Theory and Decision (2021) 91:1–80
https://doi.org/10.1007/s11238-021-09830-3

Decision-making: from neuroscience to neuroeconomics—an overview

Daniel Serra[1]

Accepted: 8 June 2021 / Published online: 28 June 2021

Abstract

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". Without the availability of an extensive variety of experimental designs for dealing with individual and social decision-making provided by experimental economics and psychology, many neuroeconomics studies could not have been developed. At the same time, without the significant progress made in neuroscience for grasping and understanding brain functioning, neuroeconomics would have never seen the light of day. The paper is an overview of the main significant advances in the knowledge of brain functioning by 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 sub-fields of cognitive neuroscience, such as affective, social, and more generally, decision neuroscience, are rather porous.

Keywords Decision-making · Brain · Neuroscience · Neuroeconomics · Experiments · Behavioral economics

JEL Classifications D87 · A12 · B41 · C7 · C9 · D81

✉ Daniel Serra
daniel.serra@univ-montp.fr

[1] CEE-M, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France

1 Introduction

The question of how we make, and how we should make, decisions has occupied researchers for many centuries, with different disciplines approaching the topic by their characteristically methods and techniques. By the late 1990s, several converging trends in economics, psychology, and neuroscience had set the stage for the birth of a new field known as "neuroeconomics" for dealing with the decision-making problem by integrating insights from these three disciplines.

The take-off of experimental and behavioral economics in the 1980s undoubtedly favored the emergence of the first studies in neuroeconomics by offering a set of well-codified experimental designs for dealing with individual and social decision-making in economic environments. 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, for visualizing what is happening inside the brain when humans make decisions. In a more fundamental way, neuroeconomics has largely been 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/cellular neuroscience* and *cognitive 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 (e.g., Gazzaniga & Mangun, 2014). Neuroeconomics is closely, but not exclusively, associated with cognitive neuroscience.

More specifically, within the mosaic of neuroscience, several branches support the field of neuroeconomics, including neurobiology, neuroimaging, neuroanatomy, neuropsychology, neurophysiology, neuroendocrinology, and computational/theoretical neuroscience. On the other hand, cognitive neuroscience includes 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.

Although neuroeconomics is still a nascent scientific field, 2 decades old at the most,[1] its domain is gigantic. It has already become the subject of a large number of

---

[1] It is commonly admitted today that the birth of neuroeconomics coincides with the publication by the neurobiologist Michael Platt and the neurophysiologist Paul Glimcher in *Nature* of a study on behavior of

papers, chapters in collective books, and of monographs providing overviews of the entire field, either as some of its parts, or as some major topics covered therein. Just a few years after the emergence of this new interdisciplinary domain, the first wave of surveys testified its double origin, with reviews from the perspective of economics on one hand (e.g., Glimcher & Rustichini, 2004; Camerer et al., 2005; Kennin & Plassman 2005) and from the perspective of neuroscience on the other (e.g., Glimcher, 2003; Montague & Berns, 2002; Montague, 2007). Over the past decade, a second wave of surveys on neuroeconomics has put forward the complementary strengths of its contributing disciplines (e.g., Sanfey et al., 2006; Sanfey, 2007; Camerer, 2008b; McCabe, 2008; Rangel et al., 2008; Loewenstein et al., 2008; Clithero, Tankersley, & Huettel 2008; Schultz, 2008; Rustichini, 2009). Last but not least, before the end of the decade, the first handbook of neuroeconomics was published by a neuroscientist (Paul Glimcher), a psychologist (Russell Poldrack), and two economists (Colin Camerer and Ernst Fehr), who grouped together important reviews on many topics, highlighting the strong interdisciplinary background of this new approach (Glimcher et al., 2009).

The field of neuroeconomics matured intellectually after this first wide synthesis. In the 2010s, a third wave of surveys outlined new advances for understanding how, where, and when decision-making was accomplished in the brain and what remained to be done in the discipline as a whole (e.g., Huettel, 2010; Camerer, 2013; Dean, 2013; Rose & Abi-Rached, 2013; Smith et al., 2014; Krajbich & Dean, 2015) or more specially in some sub-fields, including neuroeconomics for social decisions (e.g., Rilling & Sanfey, 2011; Sanfey & Rilling, 2011; Singer, 2012; Declerck & Boone, 2016; Engelmann & Fehr 2017; Allos-Ferrer, 2018), neuroeconomics of emotion (e.g., Pessoa, 2013, 2017; Kragel and LaBar, 2016; Saarimäki et al., 2018; Adolphs & Anderson, 2018), or neural reward, learning, and computational models (e.g., Kable & Glimcher, 2009; Fehr & Rangel, 2011; Schultz, 2016; Padoa-Schioppa & Conen, 2017; Konovalov & Krajbich, 2019).

In the meantime, the second edition of the Handbook edited by Glimcher and Fehr (2014a, 2014b) has widely reported many of these great advances. It includes well-documented specialized contributions (but often rather technical) on core concepts, methods, and tools used in neuroeconomic research (e.g., Rangel & Clithero 2014; Tobler & Weber, 2014; Kable, 2014; Fehr & Krajbich, 2014; Lempert & Phelps, 2014; Rustichini, 2005; Platt & Plassmann 2014; Crockett & Fehr, 2014) and a deal of new advances on how we learn and represent "value" in the brain (i.e., what is supposed to guide "economic behavior") (e.g., Berridge & O'Doherty, 2014; Daw & Tobler, 2014; Daw, 2014), on the neural process of choice itself (e.g., Glimcher, 2014a, c; Wallis & Rushworth, 2014; Wang, 2014), and on

---

Footnote 1 continued
monkey linked to anticipated "rewards" (in this case, food rewards) (Platt & Glimcher 1999). For the first time, an electrophysiological experiment on a monkey proved that the brain "value" stimuli independently of sensory or motor processes. Thanks to cerebral imaging, this finding was extended to humans in the early 2000s (Berns et al., 2001; Breiter et al., 2001; Delgado et al., 2000; Elliot et al., 2000; Knutson et al., 2000; Knutson et al., 2001). For a first brief history of neuroeconomics, refer to Glimcher & Fehr (2014b) and Serra (2022), chap. 3.

social decision-making within the framework of game theory (Camerer & Hare 2014; Singer & Tusche, 2014).

This paper is designed to be complementary to these previous studies, which are often referred to in the literature. Its particular edge is to focus mainly on the neuroscientific foundations of neuroeconomics to provide economists, particularly decision-making specialized economists, with a concise synthesis of these unfamiliar neuroscientific works the knowledge of which hopefully will allow them to easily get well versed with the neuroeconomic literature. This is not a survey of the great deal of neuroeconomic results available as on date; it just mentions some findings for illustrating applications to questions that are important to economists and how this new interdisciplinary approach can potentially improve our understanding of various economic decisions (several surveys cited above provide more thorough insights into neuroeconomic findings). The paper also does not address the thorny methodological and epistemological issues that neuroeconomic approach is likely to raise, including the "mindless economics" argument that non-behavioral data (such that neural and physiological data) are not "relevant" in economics (Gul & Pesendorfer 2008) (on this issue, refer to Harrison & Ross, 2010; Mäki, 2010; Serra, 2021).

Another singularity of the paper is to group approximately three non-independent broad topics as the main neuroscientific progress from which neuroeconomics has benefited. 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. 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 in particular prosocial behaviors. 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 respectively.

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 that 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 respectively. These works are mainly rooted in neurobiology and computational/theoretical neuroscience while taking into account findings from studies relating to the other two topics.

The paper is organized as follows. After a brief mention of the aim of neuroeconomics from a historical perspective (Sect. 2), it will deal with these three topics by talking about the “emo-rational” brain (Sect. 3), “social” brain (Sect. 4), and “computational” brain respectively (Sect. 5). To make the paper accessible to a large audience, the various neuroscientific notions used are defined and briefly explained accordingly. In the same way, the definition of the main economic models referred to in the text is recalled for researchers not specialized in experimental and behavioral economics,

2 The aims of neuroeconomics

2.1 From the two early distinct “behavioral-economics-in-the-scanner” and “neural economics” programs to the new unified “neural-and-behavioral economics” program

The pioneers of this young scientific field have different definitions for neuroeconomics. Originally, several definitions stood out from two separate communities: one mainly (although not exclusively) behavioral economic—that will be called the “behavioral-economics-in-the-scanner” program—and the other mainly (although not exclusively) neuroscientific—called the “neural economics” program, respectively. However, it can be argued that these two trends recently are converging on a largely shared research agenda that we suggest identifying to a “neural-and-behavioral” program accordingly.

2.1.1 The “behavioral-economics-in-the-scanner” program

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, 2004, 2005; Camerer et al., 2005; Camerer, 2007, 2008a, b). Neuroeconomics is recognized as a transdisciplinary domain using neuroscientific tools of measure 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 the advantage of each domain (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—first, cerebral imaging—for studying economic behaviors via well-established paradigms of experimental economics. However, the whole range of tools used in cognitive neuroscience is really much broader. It 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 change cognitive functions or behavior. Single-unit recording,[2] electroencephalography (EEG), magnetoencephalography (MEG),[3] positron emission tomography (PET),[4] and functional magnetic resonance imaging (fMRI)[5] are measurement techniques commonly used in neuroscience studies. Manipulation techniques can be grouped into three classes: (1) brain stimulation techniques, including transcranial magnetic stimulation (TMS)[6] and transcranial direct current stimulation (tDCS)[7]; (2)

---

[2] The most basic element of nervous system function is the "action potential" (or "spike") that arises when a voltage of a neuron's cell body rises above a particular threshold. Neurophysiologists use changes in firing rate of a neuron as an index of whether a stimulus changes the ongoing information processing with which that neuron is associated. Single-unit recording is a direct measurement of action potentials requiring the insertion of very fine electrodes into the neural tissue immediately adjacent to the neurons of interest. The invasive nature of this technique limits its use to non-human animals (except in the rare cases of human patients with clinically indicated electrodes).

[3] EEG and MEG are non-invasive neurophysiologic techniques. Input to a neuron changes the electrical potential of its cell membrane. If many neurons evince similar changes in their membrane potential, the collective electrical current they generate can be detected by electrodes positioned on the scalp. EEG provides high-temporal-resolution access to the electrical activity of the brain. However, electrical currents, like those generated by dendritic activity of neurons, also give rise to magnetic fields that MEG is able to measure thanks to external sensors.

[4] PET was the first functional imaging technique to gain wide-spread acceptance. It allows measuring brain metabolic activity thanks to emissions made by positrons coming from a radioactive isotope that is injected before or during scanning, depending on the isotope being used. The most salient disadvantage of PET is its invasiveness: safety guidelines restrict how that radioactive material can be created, handled, and administrated. This technique also has very limited temporal resolution.

[5] Since its development in the early 1990s, fMRI has grown to become the dominant functional imaging technique in cognitive neuroscience. Its success comes from the intertwining of the image creation process from MRI with new insights into the metabolic changes associated with brain activity. It is based on magnetic properties of hemoglobin: neural activity in a particular zone induces a stronger demand for oxygenated hemoglobin, and then generates a higher BOLD (blood oxygenation-level-dependent) signal. This technique is a good combination of spatial and temporal resolution. Much of the growth of fMRI in research has been facilitated by the prevalence of high-field scanners for clinical applications. Structural MRI (morphometry), which is effective in discriminating between gray and white matter in the brain, and diffusion tension imaging (DTI), which measures the direction and magnitude of water diffusion in brain tissue, are also used in a few neuroeconomic experiments. Near-infrared spectography is another method recently introduced in experiments.

[6] TMS stimulates neurons by means of electromagnetic induction. It uses a magnetic field which can pass easily through the skull, to generate an electrical current inside the brain. This electric current acts on the underlying neurons and triggers action potentials in axons that cross the field at appropriate orientations (e.g., perpendicular). This means that some locations in the cortex are easier to stimulate than others using this technique. The artificial and temporary lesion of the target zone allows identifying the behavioral effect. TMS is often applied repeatedly for changing induced neuronal excitability beyond the moment of stimulation (rTMS).

[7] tDCS is a more recent non-invasive electrostimulation tool able to change cortical excitability thanks to electrodes that are wrapped in sponges soaked in saline solution and mounted to the head. It can be used in two modes: anodal tDCS to upregulate and cathodal tDCS to downregulate neural processing in a brain region. tDCS has an additional advantage: it helps to avoid a problem that may arise when using rTMS in social neuroeconomic experiments; e.g., to study "social preferences". The issue is that each player must face a series of one-shot stranger-matching games sequentially with the behavioral study focusing on the participant playing second. This poses an implementation problem, because each participant will be faced with a high number of protagonists and there is a great temptation to deceive the participants and to confront them with prefabricated options. Yet, in experimental economics, it is well known that it is

neuropharmacological interventions: classically, manipulation of neuromodulators, including dopamine, serotonin, and noradrenaline/norepinephrine,[8] as well as hormones such as oxytocin and testosterone; and (3) approaches that study the consequences of brain lesions and neurological or psychiatric disorders, known as "clinical studies".

Until recently, fMRI was by far the most frequently used tool in neuroeconomic experiments. However, there is increasing emphasis on brain stimulation techniques where mainly magnetic or electric stimulation is used to manipulate the activity of specific regions of the brain, resulting in behavior changes (e.g., Knoch et al., 2006a, 2006b, 2008, 2009; Figner et al., 2010; Karton & Bachmann, 2011; Baumgartner et al., 2011; Ruff et al., 2013; Baumgartner et al., 2014). At the same time, a lot of experiments involve neuropharmacological interventions, mainly concerning hormones (see Crockett & Fehr, 2014, for a survey).

The strengths and limitations of these various neuroscience approaches are evaluated in Ruff & Huettel (2014) and Genon et al. (2018). They were added to older psychophysiology tools and methods for experimentally studying emotional and affective responses. These included 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 (Fernandez-Dols & Russell 2017)[9]; methods like these are also commonly used today in neuroeconomic experiments in addition to neuro-imagery, brain stimulation techniques, or neuropharmacological interventions for understanding the biological determinants of decisions.[10]

2.1.2 The "neural economics" program

A different neuroeconomics definition was initially adopted by other researchers. It was the philosopher Don Ross who introduced the name "behavioral economics in the scanner" for the works corresponding to previous definitions. He identified

---

Footnote 7 continued
strongly recommended not to deceive participants to keep their trust in the experimentalist. As tDCS is inexpensive, it can be administered *simultaneously* too many interacting subjects. Deep brain stimulation, microstimulation, and optogenetic are invasive stimulation methods reserved for animal experiments or for patients with chronic and severe neurological disorders (Parkinson's disease, epilepsy, and obsessive compulsive disorder).

[8] At least 60 different neurotransmitters have been identified. Some of them increase the probability that the postsynaptic cell will transmit an action potential ("excitatory" neurotransmitters), while others decrease this probability ("inhibitory" neurotransmitters). The main excitatory neurotransmitter is glutamate and the main inhibitory one is GABA. Some neurotransmitters, known as neuromodulators, act mainly by modulating the activity of glutamate and GABA releasing neurons. Examples of neuromodulators include dopamine, serotonin, and noradrenaline/norepinephrine.

[9] Charles Darwin was one of the first scholars to study emotions through facial expressions (Darwin 1872).

[10] Using these relatively simple and inexpensive tools in neuroeconomic experiments rather than the complex and very expensive neuroimaging is actively encouraged par Axel Rubinstein, an economist rather skeptical about usefulness of neuroeconomics for economists without totaling rejecting this approach (Rubinstein 2008). Reuter & Montag (2016, Part VII) give a scholarly introduction into the constellation of methods and techniques relevant to neuroeconomics.

another trend in the literature, which he named “neurocellular economics” (Ross, 2005, 2008) (Vromen, 2007, also identified these two distinct programs). Here, the goal was 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 agents such as the neuron, neurotransmitter system, or quasi-modular circuit (Ross, 2008). In other words, this author rejects the idea of anthropomorphism as an initial necessary condition for neuroeconomics. Thus, neurocellular economics, which we prefer to call “neural economics” by following 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 during 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 (e.g., Glimcher & Fehr, 2014; Glimcher et al., 2009). This distinction is useful, because in particular much of the early criticism toward neuroeconomics expressed by some economists was aimed at the behavioral-economics-in-the-scanner program.[11] 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.

---

[11] 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., Harrison 2008a, 2008b; Rubinstein 2008; Smith 2008). In addition to this issue of interest for economists and beyond the philosophical issue of the “mindless economics” argument (Gul & Pesendorfer 2008), controversial debates about neuroeconomics bear on reliability of findings, in relation to the non-trivial statistical analysis of fMRI data and particularly with the so-called reverse inference “fallacy”. The reverse inference problem, which questions the validity of the rationale underpinning neuroimaging methods—namely inferring thought processes from brain activity—is a practical issue also found in cognitive psychology experiments that rely on neuroimaging to infer particular cognitive functions (memory, attentiveness, language…). On this topic, see Poldrack (2006, 2011, 2018); Harrison (2008b); Harrison & Ross (2010); Ross (2010); Bourgeois-Gironde (2010); Poldrack et al. (2017); Serra (2021). Remark that recent progress in the development of methods for decoding human neural activity as measured with fMRI should lead to bypassing the reverse inference problem. We know that fMRI studies focused on associating brain zones with mental functions. The introduction of decoding using the so-called “multivariate pattern analysis” (MVPA) has revolutionized fMRI research by changing the questions that are asked. Instead of asking what a zone’s function is, in terms of a single brain state associated with global activity, we can now ask what information is represented in a zone, in terms of brain states associated with distinct patterns of activity, and how that information is encoded and organized (see, e.g., Normann, Polyn, & Haxby 2006; Haxby, Connoly, & Guntupalli 2014; Efron & Hastie 2016).

2.1.3 A "neural-and-behavioral economics" program

For many scholars, 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. Ultimately, the final goal of neuroeconomics is first and foremost improving the predictive power of economic theory.[12]

How do we do this? By inserting in economic theory, the biological constraints that brain functioning poses to behaviors (Glimcher, 2011a, 2014c). The common ambition of many neuroeconomists now is to develop a computational model of economic choice that would integrate the constraints applied to economic decision-making process by the brain function to understand the precise mechanism behind links between the biological or neural markers and the economic behavioral outcomes.

This evolution—which could be seen as the emergence of a "neural-and-behavioral-economics" research program—is clear in the writings of Paul Glimcher, a pioneer in the neural economics program, between the beginning and the end of the 2000s (Glimcher, 2003, 2011a) (see Vromen, 2011). However, it can also be detected over the last decade in the works of many leading neuroeconomists, including but not limited to, Colin Camerer, John Clithero, Ernst Fehr, Scott Huettel, Joseph Kable, or Antonio Rangel (Serra, 2016).

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

2.2.1 Perceptual decisions and value-based decisions

In decision-making, how does the brain choose among options? To answer this question, we have to first make the decision nature precisely 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

---

[12] In the same time, neuroeconomics results are viewed as useful in psychiatry for analyzing a constellation of mental and neurological disorders including frontotemporal dementia, obsessive–compulsive disorder, and drug addiction (see, e.g., Millan, 2013; Schutt et al., 2015; Conn 2016; Lis & Kirsch 2016; Dreher & Tremblay, 2017; Alos-Ferrer 2018).

of these approaches is clearly distinct: perceptual decisions are based on the standard neurobiological theory of perceptual categorization, i.e., the "signal detection theory", whereas VBD refers originally to the standard microeconomic theory, i.e., 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 nineteenth century German physiologist Ernst Weber (Weber, 1834/ 1996). The key assumptions of the signal detection theory are that real-world stimuli give rise to percepts through a random process like drawing from a Gaussian distribution (Green & Sweets, 1966; Macmillan & Creelmann, 2004). This notion that subjects have variable internal experiences from the same stimulus is indeed far removed from expected utility theory. Standard economic theory and perceptual psychology are quite different (Gold & 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 e.g., McFadden, 2005; Gul & Pesendorfer, 2006). The computational model arising from the neuroeconomics literature (see Sect. 5) should be regarded as providing a neurobiological foundation for these random utility models (Fehr & Rangel, 2011; Glimcher, 2011a; Krajbich, Ouf, & Fehr 2014).[13]

2.2.2 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 between the two. 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

---

[13] However, there is a significant difference between neuroeconomic choice models and random utility models. While the latter posit that *preferences* are in essence stochastic and that choices always reflect these underlying preferences, neuroscience research suggests that the *choice process* itself might be systematically biased and sub-optimal (we shall return to this point in Sect. 5).

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. Therefore, the assertion that "choices maximize values" becomes potentially falsifiable and thus truly scientific (Padoa-Schioppa, 2011).

In brief, for most neuroeconomics researchers today, 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. In other words, neuroeconomics intends to discover the neurobiological and computational bases underpinning these kinds of decisions that we can identify with "economic behaviors"[14] (Montague, 2007; Rangel et al., 2008; Schultz, 2009; Balleine, Daw & O'Doherty, 2009; Kable & Glimcher, 2009; Fehr & Rangel, 2011; Padoa-Schioppa, 2011; Rangel & Clithero 2014; Glimcher, 2014a; Padoa-Schioppa & Conen, 2017).

3 The emo-rational brain

3.1 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 nineteenth 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 & 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 (e.g., Loewenstein & O'Donoghue, 2004; Bernheim & Rangel, 2005; Benhabib & Bisin, 2005; Fudenberg & Lenine 2006). 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 Sloman, 2002; Evans 2010; Kahneman, 2011; Evans & 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

---

[14] In this respect, as suggested by Huettel (2010), neuroeconomics may be viewed as a subfield of decision neuroscience which deals with *both* perceptual and VBD decisions. Yet, some scholars do not distinguish between neuroeconomics and decision neuroscience by opposing them to molecular neuroscience (e.g., Montague 2007).

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. 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., Cartwright, 2016; Serra, 2017).

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 et al. (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 Sect. 5.

3.2 Cognitive systems

3.2.1 A brief introduction to the brain

Anatomical structures of the brain: Broadly speaking, the primate (and hence human) brain can be divided into four main divisions: (1) the telencephalon or forebrain (the cerebral cortex, basal ganglia, amygdala, and hippocampus); (2) the diencephalon (essentially the thalamus); (3) the mesencephalon or midbrain (the superior and inferior colliculus, substantia nigra, and tegmental area); (4) the brainstem or hindbrain (the pons and medulla). In addition, from a macroscopic viewpoint, cerebral hemispheres are divided into lobes: in the main, frontal, parietal, occipital, and temporal lobes, respectively. These are not functional subdivisions but rather names of convenience. Other subdivisions are usually used: lobules (internal divisions in some lobes), sulcus (troughs in the cortex, the deepest ones are called fissure), and gyrus (convolutions in lobes bounded by some sulcus). Furthermore, it should be noted that various terms used for specifying the position of the brain and its regions are the same that are used for describing the macroscopic anatomy of the rest of the body: the terms "anterior" and "posterior" indicate front and back of the head, "inferior" and "superior" indicate above and below; "medial" and "lateral" indicate toward the center or to the side, respectively. Other terms come from drawing the axis of the body and the brain. "Dorsal" refers to the back or upper half depending on whether the focus is the body or the brain; "ventral" refers to the front or the lower half. "Rostral" indicates the direction

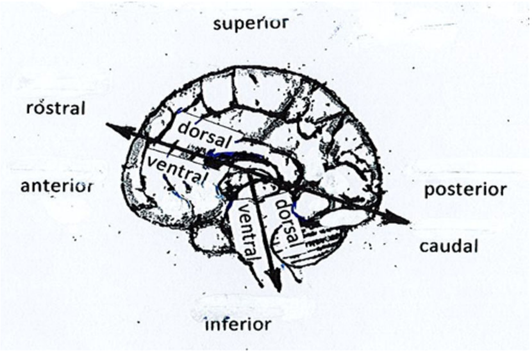


Fig. 1 Terminology to describe location in the brain

towards the nose and "caudal" towards the back (these terms will be frequently used in the paper) (Purves et al. 2011) (see Fig. 1).

The cerebral cortex. 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 twentieth century is the most well-known (Brodmann, 1909/1994). Yet, this cytoarchitectonic organization—each Brodmann area (BA thereafter) 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 (Talairac & Tournoux 1988) allows responding to this problem. Frequently used in brain imaging experiments, the method suggests an atlas describing a "standard" brain anatomically and cytoarchitectonically.[15] Most brain localization studies in neuroscience are based on this idea of a standard brain (see Fig. 2 for the map of the main BA[16]). We need to be aware, however, 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 on this issue, see Schutt et al., 2015).

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), and the other part corresponds to the sensorial and motor cortices, which encode sensory information (mechanical or thermal sensitivity, vision, audition, sense of smell, and taste) as well as movements control (see, e.g., Fuster, 2008; Purves et al. 2011).

The frontal cortex. 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

---

[15] 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, e.g., Poldrack et al., 2011, 2017).

[16] Notice that the different neural regions referred to in the text often include only a part of the BAs mentioned in bracket.

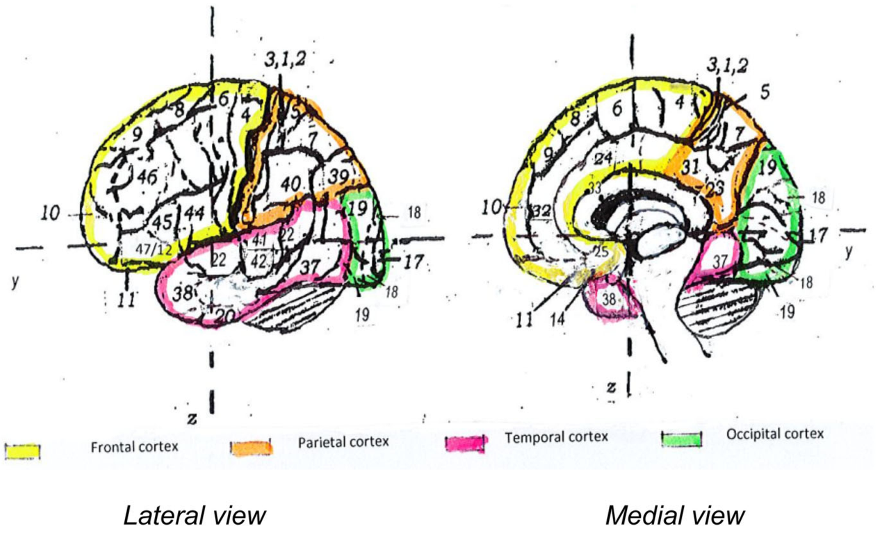


Fig. 2 Broodmare's cytoarchitectonic map of the human cerebral cortex

volume). Despite this human singularity, our information about frontal cortex connectivity in humans has been essentially derived from studies in monkeys (for comparisons between the frontal cortex of humans and monkeys, see Ongür & Price, 2000; Ongür et al., 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 and, among other mammals with humans (Dehaene et al., 2005). 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. The localization of neural areas in humans similar to those in monkeys remains to be studied in the neuroanatomy research agenda (Mackey & Petrides, 2014).

The human frontal cortex plays a fundamental function. It collects complex perceptive information from the sensory and motor cortices as well as from the parietal and temporal associate cortices, respectively. It is view as the "executive" region of the brain (Pribram, 1973) that is located In front of the motor and premotor areas in the frontal cortex stands the prefrontal cortex—a wide cortical region involved in cognitive processes.

3.2.2 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. However, a better understanding of the neural mechanisms is through ding of information sending back to loops, networks or circuitries (we will address this important point below).

In this respect, it is generally agreed that highly complex cognitive processes (i.e., logic reasoning, planning, problem-solving, and decision-making) strongly involve some anterior and lateral zones in PFC, namely the anterior cortex and dorsolateral PFC.[17] However, nowadays, it is recognized that other neural areas engaged in high-level cognitive processes are in the more posterior cortical zones, including the anterior cingulate cortex (ACC), posterior cingulate cortex (PCC), temporo-parietal junction (TPJ), and posterior parietal cortex (PPC)[18] (Miller, 2000; Miller & Cohen, 2001; Semendeferi et al., 2001; Ramnani & Owen, 2004; Zald & Rauch 2006; Fuster, 2008; Passingham & Wise, 2012). Furthermore, 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 de sense to speak of a "cognitive control network" in the brain, including primarily the dorsolateral PFC and ACC.

3.3 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".

3.3.1 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

---

[17] The anterior cortex (or frontopolar cortex) (BA 10) 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). The dorsolateral PFC (BA 8, 9, 46) corresponds to the superior part of the frontal lobe exterior. It is seen as the most "rational" part of the brain.

[18] The cingulate cortex is an internal zone located along the interhemispheric fissure above the corpus callosum. It is divided into an anterior (ACC) (BA 24, 32, 25) and a posterior (PCC) (BA 23, 31) 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. The rostral ACC is known as the paracingulate cortex. The PCC (BA 7, 40) 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". The TPJ (BA 22, 40) is a part of the temporal cortex at the edge of the parietal cortex. It is implicated both in reorienting of attention and social cognition.

structure in terms of evolution (i.e., the basal ganglia[19]), 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[20]; (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, including some cortical areas in particular (see Dalgleish 2004).

It is commonly recognized today in affective neuroscience 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 cortical zones are also viewed as engaged in emotion processing: the orbitofrontal cortex (OFC), ventromedial PFC,[21] and anterior parts of insula,[22] cingulate cortex (ACC), and temporal cortex (ATC). Both these subcortical structures (i.e., the "classic" limbic system) and cortical structures are now thought to be involved in emotion expression and processing (Dalgleish 2004; LeDoux, 1996; Phelps & LeDoux, 2005; Whalen & Phelps, 2009; Pessoa, 2010; Purves et al. 2011; Lempert & Phelps,

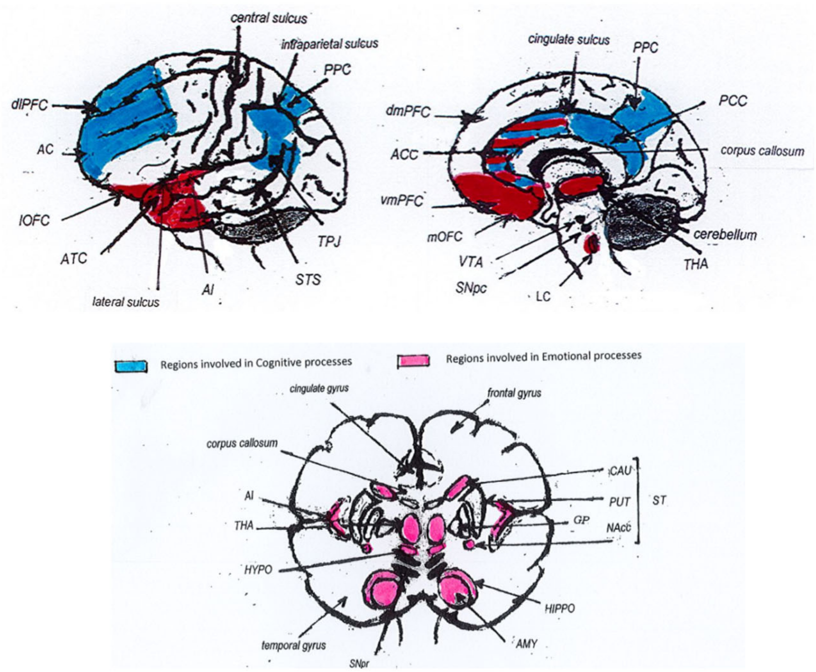
---

[19] 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). Today, many researchers simply divide the striatum into two sections: the ventral striatum (the NAcc and lower parts of the caudate and putamen), interacting with regions engaged mainly in emotion and motivation, and the dorsal striatum (the upper parts of the caudate and putamen), interacting with regions implicated in movement and memory.

[20] 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".

[21] 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/lateral parts are differentiated) (BA 11, 14 / 13 / 47/12), ventromedial PFC (BA 10, 11, 14, 32), and dorsomedial PFC (BA 9, 8, 32) (sometimes named globally medial PFC). The ventromedial PFC very often is defined as including the medial OFC.

[22] 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". 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.



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, NPR: substantia nigra pars reticula, VTA: ventral tegmental area, LP: locus coerulus

Fig. 3 Main neural regions involved in cognition and/or emotion processing: lateral view (left), medial view (right), and internal structures (axial plan) (below)

2014; Engelman & Hare 2018; Fox et al., 2018) (Fig. 3 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 as mentioned in the text accordingly).

3.3.2 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 depend on 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; LeDoux, 1996; Rolls, 2014; Saarimäki et al., 2018). The idea that a link would exist between a body function and a brain zone dates from the nineteenth 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 accordingly. Yet, we know to day that even though some neural zones are devoted to specific functions, networks, circuits, and loops are engaged most of the time. Although the central principle for understanding representation in the brain does remain as the notion of “modularity” (Fodor, 1983; Glimcher, 2014b), interpretation of network connectivity outweighs interpretation of brain localization, particularly for high-order complex functions (see Fuster, 2009; Fuster & Bressler, 2012; Lindquist & Barrett, 2012; Fehr, 2013; Pessoa, 2017).

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, and 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; Ekmann, 1982, 2003; Panksepp, 1998; Plutchik, 1980). For others, subjective experience of emotions formulated by individuals may be described according to some underlying dimensions, such that the pleasant/unpleasant dimension (positive/negative dimension) and the awareness/depression dimension, respectively (low/high activation dimension). According to some authors, it is critical to distinguish these conscious subjective experiences of emotions, often called “feelings”, from emotions as internal functional states (Adolphs & Anderson, 2018; Adolphs, 2017; Damasio, 2003, 2017; Scherer, 2005).[23]

By crossing both approaches, one gets a certain image of some neural zones specialization (Lindquist et al., 2016; Cowen & Keltner, 2017; Clark-Polner 2017; Barrett, 2017). Clearly, several distinguishable neural structures are associated with some negative emotions: the amygdala (fear, anxiety, aggressiveness, stress, and sadness), hypothalamus (anger, fear, and aggressiveness), anterior insula (disgust, distress, and anger), locus coerulus[24] (fear and anxiety), and anterior cingulate cortex (ACC) (sadness, pain, and anxiety), respectively. However, the amygdala and the ACC are also activated in perception of some positive emotions. Several areas in the ventral striatum, including the nucleus accumbens (NAcc), and caudate are strongly associated with positive emotions. And several neural areas in the wide orbitomedial region, such as the orbitofrontal cortex (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

---

[23] 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 (e.g., Scherer 2005).

[24] 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.

plays a different role[25] (for recent reviews, see Kragel & Labar, 2016; Saarimäki et al., 2018).

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. For them, 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 & Anderson, 2018; Lindquist & Barrett, 2012; Pessoa, 2013, 2016).

3.4 The interplay of emotion and rationality

It is thus clear from anatomical, physiological, and neuropsychological observations that emotion processing and complex cognitive processing interact. And this applies in particular to "rational" decision-making and emotions. A main tenet of contemporary neuroscience is that, contrary to what has typically been assumed, emotion and rationality are complementary.[26] The image of opposing "cold" regions in the frontoparietal cortex to "hot" regions in the classic limbic system regarding emotional state processing is no longer accepted (e.g., Kelso & Engstrom, 2006; Lempert & Phelps, 2014; Okon-Singer et al., 2015, 2018; Pessoa, 2013; Phelps, 2009; Richter et al., 2017).

In the 1990s, several neuropsychology studies (Bechara et al., 1994, 1996, 1997) go even farther by arguing that emotion processing often would be *necessary* for making rational decisions. From an economic viewpoint, the result of these experiments is important, because it can be interpreted in terms of the emotion of regret.[27] There are numerous neuroeconomic experiments dealing with the role of regret in decisions that have explored further into this topic. More broadly, a

---

[25] However, several meta-analyses showed that often there are differences in response intensity of a same structure depending on the emotion: e.g., 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, e.g., the right amygdala is more involved in negative emotions and the left in positive.

[26] Consider Plato's famous metaphor where the mind is seen as a chariot pulled by two horses. The rational brain is the charioteer who guides the horses. One of the horses is well bred and well behaved, while even the best charioteer has difficulty controlling the other horse; this obstinate horse represents negative, destructive emotions. The charioteer's task is to keep both horses moving forward. Through that simple metaphor, the mind was seen as conflicted, torn between reason and emotion. This dual division of the mind is one of the most enshrined ideas in Western culture. A large set of influential philosophers, from René Descartes to Sigmund Freud, and including Francis Bacon, Auguste Comte, and Emmanuel Kant, all embraced various forms of this duality, which continues through to the modern brain–computer metaphor proposed by cognitive psychology that sees emotions as antagonists of rationality. Aristotle in *The Nicomachean Ethics* is seen as an exception by claiming that rationality is not always in conflict with emotion. Another widely known exception is Spinoza, a contemporary of Descartes, Antonio Damasio highlights this opposition between Descartes and Spinoza in the titles of two of his books. *Descartes' Error: Emotion, Reason, and the Human Brain* (Damasio 1994) and *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain* (Damasio 2003).

[27] We know that in economics, the experience of regret in decision-making was initially introduced by Bell (1982) and Loomes & Sugden (1982). In this theory, we suppose that, for each decision, the agent is taking account her/his utility and the potential degree of regret/satisfaction, i.e., the comparison with what she/he could have obtained.

multitude of neuroeconomic experiments prove that, in cerebral regulation of behaviors, the coordination between emotional and cognitive systems is necessary.

3.4.1 Emotions and rational decisions: the role of regret

The "Iowa Gambling task" experiment. 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).[28] 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, 1997). Studies on subjects with damage to the amygdala reached the same result (Bechara et al. 1999).

What these pioneer experiments prove is that higher cognitive abilities do not universally govern behavior. Emotions are not always harmful to those which are feeling them.[29] On the contrary, without emotional substrate, a rational choice is hard if not impossible. The "somatic marker" hypothesis (Bechara & Damasio, 2005; Damasio, 1994, 1996; Reiman & Bechara, 2010), where emotions should be physiological states before being psychological states for making good decisions easier, is widely confirmed in the literature.[30]

---

[28] 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 detail 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).

[29] 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. There is compelling evidence that the perception of emotionally salient stimuli and the experience of emotional states can profoundly alter cognition and promote specific harmful behavioral tendencies (see, e.g., Okon-Singer et al., 2015; Engelman & Hare 2018).

[30] Over the years, several studies have questioned the somatic marker hypothesis (e.g., Dunn et al., 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.

Interestingly, the finding of these clinical studies can be interpreted in terms of the emotion of regret. Damasio remarked that his patients, even though they could not adopt an adaptive behavior, had a certain consciousness of better choices. Yet lacking any emotional feeling, they did not expect the regret of not choosing the good decision.

Neuroeconomic experiments on regret. Numerous neuroeconomic experiments have addressed the role of regret. Camille et al. (2004) extended the pioneering work of Damasio and Bechara and tested their hypotheses. Their study evaluated the ability to experience regret and disappointment in normal individuals compared with patients experiencing ventromedial prefrontal cortex (vmPFC) lesions. The hypothesis tested was that, in particular, the inability to feel regret could lead patients with vmPFC lesions to the disastrous consequences first identified by these neurologists and subsequently observed on many occasions. To undertake the test, the authors designed an experimental protocol based on a succession of choices between two lotteries with different risk levels that were capable of artificially generating regret.

The subjects had to choose between two “wheels of fortune” that produced monetary gains or losses with different pre-determined probabilities that were unknown to the subjects. This was repeated several times, with the gains or losses being cumulated accordingly. The subjects had to self-evaluate their level of satisfaction or disappointment with the outcome using a graduated scale. At the same time, a device recorded their skin conductance to provide an index of the intensity of the perceived emotion. Presenting the subjects with the outcome they would have attained had they had chosen the other wheel was sufficient to cause regret. In this regard, regret was different from the disappointment that resulted simply from noting the difference between the actual and expected gains. In normal subjects, the effect was unequivocal, that is, for a given gain, satisfaction was lower if the other wheel led to a larger gain and higher if it led to a loss. Furthermore, the emotion expressed by normal subjects on the graduated scale was more intense once they knew the other result, and so was their physiological response (i.e. their skin conductance). On the contrary, knowing the outcome that they would have obtained with a different choice had no impact on the level of satisfaction or disappointment in the subjects with vmPFC lesions, as if they were impervious to regret. This result was confirmed by conductance measurements. These observations, therefore, highlighted the role of this area of the PFC in regret perception.

The experiment further explored the differences between disappointment and regret: the authors biased the game, so that, on an average, the bet leading to the highest gain won less frequently than the alternative choice, resulting in a smaller win. The hypothesis here was that experiencing a greater sense of regret for the first time would progressively lead normal subjects to choose the least risky option. This was indeed the case. At the end of the experiment, they had all accumulated a positive gain that was, on an average, greater than the highest potential gain from a single attempt. Conversely, patients with lesions more often made the most unfavorable bet that resulted in a significant average loss.

This study tends to show that regret generates greater physiological responses and that normal subjects find it more intense than disappointment. However, this was not the case in patients with vmPFC lesions. Two distinct neural processes seem to generate two negative emotions: regret and disappointment. More precisely, the two emotions appeared to be cumulative: one may be upset with both the result and with making the wrong choice. The vmPFC integrates both the cognitive and emotional components of the entire decision-making process, and its dysfunction prevents the emergence of specific emotions such as anticipated regret, which plays an essential role in regulating individual and social behavior as it expresses a sense of responsibility for one's own choices. In the wake of clinical studies conducted on patients with vmPFC lesions, this experiment on regret comparing normal subjects with patients affected by similar lesions further demonstrated the vital role played by emotions in "rational" decision-making.

Some fMRI studies have also investigated brain activity involved in regret and disappointment by manipulating the feedback participants see after deciding to choose a certain risk (e.g., Coricelli et al., 2005). They also showed that 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. These fMRI studies supplement Camille's initial experiment, but the authors propose a somewhat different interpretation. While in the clinical study, in accordance with Damasio's original thesis, the role of emotions in "rational" decisions is highlighted, in these experiments, they stress rather on the role of cognitive processes in the determination of some emotional states (for other fMRI studies involving emotion of regret in neuroeconomic experiments, see, e.g., Coricelli et al., 2007; Chua et al, 2009; and for a magnetoencephalography (MEG) study, see Giorgetta et al., 2013).[31]

3.4.2 The necessary coordination between emotional and cognitive systems in individual economic decision-making

A multitude of neuroeconomic experiments prove that, in cerebral regulation of economic 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 "biases" or "anomalies" 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 economic contexts, such as purchasing consumer goods, financial decisions (i.e., risky and ambiguous decisions), intertemporal choices (decisions involving trade-offs among

---

[31] A lot of neuroscientific studies show that the emotion of regret also is implicated in several clinical disorders such as schizophrenia, depression, obsessive–compulsive disorder, and "chasing" behavior in pathological gambling.

payoffs available at different points in time), or social decisions (interactive decisions between several individuals). Yet, considering a more complete explanation of social decisions from a neuroeconomic perspective requires also taking account social cognition processes such as "mentalizing" and empathy (see Sect. 4). For an extensive review of this literature, see Reuter & Montage 2016; Serra, 2016. Let us just consider some examples of neuroeconomic experiments for concrete individual decision-making.

Consumer purchasing decisions. Various experiments relating to purchasing decisions, such as buying an item in a retail store or choosing cereals, have shown that during the decision process, several neural areas are activated for different reasons, but are led to collaborate accordingly. The brain undertakes an implicit arbitrage between the anticipation of the "reward" attached to purchasing the good and of the "punishment" generated by paying for it (as we usually avoid spending too much money). In the fMRI study of Knutson et al. (2007), for example, subjects' brains were scanned, while they took part in a three-phase dynamic purchasing process: first, they saw the product, then its price, and finally, they had to decide whether to buy it or not. The neural images indicate that the nucleus accumbens (NAcc) is activated during the first phase,[32] which tends to prove that it is this area of the brain, that is crucial in the "reward system" (refer to Sect. 5), which is activated first when the subjects become aware of the product. The more intense is the activation of the zone, the stronger is the desire of the agent to purchase the product. Thereafter, the reactions to the displayed price activate two other neural areas. More precisely, inflated prices activate the anterior part of the insula, an area associated with negative emotions. Thereafter, the dorsolateral PFC, one of the regions involved in deliberation, is activated, in relation with the subject calculating whether the purchase is a "good deal." The activation level is greater if the cost of the item is considered to be significantly lower than normal. Hence, during the buying process, distinct parts of the brain are activated with varying degrees of intensity at different times in response to different positive or negative stimuli. By measuring the relative level of activity in each neural zone, the authors were able to precisely predict the subjects' buying decisions. This result clearly contradicts the standard choice theory, because it shows that consumers are not always guided exclusively by price or expected utility.

The fMRI experiment undertaken by McClure et al. (2004b) is another good illustration of the role that neuroeconomics may play in clarifying the factors involved in consumer choice mechanisms. This study aimed to comprehend the mechanisms governing the choice between two brands of fizzy cola drinks, Pepsi and Coca-Cola, independent of the preferences identified when simply tasting them. In particular, it showed that the utility felt by consumers did not just depend on the taste of the fizzy cola drinks (immediate sensorial expression linked to activation of

---

[32] This example indirectly refers to understanding consumer behavior in terms of "mental accounting" as proposed in behavioral economics (Thaler 1985, 1999). This very general mental process is analyzed by distinguishing two often simultaneous phases: a "framing" phase, which is concerned with the external description of events that is given to an agent, and an "editing" phase, which is concerned with the internal process whereby the agent analyses the information. These neuroeconomic experiments focus on the editing phase.

the vmPFC), but that the knowledge of the brand generated a cognitive neural response (control of actions, recall of memories, and one's own image associated with the activation of the lateral PFC and hippocampus), which triggered a declared preference for a particular brand accordingly.

It has already been mentioned that, alongside brain imaging experiments that merely establish correlations, an increasing number of researchers are using transcranial stimulation tools in their experiments (usually TMS) to induce an exogenous behavioral effect and hence establish genuine causalities. For example, in the case of consumer purchasing decisions, studies have shown that an experimentally induced variation of the dorsolateral PFC may modify participants' willingness to pay to acquire certain goods (Camus et al. 2009).

Financial decisions: Risk, ambiguity, and uncertainty. Numerous neuroeconomic experiments on financial choices are available. Here, we only mention a few examples (Taya, 2012 and Tobler & Weber, 2014 provide a fuller review).[33]

In an fMRI study centered on the portfolio selection strategy, Kuhnen & Knutson (2005) used an inventive experimental protocol that offers various advantages in eliciting choice behavior between risky assets (two kinds of shares) and an asset with a guaranteed return (a bond); it uses monetary incentives within a dynamic framework to identify both optimal and sub-optimal choices accordingly. The authors showed that observed deviations from an "optimal" investment strategy for a risk-neutral agent can be explained by the identification of neural mechanisms involving two distinct zones: the nucleus accumbens (NAcc) where the activation preceded risky choices as well as risk-seeking mistakes ("excessive risk-seeking" behavior) and the anterior insula where the activation preceded riskless choices as well as risk-aversion mistakes ("excessive risk-averse" behavior) The authors concluded that financial decision-making appeared to require a delicate balance between two antagonistic mechanisms: one promoting risk-taking (NAcc activation) and the other dampening it (anterior insula activation). Both mechanisms were required to take or avoid risks; however, the over-activation of either could lead to mistakes. They also noted that when subjects found it difficult to identify the advantages of choosing between risky assets and guaranteed-return securities, the anterior cingulate cortex (ACC) was activated; this neural area is particularly involved in resolving conflicts.

Other experiments have focused on the distinction between behavior in risky situations (when the probability of possible events occurring can be estimated) and ambiguous situations (when this probability is vague, doubtful, or uncertain). It has been known since the famous Ellsberg (1961) experiment that individuals tend to avoid the unknown and are averse to ambiguity. However, by choosing the unknown, it may be possible to obtain missing information and improve the overall long-term performance. This is the classic exploration–exploitation dilemma; it is in an individual's interest to "exploit" the known to obtain a reward, but it is also in his/her interest to "explore" the unknown to improve the choice of future actions

---

[33] Furthermore, the work of the American financial journalist Jason Zweig (Zweig 2007) aimed at the general public uses a broad range of examples from the history of finance to show the potential of neuroeconomics to elucidate and guide financial choices.

and obtain improved returns over time. Loss aversion and ambiguity aversion, two common emotions frequently observed in financial decisions in particular, are two very different phenomena: the former refers to a simple trade-off between certain and random net gains depending on personal taste, whereas the latter concerns disturbing situations with unclear sources (e.g., Camerer & Weber, 1992; Starmer, 2000) and it is of interest to see if this difference also exists at the brain level. The comparative study of neural circuits involved in risky or ambiguous situations yields various interesting results that also reflects the need for collaboration between certain zones associated with emotions and others specializing in the higher cognitive processes.

The fMRI study undertaken by Hsu et al. (2005) showed, in particular, that identical areas—the dorsolateral PFC, the orbitofrontal cortex (OFC) and the amygdala—are involved in processing both these phenomena. However, these neural areas do not react in the same way in risky and ambiguous situations; the activation is much more intense in the case of ambiguity. According to the authors, this would suggest that these neural areas might intervene to signal the magnitude of the uncertainty: the greater the uncertainty, the more intense their activation. The fMRI experiment by Huettel et al. (2006) clarified this result accordingly. First, decisions made in risky or ambiguous situations might be supported by distinct neural mechanisms. In their experiment, the activation of a part of the dorsolateral PFC was found to be correlated mainly with the preference for ambiguity, whereas the activation of the posterior parietal cortex (PPC), at the back of the brain, was found to be correlated mainly with the preference for risk. It also suggested that the dorsolateral PFC might play a specific role in resolving ambiguity (i.e., when ambiguous situations progressively transform into risky ones). This activity of a neural zone that is heavily involved in the highest cognitive functions might reflect actions that attempt to acquire information to reduce ambiguity, a situation that creates a negative emotional feeling. This is in fact one of the conclusions of the PET-scan study of Rustichini et al. (2005), who compared four types of lottery: certain, risky, ambiguous, and uncertain in the sense of Knight (i.e., no objective quantification of probabilities is possible). Only a context of ambiguity may significantly activate the PFC neurons; the brain has various pieces of information for its computation, but this is insufficient for a conclusion to be reached, resulting in a degree of perplexity at the cognitive level.

It should be noted that the role of the dorsolateral PFC in decisions in which proper weighting of the risk is involved was also studied in several studies by brain stimulation. For instance, Knoch et al. (2006a) showed in a TMS study that participants who underwent a transiently, experimentally induced disruption of their right (but not the left) dorsolateral PFC displayed significantly riskier decision-making. Fecteau et al. (2007) established in a tDCS study that upregulation of activity in the dorsolateral PFC reduced appetite for risk during ambiguous decision-making. These findings suggest that dorsolateral PFC activity is critical for adaptive decision-making, possibly by suppressing riskier responses.

Time preference and impulsivity. The neuroeconomic experiments mentioned so far show results suggesting, on an anatomo-functional basis, that a multitude of neural circuits interact during the decision-making process. The examples given show that decisions are the result of trade-offs between various cognitive and emotional processes acting jointly. However, some authors suggest that, in certain situations, the experimental results tend to show a conflict between these two systems. Introducing temporality, discounting of the future, and impulsivity into the choice process might provide the best framework to interpret economic decisions as the outcome of the conflict between reason and emotion.

In the pioneering fMRI experiment of McClure et al. (2004a), the protocol was directly inspired by classic research on intertemporal choice in experimental economics, where the subjects had to make trade-offs between two sums of money available at different dates according to variable deadlines (measured in weeks or months).[34] According to these authors, two distinct neural circuits intervene in the anticipation of monetary gains: the first concerns the regions of the limbic and paralimbic systems involved in the "reward network," such as the ventral striatum and the ventromedial PFC (see Sect. 5), while the second concerns peripheral regions such as the dorsolateral PFC and the posterior parietal cortex (PPC). Not surprisingly, the first system, associated with involuntary emotions, tends to dominate when deadlines are short, while the second one, mainly involved in reasoning, may be intensely activated when choices have distant deadlines. The two systems may evolve in opposite directions; significant involvement of the emotional system and weak involvement of the cognitive system occurs when the time scale is very close to the moment of anticipation; thereafter, the temporal profiles may develop in opposite directions, with the former becoming less intensely involved and the latter progressively taking over. The temporal proximity of the emotion may explain the initial impulsion of the emotional system, whereas with time, reflective, conscious, and controlled processes take over.[35] Furthermore, the role of the dorsolateral PFC in self-control of intertemporal choices was confirmed by several TMS studies, which showed that experimentally induced disruption of the cortical zone in the left hemisphere makes the participants more impatient and less able to resist temptation (Figner et al., 2010; Knoch & Fehr, 2007).

The experiment of McClure et al. (2004a) is significant as it appears to provide an objective basis for the quasi-hyperbolic (or hyperbolic) discount rate hypothesis, which tends to replace the standard exponential discounting model in behavioral economics; in essence, preferences reflect a high discount rate for results expected in the near future, while implicit discount rates are much lower for results expected in the distant future (e.g., Laibson, 1997; Loewenstein & Elster, 1992; Loewenstein & Prelec, 1992).[36]

---

[34] See Frederick, Loewenstein & O'Donoghue (2002) and Camerer & Loewenstein (2004) who distinguish this "choice tasks" method from other popular experimental methods such as the "matching tasks" method.

[35] The experiment was repeated with food rewards in McClure et al. (2007) with the consumption of a fruit juice being either immediate or delayed (offset by 10 min or several minutes more). Unlike financial rewards, the emotional mechanism was activated only in the *immediate* consumption option, suggesting that time scales are perceived differently by the brain according to the nature of the reward.

[36] The Laibson model (Laibson 1997) that uses quasi-hyperbolic discounting is however criticized, because it is incompatible with the notion of self-control. Thus, Ainslie (2012) prefers the original hyperbolic approach (Ainslie 1975, 1991), but introduces a recursive process of self-prediction by the

It should be noted that this neural interpretation of quasi-hyperbolic discounting might seem paradoxical considering the role assigned to emotions in making rational choices by researchers in the Damasio camp who defend the “somatic marker” hypothesis. Emotions might be necessary to build anticipation (without emotions, it is impossible to imagine the future adverse consequences of decisions made in the present) and, at the same time, they are responsible for a kind of temporal short-sightedness that leads to favoring the very short time term. It could be argued, however, that this contradiction is only apparent if the brain circuits involved in these two types of experiments are different. However, if we leave aside the issue of locating cerebral activities and instead focus on brain function mechanisms via the reward system from an evolutionary perspective, an alternative interpretation of impulsivity in intertemporal choices seems to prevail (see Sect. 5).

4 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 (Bowles & Gintis, 2011; Tomasello, 2000). Charles Darwin was the first scientist in the nineteenth 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). Therefore, 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.[37]

At the source of these studies, there is a theory initially built in social psychology known as “theory of mind” (Premack & Woodruff, 1978) or “mentalizing” (Frith & Frith, 2003); some authors also refer to this as “mind-reading” or “cognitive perspective-taking”. 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 mainly to emotional character called the empathic system (some authors also refer to “emotional perspective taking”) (for

---

Footnote 36 continued
subjects themselves at the different expected timeframes, which may imply stronger commitment from the subjects towards themselves or, on the contrary, a progressive disengagement.

[37] In the beginning, much research in social neuroscience 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 15 years, research in social neuroscience has increasingly focused on the social behavior of mentally healthy decision-makers, encompassing many social phenomena as social interactions.

comprehensive reviews, refer to Cacioppo et al., 2006; Decety & Cacioppo, 2011; Singer, 2012; Lieberman, 2012; Schutt et al., 2015; Alos-Ferrer 2018).

Even though the field of social neuroscience and the subfield of neuroeconomics known as "social neuroeconomics" (the term was coined by Fehr & Camerer, 2007) are still perceived as two distinct fields, we have to agree that the topics they are concerned 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—has come to provide an effective quantitative framework for studying how information, incentives, and social knowledge influence optimal strategies for social interaction (Singer & Tusche, 2014).

4.1 Theory-of-mind in social psychology

Theory-of-mind refers to the human cognitive capacity to distinguish one's own mental states of self (beliefs, intentions, and 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).[38]

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 at age 4 starts to correctly attribute false beliefs to others and give verbal explanations when asked. At age 5, over 90% of children understand this task, and all do by at age 7 (Frith & Frith, 2003, 2007; Hyde et al., 2018).[39]

Several theories were proposed to explain mentalizing. The "theory of simulation" seems to account for most results in social neuroscience, proposing

---

[38] 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 (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).

[39] 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 (Call & Tomasello 2008; Krupenye et al., 2016).

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 (see Goldman, 2006; Decety & 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 imply 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 & 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.

4.2 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: the "mirror-neuron" system or "mirror system" (the term was first used by Gallese et al., 1996, and Rizzolatti et al., 1996a) and 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).

The 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 & Craighero, 2004; Rizzolatti & Sinigaglia, 2006; Rizzolatti, Fogassi & Galese, 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 (fMRI, PET) (Fadiga et al., 1995; Kilner et al., 2009; Mukamel et al., 2010; Rizzolatti et al., 1996b). 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 (Rizzolatti 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 & 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 and the posterior inferior frontal gyrus and (2) a zone located at the back of the brain formed by the inferior parietal lobule. Although it does not possess neurons with strictly speaking mirror property, a third region, in the posterior superior temporal sulcus (pasts), 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 & Sinigaglia, 2006; Mollenberghs 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 paracingulate cortex (rostral ACC) and (2) a set of neural areas in the temporal cortex including the anterior temporal cortex (ATC) and posterior superior temporal sulcus (pSTS), extending up to the temporo-parietal junction (TPJ) (Farrer & Frith, 2002; Gallagher & Frith, 2003; Amodio & Frith, 2006; Frith & Frith, 2006; Young et al., 2010; Bzdok et al., 2012; Carter et al., 2012; Olson et al., 2013; Mahy et al., 2014; Mollenberghs et al. 2016). 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 & 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 7 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 & 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 & 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 (Zak & 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 the 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 et 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 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 & 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 the theory of mind, this is referred to by individuals' general ability for mentally simulating other people's mode of reasoning (Hsu & Zhu, 2012).[40]

---

[40] For a systematic confrontation between theory of mind and game theory, see Schmidt & Livet (2014). It would also be interesting to parallel the mentalizing approach with the various informational requirements posit by normative economic in which ethical principles are conditioned by the existence of either *inter*personal comparisons of utility (i.e., ability to put yourself in others' shoes, with *their* preferences)—e.g., utilitarianism, welfarist social choice—or only *intra*personal comparisons of utility (ability to put yourself in others' place, with our *own* preferences)—e.g., theories of equity and fairness, non-welfarist social choice (on this literature on theory of utility and ethics, see, e.g., Roemer 1996; Mongin & d'Aspremont 1998).

4.3 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 twentieth 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 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 also involved in empathy.

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 Sect. 5 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 these people. In empathic situations, several regions involved in behavioral and sensorial perception regulation—the secondary somatosensory cortex and the medial anterior cingulate cortex (ACC)—were also activated (Singer & 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 & Tusche, 2014). 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 & Cowel, 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 & Singer, 2008; Singer & 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; Zak & 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 (Oliver et al., 2018; Schnell et al., 2011; Zak & Ochsner, 2012)[41] (see Fig. 4 for a schematic representation of the brain areas typically involved in cognitive and affective empathy).

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.[42] However, in economics, the strategic anticipation of others' intentions has 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. Today, social neuroeconomics parallels social neuroscience studies by looking for neural foundations of other-regarding behaviors in studies employing well-established paradigms of experimental games in behavioral economics (i.e., what is known today as "behavioral game theory") (for an overview of these paradigms, see, e.g., Camerer 2003; Montet & Serra 2003; Houser & McCabe, 2009, 2014; Cartwright, 2016).

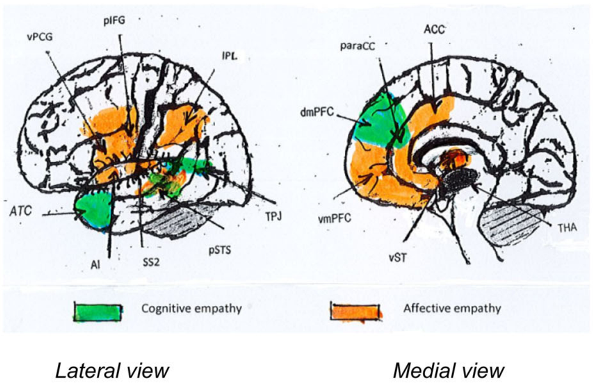
4.4 Neuroeconomics and social decisions

Many neuroeconomic experiments have shown that mutual cooperation in humans involves social cognition networks. More broadly, a lot of neuroeconomic experiments confirm an objective basis of the "rational" character of several "prosocial" behaviors that are assumed in behavioral game theory from a psychological viewpoint. In addition, to date, social neuroeconomics is the domain in which were conducted the largest number of experiments with brain stimulation or pharmacological intervention.

---

[41] Some authors introduce additional distinctions. For example, Blomm (2017) 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 & Tusche, 2014).

[42] Although the unique features of human social cognition are often emphasized, there is now evidence that they may depend on more basic social cognitive processes present in other primates and sometimes even in other mammals, including monitoring the actions of others, assigning importance to others, and orienting behavior toward or away from others (for a survey, see Rushworth, Mars, & Sallet 2013).



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) (BA 38), IPL: inferior parietal lobule (BA 40), pSTS: posterior superior temporal sulcus, TPJ: temporoparietal junction

Fig. 4 Main neural regions involved in social cognition

4.4.1 Mutual cooperation and social cognition

Many fMRI experiments show 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*[43] implies anticipating others' behavior and activating neural areas of the mentalizing network; the authors showed that a stronger activation of the dorsomedial PFC is observed in the brains of cooperative players when interacting with players localized in the lab rather than against a computer. It is further observed that for the most-cooperating subjects, the thalamus (an area often involved in emotions) is also activated (see also King-Casas et al., 2005; Delgado, Franck & Phelps, 2005). On their side, Rilling et al. (2004) examined subjects playing another well-defined paradigm, the *prisoner's dilemma games*,[44] with both human and computer partners and observed stronger activation for human partners in typical mentalizing areas.

---

[43] 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 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.

[44] 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.

A similar finding was found in the *P-beauty contest game* (or *guessing game*) (Coricelli & Nagel, 2009),[45] which examined the subject's strategic reasoning "depth" in a framework in which a psychological problem of coordination arose, but that did not refer to any social or moral motivation (like previously mentioned games) This study showed that activation of the paracingulate cortex (rostral ACC) and superior temporal sulcus (STS) was strengthened when subjects were playing against human partners rather than against a computer.[46]

Not surprisingly, thus, the regions of interest in neuroeconomic studies using classical game-theoretic paradigms coincide partly with the social cognition network.[47]

4.4.2 Neural foundation of "social" preference

Prosocial behaviors. Building on a large body of evidence that many people exhibit "social" preference (say, roughly, other-regarding preference), behavioral game theory proposed several models to explain observed behaviors inconsistent with standard game theory (voluntary participation in collective actions, help given to perfectible foreign people, costly punishment of defectors in social exchanges…) by referring to psychological factors including trust, altruism, reciprocity or inequity aversion (for a review, see Fehr & Schmidt, 2006; Serra, 2017). A lot of neuroeconomic experiments provide insights into the different neural pathways that lead to prosocial decisions and reveal explanatory mechanisms for why humans deviate from the canonical self-interest model.

---

[45] In the simplest variant of the game, each player simultaneously chooses a number P 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.

[46] This game, originally discussed as "guessing game" by Moulin (1986), is an ideal tool for assessing where the chain of iterated dominance reasoning breaks down in a strategic-form game. It was studied experimentally by Nagel (1995). This game is also called a "beauty contest" (Camerer 1997), because it captures the importance of iterated reasoning that John Maynard Keynes (1936) described in his famous analogy for stock market investment. Keynes speaks about a newspaper contest in which people guess what faces others will guess are most beautiful, and compares that contest with the stock market investment. Like people selecting the prettiest picture, each subject in the beauty contest game must guess what average number other subjects will prefer, then pick the fraction P of that average, knowing that everybody is doing the same as her/him. The P-beauty contest game is a workhorse example for the cognitive hierarchy approach in strategic thinking, such that developed by several models of bounded rationality in behavioral game theory, including *rationalizability*, *level-K,* or *cognitive hierarchy* models (Camerer, Ho, & Chong 2004a, 2004b). In these models, players use various levels of strategic thinking, and high-level thinkers distinguish themselves by correctly anticipating what players using fewer levels of thinking will do. It seems that limits of strategic thinking arise in particular from limits on working memory. For an overview of these models, see Cartwright (2016); Serra (2017).

[47] 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 PET experiments, Gallagher et al. (2002) showed that in the well-known *rock-paper-scissors game*, the paracingulate cortex (rostral ACC) 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 neuroeconomic works dealing with strategic thinking, see Camerer & Hare (2014).

Behaviors, such as altruism in *charitable donations game* (e.g., Moll et al., 2006; Harbaugh et al., 2007; Mary et al. 2009; Hare et al., 2010) and mutual cooperation in *prisoner dilemma* or *public good games*[48](e.g., McCabe et al., 2001; Rilling et al., 2002, 2004, 2008), would be viewed as rational when accounting for the "emotional" utility that these behaviors incite and elicit neural traces observed in brain regions involved in the "reward network" or in positive emotion processing, including the thalamus and NAcc or caudate (in the striatum). However, "altruistic" punishment is another form of prosocial behavior also widely studied in neuroeconomic experiments.

"Altruist" punishment behavior. We know that social punition is a pervasive feature of human society and it strongly shapes the enforcement of social norms. "Altruist" punishment behavior against "defectors" in classical games of cooperation or bargaining is one of the strong results of behavioral economics, whether it is direct punishment in public good or trust games (when the game rule allows, it is possible to punish subjects breaking the social norm of cooperation or reciprocity) or indirect punishment in ultimatum games (in which responders can take revenge by refusing an offer they consider inequitable). From a strict logical viewpoint, this behavior seems "irrational" in the two cases, because it is costly for each player: monetary cost in one case and opportunity cost in the other, because subjects could have won some money. A lot of neuroeconomic experiments focus on the neural representation of these punishments, whether on direct punishment of "defectors" in *public good with punishment*[49] or *trust game* (e.g., de Quervain et al., 2004; Joffily et al., 2014; Singer et al., 2006) or indirect punishment in *ultimatum games*[50]

---

[48] The structure of public good (PG) games is similar to that of prisoner's dilemma (PD) 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). In on another side, 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. The goal of *charitable donations games* is to experimentally study altruistic giving in a PG framework.

[49] 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.

[50] Two participants are randomly and anonymously matched, one as *proposer* (player P) and one as r*esponder* (player R), and told that 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. Remark that 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.

(e.g., Sanfey et al., 2003; Knoch et al., 2006b, 2008; Baumgartner et al., 2011, 2012) (for a recent meta-analysis, see Belluci et al., 2020).[51]

Let us consider two neuroeconomic experiments of direct punishment as an illustration of these "altruist" punishments. In de Quervain et al. (2004), a PET experiment using a variant of the trust game, investors were confronted with subjects who did not follow the reciprocity norm and had the opportunity to punish them. Such punishment could be *symbolic* (the other player's gain is not reduced) or *effective* (the other player's gain is reduced). The authors scanned the investors' brains at the moment they found out that the other player did not return any money on the investment. Compared with symbolic punishment, effective punishment activated the lower part of the caudate nucleus (an area known to be heavily involved in positive emotions) in the cheated investor. Furthermore, the greater the activation in the subjects, the more they were prepared to accept higher costs to inflict their penalty. This result supported the hypothesis that players obtained "emotional" utility from the punishment they inflicted on deviants, i.e., those who betrayed their trust. This "altruistic" punishment gave individuals the gratifying feeling of having done their social duty; a psychological experience that could be traced at the neural level.

The second example concerns an experiment that tends to show that punishment behavior does not activate the same neural areas in men and women. In Singer et al.'s study (2006), subjects of both genders first a sequential Prisoner's Dilemma game with partners whose role was actually played by the experimenter's accomplices. This first phase aimed to establish a reputation of "cooperative" or "uncooperative" players in these partners, given that it became rational to cooperate in return only with the former. The brains of the tested subjects were then scanned by fMRI when a slight electric shock was administered either to the subjects themselves or to their partners (the accomplices) considered cooperative or not. When the cooperative partners received an electric shock, the anterior cingulate cortex (ACC) and the anterior insula were activated—two areas associated with certain positive emotions, such as empathy, or negative ones, such as sorrow or disgust. All the subjects, both men and women, appeared to be able to withstand an "emotional" disutility. However, curiously with uncooperative partners, only men's brains showed activation of the nucleus accumbens (NAcc) and the orbitofrontal cortex (OFC)—two areas heavily involved in positive emotions. Therefore, only men seemed to find some "emotional" utility in the punishment inflicted on uncooperative partners in this context, a behavior interpreted by the authors as the display of a feeling of revenge.

---

[51] Several forms of social punishment are identified, including second-party or third-party punishment. "Parochial "altruism, namely a preference for altruistic behavior towards in-group members and mistrust or even hostility towards out-group members (e.g., one's ethnic, racial, or any other social group), is a pervasive feature in human society. Parochial altruism involves a third-party punishment behavior. Recent evidence from fMRI studies suggested that areas involved in social cognition (including dorsomedial PFC and bilateral TPJ) must play a role in differentiating in-group and out-group members in behavior (Baumgartner et al., 2012), while Baumgartner et al. (2014) showed that the transient disruption of the right (but not the left) TPJ reduces parochial punishment with real social group.

Social neuroeconomics, brain stimulation, and pharmacological intervention. So far, social neuroeconomics is the domain in which were conducted the largest number of experiments with tools that generate significant causal effects. It was shown in particular that a transiently experimentally induced variation in the activity of the dorsolateral PFC by rTMS or tDCS[52] may make the participants more or less receptive to fairness concerns (Baumgartner et al., 2011; Knoch et al., 2006b, 2008), more or less inclined to disregard certain social norms (Ruff et al., 2013) or to build up a favorable reputation (Knoch et al., 2009) and may even influence their spontaneous propensity to be honest or to lie accordingly (Karton & Bachmann, 2011; Baumgartner et al. 2013).[53]

Let us take as an example of the important social mechanism of reputation formation. Knoch et al. (2009) showed that disrupting the right dorsolateral PFC diminishes subjects' ability to build a favorable reputation.[54] This effect occurs even though the subjects' ability to behave altruistically in the absence of reputation incentives remains intact, and even though they are still able to recognize both the fairness standard necessary for acquiring and the future benefits of a good reputation. Thus, this result suggests an important dissociation between knowledge about one's own best interests and the ability to act accordingly in social contexts. This may help explain why reputation formation remains less prominent in most other species with less developed prefrontal cortices.

Concerning pharmacological interventions, to date, oxytocin is by far the chemical substance subject to most experimental inquiries. This neuropeptide, mainly produced in the hypothalamus, is often considered to be a "female hormone." It plays a role in promoting mother–child bonding and directly affects breastfeeding. However, it also has a broader socializing function.

Using the trust game as a behavioral paradigm to test the trust of investors following intranasal administration of oxytocin, Kosfeld et al. (2005) showed that oxytocin increases the willingness to trust by raising individuals' readiness to take "social" risks (see also Baumgartner et al., 2008; Mikolajczak et al., 2010). One of the properties of oxytocin is the inhibition of amygdala activation, a zone known to play a role in processing the negative emotions caused by fear and anxiety.[55]

Several experiments have also studied the effects of nasal inhalation of oxytocin on generosity, betrayal aversion, empathy, positive and negative reciprocity, and inequality aversion with various classical behavioral paradigms (ultimatum game, repeated prisoner's dilemma game, charitable donations game, etc.).

---

[52] For a brief presentation of these tools, refer to paragraph 2.1.1.

[53] These studies complete the rare experiments that study in a game-theoretic framework the social behavior of patients with prefrontal damage. Krajbich et al. (2009), in particular, found that patients with damage to the ventromedial PFC show a specific insensibility to guilt.

[54] We know that reputation was broadly studied in repeated game theory with private information. Several fMRI experiments directly or indirectly tap into aspects of reputation (e.g., Delgado, Franck & Phelps, 2005; Singer et al., 2004).

[55] However, it turns out that oxytocin inhalation does not affect the loyalty of the trustees. To explain this asymmetry between investors and trustees, the authors highlight the difference between "pure" trust found in investors (that can only be generated by a certain empathy) and the "calculated" trust of trustees (as it is a function of their experience during the game).

First surveys on the role of oxytocin on behavior suggest that this neuroactive hormone is closely related with social interactions and essential for prosocial behavior, in particular for trust (see, e.g., MacDonald & MacDonald, 2010; Zak, 2011; Riedl & Javor, 2012). Yet, recent literature on this issue has generally produced rather contrasted results. Even though oxytocin treatment has been suggested to have implications for at least patients suffering from disturbed social behavior (e.g., autism or social anxiety disorder), the relationship between oxytocin and human trust behavior is still an open question. Evidence that oxytocin directly influences trust behavior remains unclear. If there is an effect, it is likely to be moderated by a variety of factors (Koppel et al., 2017; Nave et al., 2015). For a recent overview of some results about the effects on social behaviors of various chemical substances, including oxytocin, vasopressin, testosterone, and serotonin, see Reuter & Montag, 2016)[56] confirms this accordingly.

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 affective and cognitive empathy—are consistently recruited when people face social dilemmas and economic exchanges in cooperative and bargaining environments.[57] These neural networks act together or in competition with those dedicated to cognitive control (see above Sect. 2) and reward processing (for reviews on social neuroeconomics, see Fehr, 2009; Sanfey & Rilling, 2011; Rilling & Sanfey, 2011; Fehr & Krajbich, 2014; Declerck & Boone, 2016; Serra, 2016; Engelmann & Fehr 2017; Dreher & Tremblay, 2017; Alos-Ferrer, 2018).

5 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 & 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 the interest in experiments with animals (e.g., non-human primates or rodents) for understanding the function of the human brain (see Santos & Platt, 2014).

---

[56] 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 & Fehr (2014).

[57] It should be noted that social neuroscience literature covers a much broader thematic domain than questions of social cognition. A lot of studies concern in particular 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 (e.g., the famous "trolley problem") (Christensen & Gomila 2012).

5.1 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 (search for food, reproduction, and danger avoidance) to enable preservation of the individual and the 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 determine most rewards. In modern humans, the physiological model of threat to the 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).

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 (Niv & Montague, 2009; Balleine et al., 2009;

Balleine & O'Doherty, 2010; Daw & Tobler, 2014). This is one of the main findings of modern neuroscience that will be briefly outlined later 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, it turns out that experimental works with rats and pigeons by some "neo-behavioral" psychologists carried the seeds for this revolution since the 1960s (Herrnstein, 1961; see also Herrnstein & 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. 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 Paul Glimcher. These studies, sometimes named the "quantitative approach of behavior" (Richard Herrnstein, then Howard Rachlin, Georges Ainslie, Drazen Prelec), were among the first to shed light on impulsivity and intertemporal choice (see Commons, 2001). The theoretical debt of neuroeconomics to these neo-behavioral psychologists is hence fully recognized (Glimcher, 2003). Importantly, there are significant differences between this branch of psychology, which is heir to the "behavioral school" born in the early twentieth century in opposition to introspective psychology (Watson, 1913), and modern behavioral economics born in the 1980s fueled by studies in cognitive psychology, primarily by Daniel Kahneman and Amos Tversky's works (Heukelom, 2014).

In the quest for rewards and to avoid punishments, how is the nervous system 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 modern 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 the standard neuroeconomics model, we focus on a major discovery of contemporary neurobiology, namely the leading role of "dopamine" in the reward system and regulation of neural activity.

5.2 Dopamine and the reinforcement learning system

For several years, most neuroscientists suspected that dopamine—one of the main neuromodulators—as well as the neurons using it (i.e., "dopaminergic" neurons) played 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 neurochemistry researchers showed that this was incorrect; their studies revealed that highly localized clusters of cell bodies synthesized 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 (Schultz, 2009; Tobler & Weber, 2014).

5.2.1 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, Adam, & Barto 1995; Montague et al., 1996; Schultz, Dayan & Montague, 1997) (this last experiment is by far the most well-known in this literature). For a more formal presentation of the reinforcement learning system, refer to Daw &Tobler (2014).[58]

These pioneering studies on the role of dopamine in reward learning date from the 1990s before the paper written by Platt & Glimcher (1999) that generally is accepted as the first neuroeconomic study. Therefore, we may consider them as a kind of “theoretical prehistory” of neuroeconomics.

The research of Platt and Glimcher is the first work using single-unit recording that actually 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. So as to drawing an analogy between reward expectation in neurobiology and expected utility in economics, Glimcher coined later the term “physiological” expected utility (Glimcher et al., 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. It is a 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

---

[58] These experiments revealing the role of dopamine in reward system were carried out in non-human primates. However, a similar mechanism was shown to also 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 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).

involved mechanism: we can understand how the monkey, conditioned to make eye movements, is constructing and assigning "subjective value" to each option offered.

Yet, we must emphasize the strong specificity of this pioneering experiment: results exclusively concern modeling of brain functioning in *non-human primates* undertaking an "action", namely an eye-tracking task. For a neurobiologist, such an experimental context may be relevant. For an economist, it must be expanded to human experiments 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 we will come back to later in the manuscript.

5.2.2 The dopamine circuit and the centralized regulation of neural activity

For a long time, there were very few human studies on the role of dopamine in reward learning as it is technically difficult to scan target areas of dopamine in the mesencephalon using an fMRI (see, however, McClure et al., 2003, and O'Doherty et al., 2003). The first experiments on humans date from the end of 2000s with the work of Ardenne et al. (2008), an experiment that adopts an original protocol mixing electroencephalogram (EEG) data and fMRI images. Zaghloul et al. (2009) performed the first electrophysiology study on humans during deep brain stimulation therapy in patients with Parkinson's disease, and Pessiglione et al. (2006) carried out the first study of pharmacological manipulation that established the causal role of dopamine. Ultimately, even if some cytoarchitectonic differences between the monkey brain and the human brain are observed, it appears that the *cerebral processes* are similar.

Today, there is a vast literature, referring to neurophysiological experiments in animals alongside neuropharmacological and neuroimaging research in human, dealing with the role of dopamine in the reinforcement learning process related to reward prediction error (RPE) and neural activity regulation (Niv & Montague, 2009; Schultz, 2010, 2013, 2016; Cools, 2011; Glimcher, 2011a, 2011b; Daw & Tobler, 2014; Daw, 2014; Doya & Kimura, 2014; Rutledge et al., 2015; Martin, Mehta, & Prata 2017).

It should be remarked however that the role of this neuromodulator 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 (e.g., Arnsten et al., 2012).[59] In

---

[59] 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. Therefore, dissociating effects of attention from those of working memory is difficult, and in practice, the two processes are interactive (Awh & Jonides 2001). The dopaminergic system is a primary pharmacological target for psychiatric disorders which are associated with attention deficits such as attention deficit, hyperactivity

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.[60]

Whether it be with the human or monkey brain, neurobiologists identified the "dopamine circuit"—the brain zones in which neural information circulates mostly due to this neuromodulator. 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, namely the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNpc) (Bjorklund & Dunnett, 2007). DNs irradiate in the brain by following two main networks: (1) from the VTA clusters up to the ventral striatum (the Nacc and lower parts of caudate nucleus and putamen) and the PFC, and (2) from the SNpc cluster up to the dorsal striatum (upper parts of caudate nucleus and putamen). Functional experiments support the preferential link of VTA DNs to reward (the "reward network") and SNpc DNs to motor action, respectively (Howe & Dombeck, 2016). 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 forming a cortico-striato-thalamic network also are involved (Chase et al., 2015; Glimcher, 2011a, for a meta-analysis) (see below a schematic representation of the dopamine circuit in Fig. 5).

Two additional results also should be reported. First, it should be noted that all regions targeted by dopamine in the reward network do not verify the encoding conditions of RPE that might be stated (Caplin & Dean, 2008, 2009). Learning from the axiomatic method commonly employed by economists in modern theoretical economics, these authors wondered whether it was 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 nucleus accumbens (NAcc) verifies the three axioms. Interestingly, Caplin & Glimcher (2014) suggested a generalization of this approach in the field of neuroeconomics for building an "axiomatic" neuroeconomics, which may have the

---

Footnote 59 continued
disorder, schizophrenia, and Parkinson's disease (e.g., Arnsten & Rubia 2012). Note that 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 & Bellgrove 2018).

[60] Rolls (2014), particularly, 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 & 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 lot 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.

advantage of solving some disagreements revealed by the literature (see also Dean, 2013).

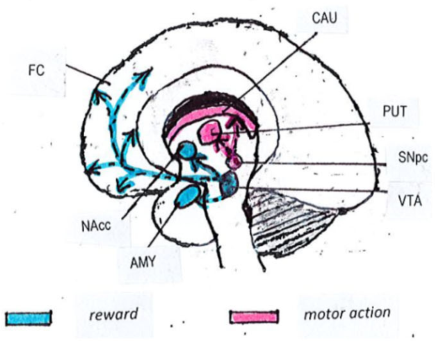
It should also be noted that some areas belonging to the reward network are actually targeted by other "afferents"—specifically serotonin. Several scholars suggest that serotonin, producing feelings of satiety and inhibitions, is active in the field of *negative* RPE (e.g., Daw et al., 2002; Niv & Montague, 2009).[61] This hypothesis is rather attractive, because it could resolve 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 & 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 (e.g., Kahneman & Tversky, 1979; Tversky & Kahneman, 1992, for humans,[62] and Chen et al., 2006, for non-human primates). Therefore, 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, yet, it is acknowledged that understanding is still rudimentary and contested concerning encoding of negative RPE (Daw, 2014; Glimcher, 2011a).

Neuroeconomics has integrated these neurobiological discoveries on the brain function of both humans and non-human primates into the conceptual framework of standard economic theory and developed a set of major theoretical innovations toward understanding the decision-making process with preliminary innovative implications in the interpretation of economic behavior.

---

[61] 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 et al., 2011).

[62] The fact that the subjective impact of a loss is greater than that of an equivalent gain is one of the general principles underlying the famous prospect theory. This theory has been tested in recent years by numerous neuroeconomic experiments that have corroborated its main hypotheses such as loss aversion and the non-linearity of the probability-weighting function, but reference-dependence in decision-making and framing effects remain unclear (refer to Fox & Poldrack 2014; Louie & De Martino 2014). Glimcher (2011a, 2011b) established a parallel between the idea of reference point introduced by Kahneman and Tversky and a similar concept in neurobiology. It is interesting to note that Kahneman himself was involved in one of the first experiments in neuroeconomics (Breiter et al., 2001). However, the status of the neural data in this experiment is ambiguous. As with all pioneering experiments in the early 2000s, it is claimed that the experiment is set within reward learning theory, yet it is clear that the prospect theory also plays the role of experimental paradigm. Neural data are alternately considered as parameters of the Kahneman–Tversky model (*exogenous* variables that must be estimated to "calibrate" the model) or explanatory variables (*endogenous* variables that are progressively corrected by the neural-learning process). This experiment shows clearly the difficulty that must be faced when transposing the "anomalies", namely the disparities between "ideal" economic rational and observed behavior, into the theoretical framework of reward learning. In neurobiology, irrational behavior is appraised against learning dynamics (Fox & Poldrack 2014).



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

Fig. 5 The dopamine circuit

5.2.3 Major neuroeconomic theoretical innovations and their implications for the interpretation of economic behavior

As has already been said, *unicit*y is viewed as a prominent property of the choice mechanism in neural terms. Acknowledging that *uncertainty* is a fundamental element to understand how the brain functions is also a major component of neuroeconomics. These two fundamental properties shed a new light on economic decisions proposed by neural economics.

Unicity of the choice mechanism: the "common currency" hypothesis. The discoveries of modern neurosciences tend to support the existence of a *unique* reward enhancement mechanism, i.e., a mechanism that departs a priori from the dualist view traditionally adopted by cognitive psychologists, behavioral economists, and the neuro-economists who follow in their footsteps. A central message 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 & Berns, 2002, p. 276). For choosing an appropriate behavior, the nervous system must estimate the value of each potential actions, and 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 predicted 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 & Berns, 2002; Landreh & Bickle, 2008; Schultz, 2009; Balleine, Daw & O'Doherty, 2009; Niv & Montague, 2009; Rangel, 2009; Chib et al., 2009; Levy & Glimcher, 2012).[63]

Neuroeconomics is now anchored in a new theoretical framework that posits the existence of a centralized neural mechanism that enables the a priori incomparable subjective values of multiple stimuli to be compared using a single scale, within which the production of dopamine appears to be one of the major components. In so doing, neuroeconomics is distancing itself from behavioral economics.

New discussion of intertemporal choice in economics. The economic issue of intertemporal choice and impulsivity provides a good example of the divergence, which must be stressed, between this new theoretical substructure and that originally adopted by behavioral economics in the scanner. According to the classic interpretation suggested by McClure et al. (2004a), intertemporal trade-offs reveal the rivalry between an "impulsive" limbic system and a more "farsighted" PFC, and this temporal inconsistency is interpreted as a reasoning bias (refer to the passage relating to this experiment in Sect. 3). In the reinforcement learning paradigm, it is no longer simply a logical consequence of the behavior of an individual adapted to his/her environment.[64] In an evolutionary perspective, impulsivity resists a simple interpretation in terms of "anomaly." For the heirs of the 1960s neo-behaviorism psychology, a rational agent is not characterized by the adoption of a constant discount rate, but rather by the ability to maintain a sufficiently stable balance between the various time scales. The dualist interpretation of intertemporal choice, originally proposed by McClure et al. (2004a), may thus be called into question. Kable and Glimcher (2007) conducted an experiment using a protocol similar to that of McClure et al. (2004a), but kept the amount of the immediate reward constant and only adjusted the amount and/or the date of the delayed reward. In showing that the activity of the ventral striatum, medial PFC, and orbitofrontal cortex (OFC) was positively correlated with the variable amount of the reward and negatively with its delay, the results of this experiment ran counter to the interpretation of McClure et al. (2004a), because some of these regions were precisely the ones identified by the latter as the components of an impulsive evaluation system, devoted to the evaluation of immediate rewards. Is it possible to reconcile these two interpretations of intertemporal choices? This was attempted by

---

[63] Today, the "common currency" hypothesis is widely accepted in the neuroscientific community. Yet, there are some rare researchers who do not fully agree with it. They argue 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 & Rolls 2011; Rolls 2014).

[64] Of course, this evolutionary advantage may become a disadvantage in some environments where the structure emphasizes likely utilities and rewards in the very short term. However, the flaw lies in the *environment* and not in the individual (Ainslie 1992).

Hare, Camerer, & Rangel (2009), who dissociated the specific functional role of different regions within the reward network and introduced self-control as an additional variable in the choice mechanism.

The key feature of this interesting experiment lies in the behavioral discrimination between two groups of subjects—obese subjects regarded as impulsive and "healthy" subjects regarded as self-controlled. An original element of the study was to use a symbolic delay in granting a food reward: the positive long-term effect on health (dietary quality of the food chosen). The results showed that the ventromedial PFC was active in all the subjects when the food item was selected, reflecting the value given to it by the brain. However, while the vmPFC activity was correlated with both taste and dietary quality in self-controlled subjects, it was only correlated with taste in impulsive subjects. Furthermore, the dorsolateral PFC was also activated when the choice implied self-control (i.e., choosing to reject an appetizing but unhealthy food item). In addition, although the neurons in this region were activated in all subjects, the activity was relatively more intense in those with self-control. Modulation by the dorsolateral PFC of the signal of the value encoded in the vmPFC may thus help to incorporate long-term objectives in the evaluation of stimuli.

Where is the fundamental difference between this analysis and that initially suggested by McClure et al. (2004a)? It lies mainly in the interpretation of the executive function fulfilled by the dorsolateral PFC, which is more of a "modulation" than an "inhibition." The opposition between an "impulsive" system (located in the limbic and paralimbic regions) and a "patient" system (in the PFC) gives way to an interpretation that is consistent with the evolutionary hypothesis: the executive function of the dorsolateral PFC is not seen as competing with the evaluation of immediate responses by the limbic system (i.e., an inhibitory function that might counteract the reward system) but rather as its *evolutionary extension.* This justifies the continuity between the evaluation of primary rewards and more complex, specifically human forms of reward (monetary or symbolic/social) in line with the "neuronal recycling" hypothesis of Dehaene & Cohen (2007) (refer to Bourgeois-Gironde, 2010).

Stochasticity of choice mechanism. According to modern neuroscience findings, *stochasticity* is also a prominent choice mechanism in neural terms. Owing to the uncertainty that affects the period of time between making a decision and obtaining the reward, it is never possible to exactly estimate a future reward. Such uncertainty must be managed, because future reward predictions incur real costs for the organism (Montague & Berns, 2002). Fundamentally, the random nature of neural signals stems from the need for the brain to develop reward learning strategies. Reward prediction is designed to be applied to choices between options where gain probabilities are unknown at the start but are learned progressively. The main property is that the neural signal does not concern isolated choices, but rather the sequences of repeated choices. It can, therefore, take different values during the repetitions, even if the choice remains exactly the same (Glimcher, 2003).

For economists, the stochastic property of the neural choice mechanism is just as fundamental as the "common currency" hypothesis. In particular, it shows that there is a significant difference between the way the brain processes uncertainty and the way economists generally address the issue of risk. Nearly all economic models, in risky situations, separate the occurrence probabilities of various potential states and the expected utilities in these states. These probabilities and preferences are assumed to be given. In neurobiology, this separation no longer stands: the brain functions in a way that suggests a separation between the encoding of *anticipated reinforcements* and the *risk* that inevitably affects the choice process, as the nature of the reinforcements is necessarily stochastic. The learning process relates to both the amount of reward and its probability of occurrence, two inseparable dimensions in the choice process (e.g., Bossaerts, Preuschoff, & Hsu 2009). Uncertainty is present during the entire decision-making process, a finding that is a major contribution of neurosciences to the understanding of decision-making mechanisms (Fehr & Rangel, 2011; Glimcher, 2011a; Rolls, 2014). Incidentally, this implies that the reinforcements identified in neurosciences should not be confused with expected utilities traditionally calculated by economists—a confusion that may have been nurtured by Glimcher who, for a while, used the term "physiological expected utility" (Glimcher et al., 2005), especially because some reinforcements are not under the control of higher cognitive functions, but are generated mostly by the involuntary emotional system. The brain function in terms of rewards/punishments eclipses the distinction between "rational" and "emotional" systems accordingly.

New discussion of financial choices in economics. Some neuroeconomic experiments relating to financial choices illustrate the division of the brain between rewards/punishments and risk. They also show that it is worth addressing the issue of "when" and not just "where" the neural signal intervenes. This work also clarifies how risk is encoded by the brain and it distinguishes clearly between anticipated return (gain anticipation) and risk perception (gain variance). On the contrary, although the same area, the NAcc (within the ventral striatum), is involved, neurons react to return without delay, whereas they take some time to respond to risk, although more intensely (Knutson et al. 2003, as regards the correlation between this area and return; Preuschoff et al., 2006 and Preuschoff, Quartz, & Bossaerts 2008a, as regards the correlation between return and risk). This result suggests that, rather than addressing risk through a unidimensional indicator (i.e., expected utility or a modified version of expected utility) as in the standard decision theory, the brain uses the alternative approach instead, developed traditionally in behavioral finance, which involves breaking down this subjective measurement using a Taylor series to clarify the different facets of risk (mathematical expectation, variance, skewness, kurtosis, etc.) and focusing on the first two statistical moments following the Markowitz theory(1952).

Should we therefore accept, as do some authors (Quartz, 2008; Schultz et al., 2008), the neural "validation" of the mean–variance model rather than the expected utility model with regard to financial choices? This interpretation, although interesting, seems premature. In particular, these two models are representative of two quite different approaches, and we know that this is not the case because of their conceptual affinity: the two models actually coincide in the case of a quadratic utility function (Schmidt, 2008).

"Actions" versus "economic choices". In the standard neuroscientific approach and typically in neural economics, choices are fundamentally choices between

"actions". Intuitively, it is a less than satisfactory hypothesis; we have a strong intuition that, for example, when consumers choose durable household goods such as 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 & Assad, 2006, 2008; Padoa-Schioppa, 2011).

In any case yet 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 & Glimcher, 2009; Rangel & Hare, 2010). But importantly, 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 though, making remains in fact a controversial issue (see Glimcher, 2014a; Platt & Plassmann 2014; Padoa-Schioppa & Conen, 2017).

5.3 Learning and valuation

To overcome this uncertainty, the brain develops learning mechanisms. As mentioned above, there is agreement in neuroscience literature that several distinct learning processes coexist in the brain—at least three systems. On the other hand, behavioral game theory has developed several learning models, and it is interesting to know whether they match these neural algorithms.

5.3.1 The three learning system

Psychologists and neuroscientists nowadays agree that the brain employs at least three different systems (see Bouton, 2007; Rangel et al., 2008; Balleine, Daw & O'Doherty, 2009; Balleine & O'Doherty, 2010; Dolan & Dayan, 2013; Daw & 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.[65] With this mechanism, an organism can learn to make predictions about when biologically significant events are likely to occur and to learn the stimuli that tend to precede them (Pavlov, 1927/1960). Pavlovian behaviors are more flexible than simple

---

[65] For example, where reward is concerned, eat any food within reach in a buffet regardless of how hungry you are; where punishment is concerned, cross the road at the sight of a suspicious-looking individual to avoid a possible attack.

reflexes to emit behaviors shaped by predictive learning, but they also are inflexible since the responses themselves are stereotyped.[66]

(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 foundation in the stimuli–response reflex. Habitual systems are relevant only for stable enough environments, without sudden change, and at slow rates.[67]

(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.[68]

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 & 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, learning is merely passive—behaviors are independent of their outcomes. The same is not true of behaviors acquired in other conditioning preparations, notably *instrumenta*l conditioning (Thorndike, 1911). However, there is now considerable evidence confirming that goal values and instrumentally conditioned reflex values are mediated by distinct neural processes (Balleine, Daw & 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.

It turns out, in fact, that reinforcement learning has been primarily developed within the field of artificial intelligence and machine learning. 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) (Dayan, 2008; Sutton & Barto, 1998). Importantly, this implies correspondence between these two classes

---

[66] Pavlovian learning is known to be present in vertebrates, including humans, as well as many invertebrates, including insects such as drosophila.

[67] For example, where reward is concerned, drink a cup of coffee every morning as a stimulant regardless of the specific need felt on that particular day; where punishment is concerned, select the same route every day to drive to work regardless of any foreseeable traffic jam on that particular day.

[68] For example, where reward is concerned, select the film at the cinema according to your taste to make it the most pleasurable experience possible; where punishment is concerned, decide to jog regularly to minimize the risk of obesity.

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. The "temporal difference" algorithm, which supposes that reward predictions are continuously improved by comparing them with current rewards, is an often-quoted example of model-free RL in computer science (the most known variant is the "actor-critic" model). Although model-free RL has received the majority of attention, there has been more recently an increasing understanding for how the brain uses model-based methods (Daw et al., 2005; Balleine, Daw & O'Doherty, 2009; Dolan & Dayan, 2013; Daw & O'Doherty, 2014).

Nowadays, concerning learning many questions persist. First, we do not know whether only three kinds of learning systems exist or whether there exist multiple Pavlovian, habitual, and goal-directed systems, each one specialized in a class of actions or outcomes (Rangel, Camerer & Montague, 2008).[69] Second, the question of status assigned to each system in terms of state of consciousness is still open. Although Rolls (2014) 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 & Changeux, 2011; Dehaene et al., 2006). 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, i.e., to minimize error risks. It is typically observed that the brain mobilizes a mixture of algorithms, though the weights of model-free or model-based strategy vary across individuals and contexts (Daw & O'Doherty, 2014; Daw et al., 2005; Gläscher et al., 2010; Rolls, 2014).[70] In this respect, from the study of neural mechanisms involved in the treatment of uncertainty, we find a general qualitative conclusion that converges to the one we highlighted in the anatomical–functional identification of neural regions specialized in emotional and cognitive treatments (see Sect. 3). 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.

---

[69] Lengyel & Dayan (2007) advance the hypothesis of a fourth "episodic" system managed by the hippocampus. More recently, O'Doherty et al (2017) review evidence that an additional system would guide inference concerning the hidden states of other agents, such as their beliefs, preference, and intentions, in a social context.

[70] For Pavlovian systems, Dayan et al. (2006) have proposed some hypothesis. More recently, Clark et al. (2012) review first evidence of the existence of multiple parallel Pavlovian valuation systems. Interaction between habitual and goal-directed 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 (Daw & O'Doherty 2014).

5.3.2 Learning economic models and neural algorithms

Economic models are organized around two forms of learning: reinforcement-based learning and belief-based learning.

Originally, reinforcement learning models are derived from behavioral psychology: people are assumed to select strategies based on what worked in the past, i.e., their decisions that were reinforced by their experience. This principle, originally applied to individual decision-making, illustrates the influence of *routine* on behavior (Bush & Mosteller, 1955), and through the *imitation* game can also be extended to game situations (e.g., Roth & Erev, 1995).

Belief learning models are more sophisticated and seem to be more efficient in keeping with the usual mode of reasoning of economists: individuals are assumed to establish beliefs about what other people will do based on past choices and they tend to adopt the best response based on these beliefs, i.e., they choose a strategy that maximizes their expected utility given their beliefs. Various models are possible depending on the rules of belief formation. In the simplest model, it is assumed that players select the best response based on the behavior of their competitors during the most recent period only (Cournot, 1838); in theories of *fictitious play*, players are assumed to form their beliefs from the frequency of observed decisions made by their competitors in the past rather than just during the previous period (Brown, 1951); in *weighted fictitious play*, importance is given to beliefs acquired in the recent rather than remote past (Cheung & Friedman, 1997). The latter model has the advantage of including as special cases both the Cournot model and the traditional *fictitious play* model, where past observations are equally weighted (refer to Fudenberg & Levine, 1998 for a presentation of the different belief-based learning models).

These two approaches have long been considered irreducible. However, it has since been shown that it is possible to combine them by developing hybrid models, of which the experience-weighted attraction model (EWA) developed by Camerer & Ho (1999) is the best known (see also Camerer, Ho, & Chong 2008). The strong advantage of the EWA model is that it combines both reinforcement- and belief-based learning (for surveys of economic learning models refer to, e.g., Camerer & Hare 2014; Serra, 2016).

The question arises as to whether there is any correspondence between these various economic models of learning and the main classes of neural-learning algorithms identified in neuroscience. From the outset, it should be stressed that the expression "reinforcement-based learning" does not have exactly the same meaning in artificial intelligence and neuroscience on one hand and in economics on the other. In the former disciplines, as mentioned previously, it refers broadly to any learning process within a decision-making context, including both the model-free approach (habitual decisions) and the model-based approach (goal-directed decisions), whereas in economics, it only relates to the model-free approach (Daw, 2014, p. 302). However, some concepts from the theory of behavioral games, incorporated as specific cases by the EWA model, are similar to the two main classes of neural-learning algorithms. Some neuroeconomic fMRI experiments have studied this model without drawing firm conclusions on the neural areas involved accordingly. Thus, Zhu et al. (2012) observed an activation of the paracingulate cortex (rostral ACC)—which is frequently involved in strategic decision-making and social cognition issues (refer to Sect. 4)—in belief-based learning and an activation of the ventral striatum in reinforcement-based learning. However, for their part, Lohrenz et al. (2007) concluded that there are neural correlates in the striatum in belief-based learning (for a more thorough discussion on this topic, refer to Camerer & Hare 2014).

5.4 Towards a computational model of 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 (Fehr & Rangel, 2011; Glimcher, 2011a). The contours of such a model emerge today in neuroeconomic literature. They arise from the theoretical/computational neuroscience teachings coupled with the various experimental findings of the past 20 years.

This model applies to goal-directed decisions[71] 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. 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 (for risky decisions, see Levy et al., 2010; for intertemporal decisions, see Hare et al., 2009; Fehr & Rangel, 2011; Kable, 2014). 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 (Hutcherson et al., 2015; Krajbich et al., 2015; Ruff & Fehr, 2014; Wake & Izuma, 2017).

Schematically, the key components of the model—which can be described as "standard" in that it covers the various variants proposed in the literature—are based on two mechanisms that are supposed to proceed sequentially: (1) a *valuation* mechanism that learns, stores, 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 in so far as

---

[71] 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.

these resources are scarce and costly in terms of consuming energy (Fehr & Rangel, 2011; Glimcher, 2014a; Kable & Glimcher, 2009).

5.4.1 The valuation mechanism

The concept of "subjective value" (SV) of a decision is at the heart of this device (see Sect. 2) 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 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 & Glimcher, 2009; Rangel & Hare, 2010; Berridge & O'Doherty, 2014; Fehr & Rangel, 2011; Padoa-Schioppa, 2011; Padoa-Schioppa & 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, 2003, 2011; Kahneman et al., 1997).[72]

It should be noted that some authors do not explicitly refer to this notion of "experienced utility". Rangel & Clithero (2014) are talking 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 "physiological" expected utility proposed earlier in Glimcher, Dorris & Mayer, (2005)—and neural regions where this value is "learnt and stocked" (i.e., where it is "constructed"). We will follow this way of presenting the valuation mechanism.

The "phase" of decision subjective value. A lot of 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 Sect. 2). Furthermore, experiments using manipulation techniques (i.e., experiments able to prove causality and not just correlation), such as brain stimulation, ended 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 qualitative reviews of these many studies, see Fehr & Rangel, 2011; Rushworth et al., 2011; Wallis 2012; Levy & Glimcher, 2012; Platt & Plassmann 2014; and Clithero & Rangel, 2014).

---

[72] Other distinctions are developed in the neuroeconomic literature. Bossaerts, Preuschoff & Hsu, (2009), in particular, mention "true" preferences (what individuals want) and "revealed" preferences (what individuals do), while Berridge & O'Doherty (2014) separate 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" (Berridge & O'Doherty 2014, p. 242).

It turns out that several authors add the ventral striatum (the NAcc and lower parts of the caudate nucleus and putamen) to the ventromedial PFC as contributing to SV (e.g., Glimcher, 2011a; Knutson et al., 2009; Levy & 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) make 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).

Moreover, in meta-analyses of several hundred fMRI studies, Bartra et al. (2013) and Clithero & Rangel (2014) found evidence that the posterior cingulate cortex (PCC) also correlates with SV in different studies with human subjects. However, Kanayet et al. (2014) have argued that these correlations are likely due to the role of PCC in processing magnitude rather than SV per se.[73]

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 is not negligible relative to the benefits from consuming them or when the action costs of the options under consideration are very different. For Wallis & 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; Rangel & Hare, 2010; Rangel & Clithero 2014).[74]

The “constructing” phase of decision subjective value. This phase refers to processes by which the brain is using experienced utility signals supposed to be computed at the time of reward or punishment “consumptions”, i.e., 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. However, authors agree that the reward system and RL processes are likely taking a

---

[73] For instance, the PCC is more active in response to a reward of 100 cents than 1 dollar, while the ventromedial PFC and striatal responses to these rewards are indistinguishable.

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

central place in decision value construction by favoring the building of a stimuli valuation common scale and by feeding the memorization system that keeps track of relevant environmental information collected and continuously updated (see above in this section).

Abundant evidence suggests that a wide neural network takes part in the computation of experienced utility signals, including: (1) the orbitofrontal cortex (OFC) for the valuation of many sorts of primary rewards (various consumer goods: food, drinks, music, works of art, etc.) 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 anterior cingulate cortex (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 constructs a subjective value (SV) signal that guides choices. These findings are thoroughly documented in some meta-analyses of several hundred fMRI studies (Bartra et al., 2013; Chase et al., 2015; Clithero & Rangel, 2014) and many qualitative reviews (Glimcher, 2011a, 2014a; Wallis & Rushworth, 2014; Fehr & Krajbich, 2014; Platt & Plassmann 2014; Rangel & Clithero 2014; Lempert & Phelps, 2014; Jung et al., 2016; Padoa-Schioppa & Conen, 2017).

Some scholars (Glimcher, 2011a; Rolls, 2014) 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 & 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.

It should be note that a particular region, the orbitofrontal cortex (OFC), is playing a central role in the representing phase of decision SV, essentially in its medial part (often included in the ventromedial PFC by researchers), but this region also would be active in the constructing phase, if only for primary and secondary rewards (Padoa-Schioppa & Assad, 2006; Padoa-Schioppa & Conen, 2017; Plassman et al., 2007; Rolls, 2014; Rushworth et al., 2011).

5.4.2 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 are the neural structures taking part in this process of choice implementation, and what kind of choice mechanism is likely to be implemented?

The neural substrates involved in the 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) (see Glimcher, 2011a, 2014a).[75] Knowing if these results may be transposed from "actions" to "economic choices" (i.e., decisions 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 would correspond to a specific network including the dorsomedial PFC, the posterior parietal cortex (PPC), and the nearby intraparietal sulcus (Kable & Glimcher, 2009; Hare et al., 2011; Glimcher, 2011a, 2014a). It is argued that these areas implement a comparison process, and that the output of these so-called "comparator regions" modulates activity in motor cortex to implement the choice.

The choice implementation mechanism. Concerning the mechanism of choice implementation, two basic models of brain functioning are commonly developed in theoretical neuroscience: (1) the "winner-take-all" (WTA), and (2) the "drift–diffusion" (DD) models (e.g., Deco et al., 2013; Glimcher, 2011a; Rolls, 2014). The DD model belongs to the class of sequential sampling models which have been employed for over 50 years in cognitive psychology in domain such that perception and memory. It was introduced 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

---

[75] The most obvious evidence provides from a decision system with which neurophysiologists are familiar, the monkey visio-saccadic system, which for widely technic reasons was above all studied since the 1980s for understanding the sensorimotor control in general. 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 PFC), and the superior colliculus (in the midbrain). These findings were generalized later to body movements; it has been shown that the primary motor cortex, some anterior areas of the parietal cortex, and supplementary motor area, are playing an equivalent role.

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 (Palmer et al., 2005; Ratcliff et al., 2016; Roitman & Shadlen, 2002, for a survey).

We could 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 & Glimcher, 2009; Glimcher, 2014a; Wang, 2014). However, if the DD model is widely accepted today for perceptual decisions (Palmer et al., 2005; Gold & Shadlen, 2007; Glimcher, 2011a), the relevance of the WTA model for VBD is more controversial. 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 fulfill these two different functions (perceptual decisions and VBD) or whether there is one *integrate* decision-making model (Glimcher, 2014a).

Soltani & 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 not only be possible but perhaps necessary (see Wang, 2014, for a survey). It should be remarked that, if this finding was to be confirmed, then the specificity of neuroeconomics in relation to decision neuroscience would tend to lose a 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. We briefly mention here the attentional DD model (e.g., Fehr & Rangel, 2011; Padoa-Schioppa & Conen, 2017). 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 ($\beta >$ or = 1), and ε_t being 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) suggests a parallel between this notion of "threshold" in terms of 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 (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 et al., 2014) (see Fig. 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 (see O'Doherty et al., 2008; Fehr & Rangel, 2011).

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 et al., 2010, 2012, 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 & Rangel, 2011). However, scholars agree that additional research is necessary for precisely delimiting the field in which the DD model can be useful (Krajbich et al., 2014).

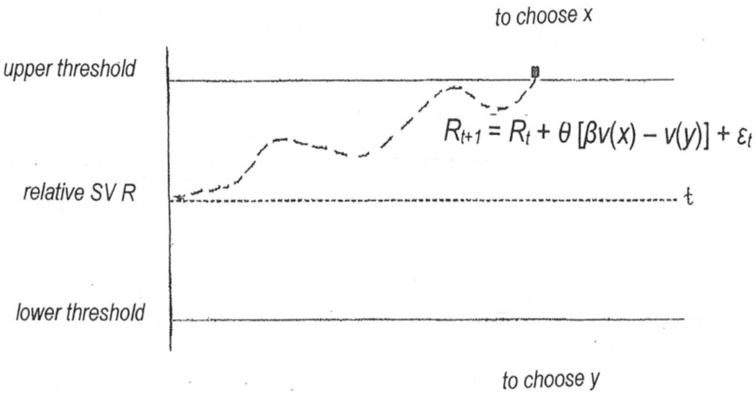


Fig. 6 The attentional drift–diffusion model (binary choice)

6 Conclusions

Neuroeconomics is a recent scientific field, but it has been very active for the past 2 decades. The annual output of neuroeconomic papers has roughly doubled since 2005, and there have been numerous surveys carried out in this new field at the crossroads of the economics, psychology, and neuroscience domains. The originality of this additional survey lies in its focus on neuroscientific foundations of neuroeconomics. Without the great progress made by various branches of modern neuroscience over the past few decades, neuroeconomics would not have been possible. This paper seeks to make these advances in modern neuroscience accessible to economists.

These neuroscientific advances have been reviewed by grouping them together into three non-independent topics referred to as the "emo-rational" brain, the "social" brain, and the "computational" brain. For each topic, we have emphasized findings that we consider critical to the birth and development of neuroeconomics while highlighting some of the key questions that need to be addressed through future research. We have also shown that the boundaries between neuroeconomics and several recent branches of cognitive neuroscience, such as affective, social, and, more generally, decision neurocience, are particularly porous.

In conclusion, we wish to broaden our thinking and discuss where neuroeconomics now stands with regard to its various goals, the usefulness of the available results to economists, and what can be expected from recent developments in the field. Without going so far as some enthusiastic economists, who perhaps a little hastily in its early days gave neuroeconomics a "revolutionary" potential (Camerer, Loewenstein & Prelec, 2005),[76] it must be recognized that many results have already been proven about how the human brain makes choices, and that these findings have provided new insights into the understanding of economic behavior in many domains.

It will be recalled that, in essence, two complementary goals can be identified for neuroeconomics. In one way, neuroeconomics tries simply to complement the behavioral economics approach by looking closely at the biological origin of behavior. It searches for links between biological markers and behavioral outcomes in human decision-making. However, more ambitiously, neuroeconomics investigates the neural mechanisms underlying behavior and seeks to uncover the computations being carried out in the brain. It attempts to identify what particular brain regions and neural networks do and how they interact to produce behavior.

To what extent have these objectives been achieved? It is clear that the first objective has been largely fulfilled over the past two decades. Neuroeconomic experiments have confirmed in biological terms the explanations proposed by behavioral economics concerning mental states and processes in various contexts. For instance, several central hypotheses posited by the prospect theory, including loss aversion and non-linearity of the probability-weighting function, have been

---

[76] It is only fair to recognize that the declared ambition of the researchers in the "behavioral economics in the scanner" program was quickly limited to "simply improving the understanding of the decision-making process" (see in particular the review by Sanfey et al., 2006, only a year after the survey of Camerer, Loewenstein, & Prelec 2005).

confirmed. Likewise, by revealing the neural bases of specific choice elements in areas as varied as desire, regret, anticipation, risk, ambiguity, or time-attitude, the works carried out within the framework of the behavioral-in-the-scanner program have contributed to improving the understanding of the motivations specific to individual decisions. Additionally, many psychological explanations put forward in behavioral economics of the so-called "irrationalities" observed in contexts of social interactions were also confirmed by neuroeconomic experiments. In particular, social neuroeconomics has helped to define the emotional content of "other-regarding" preferences as an explanatory factor. In so doing, it has clarified the neural bases of the motivations that behavioral economics generally attributes to people to explain their observed prosocial behavior, including trust, altruism, reciprocity, empathy, generosity, or concern for equity. Importantly, these neuroeconomic experiments do not just corroborate these psychological hypotheses; they provide them with a stronger *objective* justification. It must be acknowledged that these achievements are far from negligible, especially because, owing to the progress in recent years on statistical problems encountered in brain imaging (in particular, the use of multivariate machine learning techniques) and to the increasing use of brain stimulation techniques and pharmacological manipulations in experiments today, neuroeconomics studies have shifted from correlational issues to predictions accordingly. Thus, the results acquire stronger validity by proving genuine causality. Moreover, many studies are beginning to shed new light on interesting issues for economists in areas such as willingness to pay, risk aversion, impulsivity, selfishness, reputation building, guilt, honesty, and so on.

However, it is through studies that aim to discover the precise mechanisms linking biological markers to behavior, that neuroeconomics has the greatest potential to improve the economic theory and perhaps even transform economics completely in the future. Unfortunately, in the early years, the neural economics program, which pursued this broad goal, did not appear to be very useful for economists, because it was too brain-centric. It should be recognized that, initially, neuroeconomics incorporated ideas, concepts, and methods from the economic theory without genuine reciprocation. The evolutionist framework has elucidated major properties of brain functioning, in particular by providing answers to the question of how "values" (i.e., what guides preferences and choices) are learnt and represented in the brain. These results provide new insights into familiar economic questions, such as time preference and risky choices. We began to understand how to articulate neural-learning algorithms and economic learning models. However, the direct application of these findings to questions of interest in economics remains limited.

In contrast, we hope that the recent direction taken by neuroeconomics via the neural-and-behavioral program should validate the ability of neuroeconomics to improve economic models and address questions of interest to economics. By integrating into economic models, the biological constraints imposed by the way the brain works, as we are beginning to understand it (in particular, established properties such as unicity, stochasticity, and sequentiality in the brain's choice mechanism, the limited capacity of the brain to process information, and the role of non-conventional observable variables in the choice process, such as the response time or the visual attention of individuals), neuroeconomics might improve the predictive capacity of models. It is already apparent that the outline of the "standard" neuroeconomic choice model, as we are able to draft it, already shows the limits of mainstream economic models. In particular, the expected utility model is only relevant as a choice model in specific situations. Sequential sampling models of decision-making, such as the drift–diffusion model (the most popular one), where preferences are "revealed" from choices and response times, may be applied broadly in many economic areas.

Finally, even though the strongly brain-centered nature of many neuroeconomic works has undoubtedly hampered the influence of neuroeconomics on classical themes of microeconomics, developments in this field over the last decade are likely to strengthen its impact on economic decision-making in the future. However, if neuroeconomics seeks to exert a deeper influence on economics, it is probably necessary to extend its scope and, as with behavioral economics recently, address policy-oriented concerns, because economics is ultimately a discipline aimed at designing and developing policies.

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

Declarations

Conflict of interests The author declares that he has no conflict of interest.

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

Ainslie, G. W. (1975). Specious reward: A behavioral theory of impulsiveness and impulsive control. *Psychological Bulletin, 82*(4), 463–496.

Ainslie, G. W. (1991). Derivation of 'rational' economic behavior from hyperbolic discount curves. *American Economic Review, 81*(2), 334–340.

Ainslie, G. W. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person*. Cambridge University Press.

Ainslie, G. W. (2012). Pure hyperbolic discount curves predict 'eyes open' self-control. *Theory and Decision, 73*, 3–34.

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

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

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

Balleine, B.W, Daw, N.D., & O'Doherty, J.P. (2009). Multiple forms of value learning and the function of dopamine. In P.W. Glimcher, C.F., Camerer, E. Fe hr, & R.A. Polack (Eds.), *Macroeconomic. Decision making and the brain* (pp. 367–388). Amsterdam: Elsevier.

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

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

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

Basten, U., Bille, G., Heekeren, H. R., & Fiebach, C. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Science USA, 107*, 21767–21772.

Baumgartner, T., Gotte, L., Gugler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping, 33*(6), 1452–1469.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*, 639–650.

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

Baumgartner, T., Schiller, B., Rieskamp, J., Gianotti, L. R. R., & Knoch, D. (2014). Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. *Social Cognitive and Affective Neuroscience, 9*, 653–660.

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

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

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

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*, 1293–1295.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond automatically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex, 6*, 215–225.

Bell, D. (1982). Regret in decision making under uncertainty. *Operations Research, 30*, 961–981.

Belluci, G., Camilleri, J. A., Iyengar, V., & Gruger, F. (2020). The emerging neuroscience of social punishment: Meta-analytic evidence. *Neuroscience & Behavioral Reviews, 113*, 426–439.

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

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

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain responses to reward. *Journal of Neuroscience, 21*, 2793–2789.

Berridge, K.C. & O'Doherty, J.P. (2014). From experienced utility to decision utility. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 335–351). Amsterdam: Elsevier.

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

Blair, R. J. R. (2005). Responding to the emotions of others: 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 Journal of Experimental Psychology, 61*(1), 157–170.

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

Bossaerts; P., Preuschoff, K., & Hsu, M. (2009). The neurobiological foundations of valuation in human decision making under uncertainty. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 353–366). Amsterdam: Elsevier.

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

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

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

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (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. (1994). *Brodmann's "Localisation in the cerebral cortex"*. London: Smith-Gordon.

Brown, (1951). Iterative solution of games by fictitious play. In T.C. Koopmans (ed.), *Activity analysis of production and allocation*. New York: John Wiley

Bush, R., & Mosteller, R. (1955). *Stochastic models of learning*. John Wiley.

Bzdok, D., Schibach, L., Vognely, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (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., Visser, P. S., & Pickett, C. L. (Eds.). (2006). *Social neuroscience: People thinking about thinking people*. MIT Press.

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

Camerer, C. F. (1997). Progress in behavioral game theory. *Journal of Economic Perspectives, 11*, 167–188.

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 A. Caplin & A. Schotter (Eds.). *Foundation of positive and normative economics*: *A handbook*, (pp. 43–69), New York: Oxford University Press, 2008.

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

Camerer, C. F. (2013). Goals, methods, and progress in neuroeconomics. *Annual Review of Economics, 5*(1), 425–455.

Camerer, C.F. & Hare. T.A. (2014). The neural basis of strategic choice. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 479–511). Amsterdam: Elsevier.

Camerer, C. F., & Ho, T.-H. (1999). Experienced-weight attraction learning in normal form games. *Econometrica, 67*, 827–874.

Camerer, C. F., & Ho, T.-H. (2004a). A cognitive hierarchy model of games. *The Quarterly Journal of Economics, 119*(3), 961–898.

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

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

Camerer, C. F., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty, 5*(4), 325–370.

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

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

Caplin, A. & Dean, M. (2009). Axiomatic neuroeconomics. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 21–32). Amsterdam: Elsevier.

Caplin, A., Dean, M., Glimcher, P. W., & Rutledge, R. (2010). Testing the reward prediction error hypothesis with an axiomatic model. *Journal of Neuroscience, 30*, 13525–13536.

Carr, L., Iacoboni, M., Dubeau, M.-C., 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., Bowling, D. L., Reech, C., & Huettel, S. A. (2012). A distinct role of the temporo-parietal junction in predicting socially guided decisions. *Science, 337*, 109–111.

Cartwright, E. (2016). *Behavioral economics* (3rd ed.). Routledge.

Chase, H. W., Kumar, P., Eickhoff, S. R., & Dombrovski, A. Y. (2015). Reinforcement learning model and their neural correlates: An activation likelihood estimation meta-analysis. *Cognitive, Affective, and Behavioral Science, 15*, 435–459.

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

Cheung, Y. W., & Friedman, D. (1997). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior, 19*, 46–76.

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (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., & Gomila, A. (2012). Moral dilemma in cognitive neuroscience of moral decision-making: A principle review. *Neuroscience and Biobehavioral Reviews, 36*(4), 1249–1264.

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

Clark, J. J., Hollon, N. G., & Phillips, P. E. M. (2012). Pavlovian valuation systems in learning and decision making. *Current Opinion in Neurobiology, 22*, 1054–1061.

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multi-voxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex, 27*, 1844–1948.

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience, 9*(9), 1289–1302.

Clithero, J. A., Tankersley, D., & Huettel, S. A. (2008). Foundation of neuroeconomics: From philosophy to practice. *PLOS Biology, 6*(11), e298.

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

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

Commons, M. L. (2001). A short history of the Society for the quantitative analysis of behaviour. *Behavior Analyst Today, 2*(3), 275–279.

Conn P.M. (Ed.). (2016). *Conn's translational neuroscience*, Elsevier.

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

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

Coricelli, G., Dolan, R. J., & Sirigu, A. (2007). Brain, emotion and decision making: The paradigmatic example of regret. *Trends in Cognitive Science, 11*(6), 258–265.

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

Cournot, A.A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses*. Paris: Librairie des sciences politiques et sociales (English: N. Bacon (ed). *Researchs into the mathematical principles of the theory of wealth*. London: Macmillan, 1987).

Cowen, A. S., & Keltner, D. (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., Uddin, L. Q., Castellanos, F. X., Milham, M. P., & Kelly, C. (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.

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

Crockett, M.J. & Fehr, E. (2014). Pharmacology of economic and social decision making. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 259–279). Amsterdam: Elsevier.

Dalgleich, T. (2004). The emotional brain. *Nature Reviews Neuroscience, 5*, 583–589.

Dalley, J. W., Everitt, B. J., & Robbins, T. W. (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*. A. Grosset/Putnam Books.

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*. Harcourt Inc.

Damasio, A.R. (2017). *The stranger order of things. Life, feeling, and the making of cultures*. New York: Pantheon Books.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science, 319*, 1264–1267.

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 P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 299–320). Amsterdam: Elsevier.

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

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

Daw, N.D. & O'Doherty, J.P (2014). Multiple systems for value learning. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 393–410). Amsterdam: Elsevier.

Daw, N.D. & Tobler, P.N. (2014). Value learning through reinforcement: The basics of dopamine and reinforcement learning. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 283–298). Amsterdam: Elsevier.

Dayan, P. (2008). The role of value systems in decision making. In C. Engel & W. Singer (Eds.), *Better than conscious? Implications for performance and institutional analysis* (pp. 51–70). MIT Press.

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

Dean, M. (2013). What can neuroeconomics tell us about economics (and vice versa). In P. H. Crowley & T. R. Zentall (Eds.), *Comparative decision making* (pp. 163–203). Oxford University Press.

Decety, J., & Cacioppo, J. T. (Eds.). (2011). *The Oxford handbook of social neuroscience*. Oxford University Press.

Decety, J. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences, 18*(7), 337–339.

Decety, J., & Grèzes, J. (2006). The power of simulation: Imaging one's own and other's behaviour. *Brain Research, 1079*(1), 4–14.

Declerck, C. & Boone, C. (2016). The neuroanatomy of prosocial decision making. In C. Declerck & C. Boone (Eds.), *Neuroeconomics of prosocial behaviour. The compassionate egoist* (Chap. 2). London: Elsevier.

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

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

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

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron, 56*(2), 384–398.

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

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

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

de Quervain, D. J., Fishbacher, U., Treyer, V., Schellhaller, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305*, 1254–1258.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*, 176–180.

Dolan, R. J., & Dayan, P. (2013). Goals and habits. *Neuron, 80*(2), 312–325.

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

Doya, K. & Kimura, M. (2014). The basal ganglia, reinforcement learning, and the encoding of value. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 321–334). Amsterdam: Elsevier.

Dreher, J.-C., & Tremblay, L. (2017). *Decision neuroscience*. Academic Press.

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

Efron, B., & Hastie, T. (2016). *Computer-age statistical inference*. Cambridge University Press.

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

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

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics, 75*, 643–669.

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

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

Engelmann, J.B. & Hare, T.A. (2018). Emotions can bias decision-making processes by promoting specific behavior. In A.S. Fox, R.C. Lapate, A.J. Shackman, & R.J. Davidson (Eds.), *The nature of emotion. Fundamental questions* (2nd ed.) (pp. 355–359). Oxford University Press.

Engelmann; J.B. & Fehr, E. . (2017). The neurobiology of trust and social decision: The important role of emotions. In P. A. M. Van Lange, B. Rockenbach, & T. Yamagishi (Eds.), *Trust in social dilemmas* (pp. 33–56). Oxford University Press.

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

Evans, J. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.

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

Farrer, C., & Frith, C. D. (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.

Fecteau, S., Pascual-Leone, A., Zald, D. H., Liguori, P., Théoret, H., Goggio, P. S., & Fregni, F. (2007). Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making. *The Journal of Neuroscience, 27*(23), 6212–6218.

Fehr, E. (2009). Social preferences and the brain. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.), *Neuroeconomics. Decision making and the brain* (pp. 215–232). Amsterdam: Elsevier.

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

Fehr, E. & Krajbich, I. (2014). Social preferences and the brain. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain*. (2nd ed.) (pp. 193–218). Amsterdam: Elsevier.

Fehr, E., & Rangel, A. (2011). Neuroeconomic foundations of economic choices – Recent advances. *Journal of Economic Perspectives, 25*(4), 3–30.

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

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. & Russel, J.A. (Eds.) (2017). *The science of facial expression*. Oxford University Press.

Figner, B., Knoch, D., Jonhson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., & Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience, 13*(5), 538–539.

Fodor, J. A. (1983). *The modularity of mind*. MIT Press.

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

Fox, C.R. & Poldrack, R.A. (2014). Prospect theory and the brain. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain*. (2nd ed.) (pp. 533–567). Amsterdam: Elsevier.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*(June), 351–401.

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., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society b: Biological Sciences, 358*(1431), 459–473.

Frith, U., & Frith, C. D. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531–534.

Frith, U., & Frith, C. D. (2007). Social cognition in humans. *Current Biology, 21*, R724–R732.

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

Fudenberg, D., & Levine, D. K. (1998). *Theory of learning in games*. MIT Press.

Fudenberg, D., & Levine, D. K. (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., & Bressler, S. L. (2012). Cognitive activation: A mechanism enabling temporal integration in working memory. *Trends in Cognitive Sciences, 16*(4), 207–218.

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

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

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593–609.

Gazzaniga, M. S., & Mangun, G. R. (Eds.). (2014). *The cognitive neurosciences* (5th ed.). MIT Press.

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

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

Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., Paul, L. K., & Tranel, D. (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.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*(4), 585–595.

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

Glimcher, P. W. (2011a). *Foundations of neuroeconomic analysis*. 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 P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 373–391). Amsterdam: Elsevier.

Glimcher, P.W. (2014b). Introduction to neuroscience. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain*. (2nd ed.) (pp. 63–76), Amsterdam: Elsevier.

Glimcher, P.W. (2014c). Understanding the hows and whys of decision-making: From expected utility to divisive normalization. *Cold Springer Harbor Symposia on Quantitative Biology*, *LXXIX*, 169–176.

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

Glimcher, P.W., Camerer, C.F., Fehr, E., & Poldrack, R.A. (Eds.) (2009). Introduction. A brief history of neuroeconomics. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 1–12). Amsterdam: Elsevier.

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

Glimcher, P. W., & Fehr, E. (Eds.). (2014a). *Neuroeconomics. Decision making and the brain* (2nd ed.). Elsevier.

Glimcher, P.W. & Fehr, E. (2014b). Introduction: A brief history of neuroeconomics. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. xvii-xxviii). Amsterdam: Elsevier.

Glimcher, P. W., & Rustichini, A. (2004). Neuroeconomics: The consilience of brain and decision. *Science, 206*, 447–452.

Gold, J.I. & Heekeren, H.R. (2014). Neural mechanism for perceptual decision making. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 355–372). Amsterdam: Elsevier.

Gold, J. I., & Shadlen, M. N. (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 mind-reading*. Oxford University Press.

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

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

Gul, F., & Pesendorfer, W. (2006). Random expected utility. *Econometrica, 74*, 121–146.

Gul, F. & Pesendorf, W. (2008). The case of mindless economics. In A. Caplin A. & A. Schotter (Eds.). *The foundation of positive and normative economics: A handbook* (pp. 3–39). New York: Oxford University Press.

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

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

Hare, T. A., O'Doherty, J. P., Camerer, C. F., Schultz, W., & Rangel, A. (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., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (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., & Ross, D. (2010). The methodology of neuroeconomics. *Journal of Economic Methodology, 17*(2), 185–196.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience, 37*, 435–456.

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., & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives, 5*(3), 137–156.

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

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

Houser, D. & McCabe, K. (2009). Experimental neuroeconomics and non-cooperative games. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 47–62). Amsterdam: Elsevier.

Houser, D. & McCabe, K. (2014). Experimental economics and experimental game theory. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 19–34). Amsterdam: Elsevier.

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

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science, 310*, 1680–1683.

Hsu, M., & Zhu, L. (2012). Learning in games: Neural computations underlying strategic learning. *Recherches Économiques De Louvain, 78*(3), 47–72.

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

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

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

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

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

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

Joffily, M., Masclet, D., Noussair, C., & Villeval, M. C. (2014). Emotions, sanctions, and cooperation. *Southern Economic Journal, 80*, 1002–1027.

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

Kable, J.W. (2014). Valuation, intertemporal choice, and self-control. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 173–192). Amsterdam: Elsevier.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience, 10*(12), 1625–1633.

Kable, J. W., & Glimcher, P. W. (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*. FSG.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An anakysis of decision under risk. *Econometrica*, *47*, 263–291.

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

Kanayet, F. J., Opfer, J. E., & Cunningham, W. A. (2014). The value of numbers in economic rewards. *Psychological Science, 25*, 1534–1545.

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

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

Kenning, P., & Plassmann, H. (2005). Neuroeconomics: An overview from an economic perspective. *Brain Research Bulletin, 67*, 343–354.

Keynes, J. M. (1936). *The general theory of interest, employment, and money*. Macmillan.

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

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

Knoch, D., & Fehr, E. (2007). Resisting the power of temptations: The right prefrontal cortex and self-control. *Annual New York Academy of Science, 1104*, 123–134.

Knoch, D., Gianotti, L. R., Pascual-Leone, A., Treyer, V., Regard, M., Hoffman, M., & Brugger, P. (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., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (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., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006b). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*, 829–832.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., & Fehr, E. (2009). Disrupting the prefrontal cortex diminishing the human ability to build a good reputation. *Proceedings of the National Academy of Science of the USA, 106*, 20895–20899.

Knutson, B.K., Delgado, M.R., & Phillips, P.E.M. (2009). Representation of subjective value in the striatum. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 389–406). Amsterdam: Elsevier.

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

Knutson, B. K., Fong, G. W., Bennett, S. M., Adams, C. M., & Homme, D. (2003). A region of mesial prefrontal cortex tracks monetary rewarding outcomes: Characterisation with rapid event-related fMRI. *NeuroImage, 18*, 263–272.

Knutson, B. K., Ricks, S., Wimmer, G. S., Prelec, D., & Loewenstein, G. (2007). Neural predictors of purchases. *Neuron, 53*, 147–156.

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

Konovalov, A., & Krajbich, I. (2019). Over a decade of neuroeconomics: What have we learned? *Organizational Research Methods, 22*(1), 148–173.

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

Kosfeld, M., Heinrichs, M., Zak, P., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*, 673–676.

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

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience, 29*, 2188–2192.

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

Krajbich, I., & Dean, M. (2015). How can neuroscience inform economics? *Current Opinion in Behavioral Sciences, 5*, 51–57.

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

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

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

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

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

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron, 47*, 763–770.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics, 112*(2), 443–477.

Landreh, A., & Bickle, J. (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*. Simon and Schuster.

Lempert, K.M. & Phelps, E.A. (2014). Neuroeconomics of emotion and decision making. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 219–236). Amsterdam: Elsevier.

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

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

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

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

Lieberman, M. D. (2012). A geographical history of social cognitive neuroscience. *NeuroImage, 61*, 432–436.

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

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

Lis, S., & Kirsch, P. (2016). Neuroecnomic approaches in mental disorders. In M. Reuter & C. Montag (Eds.), *Neuroeconomics* (pp. 333–344). Springer-Verlag.

Lombardo, M. V., Chakravarti, B., Bullmore, E. T., Wheelwright, S. J., Sadeh, S. A., Suckling, J., MRC AIMS Consortium, & Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience, 22*(7), 1623–1639.

Loewenstein, G., & Elster, J. (1992). *Choice over time*. Russel Sage.

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

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics, 107*(2), 573–597.

Loewenstein, G., Rick, S., & Cohen, J. D. (2008). Neuroeconomics. *Annual Review of Psychology., 59*, 647–672.

Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences, 104*(22), 9493–9498.

Loomes, G., & Sudgen, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92*, 805–824.

Louie, K.L. & De Martino, B. (2014). The neurobiology of context-dependent valuation and choice. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 455–476). Amsterdam: Elsevier.

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

MacDonald, K., & MacDonald, T. M. (2010). The peptide that binds: A systematic review of oxytocin and its prosocial effects in humans. *Harvard Review of Psychiatry, 18*(1), 1–21.

Mackey, S., & Petrides, M. (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*. Plenum Press.

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

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

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

Mäki, U. (2010). When economics meet neuroscience: Hype or hope. *Journal of Economic Methodology, 17*, 107–117.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*, 77–91.

Martin, D., Mehtra, M. A., & Prata, D. (2017). The «highs and lows» of the human brain on dopaminergics: Evidence from neuropharmacology. *Neuroscience and Biobehavioral Reviews, 80*, 351–371.

Mayr, U., Harbaugh, W.T., & Tankersley, D. (2009). Neuroeconomics of charitable giving and philanthropy. In P.W. Glimcher, C.F. Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 303–320). Amsterdam: Elsevier.

McCabe, K. A. (2003). Neuroeconomics. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol. 3, pp. 294–298). Publishing Group, Macmillan Publishers Ltd.

McCabe, K. A. (2008). Neuroeconomics and the economic science. *Economics and Philosophy, 24*(3), 345–368.

McCabe, K. A., Houser, R., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Science USA, 98*, 11832–11835.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron, 38*, 339–346.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience, 27*(21), 5796–5804.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004a). Separate neural systems value immediate and delayed monetary rewards. *Science, 306*, 503–307.

McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004b). Neural correlates of behavioral preferences for culturally familiar drinks. *Neuron, 44*, 379–387.

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

McCullogh, W. S., & Pitts, W. (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 P. Zarembka (Ed.), *Frontier in econometrics* (pp. 105–142). Academic Press.

McFadden, D. (2005). Revealed stochastic preference: A synthesis. *Economic Theory, 26*, 245–264.

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

Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., De Timary, Ph., & Luminet, O. (2010). Oxytocin makes people trusting, not gullible. *Psychological Science, 21*, 1072–1075.

Millan, M. J. (2013). An epigenetic framework for neurodevelopmental disorders: From pathogenesis to potential therapy. *Neuropharmacology, 68*, 2–82.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Review Neuroscience, 1*, 59–65.

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

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

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

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

Mongin, P. & d'Aspremont, C. (1998). Utility theory and ethics. In S. Barbera, P. Hammond, & C. Seidel (Eds.), *Handbook of utility theory, vol. 1: Principles* (chap. 10), Dordrecht: Kluwer Academic Pub.

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

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

Montague, P. R., Dayan, P., Person, C., & Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive Hebbian learning. *Nature, 377*, 725–728.

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

Moulin, H. (1986). Game theory for social sciences. 2nd rev. ed., New York: New York University Press.

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

Nagel, R.-M. (1995). Unraveling in guessing games: An experimentally study. *American Economic Review, 85*, 1313–1326.

Nave, G., Camerer, C. F., & McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspectives on Psychological Science, 10*, 772–789.

Niv, Y. & Montague, P.R. (2009). Theoretical and empirical studies of learning. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 331–351). Amsterdam: Elsevier.

Normann, K. A., Polyn, S. M., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Science, 10*(9), 424–430.

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual Review of Psychology, 68*, 73–100.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38*, 329–337.

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

Okon-Singer, H., Hendler, D., Pessoa, L., & Shackman, A. J. (2015). The neurobiology of emotion-cognition interactions: Fundamental questions and strategies for future research. *Frontiers in Human Neuroscience, 9*, 1–14.

Okon-Singer, H., Stout, D.M., Stockbridge, M.D., Gamer, M.D., Fox, M., & Shackman, A.J. (2018). The interplay of emotion and cognition. In A. Fox, R. Lapate, A. Shackman, & R.J. Davidson (Eds.). *The nature of emotion. Fundamental questions* (2nd ed.) (181–185). Oxford: Oxford University Press.

Oliver, L. D., Vieira, J. B., Neufeld, R. W. J., Dziobek, I., & Mitchell, D. G. V. (2018). Greater involvement of action simulation mechanisms in emotional vs cognitive empathy. *Social Cognitive and Affective Neuroscience, 13*(4), 367–380.

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

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

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

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., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441*, 223–226.

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

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

Palmer, J., Huk, A. C., & Shadlen, M. N. (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*. Oxford University Press.

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

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

Pavlov, I.P. (1927/1960). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. New York: Dover Publications (unaltered republication of the 1927 translation in 1960 by Oxford University Press).

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

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

Pessoa, L. (2010). Emotion and cognition and the amygdala: 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. (2016). The emotional brain. In P.M. Conn (Ed.), *Conn's translational neuroscience* (pp. 635–656). Elsevier.

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

Phelps, E.A. (2009). The study of emotions in neuroeconomics. In P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 233–250). Amsterdam: Elsevier.

Phelps, E. A., & LeDoux, J. E. (2005). Contribution of the amygdala to emotion processing: From animal models to human behavior. *Neuron, 48*, 175–187.

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

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

Platt, M. L., & Huettel, S. A. (2008). Risky business: The neuroeconomics of decision making under uncertainty. *Nature Neuroscience, 11*, 398–303.

Platt, M.L. & Plassman, H. (2014). Multistage valuation signals and common neural currencies. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 237–258). Amsterdam: Elsevier.

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. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron, 72*(5), 692–697.

Poldrack, R. A. (2018). *The new mind reader*. Princeton University Press.

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

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

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a ‘theory of mind’? *Behavioral and Brain Sciences, 4*, 515–526.

Preuschoff, K., Bossaerts, P., & Quartz, S. R. (2006). Neural differentiation of expected reward and risk in human subcortical structures. *Neuron, 51*, 381–390.

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience, 28*, 2745–2752.

Pribram, K. H. (1973). The primate frontal cortex – Executive of the brain. In K. H. Pribam & A. R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 293–314). Academic Press.

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

Quartz, S. R. (2008). From cognitive science to cognitive neuroscience to neuroeconomics. *Economics & Philosophy, 24*, 459–471.

Ramnani, N., & Owen, A. M. (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 P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.). *Neuroeconomics. Decision making and the brain* (pp. 425–440). Amsterdam: Elsevier.

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

Rangel, A & Clithero, J.A. (2014). The computation of stimulus values in simple choice. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 125–148). Amsterdam: Elsevier.

Rangel, A., & Hare, T. (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.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Science, 20*(4), 260–281.

Reiman, M., & Bechara, A. (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.

Reuter, M., & Montag, C. (Eds.). (2016). *Neuroeconomics*. Springer-Verlag.

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

Riedl, R., & Javor, A. (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., Gutman, D., Zeh, T., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*, 395–405.

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

Rilling, J. K., & Sanfey, A. G. (2011). The neurosciences of social decision-making. *Annual Review of Psychology, 62*, 43–48.

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

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

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996a). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

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

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

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

Rizzolatti, G., & Sinigaglia, C. (2006). *So quel che fai. Il cervelloche agisce e i neuroni specchio.* Raffaello Cortina Editore.

Roemer, J. E. (1996). *Theories of distributive justice*. Harvard University Press.

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

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

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

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

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

Ross, D. (2010). Neuroeconomics and economic methodology. In J. B. Davis & D. W. Hands (Eds.), *Handbook of economic methodology* (pp. 61–93). Edward Elgar.

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior, 8*(1), 164–212.

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

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

Ruff, C.C. & Huettel, S.A. (2014). Experimental methods in cognitive neuroscience. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 77–108). Amsterdam: Elsevier.

Ruff, C. C., Ugazio, U., & Fehr, E. (2013). Changing social norm compliance with non-invasive brain stimulation. *Science, 342*, 482–484.

Rushworth, M. F. S., Mars, R. B., & Sallet, J. (2013). Are there specialized circuits for social cognition and are they unique to humans? *Current Opinion in Neurobiology, 23*, 436–442.

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

Rustichini, A. (2005). Neuroeconomics: Present and future. *Games and Economic Behavior, 52*, 201–212.

Rustichini, A. (2009). Neuroeconomics: What have we found, and what should we search for. *Current Opinion in Neurobiology, 19*, 672–677.

Rustichini, A., Dickhaut, J., Ghirardato, P., Smith, K., & Pardo, J. V. (2005). A brain imaging study of the choice procedure. *Games and Economic Behavior, 52*, 257–282.

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

Saarimäki, H., Ejtehadian, L. F., Glerean, E., Jääkeläinen, L., Vuilleumier, P., Sams, M., & Nummenmea, L. (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.* Picador.

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

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

Sanfey, A. G., & Rilling, J. K. (2011). Neural bases of social decision making. In O. Vartorian & D. R. Mandel (Eds.), *Neuroscience of decision making* (pp. 223–242). Psychology Press.

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

Santos, L.R. & Platt, M.L. (2014). Evolutionary anthropological insights into neuroeconomics: What non-human primates can tell us about human decision-making strategies. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 109–122). Amsterdam: Elsevier.

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

Schmidt, C. (2008). What neuroeconomics does really mean? *Revue D'economie Politique, 118*(1), 7–34.

Schmidt, C., & Livet, P. (2014). *Comprendre nos interactions sociales. Une perspective neuroéconomique*. Paris: Odile jacob.

Schnell, K., Bluschke, S., Konradt, B., & Walter, H. (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 P.W. Glimcher, C.F., Camerer, E. Fehr, & R.A. Poldrack (Eds.), *Neuroeconomics. Decision making and the brain* (pp. 323–329), Amsterdam: Elsevier.

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., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.

Schultz, W., Preuschoff, K., Camerer, C. F., Hsu, M., Fiorillo, C. D., Tobler, N. P., & Bossaerts, P. (2008). Explicit neural signals reflecting reward uncertainty. *Philosophical Transactions of the Royal Society b: Biological Sciences, 363*, 3801–3811.

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

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

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

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

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

Serra, D. (2021). Controversies around neuroeconomics: Empirical, methodological and philosophical issues, *Review of Economic Philosophy/ Revue de philosophie économique* (to be published).

Serra, D. (2022). *La "révolution" expérimentale en économie. Une histoire des courants de recherche qui l'incarnent. Presses Universitaires de la Méditerranée* (forthcoming).

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

Shamay-Tsoory, S.G. (2011b). Empathic processing: Its cognitive and affective dimensions and neuroanatomical basis. In J. Decety J. & W. Ickes (Eds.). *The social neuroscience of empathy* (chap. 16), MIT Press.

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. *The Quarterly Journal of Economics, 69*(1), 99–118.

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

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

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron, 41*, 653–662.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*, 466–469.

Singer, T. & Tusche, A. (2014). Understanding others: Brain mechanisms of theory of mind and empathy. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 513–532), Amsterdam: Elsevier.

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

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–398). Cambridge University Press.

Smith, A., Bernheim, B. D., Camerer, C. F., & Rangel, A. (2014). Neural activity reveals preferences without choices. *American Economic Journal: Microeconomics, 6*(2), 1–36.

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

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

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

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

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 39*, 332–338.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system – An approach to cerebral imaging*. Thieme Medical Publishers.

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. (1985). Mental accounting and consumer choice. *Marketing Science, 4*, 199–214.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12*, 183–206.

Thiele, A., & Bellgrove, M. A. (2018). Neuromodulation of attention. *Neuron, 97*, 769–785.

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

Tobler, P.N. & Weber, E.U. (2014). Valuation for risky and uncertain choices. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 149–172), Amsterdam: Elsevier.

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

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

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

Tversky, A., & Kahneman, D. (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., & Baetens, K. (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., & Izuma, K. (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. & Rushworth, M.E. (2014). Integrating benefits and costs in decision making. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 411–433). Amsterdam: Elsevier.

Wang, X-J. (2014). Neuronal circuit computation of choice. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 435–454). Amsterdam: Elsevier.

Watson, J. (1913). Psychology as the behavioural views it. *Psychological Review, 20*(2), 158–177.

Weber, E.H. (1834/1996). *E.H. Weber: On the tactile senses* (with translation of De Tactu), H.E. Ross & D.J. Murray (trans. and eds.), New York: Experimental Psychology Society.

Whalen, P. J., & Phelps, E. A. (Eds.). (2009). *The human amygdala*. Guilford.

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

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

Zaghloul, K. A., Blanco, J. A., Weidemann, C. T., McGill, K., Jaggi, J. L., Baltuch, G. H., & Kahana, M. J. (2009). Human substantia nigra neurons encode unexpected financial rewards. *Science, 323*, 1496–1499.

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

Zak, P. J. (2011). The physiology of moral sentiments. *Journal of Economic Behavior and Organization, 77*, 53–65.

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

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences., 109*(5), 1419–1424.

Zweig, J. (2007). *Your money and your brain*. Simon & Schulster.