience*, 11: 95-102.

Padoa-Schioppa C. and K.E. Conen (2017), "Orbitofrontal cortex: A neural circuit for economic decisions", *Neuron*, 96(4): 736-754.

Palmer J., A.C. Huk, and M.N. Shadlen (2005), "The effect of stimulus strength on the speed and accuracy of a perceptual decision", *Journal of Vision*, 5(5): 376-404.

Panksepp J. (1998), *Affective neuroscience. The foundations of human and animal emotions*, New York, NY: Oxford University Press.

Papez J.W. (1937), "A proposed mechanism of emotion", *Archives of Neurology and Psychiatry*, 38(4): 725-743.

Passingham R.E. and S.P. Wise (2012), *The neurobiology of prefrontal cortex: anatomy, evolution, and the origin of insight*, Oxford University Press.

Pavlov I.P. (1927), *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*, New York: Dover.

Pennycook G., W. De Neys, J. St. B.T. Evans, K.E. Stanovitch, and V.A. Thompson (2018), The mytical dual-process typology, *Trends in Cognitive Sciences*, 22: 667-668.

Pessiglione M., B. Seymour, G. Flandin, R.J. Dolan, and C.D. Frith (2006), "Dopamine-dependent prediction errors underpin reward-seeking behavior in humans, *Nature*, 442 (7106): 1042-1045.

Pessoa L. (2008), "On the relationship between emotion and cognition", *Nature Reviews Neuroscience*, 9: 148-158.

Pessoa L. (2010), "Emotion and cognition and the amygdalia: From 'what is it?' To 'what's to be done?'", *Neuropsychologia*, 48: 3416-3429.

Pessoa L. (2013), *The cognitive-emotional brain: From interactions to integration*, MIT Press.

Pessoa L. (2017a), "The emotional brain", in Conn P.M. (ed.), *Conn's translational neuroscience*, Elsevier, 635-656.

Pessoa L (2017b), "A network model of the emotional brain", *Trends in Cognitive Sciences*, 21(5): 357-371.

Petrides M. and D.N. Pandya (1994), "Comparative cytoarchitectonic analysis of the human and the macaque frontal cortex, in Boler F. and J. Grafman (eds.), *Handbook of neuropsychology*, Amsterdam: Elsevier, 17-58.

Phelps E.A. (2009), "The study of emotions in neuroeconomics", in Glimcher P. *et al*. (eds.) (2009), Chap.16: 233-250.

Phelps E.A. and J.E. LeDoux (2005), "Contribution of the amygdala to emotion processing: from animal models to human behavior", *Neuron*, 48: 175-187.

Plassman H., J. O'Doherty, and A. Rangel (2007), "Orbitofrontal cortex encodes willingness to pay in everyday economic transactions", *Journal of Neuroscience*, 27: 9984-9988.

Platt M.L. and P.W. Glimcher (1999), "Neural correlates of decision variables in parietal cortex", *Nature*, 400: 233-238.

Platt M.L. and S.A. Huettel (2008), "Risky business: the neuroeconomics of decision making under uncertainty", *Nature Neuroscience*, 11: 398-303.

Plutchik R. (1980), *Emotion: a psychoevolutionary synthesis*, Harper & Row.

Poldrack R.A. (2006), "Can cognitive processes be inferred from neuroimaging data?" *Trend in Cognitive Sciences,* 10(2): 59-63.

Poldrack R.A., C.I. Baker, J. Durnez, K.J. Gorgolewski, P.M. Matthews, M.R. Munato, T.E. Nichols, J.B. Poline, E. Vul, and T. Yakoni (2017), "Scanning the horizon: Towards transparent and reproductible neuroimaging resource", *Nature Review Neuroscience*, 18(2): 115-126.

Poldrack R.A., J.A. Mumford, and T.E. Nichols (2011), *Handbook of functional MRI data analysis*, Cambridge University Press.

Povinelli D.J. and J.M. Bering (2002), "The mentality of apes revisited", *Current directions in psychological Science*, 11(4): 115-119.

Premack D. and G. Woodruff (1978), "Does the chimpanzee have a 'theory of mind'?" *Behavioral and Brain Sciences*, 4: 515-526.

Pribam K.H. (1973), "The primate frontal cortex – Executive of the brain", in Pribam K.H. and A.R. Luria (eds.), *Psychophysiology of the frontal lobes,* New York: Academic Press, Chap. 14: 293-314.

Purves D., G.J. Augustine, D. Fitzpatrick, W.C. Hall, A-S. LaMantia, J.O. McNamara, and L.E. White (eds.) (2011), *Neuroscience,* 5th ed., Sinauer Associates, Inc.

Raichle M.E (2015), "The brain's default mode network", *Annual Review of Neuroscience*, 38: 433-447.

Ramnani N. and A.M. Owen (2004), "Anterior prefrontal cortex: insights into function from anatomy and neuroimaging", *Nature Reviews Neuroscience*, 5:185-194.

Rangel A. (2009), "The computation and comparison value in goal-directed choice", in Glimcher P. et al. (eds.) (2009), Chap. 28: 425-440.

Rangel A., C. Camerer, and P.R. Montague (2008), "A framework for studying the neurobiology of value-based decision making", *Nature Reviews Neuroscience*, 9: 545-556.

Rangel A and J.A. Clithero (2014), "The computation of stimulus values in simple choice", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 8: 125-148.

Rangel A. and T. Hare (2010), "Neural computations associated with goal-directed choice", *Current Opinion in Neurobiology,* 20: 262-270.

Ratcliff R. (1978), "A theory of memory retrieval", *Psychological Review*, 85(2): 59-108.

Real L.A. (1991), "Animal choice behavior and the evolution of cognitive architecture", *Science*, 253: 980-986.

Reiman M. and A. Bechara (2010), "The somatic marker framework as a neurological theory of decision-making: Review, conceptual comparisons, and future neuroeconomics research", *Journal of Economic Psychology*, 31: 767-776.

Rescorla R.A. and A.R. Wagner (1972), "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement", in Black A.H. and W.F. Prokasy (eds.), *Classical conditioning II: current research and theory*, New York: Appleton Century Crofts, 64-99.

Richter T., A.J. Shackman, and H. Okon-Singer (2017), "The neurobiology of emotion-cognition interactions", in Baune B. and C. Harmer (eds.), *Cognitive dimensions of major depressive disorder: Cognitive, emotional and social cognitive processes*, New York: Oxford University Press.

Riedl R.and A. Javor (2012), "The biology of trust: Integrating evidence from genetics, endocrinology and functional brain imaging", *Journal of Neuroscience, Psychology, and Economics*, 5(2): 63-91

Rilling J.K, D. Gutman, T. Zeh, G. Pagnoni, G.S. Berns, and C.D. Kilts (2002), "A neural basis for social cooperation", *Neuron*, 35: 395-405.

Rilling J.K, B. King-Casas, and A.G. Sanfey (2008), "The neurobiology of social decision-making", *Current Opinion in Neurobiology*, 18: 159-165.

Rilling J.K. and A.C. Sanfey (2011), "The neurosciences of social decision-making", *Annual Review of Psychology,* 62: 43-48.

Rilling J.K., A.G. Sanfey, J.A. Aronso, L.E. Nystrom, and J.D. Dohen (2004), "The neural correlate of theory of mind within interpersonal interactions", *NeuroImage*, 22(4): 1694-1703.

Rizzolatti G. and L. Craighero (2004), "The mirror-neuron system", *Annual Review of Neuroscience*, 27: 169-192.

Rizzolatti G., L. Fadiga, V. Gallese, and L. Fogassi (1996a), "Premotor cortex and the regognition of motor actions", *Cognitive Brain Research*, 3: 131-141.

Rizzolatti G., L. Fogassi, and V. Gallese (2001), "Neurophysiological mechanisms underlying the understanding and imitation of action", *Nature Reviews Neuroscience*, 2: 661-670.

Rizzolatti G., L. Fogassi, and V. Gallese (2009), "The mirror neuron system: A motor-based mechanism for action and intention understanding", in M.S. Gazzaniga (ed.), *The cognitive neurosciences*, 4th ed., Cambridge: The MIT Press, 635-640.

Rizzolatti G., L. Fogassi, M. Matteli, V. Bettirnadi, E. Paulesu, D. Perani, and F. Fazio (1996b), "Localization of grasp representations in humans by PET: 1. Observation versus execution", *Experimental Brain Research*, 111: 246-252.

Rizzolatti G. and C. Sinigaglia (2006), *So quel che fai. Il cervelloche agisce e i neuroni specchio*, Milano : Raffaello Cortina Editore (French translation: *Les neurones miroirs*, Paris : Odile Jacob, 2008)

Roitman J.D. and M.N. Shadlen (2002), "Response of neurons in the lateral intraparietal aere during a combined visual discrimination reaction time task", *Journal of Neuroscience*, 22: 9475-9489.

Rolls E.T. (1999), *The brain and emotion*, Oxford, UK: Oxford University Press.

Rolls E.T. (2007), *Emotion explained*, Oxford, UK: Oxford University Press.

Rolls E.T. (2014), *Emotion and decision-making explained*, Oxford, UK: Oxford University Press.

Rolls E.T. and Deco G. (2010), *The noisy brain: Stochastic dynamics as a principle of brain function*, Oxford: Oxford University Press.

Rose N. and J.M. Abi-Rached (2013), *Neuro: the new brain sciences and the management of the mind*, Harvard University Press.

Ross D. (2005), *Economic theory and cognitive science: microexplanation*, Cambridge, MA: MIT Press.

Ross D. (2008), "Two styles of neuroeconomics", *Economics and Philosophy*, 24(3): 473-483.

Rubinstein A. (2008), "Comments on neuroeconomics", *Economics and Philosophy,* 24(3): 485-494.

Rubinstein A. (2013), "Response time and decision making: an experimental study", *Judgment and Decision Making*, 8(5): 540-551.

Rudebeck P.H., J.A. Ripple, A.R. Mitz, B.B. Averbeck, and E.A. Murray (2017), "Amygdala contribution to stimumus-reward encoding in the macaque medial and orbital frontal cortex during harming", *Journal of neuroscience*, 37: 2186-2202.

Ruff C.C. and E. Fehr (2014), "The neurobiology of rewards and values in social decision", *Nature Review Neuroscience*, 15: 549-562.

Ruff C.C. and S.A. Huettel (2014), "Experimental methods in cognitive neuroscience", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 6: 77-108.

Ruff C.C., G. Ugazio, and E. Fehr (2013), "Changing social norm compliance with noninvasive brain stimulation", Science, 342: 482-484.

Rushworth M.F.S., M.A.P. Noonan, E.D. Boorman, M.E. Walton, and T.E. Behren (2011), "Frontal cortex and reward-guided learning and decision-making", *Neuron*, 70: 1054-1069.

Rutledge R.B., N. Skandali, P. Dayan, R.J. Dolan (2015), "Dopaminergic modulation of decion making and subjective well-being", *Journal of Neurosciences*, 35: 9811-9822.

Saarimäki H., L.F. Ejtehadian, E. Glerean, L. Jääkeläinen, P. Vuilleumier, M. Sams, and L. Nummenmea (2018), "Distributed affective space represents multiple emotion categories across the human brain", *Social Cognitive and Affective Neuroscience*, 13(5): 471-482.

Sacks O. (1995), *An anthropologist on Mars*. *Seven paradoxical tales*, London: Picador (French translation: *Un anthropologue sur Mars*, Paris: Seuil, 1996).

Sanfey A. (2007), "Social decision-making: Insights from game theory and neuroscience", *Science*, 318: 598-602.

Sanfey A.G., G. Loewenstein, S.M. McClure, and J.D. Cohen (2006), "Neuroeconomics: cross-currents in research on decision-making", *Trends in Cognitive Sciences*, 10(3): 108-116.

Sanfey A.G. and J.K. Rilling (2011), "Neural bases of social decision making", in Vartorian O.and D.R. Mandel (eds.), *Neuroscience of decision making*, New York: Psychology Press, 2011, Chap. 9: 223-242.

Sanfey A., J. Rilling, J. Aronso, L.E. Nystrom, and J.D. Cohen (2003), "The neural basis of economic decision-making in the ultimatum game", *Science*, 300: 1755-1758.

Santos L.R. and M. L. Platt (2014), "Evolutionary anthropological insights into neuroeconomics: What non-human primates can tell us about humans decision-making strategies", in Glimcher and Fehr (eds.) (2014), Chap. 7: 109-122.

Scherer K.R. (2005), "What ar emotions? And how can they be measured?", *Social Science Information*, 44: 695-729.

Schinger H.D. (2009), "Theory of mind: an overwiew and behavioural perspective", *Psychological Record*, 59: 535-448.

Schnell K., S. Bluschke, B. Konradt, and H. Walter (2011), "Functional relations of empathy and mentalizing: An fMRI study on the neural basis of cognitive empathy", *NeuroImage*, 54(2): 1743-1754.

Schultz W. (2008), "Introduction. Neuroeconomics: The promise and the profit", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363: 3767-3769.

Schultz W. (2009), "Midbrain dopamine neurons: A retina of the reward system?" in Glimcher P. *et al.* (eds.) (2009), Chap. 21: 323-329.

Schultz W. (2010), "Dopamine signals for reward value and risk: Basic and recent data", *Behavioral and Brain Function*, 6:24.

Schultz W. (2013), "Updating dopamine reward signals", *Current Opinion in Neurobiology*, 23: 229-238.

Schultz W. (2016), "Neural reward and decision signals: From theories to data, *Physiological Review*, 95: 853-951.

Schultz W., P. Dayan, and P.R. Montague (1997), "A neural substrate of prediction and reward", *Science*, 275 (5306): 1593-1599.

Schurz M., J. Radua, M. Aichhorn, F. Richlan, and J. Perner (2014), "Fractionating theory of mind: A meta-analysis of functional brain imaging studies", *Neuroscience and Biobehavioral Reviews*, 42C: 9-34.

Schutt R.K., L.J. Seidman, and M.S. Keshavan (eds.) (2015), *Social neuroscience: Brain, mind, and society*, Cambridge and London: Harvard University Press.

Semendeferi K., E. Armstrong, A. Schleicher, K. Zilles, and G.W.V. Hoesen (2001), "Prefrontal cortex in humans and apes: a comparative study of aera 10", *American Journal of Physical Anthropology*, 114: 224-241.

Serra D. (2012), « Un aperçu historique de l'économie expérimentale : des origines aux évolutions récentes », *Revue d'Economie Politique*, 122(5) : 749-786.

Serra D. (2016), *Neuroeconomie*, Paris : Economica.

Serra D. (2017), *Economie comportementale*, Paris : Economica.

Shamay-Tsoory S.G. (2011a), "The neural basis for empathy", *Neuroscientist*, 17(1): 18-24.

Shamay-Tsoory S.G. (2011b), "Empathic processing: iIs cognitive and affective dimensions and neuroanatomical basis", in Decety J. and W. Ickes (eds.), *The social neuroscience of empathy*, The MIT Press, p. 215.

Sherrington C.S. (1906), *The integrative action of the nervous system*, New Haven: Yale University Press (re-issued in 1947 with a new foreword by Sherrington).

Simon H.A. (1955), "A behavioral model of rational choice", *Quarterley Journal of Economics*, 69(1): 99-118.

Simon H.A. (1979), *Models of thought*, New Haven: Yale University Press.

Singer T. (2009), "Understanding others: Brain mechanisms of theory of mind and empathy", in Glimcher P. *et al.* (eds.), Chap. 17: 251-268.

Singer T. (2012), "The past, present and future of social neuroscience: A European perspective", *NeuroImage*, 61: 437-449.

Singer T. and A. Tusche (2014), "Understanding others: Brain mechanisms of theory of mind and empathy", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 27: 513-532.

Skinner B.F. (1945), "The operational analysis of psychological terms", *Psychological Review*, 52: 270-277

Skinner B.F. (1953), *Science and human behaviour*, New York: Macmillan.

Sloman, S.A. (2002), "Two systems of reasoning", *in* Gilovich T., D. Griffin, and D. Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment,* Cambridge: Cambridge University Press, 379-398.

Smith V.L. (2008), *Rationality in economics: constructivist and ecological forms*, Cambridge: Cambridge University Press.

Soltani A. and X.J. Wang (2008), "From biophysics to cognition: reward-dependant adaptive choice behaviour", *Current Opinion in Neurobiology*, 18: 209-216.

Sperdutti M., S. Guionnet, P. Fossati, and J. Nadel (2014), "Mirror neuron system and mentalizing system connect during online social interaction", *Cognitive Processing*, 15(3): 307-316.

Spunt R.P. and R. Adolphs (2014), "Validating the why/how contrast for functional MRI studies of theory of mind", *NeuroImage*, 99(1): 301-311.

Stanovich K.E. and R.F. West (2000), "Advancing the rationality debate", *Behavioral and Brain Sciences*, 23: 701-726.

Strotz R.H. (1955), "Myopia and inconsistency in dynamic utility maximization", *Review of Economic Studies,* 23: 165-180.

Sunstein C. (2014), *Why nudge? The politics of libertarian paternalism*, Yale University Press.

Sunstein C. et R. Thaler (2003), "Libertarian paternalism is not an oxymoron", *The University of Chicago Law Review*, 70(4): 1159-1202.

Sutton R.S. (1988), "Learning to predict by the method of temporal differences", *Machine Learning*, 3: 9-44.

Sutton R.S. and A.G. Barto (1998), *Reinforcement learning: an introduction*, Cambridge, MA: MIT Press.

Svorencik A. (2015), *The experimental turn in economics. A history of experimental economics*, PhD Thesis, University of Utrecht.

Taya F. (2012), "Seeking ambiguity: A review on neuroimaging studies on decision making under ambiguity", *Recherches Economiques de Louvain*, 78(3): 83-100.

Thaler R.H. and C.R. Sunstein (2003), "Libertarian paternalism', *American Economic Review*, 93(2): 175-179.

Thaler R.H. and C.R. Sunstein (2008), *Nudge. Improving decisions about health, wealth, and hapiness*, Penguin Books (revised and expanded edition, 2009).

Thiele A. and M.A. Bellgrove (2018), "Neuromodulation of attention", *Neuron*, 97: 769-785.

Thorndike E.L. (1911), *Animal intelligence: Experimental studies*, New York: Macmillan.

Tobler P.N. and E.U. Weber (2014), "Valuation for risky and uncertain choices", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 9: 149-172.

Tom S.M., C.R. Fox, C. Trepel, and R.A. Poldrack (2007), "The neural basis of loss aversion in decision making under risk", *Science*, 315(5811): 515-518.

Tomasello M. (2000), *The cultural origin of human cognition*, Cambridge, MA: Harvard University Press.

Trope Y. and R. Gaunt (2000), "Processing alternative explanations of behavior: Correction or integration? *Journal of Personality and Social Psychology*, 79(3): 344.

Tversky A. and D. Kahneman (1981) "The framing of decision and the psychology of choice", *Science*, 211: 453-458.

Tversky A. and D. Kahneman (1986), "Rational choice and the framing of decisions", *Journal of Business*, 59: 5251-5278.

Van Overwalle F. (2009), "Social cognition and the brain: A meta-analysis", *Human Brain Mapping*, 30(3): 829-858.

Van Overwalle F. and K. Baetens (2009), "Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis", *NeuroImage*, 48(3): 564-584.

Vromen J. (2007), "Neuroeconomics as a natural extension of bioeconomics: The shifting scope of standard economic theory", *Journal of Bioeconomics*, 9(2): 145-167.

Vromen J. (2011), "Neuroeconomics: Two camps gradually converging. What can economics gain from it?" *International Review of Economics*, 58: 267-285.

Wake S.J. and K. Izuma (2017), "A common neural code for social and monetary rewards in the human striatum", *Social Cognitive and Affective Sciences*, 12(10): 1558-1564

Wallis J.D. (2012), "Cross-species studies of orbitofrontal cortex and value-based decision-making", *Nature Neuroscience*, 15(1): 13-19

Wallis J.D. and M.E. Rushworth (2014), "Integrating benefits and costs in decision making", in Glimcher P. and E. Fehr (2014), Chap. 22, 411-433.

Wang X-J. (2014), "Neuronal circuit computation of choice", in Glimcher P. and E. Fehr (eds.) (2014), Chap. 23: 435-454.

Watson J. (1913), "Psychology as the behavioural views it", *Psychological Review*, 20.

Whalen P.J. and E.A. Phelps (eds.) (2009), *The human amygdala*, New York: Guilford.

Wimmer H. and J. Perner (1983), "Beliefs abiout beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception", *Cognition*, 13: 103-128.

Wise R.A. (1982), "Neuroleptics and operant behaviour: The anhedonia hypothesis", *Behavioral and Brain Sciences*, 5: 39-53.

Wong C. and J. Gallate (2012), "The function of the anterior temporal lobe: A review of the empirical evidence", *Brain Research*, 1449: 94-116.

Woodford M. (2014), "Stochastic choice: An optimizing neuroeconomic model", *American Economic Rewiew*, 104(5): 495-500.

Young L., D. Dodell-Feder, and R. Saxe (2010), "What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind", *NeuroPsychologia*, 48: 2658-2664.

Zak P.J. (2004), "Neuroeconomics", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1451): 1737-1748.

Zaki J. and K. Ochsner (2012), "The neuroscience of empathy: Progress, pitfalls and promise", *Nature Neuroscience,* 15(5): 675-680.

Zald D.H. and S.L. Rauch (eds.) (2006), *The orbitofrontal cortex*, Oxford University Press.

Zalocusky K.A. and K. Deisseroth (2014), "Optogenetics in the study of mammalian behaviour", in Gazzaniga M.S. and G.-R. Mangun, *The cognitive neuroscience*, 5th edition, Chap. 82, The MIT Press.

CEE-M Working Papers[1] - 2019

WP 2019 - 01	**Adrien Nguyen-Huu**. & Antonin Pottier «Hicksian Traverse Revisited: Conditions for the Energy Transition»
WP 2019 - 02	Oumar Mbodji, **Adrien Nguyen-Huu** & Traian A. Pirvu «Optimal Sharing Rule for a Household with a Portfolio Management Problem»
WP 2019 - 03	Tristan Cotty, Tristan, Elodie Maitre d'Hotel &**Julie Subervie** «Inventory credit to enhance food security in Africa»
WP 2019 - 04	**Raphael Soubeyran** «Incentives, Pro-social Preferences and Discrimination»
WP 2019 - 05	Adrien Fabre, Mouez Fodhaz & **Francesco Ricci** «Optimal Timing of Energy Production»
WP 2019 - 06	**Claude Bismut** & **Ismaël Rmajo** «A world of low interest rates»
WP 2019 – 07	**Gabriela Demarchi**, **Julie Subervie**, Fernando Palha Leite & Jean-Paul Laclau « Farmers' preferences for water-saving strategies in brazilian eucalypt plantation »
WP 2019 - 08	Arnaud Tognetti, David Doat, **Dimitri Dubois** & Rustam Romaniuc « Does the presence of a physically disabled erson in the group increase cooperation? An experimental test of the empathyaltruism hypothesis »
WP 2019 – 09	Isabelle Tritsch, **Gwenole Le Velly**, Benoit Mertens, Patrick, Meyfroidt, Christophe Sannier, Jean-Sylvestre Makak & Kenneth Houngbedji « Do Forest-Management Plans and FSC Certification Reduce deforestation in the Congo Basin? »
WP 2019 - 10	**Jean-Michel Salles** « Valuing the loss and damage from climate change: a review of some current issues »
WP 2019 - 11	**Can Askan Mavi** « Can harmful events be another source of environmental traps? »
WP 2019 - 12	**Daniel Serra** « Neuroeconomics and Modern Neuroscience»

[1] **CEE-M Working Papers**
Contact : laurent.garnier@inra.fr/ 04 99 61 31 21