Neuroeconomics: reliable, scientifically legitimate and useful knowledge for economists?

Daniel Serra

To cite this version:

Daniel Serra. Neuroeconomics: reliable, scientifically legitimate and useful knowledge for economists?. 2020. hal-02956441

HAL Id: hal-02956441
https://hal.inrae.fr/hal-02956441v1

Preprint submitted on 2 Oct 2020

Neuroeconomics:
reliable, scientifically legitimate and useful knowledge for economists?

Daniel Serra

CEE-M Working Paper 2020-13

NEUROECONOMICS: RELIABLE, SCIENTIFICALLY LEGITIMATE AND USEFUL KNOWLEDGE FOR ECONOMISTS?

Daniel Serra* °

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

September, 20, 2020

Abstract

Thanks to the joint collaboration of economics, psychology, and neuroscience from the late 1990s, "neuroeconomics" sheds new light on decision-making analysis. As with any emerging discipline, however, neuroeconomics raises many practical and methodological questions resulting in debates and controversies that this article discusses by addressing three major issues concerning the contribution made so far to knowledge: Is it reliable? Is it scientifically legitimate? Is it useful for the economist? Without claiming to be exhaustive, this analytical framework enables understanding of the thrust of the major criticisms of neuroeconomics and at the same time the nature of the likely responses.

Key words: *Decision-making processes; Neuroscience methods; Brain data; Design of experiments; Economic methodology; Philosophy of science; Computational models; Quantitative research*

JEL classification: *D87, B41, C7; C9*

---

* Corresponding author: Daniel Serra, CEE-M, *CEE-M, Univ. Montpellier, CNRS, INRAE, Montpellier SupAgro, Montpellier, France, Faculté d'économie, Site Richter, Avenue de la Mer-Raymond Dugrand, 34960, Montpellier, CS : 79606 ; e-mail : daniel.serra@umontpellier.fr.*

° The author thanks Sacha Bourgeois-Gironde, Philippe Mongin, Pierre Livet, Nicolas Vallois and Thierry Blayac for very useful comments or discussions on topics covered in the paper.

1. INTRODUCTION

The brain has always fascinated mankind. “Neuroeconomics” first emerged in the United States in the late 1990s and is the most recent scientific discipline to place this mysterious organ at the heart of its investigations. Neuroeconomics, at the junction between economics, psychology and neurosciences, is a natural extension of bioeconomics which uses evolutionary biology to build models that predict human behavior. Undoubtedly, the development of experimental and behavioral economics in the 1980s aided the initial neuroeconomic studies by providing a set of well-codified experimental designs and useful economic paradigms to address problems in economic decision-making whether in individual or interactive situations. But it is acknowledged that this emerging discipline relies heavily on the revolution in cognitive neurosciences that occurred some ten years earlier and on the availability of a wide range of sophisticated investigative tools for its experiments, especially functional magnetic resonance imaging (fMRI), a non-invasive technique used to measure the metabolic correlates of neural activity.

Whilst neuroeconomics has generated great enthusiasm in the scientific community, it remains controversial with economists. Neuroeconomics generates both fascination and skepticism, especially as debates between experts are intense, if not bitter. This article discusses these debates and controversies by organizing them around three major issues. One set of criticisms focuses on the reliability of the results. This concerns practical issues such that the scarcity of the data on which the results depend, the reliability of the tools used, the difficulties encountered in interpreting the neural data but also the rationale behind some of the experiments (reliable knowledge?) (section 2). A second category of criticisms, of a more philosophical nature, focuses on the relevance of non-behavioral data (cognitive, physiological and neural data) for the testing of economic choice models: a thorny issue of scientific legitimacy that also concerns behavioral economic studies (scientifically-justified knowledge?) (section 3). As well as questioning the reliability of the neuroeconomic experiments’ results and the legitimacy of the approach, some authors query the interest of the lessons drawn from neuroeconomic experiments for economists, in view of already established results from behavioral economics (useful knowledge?) (section 4).

2. THE RELIABILITY OF THE RESULTS?

Among the numerous criticisms of neuroeconomics, many focus on the practices or “methods” (as opposed to issues of “methodology” or epistemology that will be addressed in section 2). They question the reliability and trustworthiness of the results. The issues raised concern both data production (too limited and unreliable) and interpretation (correlation or causality?). Some critics target more specifically neuroimaging studies, questioning both the reliability of the tool and the validity of the rationale underpinning the results ("reverse inference").

2.1. Data production

Some economists criticized neuroeconomics for drawing somewhat hasty conclusions on the basis of scanty data (Rubinstein, 2008; Harrison, 2008a) and unreliable findings (Harrison, 2008a, Harrison, 2008b). Although criticisms concerning inadequate sample size are not unfounded, they are gradually becoming less pertinent with experiments now using more substantial samples relating to a wider range of tools and, most importantly, with the development of field experiments in natural contexts. As regards "statistical simplifications" that undermine the reliability of results, criticisms focus mainly on the neuroimaging technique in neurosciences in general.

2.1.1. The small number of observations

If the brain is counted as the unit of analysis, the very first experiments were indeed based on a very limited data set: about a dozen observations. Some studies subsequently concerned slightly larger samples: for example, experiments carried out by Lorhens et al*.,* (2007) and Smith et al. (2014) concerned 54 and 35 participants respectively. But it is true that, even recently, many experiments using neuroimaging consist of relatively small samples, mainly for financial reasons[1]. That being so, we should not lose sight of the fact that the unit of analysis is not in fact the brain: it is the activation per unit of time of a certain area located in the brain, so that in the end, each study has a relatively rich data set from a few brains, each contributing many observations at each point in time. As acknowledged by Harrison (2008a), this undermines the criticisms to some extent (see also Harrison and Ross, 2010; Ross, 2011).

It is also important to note the development of hyper-scanning technology, a method by which two (or more) participants, each in a separate fMRI scanner, can interact with one another while their brains are simultaneously scanned (Montague et al*.,* 2002). This innovative technology, which produces at least twice as much data in the same time, has opened up new perspectives for "social neuroeconomics", i.e., studies focusing on social interaction (a term introduced by Fehr and Camerer, 2007). However, the increase in processed data should also involve the development of field experiments.

---

[1] We know that in experimental economics and psychology, first generation experiments were also based on quite small samples.

2.1.2. Field experiments: a way forward

In situ neuroeconomics draws on contributions from neuroendocrinology and psychophysiology that provide access to an extensive battery of tools that are both inexpensive and may be used in the field (administration of substances, electrodermal response, oculometry, etc.). It is an inexpensive response to the issue of sample size. It also provides part of the answer to the issue of "external validity" (or "ecological validity") in neuroeconomic experiments (i.e., the generalization of the results beyond the laboratory in subjects' natural environment), an issue that has yet to be properly addressed despite being explored in depth in experimental economics (see in particular Bardsley et al., 2010; Serra 2012b).

One of the first experiments of this type was undertaken in 2007 by John Coates (a *Wall Street* trader turned neurophysiologist) on a London trading floor (Coates and Herbert, 2008). He sampled certain hormonal levels in traders as they worked and established how these levels fluctuated with both their financial performance and market volatility over the course of several days. Hormones found in the blood are a medium through which the brain communicates with various organs; in that regard, they provide information on the way in which the brain reacts in a certain context. During this experiment, the participants were in a "natural" work environment with the potential to make or lose money for real[2]. Lo & Repin (2002) also studied the emotional reactions of traders to market fluctuations using some psychophysiological tools (skin conductance, heart rate, and so on). This pilot study undertaken with only 10 people was extended by Lo, Repin, & Steenbarger (2005) with a larger sample of 80 volunteers using a complementary method to gauge their emotional response (questionnaires completed before and after a five-week day-trader training program).

Generally speaking, these field experiments concern much larger samples than laboratory experiments. For example, the well-known experiment on trust carried out by Kosfeld et al*.,* (2005) had a sample of 194 participants (see below for a fuller discussion of how hormones may affect behavior).

2.1.3. Debatable statistical simplifications?

A sufficient number of observations underpinning the experimental work does not necessarily solve the data issue – at least in neuroimaging experiments. This is because both the methods used in the statistical analysis of these neural images in the form of chronological series and the pooling

---

[2] The study showed that levels of testosterone (a hormone that plays a role in not only sexual but also impulsive and risky behavior) and cortisol (a stress-related hormone) were highly affected by volatile market conditions. A major lesson from this study is that these two hormones – one leading to impulsivity and the other related to stress – blow "hot and cold" in the sense that they must remain within a certain concentration range for a trader to behave efficiently on the market.

procedure employed to construct an “average brain” raise a number of issues. Val et al. (2009) in particular argued that much of the use of neuroimagery techniques in research involves fundamental errors in data handling and statistics. Harrison (2008b) noted these “statistical simplifications”, which, according to him, significantly reduce the reliability of the results. He claims most notably that data standardization leads to the overestimation of the significance thresholds, the minimization of individual variations and hence to a more ready identification of the neural zones activated during the completion of tasks by the participants. Harrison (2008a), with his very colorful language, talks of a genuine “sausage-making factory” (when he talks about the statistical analysis of spatio-temporal data collected on a brain), about “a can of worms” (with respect to the inferences required for pooling across several brains), describing as “MacGyver’s econometrics” (using the name of the hero of an American TV series, known for his ingenuity in dealing with any practical problem) the wider problem of neural data statistical modelling used in neuroeconomics experiments! Moreover, he rebukes the authors of these experiments for never publishing their neural data, a traditional practice in neurosciences that contrasts with experimental economists who always share their data on request. “Replicability” is of course the issue here. Similar criticisms can be found – about the same issues although not in the same style! – in Platt & Huettel (2008), Huettel (2010), Kable (2011) and Van Rooij & Van Orden (2011).

Although this criticism is not specific to neuroeconomics – it also addresses experiments using neuroimaging in general – it should not be underestimated. It contributes to the general conclusion that can be drawn from studying the criticisms concerning the reliability of outcomes in neuroeconomics experiments, which is the need to encourage the use of a wide range of investigative tools.

2.2. Data interpretation

Other criticisms focused on the interpretation of the data provided by different investigative tools (Fudenberg, 2006; Poldrack, 2006; Harrison, 2008a; Harrison, 2008b; Harrison & Ross, 2010; van Rooij & van Orden, 2011). Neuroeconomics, once again, experiences difficulties that are not unique to it but concern cognitive neurosciences in general; every tool used in neuroscientific experiments has advantages and limits (Cravel, 2007). The interpretation of images in fMRI experiments, in particular, raises many difficulties related to the specificity of this tool. Moreover, while neuroimaging tools help to *localize* the areas of the brain activated in decision-making, it is clear that they are not sufficient to establish a *causal* relationship. Identifying neural zones is one thing; understanding what is happening and their role in brain function is what really matters. Fortunately, neuroeconomics now has a battery of intervention tools that complement those concerning localization, enabling genuine causality to be established rather than mere correlations. In addition, recently the focus in neuroimaging has somewhat shifted from correlational techniques (i.e., regressions) to out-of-sample predictions.

2.2.1. Difficulties concerning the interpretation of brain images

The interpretation of the cerebral "default mode" network features among the many difficulties raised by the interpretation of fMRI. Even at rest, the brain shows some activity (Raichle, 2015). Brain imaging by fMRI is also affected by a technical constraint: the BOLD signal measured by the voxels which make up the image is very weak. This explains in particular why the neuroeconomic experiments undertaken using this technology require that the same task be repeated several times in order to recover the *average* signal. As a result, the study of interactive decisions in one-shot games cannot be undertaken with traditional experimental protocols, they have to be modified to take account of this technical constraint. Multi-shot games are used: the game is played several times using so-called "strangers" matching (with participants rewarded on the basis of a random draw from one of the games) (Houser & McCabe, 2009, p. 55). Similarly, any other cause of signal variation, such as a head movement in the scanner must be strictly controlled. In order to minimize this bias, participants' heads are often stabilized using a foam cushion, the visual stimuli are projected onto a screen behind the scanner so as to be visible in a mirror. Some authors go so far as to give an additional financial bonus to subjects who fully comply with the condition of keeping their head still (e.g., Smith et al*.,* 2014). Despite the precautions taken during fMRI experiments, it is still common to have to exclude some participants from the analysis as a result of excessive head movements during image acquisition (see for instance the recent study of Nagel et al*.,* 2018, which tests the hypothesis that choices made under strategic uncertainty might be assimilated to choices made in risky situations, where the observations of two out of the twenty participants had to be excluded from the analysis for this reason).

Apart from criticisms that target the limits of the tool used, an additional criticism may concern the actual design of some neuroeconomic experiments. Hence for Harrison (2008a), in social-decision or intertemporal-choice experiments many researchers probably neglected the existence of possible artefacts or confounds in their interpretation of the observed behavioral data. Yet this problem is well-known in experimental economics (see in particular Bardsley et al., 2010, and Serra, 2012b, for a discussion of this issue in experimental economics). "It is clear" writes Harrison "that we have conceptual work to do before we fire up the scanner" (2008a, p. 321). Otherwise, there is a great risk that the researcher will simply push, with no real experimental proof, the explanation they held when designing the experiment rather than any of the alternatives, Harrison takes the example of the trust game in which neuroeconomists, without reservation according to him, put forward dubious explanations. For example, they interpret observed cooperative behavior as a sign of the trust investors put in their partners while their motivation might be extremely different. Either player might be altruistic toward the other, spiteful towards the experimenter, wanting to inflict high monetary losses to him. Or they might be risk loving in the case of trusting behavior. They also interpret "punitive" behavior as "emotionally" useful whilst it might be considered as perfectly rational to punish those who deviate from a profitable strategy if account is taken of the fact that the participants may perceive the situation as a repetitive game, even though it is not, as it is played as perfect *stranger* matching. Participants' deception, against which experimental economists have developed compelling arguments, may also lead to confusion in the interpretation of some neuroeconomics experiments. However, it is true that in this area economics diverges with psychology (see e.g., Serra, 2012b).

2.2.2. Localization techniques: simple correlations

Insofar as there is a high risk of interaction between the different areas of the brain and given the fact that the activation of a specific brain area is not always strictly correlated with a motivation or a clearly defined cognitive mechanism, it seems difficult to infer a causal relationship from the effects produced in terms of neural processes and functions

Whilst neuroimaging experiments do not allow pronouncements on how mental functions and states are implemented in the brain, it is worth noting that there is a particular category of experiments – the model-based experiments, – that have an advantage over the usual fMRI approaches. In these experiments, rather than correlating directly the neural and behavioral data ("direct approach"), the correlations are conveyed via an explicit computational model. In the simplest cases, the distinctive feature of this approach is to begin with a computational model describing the transformation of a set of stimuli into a set of behavioral responses. The specific "internal" operations required to effect such a transformation are the variables of interest in the neuroimaging study as it is these variables that will ultimately be correlated with the neuroimaging data. By incorporating the behavioral parameters estimated in an explicit computational model into neuroimaging analyses, it becomes possible to know not only "where" but also "when" a particular function is implemented in the brain. Neural data, by complementing behavioral parameters, becomes a powerful source of constraint for neural computational models (O'Doherty, Hampton, & Kim, 2007, Corrado & Doya, 2007 and Corrado et al., 2009 provide detailed descriptions of model-based experiments in neuroeconomics). As stressed by O'Doherty, Hampton, & Kim (2007), the drawback is that these experiments may limit the possibility of detecting unexpected results. Hence, "a reasonable policy most probably consists in using more traditional analyses based on a series of trials in parallel with a model-based approach" (p. 50). Furthermore, it is true to say that these experiments cannot always establish causal explanations.

Fehr (2009, p 226) also mentioned the possibility of generating prediction in experiments solely through the choice of appropriate experimental designs. Thus, the neural activity observed during a specific "treatment" may be used to predict behavior during another "treatment" ("off-treatment" projection). He gives the example of the pioneering experiment undertaken in social neuroeconomy by de Quervain et al. (2004) in which the authors use the individual differences observed on activation of the caudate nucleus (in the striatum) when the punishment of uncooperative participants does not cost anything to those inflicting it (initial treatment) to predict how many individuals will pay to punish when it becomes costly (different treatment).

A step further for improving predictive inference in neuroeconomic experiments is to shift from study brain activity to measure anatomical features of the brain. Structural fMRI (morphometry) allows measuring the size (i.e., the volume of grey matter) of target brain regions or the quality of connections (i.e., white matter tracks) between these regions. These last years, a few pioneer neuroeconomic experiments used this brain imaging tool for studying risk tolerance (Gilaie-Dotan et al., 2014), self-control in intertemporal choice (Bjork, Momenan, & Hommer, 2009; Olson et al., 2009) and social behavior (Bickart et al., 2011; Morishima et al., 2012).

The results of some of these experiments, besides, may sometimes be unsettling. Consider one example by way of illustration. It concerns altruism – the development of which is often explained through active learning or social influences – which might reflect changes in the size and activation of a particular neural structure. This is what Morishima et al. (2012) claim. They examined the neuro-anatomical bases of observable individual altruistic differences through the use of a morphological MRI tool to scan the brains of participants whose social preferences were estimated using dictator and reciprocity games. They establish that the gray matter volume in the temporo-parietal junction is strongly associated with both individuals' altruism and the individual-specific conditions under which this brain region is recruited during altruistic decision making. Thus, individual differences in grey matter volume in the temporoparietal junction not only translate into individual differences in the general propensity to behave altruistically, but they also create a link between brain structure and brain function by indicating the conditions under which individuals are likely to recruit this region when they face a conflict between altruistic and selfish acts. This result needs, of course, to be confirmed by further studies. Nonetheless, it can be linked with some experiments in cognitive neuroscience which show that structural brain changes are connected with prolonged learning[3].

Although these morphometric studies, by providing evidence for specific anatomical markers of individual preferences, may have interesting implications for economic models (Konovalof & Krajbich, 2019), it is nevertheless clear that the kind of predictive inferences they make do not provide genuine causality.

2.2.3. Intervention techniques: genuine causality

Most of the difficulties related to the interpretation of brain data in fMRI experiments disappear with the introduction of experimental procedures, such as transcranial stimulation or substance administration, designed to induce a behavioral effect exogenously. Using these methods, which are beginning to develop over the past decade, helps establish genuine causality in neuroeconomics rather than simple correlations. They complement lessons learned from studies involving patients suffering from lesions or certain diseases which remain a major source of knowledge in the field[4].

Transcranial stimulation

Repetitive transcranial magnetic stimulation (rTMS) and transcranial direct-current stimulation (tDCS) both give the option to transiently increase (excitatory effect) or to decrease (inhibitory effect) neural activity in specific cortical areas before the participants make a decision when they are performing an experimental task designed to reveal their preference.

tDCS has an additional advantage: it helps to avoid a problem that may arise when using brain imaging tools or rTMS in social neuroeconomics to study “social preferences”, i.e., when it is assumed that individuals trade-off other-regarding behavior with selfish goals (altruism, benevolent reciprocity, inequity aversion…). 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, as previously mentioned, it is

---

[3] For example, following several days of complex motor skill learning, the motor map size of the fingers engaged in the task increases (Dayan & Cohen, 2011). Likewise, higher volumes can be recorded in the motor and auditory cortex of musicians compared with the same neural areas in non-musicians (Münte et al., 2002).

[4] It was back in 1994 that the neurologist Antonio 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. 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).

strongly recommended not to deceive participants in order to keep their trust in the experimentalist. As tDCS is inexpensive, it can be administered simultaneously to many interacting subjects (see e.g., Knoch et al., 2008).

During brain imaging experiments, however, there is another way to deal with this problem, which is to present each participant with the choices made *previously* by interacting partners during identical games. This experimental strategy was used in particular in the seminal experiment carried out by de Quervain et al. (2004). This experiment used as a tool positron emission tomography (PET). Each participant was presented with the choices made previously by certain participants in the pilot experiment; at the end of the latter, the participants agreed to the collected data being used in a future experiment in the knowledge that they would be rewarded according to their own choices (see Fehr, 2009, for further discussion of these different methodological aspects).

Some results of transcranial stimulation's experiments can be given. It was possible to show, inter alia, that an experimentally-induced variation of the dorsolateral prefrontal cortex (the most usually concerned neural area) may modify participants' willingness to pay to acquire certain goods (Camus et al., 2009), their risk-aversion (Knoch et al, 2006a; Fecteau et al., 2007) or their capacity to resist temptation (self-control problem) (Knoch & Fehr, 2007; Figner et al., 2010). It was also shown that an experimentally-induced variation in the activity of this neural area may make the participants more or less receptive to fairness concerns (Knoch et al, 2006b; Knoch et al., 2008; Baumgartner et al., 2011), to guilt feelings (Krajbich et al., 2009), more or less inclined to disregard certain social norms (Ruff, Ugazio, & Fehr, 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 (Karton & Bachmann, 2011; Baumgartner, Gianotti, & Knoch, 2013).

Consider as an illustration the study of Karton & Bachmann (2011) according to whom honesty (i.e., spontaneous truth-telling), an individual feature of a profoundly moral nature, might be associated with the activation of a specific brain area. By combining rTMS and fMRI neuroimaging, the authors show that spontaneous propensity to lie can be manipulated by stimulation of the dorsolateral prefrontal cortex. Their experiment used a visual object recognition task without penalty. They show more precisely that stimulation of this zone in the right hemisphere tends to decrease the propensity to lie whereas left hemisphere stimulation tends to have the opposite effect. Hence, our sense of right and wrong does not just depend on education and religious and philosophical beliefs but also on the biology of our brain… Of course, the results of this exploratory study would benefit from further confirmation, especially with more realistic experimental paradigms that would introduce some reasons to tell the truth or to lie[5].

Pharmacological manipulations

We now know that some pharmacological manipulations also affect the brain and generate significant causal effects. Endocrinological economics, drawing on certain established physiological mechanisms, is without doubt emerging as one of the main directions in neuroeconomics. The main chemical substances used are hormones (cortisol, testosterone, adrenaline…) or neuropeptides (oxytocin, arginine, vasopressin…) and more rarely psychotropic substances (benzodiazepines, nicotine, amphetamines…). All these substances are thought to affect behavior either by acting as neurotransmitters or through their action on some of the central nervous system's neurotransmitters (dopamine, serotonin, noradrenaline, acetylcholine, glutamate, GABA…).

To date oxytocin is by far the substance most frequently investigated in experiments. This neuropeptide, mainly produced in the hypothalamus, is often considered to be a "female hormone". It plays a role in promoting mother-child bonding, directly affecting breastfeeding. But 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 oxytocin's properties is to inhibit amygdala activation. However, oxytocin inhalation does not appear to affect the loyalty of the trustees. In order 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). However, this neuropeptide may affect reciprocity. Oxytocin release may increase in trustees when they interpret the behavior of investors as a signal of their trust in them (Zak, Kurzban, & Matzner, 2005b). It should be noted that this is not the result of an experiment wherein this hormone is *administered* to the participants (i.e., the *exogenous* modification of their physiological status) where the results are then compared with a "placebo" group; it is instead the comparison of the *endogenous* oxytocin level in the blood of the same participants playing the trust game and an

---

[5] Several studies concerning honest and dishonest behavior can be found in the experimental literature (see e.g. Fioretti & Marden, 2015). These studies reflect an emerging "neurolaw" literature linked to the increasing use of functional imaging in criminal trials in the United States and the emergence of a "truth business" with the creation of powerful companies (such as *NoLieMRI* and *Cephos*) that offer methods of lie detection using fMRI (for a debate on neurolaw and the ethical problems raised by this approach, see in particular Sinnott-Armstrong et al*.,* 2008; Forest, 2014).

impersonal standard game. In Zak, Kurzban, & Matzner (2004), a similar protocol shows a positive correlation between oxytocin level and the level of investor trust.

Oxytocin may also have other effects revealed by the ultimatum game – another famous behavioral paradigm: it might increase generosity (Zak, Stanton, & Amadi, 2007; Vercoe & Zak, 2010) and reduce betrayal aversion (De Dreu, 2012a). It may also correlate with the subjective experience of empathy (Barraza & Zak, 2009). The effects of intranasal administration of this neuropeptide were also studied in the repeated prisoner's dilemma game. As it is a symmetrical game (the participants take turns going first and second), no difference is found between the two roles as a result of this administration. This is no longer the case for reciprocity: oxytocin promotes positive reciprocity manifestations without significantly affecting negative reciprocity reactions (Rilling & Sanfey 2011; Rilling et al., 2012). The effects of arginine vasopressin (AVP), another neuropeptide, on mutual cooperation were also examined in these two experiments. AVP is well known for its anxiogenic action (it tends to cause anxiety, stress and anguish) (Heinrichs, von Dawans, & Domes, 2009). The authors therefore expected the administration of this neuropeptide to increase the activation of the amygdala and reduce cooperation from a behavioral viewpoint. Instead the results showed that subjects treated with AVP have higher positive reciprocity levels than those treated with oxytocin or a placebo. The authors suggest an explanation based on the fact that AVP might act on the connection between the amygdala and the anterior section of the insula, such that cooperative behavior could be seen as a response intended to minimize the risk of social unpleasantness. The influence of oxytocin on pure altruism was also tested in a charitable donation game (Barraza et al., 2011): oxytocin causes individuals to engage in acts of indirect generosity, increasing the size of monetary donations to charitable organizations.

In syntheses of many experimental studies undertaken on the role of oxytocin on behavior, this neuropeptide appears to be closely related with social interactions and prosocial behavior (MacDonald & MacDonald, 2010; Zak, 2011, 2012; Riedl & Javor, 2012; Shamay-Tsoory et al., 2013). Yet more recent literature on this issue has generally produced rather contrasted results. Hence for some authors, oxytocin is mainly presented as an anxiolytic, pre-supposing that social interactions are responsible for triggering anxiety (Evans et al., 2013). For others, it would only have a significant impact on trust (or reduce mistrust) in close relatives or members of the same group (De Dreu, 2012b). Still others think it might have an opposite effect on anonymous interactors (Pfeiffer, 2013; Lambert, Declerck, & Boone, 2014). Overall, even if it seems that taking oxytocin may have an effect on patients with social behavior disorders (such as autism or social anxiety issues), the relationship between oxytocin and trust in humans remains an open question. There is very little evidence showing that oxytocin directly influences a trusting attitude. If such an effect exists, it is likely to be curbed by many factors (Nave, Camerer, & McCullough, 2015; Koppel et al., 2017). Furthermore, a recent study undertaken in a social dilemma game with sanctions[6] (Aydogan et al., 2017) showed that the intranasal administration of oxytocin tends to amplify the display of negative social emotions such as anger, by enhancing the inclination to sanction free-riders. Ultimately, it is evident that the "prosocial" nature of oxytocin requires clarification.

Testosterone, the premier male hormone, has also been the subject of some pharmacological manipulation experiments. Hence, the trust game has also been studied "under testosterone". In this case, the findings are a far cry from those resulting from oxytocin nasal administration: testosterone tends to reduce the propensity to trust, in that investors' "offers" were reduced (refer to Zak et al., 2009, for an experiment involving only male participants). But contrary to the generally accepted view, testosterone does not always lead to aggressive behavior. Whilst several studies establish positive correlations between high blood testosterone levels and aggressive behavior, there is no experimental evidence of the causal impact of this hormone on aggressivity. Indeed, the experiment carried out by Eisenegger, Haushofer, & Fehr (2011) in an ultimatum game involving only female participants (as their initial blood testosterone level is *a priori* lower) gave the opposite results. These researchers' hypothesis was that if testosterone fosters aggressivity, its administration should lead to lower offers. However, in their experiment, women who received one dose of this hormone (and not a placebo) selected significantly higher offers. How can this result be explained? By the fact that the fear of the social affront resulting from the offer being refused ultimately prevails on pursuing purely financial interest. In a socially-complex human environment, prosocial, rather than aggressive, behavior is expected to give a certain social status, i.e., some importance in the eyes of others. In economics experiments, however, various precautions are usually taken to minimize the intervention of these external factors imported in the laboratory (double-blind, task repetition, perfect stranger matching, etc.). But even in a controlled laboratory environment, it is difficult to remove entirely these idiosyncratic factors that color homegrown preferences (Smith, 2010). This is therefore the basis of the interpretation of laboratory behavior by Eisenegger, Haushofer, & Fehr (2011) in their experiment on the persistence of this central element of social life intersubjectivity. This interesting experiment also showed that the physiological manipulation of the respondents has no effect on the proportion of offers accepted; this suggests that this hormone is unlikely to promote

---

[6] In some prisoner' dilemma games the rule of the game states that players violating social norm of cooperation or reciprocity may be monetary sanctioned (so-called "altruistic" punishment).

altruistic behavior[7]. But other results, of a different nature, also emerge from experiments involving testosterone administration: in particular, it seems that this hormone may make people behave more honestly (Wibral et al., 2012).

Many other substances are administered to participants in neuroeconomics experiments. In particular, it has been established that serotonin depletion raises the rejection rate of unfair offers in the ultimatum game (Crockett et al*.,* 2008) whilst benzodiazepine reduces it (Gospic et al*.,* 2011) and that serotonin, by shaping moral judgement, could increase aversion to harming others (Siegel & Crockett, 2013). Crockett and Fehr (2014) present a comprehensive overview of this literature in a study that identifies the role of these various neuropeptides, hormones or other substances in shaping time, risk aversion and social preferences.

Fehr (2009, p. 222) saw in the use of pharmacological manipulations an opportunity for the radical transformation of neuroeconomics (he even spoke of a "revolution") as the low cost of this experimental method makes it possible to undertake experiments with samples of comparable size to those in experimental economics. De facto, in recent years there have been many more behavioral study experiments with pharmacological interventions.

2.3. The problem of "reverse inference".

In neuroeconomic experiments with neuroimaging, neural activity is supposed to be associated with some cognitive functions or some mental states. It is therefore indispensable to guarantee the connection between the activation of the neural zone and the cognitive function or mental state. The validity of a typical neuroeconomic experiment using neuroimaging, based on inferring neural activity to cognitive functions or mental states relies on this hypothesis. However, this cannot be taken for granted. The conditions that govern the validity, if not the legitimacy, of "reverse inference" reasoning must be discussed[8].

2.3.1. The functional value of neural activity is indeterminate

The accumulation of neuroimaging data over the past twenty years has provided a stable functional cartography of the brain which is gradually improving. In neuroimaging experiments, in neuroeconomics, these earlier studies established a correspondence between the activation of a

---

[7] The behavioral and physiological differences between genders were also addressed by Zak et al. (2005a) who established that mistrust was associated with an increased level of dihydrotestosterone (DHT) and showed that men but not women might respond to mistrust with an increased level of this hormone.

[8] The problem of "reverse inference" is not unique to neuroeconomics; it is also found in cognitive psychology experiments that rely on neuroimaging to infer particular cognitive functions (memory, attentiveness, language…) from the activation of specific neural zones. The initial methodological discussions of the issue focused only on cognitive psychology (D' Esposito et al., 1998). Poldrack (2006) was the first author to bring in some neuroeconomic experiments.

specific neural zone and the engagement of a cognitive process that is now used to infer "in reverse" that the activity observed in this particular neural zone in new experiments reflects the engagement of the same cognitive process. Harrison (2008b) called this reasoning "fallacious" (p. 535), as it involves reasoning backwards from the current neurobiological approach which consists of first localizing a cognitive function in a neural area then inferring that the engagement of these predefined functions depends on the activation of this area (see also Harrison & Ross, 2010; Ross, 2011). This economist even labeled neuroeconomics the new "neo-phrenology" (2008b p. 536).[9]. Bourgeois-Gironde (2010, p. 245), for his part, talks about a "same-areas strategy" and emphasizes that this approach is only analogical and not deductive.

The reasoning can be broken down as follows: (1) in the study concerned, brain area Z is activated when the participant performs the task T; (2) in other studies, when cognitive process X was supposed to be engaged, area Z was activated; (3) it follows that the activity of brain area Z shows that the cognitive process X is engaged when the participant undertakes task T in this study. Strictly speaking, this reasoning is only valid if, in earlier studies, a biunivocal relationship between the engagement of the cognitive process and the activation of the neural area was established. But this is not the kind of result that is available. Beside the issue of the scale chosen for reasoning (the neuron or a functionally-defined area), it seems likely that a given neural system can "realize" several distinctive functions: hence the indeterminacy of the functional value of neural activity (e.g., Poldrack, 2006, 2011; Harrison, 2008b; Phelps, 2009; Bourgeois-Gironde, 2010; Foster, 2014).

2.3.2. Which stratagem to adopt?

Faced with this weakness of the reasoning underlying the conclusions drawn from neuroimaging in neuroeconomics, the question arises of how to deal with the issue. Several routes were suggested for legitimating reverse inference before recent progress in the imaging neuroscientist's statistical arsenal allow to mitigating this issue.

Several alternative routes for trying to legitimate reverse inference

According to Poldrack (2006), the first factor that may give credit to the reverse inference is the "selectivity" of the response of a specific brain area. The selectivity of neural activation is inversely correlated with its involvement in all the conceivable experimental tasks and therefore potential cognitive processes. The more the neural response is engaged in a wide variety of tasks and cognitive processes (i.e., the less selective it is), the less it is likely that the information provided by

---

[9] The term "new phrenology" used by Harrison had already been suggested by the psychologist William Uttal to describe the neuroscience approach which consists in acknowledging brain activity as a determinant of mental states in the light of functional neuroimaging (see in particular Uttal, 2001).

inverse inference will be useful. Evaluating neural response selectivity relies for a large part on the availability of databases that record correlations between brain anatomical structures and cognitive functions (sometimes called "cognitive ontologies" in computing terminology) (Price & Friston, 2005). Yet, as existing cognitive ontologies do not have entries specifically adapted to theoretical economic constructs, neuroeconomists must rely on a few earlier studies that have shown some correlations. Even in cognitive psychology, the discipline initially targeted by Poldrack (2006), the "cognitive ontologies" of existing databases are quite coarse in comparison with current theories of cognitive psychology – in any case when compared with those developed for example for genome informatics. According to the author, a second factor may help to legitimize reverse inference: the activation of a set of regions functioning as connected networks. At the same time, however, the smaller is the relevant area, the more confidence there will be in the reality of inverse inference.

Let's consider some examples. It was known that the amygdala (a small area in the medial temporal lobe of the brain) and the anterior insula (a rather bulky area situated in the lateral sulcus) represent two neural areas heavily involved in emotional behavior. But they are particularly non-selective (even the amygdala despite its small size) as they serve a number of functions. So, it is difficult to infer any conclusion as regards behavior or emotional displays from their sole activation; all that can be observed are some correlations (Craig, 2009). If for example, activating the amygdala is supposed to create a feeling of fear in the participant (it is one of the emotions most often attributed to it), other measures will have to comfort this hypothesis (for example via an electrodermal response, a facial expression or verbal communication). However, for another small-size area such as the caudate nucleus, one of the structures that make up the striatum with more limited functions, inverse inference might be more reliable[10]. Finally, Poldrack stresses that lessons drawn from neural data would be more useful if confirmed by a combination of data from different sources (behavioral or physiological data) (see also Phelps, 2009; Poldrack, 2011)

More recently, Hutzler (2013) provides a revised formulation of reverse inference that includes an additional conditional constraint that has been previously acknowledged, but so far not implemented: the task-setting. He shows that this revised formulation reveals that reverse inference can have high predictive power and allows an estimation of reverse inference on the basis of meta-analyses instead of large-scale databases. So, he concludes that reverse inference cannot be disregarded as a fallacy per se. Rather, the predictive power of reverse inference can even be

---

[10] It is known that the upper part of the caudate is responsible for executive and memory related functions while its lower part is engaged with emotion and motivation processes.

"decisive"-dependent on the cognitive process of interest, the specific brain region activated, and the task-setting used.

To date, one of the most elaborate logical discussions of reverse inference is found in Bourgois-Gironde (2010). This philosopher suggests some ways forward to make sense of this reasoning. To that end, he uses in particular the concept of "abduction" borrowed from Peirce (1878), which in substance can be described as an attempt "to conclude A from observation B, knowing that A implies B". However, whilst this reasoning may be useful to generate hypotheses, it does not have the capacity to prove various propositions (Bourgeois-Gironde, 2010, p. 235)[11]. On the other hand, if resorting to prior studies can help to formulate interpretive hypotheses, these must then be confirmed. In particular, the description of the cognitive function being studied must be sufficiently general so as to include all the other functions likely to depend on the relevant neural area without losing its interpretive relevance. As it is difficult to refer to earlier data in neuroeconomics to evaluate neural response selectivity, Bourgeois-Gironde suggests another route: the experimental paradigms used in the discipline should develop their own measures of neural selectivity. This method of "repetition suppression" (the reduction of neural response that can be observed when stimuli are presented several times) has already been used in cognitive psychology in particular in the case of perception. But it is not limited to the investigation of the most fundamental neural functions and may be applied to evaluate the role of relevant cognitive and affective processes in neuroeconomics.

On his side, Forest (2014) suggests a response of a very different nature to the issue of reverse inference. It consists of resorting to a different concept of scientific knowledge, an alternative epistemology that would give a prominent role to brain images. While functional imaging may not be a source of knowledge from a "reliability" viewpoint (i.e., it cannot play the role of a "reliable indicator", an indicator that "reveals the truth" regardless of the circumstances), this does not mean that it cannot be a source of knowledge at all. The philosopher refers to an alternative concept of knowledge, the "coherentist" approach: "a belief is only justified if it is in compliance with all our background beliefs already accepted or if the latter can be adjusted to be made compliant with it" (Forest, 2014, p. 56 /translated from the French). Coherence may be understood as "weak", i.e., a

---

[11] In his work on scientific reasoning, Peirce introduces the term of "abduction" to denote a type of non-deductive inference different from the familiar inductive type. Abduction relates to deduction and induction as follows. Deduction and induction come into play at the later stage of theory assessment: deduction helps to derive testable consequences from the explanatory hypotheses that abduction has helped us to conceive, and induction finally helps us to reach a verdict on the hypotheses, where the nature of the verdict is dependent on the number of testable consequences that have been verified. According to Peirce, abduction belongs to logic because it can be given a schematic characterization, to wit, the following: the surprising fact, C, is observed; but if A were true, C would be a matter of course. Hence, there is reason to suspect that A is true (see e.g., Fann, 1970).

simple joint possibility or an absence of logical contradiction, or as "strong", i.e., "a congruence" characterized by the following: "S's belief in P is justified if and only if P is part of a set of propositions that support and confirm each other" (Lewis, 1946). But congruence raises at least two issues: first, the mutual selection of beliefs does not guarantee *ipso facto* their validity and second, nothing is mentioned in the relevant formula for the choice of P: why give P the priority compared to other alternatives that are just as acceptable, in particular if they do not directly harm accepted beliefs? For this reason, Forest selects a constrained version of strong coherence that he calls "qualified congruence": "S is justified in accepting P, if and only if there are reasons to favor P (over P', P"), from a set of propositions which support and confirm each other on the basis of independent and informative sources of evidence[12] … Qualified congruence helps to generate knowledge … even if we do not have at our disposal reliable mechanisms for information acquisition" (Forest, 2014, p. 58 / translated from the French). The interpretation of an image may be validated on the grounds of its coherence using interpretations of previously obtained images and background knowledge. Consequently, following this epistemological approach, the references made to existing literature by researchers seeking to justify inference make perfect sense[13].

As we can see, the issue of reverse inference raises practical issues and delicate problems of scientific logic. Fortunately, recent developments in statistical analysis and informatics have provided new and more powerful ways to infer mental states from neuroimaging data (Poldrack, 2011; Poldrack et al., 2017; Poldrack, 2018).

Decoding techniques of brain images

We know that fMRI studies focused on associating brain zones with mental functions. The introduction of decoding techniques – including "multivariate pattern analysis" (MVPA), the most widespread of these techniques – 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. Now, many fMRI studies use decoding techniques for determining whether spatial patterns of brain activity can predict, out of sample, what a subject is seeing or doing or even predict later behavior (Normann, Polyn, & Haxby, 2006; Haxby, Connoly, & Guntupalli, 2014) (see e.g.,

---

[12] These sources of evidence are "independent", in the sense that they rely on distinct methodologies or knowledge sources, and "informative", because they have a proven track record in generating recognized knowledge.

[13] On the relationship between the different theories of knowledge such as "foundationalism", "reliabilism" and "coherentism", see e.g., Goldman (1988).

Wake & Izuma, 2017, for using of MVPA in a recent neuroeconomic experiment proving it exists a common neural code for social and monetary rewards in the human striatum).

Reverse inference consists in using brain activation data (fMRI or PET) to “informally” infer the engagement of a certain cognitive function. Decoding is just a process of applying machine learning algorithms to such brain activation patterns in order to infer something about the cognitive function that elicited this response[14]. It’s a powerful way to infer mental states from neuroimaging data.

Reverse inferences suffer from several potential pitfalls, among them the fact that: (i) a particular pattern of activity usually indicates different mental states in different contexts; (ii) patterns isolated using a statistical analysis usually do not fully describe the mental state; (iii) mental state labels themselves are context-dependent, subject-dependent, and even temporally variable; (iv) mental states are not discrete, as evidenced by all the "noise" in neuroimaging data, so subjects may identify their mental state using a variety of different labels depending on what part of their mental state they attend to. Reverse inference, which is based on a researcher’s knowledge, also is flawed because an individual’s knowledge is limited by what she/he remembers and has read; additionally, poor interpretations are compounded from one researcher to another (Poldrack, 2011).

Remark that reverse inference is subject to these issues no matter what statistical analysis is used, MVPA is not immune. Yet, many of these issues can be overcome using a variety of new techniques. In particular, MVPA is very valuable for dealing with the first issue: thanks to much higher "resolution" (level of detail, scope of data, linearity…), this decoding technique is significantly less likely to confuse different mental states that appear to have the same activity pattern at a lower resolution. Moreover, an interesting side-effect of machine-learning pattern analyzers is that they tend to remove the temptation to make unwarranted reverse inferences. In the early days, when patterns of activation were "informal" (i.e., based on gross anatomy), it was too easy to make sweeping statements such as "... the amygdala, which is associated with fear … or the anterior insula which is associated with disgust...". With MVPA, patterns of activation are practically impossible to describe in human language, so machines are required to make inferences, which are then inherently more objective[15].

At the end, what conclusion should be drawn from the limits inherent to the various methods and tools used in neuroeconomics? No doubt, they should prompt us to be cautious about the significance that may have been given to the results of some of the experiments undertaken to date. They highlight at the very least the interest in, and even the necessity for, the integration of all

---

[14] For a clear overview on machine learning, see Jordan & Mitchell (2015).

[15] For an overview on computer-age statistical inference, see e.g., Efron & Hastie (2016) or Bzdok & Yeo (2017).

available methods and tools: various neuroimaging tools using now new statistical methods for decoding brain images, morphometric studies, magnetic, electric or optical stimulation techniques, pharmacological manipulations, lesion studies, psycho-physiological measurements, as well as neurophysiological studies with non-human primates. Ultimately, the credibility of the results will only improve by combining the various techniques drawn from the rich and diverse tool box now available and comparing the different potential investigative methods[16].

3. A SCIENTIFICALLY-LEGITIMATE APPROACH?

Neuroeconomics also raises an exclusively methodological, if not philosophical, problem. The issue goes beyond the strict boundary of neuroeconomics and includes also behavioral economics in that it collects cognitive and non-behavioral data, i.e., data that concerns cognitive states and processes assumed to be responsible for certain types of behavior but not on the behavior itself.

Most of the data used in economics is behavioral, whether collected in the field or in a controlled experimental environment. From this data, models are tested and subjective variables are measured and predicted from the models. However, in some behavioral economic studies (where emotional measurement and induction methods are used) and, more particularly, in neuroeconomics, *physiological and neural data* is also collected. There is currently disagreement on whether or not this non-behavioral data is "relevant" in economics. It is necessary to examine the nature of the epistemological relationship between non-behavioral data and the choice models used in economics, whether based on the standard or new alternative models developed in behavioral economics, such as prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Most of the positions and arguments put forward in the debate are reactions to the notion of "mindless economics" introduced by Gul & Pesendorfer (2008) (see in particular the articles collected in Caplin & Schotter, 2008, and in the special issues of *Economics and Philosophy* (2008) and the *Journal of Economic Methodology* (2010).

3.1. The lack of relevant non-behavioral data to test economic choice models: the notion of "mindless economics".

Whilst not denying the "heuristic wealth" that may emerge from the lessons drawn from non-behavioral data, Gul and Pesendorfer (2008) argued that this data has no epistemological relevance in economic choice models. According to these authors, non-behavioral data cannot have a "conformational" relationship with choice models, i.e., the data can neither support nor reject choice

---

[16] An evaluation of the advantages and limits of the range of tools and methods used in neuroeconomic experiments can be found in Ruff & Huettel (2014), Reuter & Montag (2016) and Genon et al. (2018).

models as these latter cannot be called into question on the basis of such data. For these authors, according to the description used by Cozic (2012), economic choice models are "cognitively silent" (p. 138 / translated from the French). Such silence means that these choice models do not affect cognitive data, i.e., a mentalist approach is possible but they in no way constrain the cognitive data that may be collected[17].

In other words, the postulate of Gul and Pesendorfer may be summarized as follows: economics does not focus on the same issues as neurosciences or psychology. It pursues different objectives and must therefore use abstractions specific to its purpose. The biological or psychological variables responsible for choices are not relevant from an economic viewpoint. An economist does not need to know, for example, if "the choice of x over y may be motivated by the pursuit of happiness, a sense of duty or religious obligation or reflects an impulse*"* (Gul and Pesendorfer, 2008, p. 24). For an economist the only thing that matters are the preferences revealed by the choices. Yet, defending a strictly instrumentalist approach does not necessarily entail that economics is completely detached from psychological research. What it does entail is that it should only integrate relevant data, i.e., data concerning choices and behavior. In this respect, data produced by economics or experimental psychology may very well be used to calibrate an economic model or predict future choices or equilibria. However, data or variables that are uncorrelated with choices, such as data relating to neural activities, should not be integrated into the model. How seriously should Gul and Pesendorfer's critique of neuroeconomics be taken?

3.2. The scope of Gul and Pesendorfer's critique.

Several replies have been made in answer to Gul and Pesendorfer's critique (2008). We will examine some of them in increasing distance from the dominant methodology in economics.

The first approach amounts to denying that there is a problem by minimizing potential differences between standard economic theory, behavioral economics and neuroeconomics. Camerer in particular, presents the latter two as simple generalizations of rational-choice theory: behavioral economics adds *psychological* variables to the theory whereas neuroeconomics broadens the analysis further by adding *biological* components to behavior. It adds a mechanistic component to the standard economic theory without fundamentally affecting the approach. According to this viewpoint, behavioral economics and neuroeconomics might help to generalize "the rational-choice theory that incorporates limits of rationality, will power, and self-interest in a formal way" (Camerer,

---

[17] Cozic (2012) notes two implications of the cognitive silence of economic choice models: (i) they do not *contain* cognitive concepts; (ii) they do not have an *impact* on cognitive data. But he convincingly argues that the former is unacceptable, unless revealed preference theory is accepted unconditionally.

2008, p. 44). The central question (which is not answered by Gul and Pesendorfer) is to know whether this generalization "will improve the capacity to understand and predict choice, while maintaining a mathematical discipline and use of behavioral (choice) data" (p. 45).

Wilkinson (2008) suggests that there is some degree of current misunderstanding about the ambitions of behavioral economics and neuroeconomics. Instrumentalism (in essence, the reliance on "as if" arguments), which is dominant in economics, is entirely appropriate, even for a behavioral author, provided that the theory yields correct predictions. In neuroeconomics, it is not generally suggested that neural data could *falsify* a theory. The argumentation is as follows: as standard economic theories are often falsified by empirical studies based on data relating to revealed preference theory (i.e., they are falsified on their own terms), neuroeconomic studies may be useful to understand why theories are invalidated, by recourse to the underlying psychological and biological processes. Consequently, these studies may help establish that some hypotheses that are implicit in the "as if" reasoning of the standard model are bad hypotheses in as much as they give false predictions, thus suggesting possible paths towards the development of "better" economic models. Hence, a sound knowledge of the mechanisms behind the decision-making process may help improve theoretical predictions. Whilst neuroeconomics does not enable direct testing of current economic models, it should still be regarded as a relevant approach to evaluate and test the plausibility of economic conjectures indirectly (see also Wilkinson & Klaes, 2018).

Alternatively, one might agree with Aydinonat (2010) that the relevance of neuroeconomics is doubtless much broader: beyond the indirect role it might play in the empirical validation of current economic models, it should also be judged on its capacity to improve the explanation of certain specific economic phenomena for which it is often necessary to take into account some of their specific characteristics, this would entail factors supposed to be outside the current economic theoretical framework. For example, the "context" of a decision should be taken into account as the way the brain assesses rewards depends on both the personal history of the individual and the context. For the most part, neuroeconomics should contribute to improving our understanding of economic phenomena in general. Aydinonat takes the example of money illusion, it can be explained by the fact that money is not perceived by the brain as a simple tool that provides services but also as a drug that highjacks the reward system in many interesting ways. Several neuroeconomic experiments seem to suggest that the impact of money on the brain is similar to that of a drug. For the brain, money might not simply be a way of accessing consumption but a source of pleasure per se. This is why money "consumption" is often very addictive, just like primary "reinforcers" such as drugs, sex or food (Delgado et al., 2000; Knutson et al., 2001).

A slightly different argument is advanced by Clithero, Tankersley, & Huettel (2008). They do not disagree that researchers could falsify any possible economic model using behavioral data and may predict behavioral phenomena without an understanding of mechanism. Yet that supposes the data necessary to falsify or support a model can, in practice, be identified and collected, which is not often the case. No researcher can obtain all possible data about all possible behaviors. So, the thesis put forward by Gul & Pesendorfer "succeeds in principle but fails in practice" (p. 2349).

For his part, Harrison, although quite critical of neuroeconomic experiments – at any rate, of the way they were originally designed and implemented (refer to section 1) – does not necessarily endorse Gul and Pesendorfer's view. Although less than enthusiastic about the reliability of results due to the limits of the experimental laboratory setting and the difficulties encountered in interpreting the neural data, he dismisses the argument about the lack of scientific legitimacy. The potential role of neural data should not be discounted *a priori*. However, the formal mechanisms that govern the economic decision-making process should be re-formulated, stressing first and foremost the *processes* leading to the choice rather than the choice itself. According to Harrison, neuroeconomics might "bring answers to new questions" (referring to Smith, 2008), if "it is proposed to formally view economic behavior as the outcome of algorithmic processes…. The idea is familiar to many behavioral economists already but only as a metaphor" (p. 330). Neural data within this project would certainly have a role to play, but for now the priority is to develop a genuine communication program between economics and cognitive psychology. This is precisely the direction taken by research in recent years, an evolution initiated at the same time in experimental economics by Vernon Smith, one of this discipline's pioneers. Having originally defined the characteristics of the economic experimental method from earlier experiments undertaken by psychologists, the 2002 Nobel Prize in economics then worked on increasing the links between this discipline and psychology (Smith, 2008).

In a similar vein, Walliser (2012) refers to Gul and Pesendorfer's viewpoint as an "integrist position" that in effect "states that studying choice at the sole behavioral level is not restrictive" and that knowing the mental and neural bases for a choice is of no interest. On the other hand, he approves of an "ecumenical position" which "states that the decision process may benefit from observations at the mental and neural levels" (p. 18).

Quartz (2008) takes a more radical stance. Drawing a parallel with the mid-1980s controversies in cognitive sciences during the emergence of computational and cognitive sciences, he stresses the danger attached to a strict autonomy of economics with respect to neurosciences. He does not see what could justify Gul and Pesendorfer's claim that "neuroscience evidence cannot refute economic models because the latter make no assumptions and draw no conclusion about the physiology of the brain… other than by invoking a strong autonomy thesis". Yet, according to the author, "while economic theory may make no explicit assumptions about the physiology of the brain, it is not the case that it makes no predictions that could either be confirmed or disconfirmed by neuroscience" (p. 463). In this respect, see the interpretations of some authors of the results of their experiments: the data "might confirm" the mean variance model rather than the expected utility theory (Quartz, 2008, and Schultz et al., 2008, concerning in particular the experiment of Preuschoff et al., 2006), this data would "lend support to" an ambiguous option model that evaluates separately expected utility and ambiguity (Camerer, 2008b, on the experiment by Hsu et al., 2005), it "might contradict" the consequence of ambiguous option models which supposes that the processes evoked by risky decisions are subsumed within those evoked by ambiguous decisions" (Huettel et al., 2006), it "might confirm" the quasi-hyperbolic model and "might discredit" the exponential discounting model (McClure et al., 2004a), or the results might be considered as "substantiated evidence" of the social preferences model (de Quervain et al*.,* 2004; Sanfey et al*.,* 2003) (all these experiments are mentioned briefly in Serra, 2016 and Reuter & Montag, 2016). In the end, we are back to the old epistemological feud between supporters of a "realist" stance (in a nutshell, the role of a theory is to posit terms that could correspond to real world entities) and defenders of an "instrumentalist" stance (in a nutshell, only the predictive capacity of the theory matters)[18]. Citing the disappointments generated by instrumentalist interpretations of the symbolic level in cognitive science and, more generally, the way this thesis of cognitive science autonomy has stunted progress, Quartz rejects the "mindless economics" defended by Gul and Pesendorfer, as it embodies methodological isolationism and inflexibility.

There is a last a very different argument against Gul and Pesendorder's view. It was expressed by Caplin & Dean (2009) who observe that the original axiomatic approach they advocate is *a priori* "immune to such criticisms as it produces models which formally characterize whatever data is under consideration" (p. 31) (for a discussion of "axiomatic neuroeconomics", see Capln & Glimcher, 2014; Dean, 2013)[19].

---

[18] On the methodology of Friedman's positive economics and its relationship with instrumentalism, see in particular Mäki (2003). A discussion of realism, instrumentalism and other forms of anti-realism philosophy can be found in Bird (1998) and Chalmers (1999). Adopting a realist scientific stance is not the only possible basis for a critical argumentation against the methodology advocated by Friedman. Hausman (1992) developed an alternative critique which rejects the argument that "the irrelevance of realistic assumptions" would follow from an instrumental stance according to which theories should be evaluated by their predictive success.

[19] Like the economists who, since Samuelson and Houthakker, claim that saying that an individual maximizing a utility function is the same as saying that they obey the well-known axioms of preference theory, these authors question whether it would not be possible to identify a restricted set of axioms that would characterize all classes of "reward prediction error" algorithms, i.e., the mechanism that is supposed to work in reward learning, which characterizes the new paradigm adopted by neurosciences in the 1990s (see below).

Given these multiple arguments in response to Gul and Pesendorfer's critique of neuroeconomics, can it be concluded that the issue they raise can easily be avoided? Not necessarily. From a strictly epistemological viewpoint, the problem remains unresolved. It is a fact that traditional economic choice models do not explain the physiological or neurological mechanisms underlying the decision-making process. *Linking propositions* (in the sense of Teller & Pugh, 1983) are lacking between economic choice models and cognitive or neural data; we are faced with a problem of "missing links" (Cozic, 2012, p. 141). The main argument that pleads in favor of Gul and Pesendorfer's view is the fact that the only variables explicitly taken into account in current economic choice models are actually of a behavioral nature. They seldom include observable variables other than those. Of course, this does not exclude that some results may be inferred from the models given a few plausible general assumptions concerning cognition. But "it seems difficult to find general and precise results from choice models as regards cognitive data" (Cozic, 2012, p. 145 / translated from the French). This is doubtless one of the main reasons for the misgivings expressed some years ago by some economists concerning the epistemological relevance of neuroeconomics (e.g., Bernheim, 2009; Harrison, 2008a; Rubinstein, 2008[20]).

In this regard, some lines of reflection have been explored to suggest the existence of linking hypotheses between economic choice models and cognitive data and to try to give some "confirmation" relevance to cognitive data. This is in particular the case with Quartz (2008), who distinguishes between the "implicit" and "explicit" cognitive consequences of a choice model[21], or Cozic (2012), who prefers to see in some "plausible auxiliary hypotheses" added to the choice models their link to cognitive data. But then, as he acknowledges, the confirmation links become contingent: they depend on factors extrinsic to the model (i.e., the auxiliary hypotheses).

Exploring the typical case of a neuroeconomic experiment using neuroimaging in the case of a binary choice in an uncertain situation, Cozic shows that when a psycho-cerebral auxiliary hypothesis H is introduced, according to which the brain activity in neuronal area A encodes the individual's preferences (a reasonable hypothesis as it stems from another constitutive hypothesis of the elementary choice model: the individual chooses what they prefer), one obtains observable consequences relating to the neural data that are capable of *confirming* the elementary choice model (and the auxiliary hypothesis): the preferred option is reflected in the neural signal. On the other hand, the *infirmation* of the model is not clear: should we reject the hypothesis that the

---

[20] This difficulty in establishing "linking propositions" between the different relevant levels in the analysis is not unique to neuroeconomics. It is the case in all cognitive sciences. Hence, the epistemological issue of the connection between cerebral and mental levels, inter alia, remains relevant in cognitive psychology (see in particular Henson, 2005).

[21] He developed this argumentation to support the idea that the results of some financial choice experiments "might confirm" the mean-variance model because they "validate" some of their "implicit" cognitive outcomes (see Serra, 2016).

individual chooses the option they prefer or the auxiliary hypothesis H by which activation of area A encodes the preference? The answer to this question is not trivial, except for supporters of revealed preference theory who will always reject hypothesis H rather than the preferential choice hypothesis. Otherwise, there may well be reasons to preserve hypothesis H and accept that the preferential choice hypothesis may be rejected. In epistemology there are indeed strong objections towards the revealed preference theory (see e.g., Hausman, 2008).

The argumentation developed here is akin to a claim that the main difficulty raised by Gul and Pesendorfer (2008) would tend to disappear as long as the revealed preference concept is not propounded as a dogma. Fundamentally, the "mindless economics" postulate would only be an argument in defense of the standard economic methodology dating from the mid-20th century[22].

4. THE USEFULNESS OF THE RESULTS FOR ECONOMISTS?

Neuroeconomics helps to reveal the physiological and neural components and processes involved in economic decision-making. However, the neural foundations revealed by neuroeconomics tend to be already well known from a behavioral viewpoint. This is another reason for the "neuroskepticism" of some economists (e.g., Smith, 2008, chap. 14; Harrison, 2008a, 2008b; Rubinstein, 2008). Hence Harrison (2008a), when discussing the psychological phenomena highlighted by neuroeconomists, argued that "economists have known this for decades" (p. 308). If a phenomenon is already well understood at the psychological and behavioral levels, what might an economist gain from knowledge of the neural correlates and mechanisms, especially as the cost of certain neuroeconomic experiments (in particular neuroimaging) is disproportionate compared to behavioral economic experiments? Rubinstein (2008), in particular, developed this criticism and recommended restricting the use of neuroimaging tools, as their very high cost might lead researchers to overestimate their reliability, and to favor simple and cheap indicators drawn from psychology (oculometry, response time, electrodermal response, and so on)[23].

How valid is this argument which, in the end, amounts to claiming that neuroeconomic experiments bring little real beneficial economic information and are probably of little pragmatic interest? It now seems excessive, if not unjustified, at least if neuroeconomics is not just assimilated to "behavioral economics in the scanner" – the research program predominantly (although not exclusively)

---

[22] In this respect, the debate around the scientific legitimacy of neuroeconomics (and more generally behavioral economics) is reminiscent of the controversy between Gary Becker and Edward Wilson during the emergence of socio-biology in the 1970s (see in particular Wilson, 1975, and Becker, 1976).

[23] Yet, according to Rubinstein, neuroeconomics will probably influence economics because economics is a culture and not a science, by conceiving "culture" as a collection of accepted ideas and conventions that are used in our thinking.

developed by economists – and if the focus is also on “neural economics” – the research program predominantly (although not exclusively) proposed by some neuroscientifics.

It turns out that two main directions can be originally identified in neuroeconomics. In one perspective, it is presented as a transdisciplinary field using neuroscientific tools to investigate the neural bases of economic decisions. It aims simply to complement the behavioral economic approach by looking closely at the biological origins of assumptions concerning the psychological characteristics of the individuals in the models of economic phenomena proposed by this approach. In a sense, neuroscience may be considered as “serving” economics (McCabe, 2003; Zak, 2004; Camerer, Loewenstein, & Prelec, 2005; Sanfey et al., 2006; Camerer, 2007). The philosopher Don Ross, who called this work “behavioral economics in the scanner”, identified another branch that he called “neurocellular economics” (Ross, 2005, 2008). In this work, microeconomic concepts and mathematics relating to optimization and equilibrium are used to model brain cell activity and test some hypotheses on 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. In other words, these researchers reject the idea of anthropomorphism as an initial necessary condition for neuroeconomics. Thus, neurocellular economics, which we prefer calling “neural economics” after Montague & Berns (2002), borrows the conceptual arsenal of economic theory for understanding neural mechanisms through which the brain is valuing and comparing the multiple stimuli it is subjected to in decision-making. In a way, the project is in contrast to the goal of behavioral economics in the scanner; rather than seeking to improve economic theory by borrowing tools from neuroscience, the purpose is to use standard economic theory for analyzing brain functioning (Glimcher, 2003; Montague, 2007). This time, it is economics that may be considered as “serving” neuroscience (on the identification of this now widely-recognized distinction refer also to Vromen, 2007, 2011). In their introduction to the first edition of the *Handbook of neuroeconomics* (Glimcher et al., 2009), the editors do not hide this internal division within the neuroeconomist community. This distinction is useful, not least because some criticisms of neuroeconomics by economists mainly target the first program. However, it is gradually losing its relevance in the light of recent advances in research, and it appears that these two important historical camps of neuroeconomics are converging towards the adoption of a research agenda common to a large part of the scientific community (see Serra, 2016).

Let’s briefly outline beforehand the arguments traditionally put forward by certain economists in favor of neuroeconomics.

4.1. Why might economists need neuroeconomics?

It is a fact that over a few decade behavioral economics has managed to produce a number of alternatives to the standard economic models, whether in individual or in interactive choices. Such success is due to both theoretical advances and the use of experiments to generate data. In these circumstances, why would behavioral economics need neuroeconomics as a “complement”?

A first answer may be found in the scope and power of the new investigative tools now available to neuroscience specialists. These tools are expected to refine the psychological interpretations that the behavioral economic models provide of observed behavior (Camerer, Loewenstein, & Prelec, 2005). This stance, expressed by three enthusiastic advocates of the neuroeconomic approach, appears to be shared by most participating economists as well as by some leading economists who are less involved in this new scientific approach such as Jean Tirole, the 2014 Nobel Prize in economics, who sees in neuroeconomics a tool designed to “finetune the models” (Tirole, 2007).

Thanks to the use of these new tools, neuroeconomics may help to ensure that behavioral economic models “are on the right track” verifying, for example, whether distinct variables correspond to distinct components or processes. It may confirm the explanations provided by behavioral economics concerning the gaps between experimental results and standard economic theory (Camerer, 2008).

But another response may be put forward concerning the perverse effects of the relationship between the observation and the observer mechanism in laboratory economic experiments, such as the “experimenter effect”, the “reputation effect” and the “volunteer bias”. This is another reason that is often put forward by some economists to justify using neuroscientific tools in their experiments. By direct investigation on the brain, i.e., on the cerebral activities underlying behavior, neuroscience tools enable far more objective analyses, unaffected by the unwanted effects that may be involved when a person is being studied (Camerer, Loewenstein, & Prelec, 2004; Kenning & Plassman, 2005).

A final strong argument in favor of neuroeconomics cannot be ignored; it concerns the very foundations of rational choice in economics. Thanks to neuroeconomics, the assumption that the choice results from maximizing the value attributed by the agent to the various possible options in the decision-making process becomes potentially falsifiable and therefore genuinely scientific. The idea of the subjective value of a decision is a central concept in neuroeconomics as it is in economics. But there is a crucial difference. In economics, the concept of value is behavioral and analytical, not psychological. There is a fundamental limit to the "as if" approach: on the basis of behavior alone, values cannot be measured independently of the choice. The assertion that "the choice maximizes the value" is circular. Neuroeconomics breaks this circular reasoning by showing that values are actually computed by the brain. By establishing a correspondence between a neural signal and a behavioral measurement of the value of a decision, the signal observed in the brain through neuroimaging provides an independent measure of the value that may in principle be dissociated from the choice (Padoa-Schioppa, 2011; Padoa-Schioppa & Conen, 2017).

4.2. Overall assessment of the critique

Whether convinced or not by these arguments in favor of neuroeconomics, how should the criticism that neuroeconomic experiments add nothing essential to the knowledge of economic behavior be addressed? In order to answer this question, the results of these experiments and the lessons that can be drawn from them must be examined. It is not the intention to review in detail here all the lessons learned so far from neuroeconomics (for more extended overviews see Glimcher et al., 2009; Glimcher & Fehr, 2014; Serra, 2016; Reuter and Montag, 2016). The aim is only to evaluate the criticisms on the basis of the most significant findings, distinguishing between the two historical research programs, "behavioral economics in the scanner" and "neural economics", including recent developments towards a "neural and behavioral economics" program.

4.2.1. Behavioral economics in the scanner

It must immediately be acknowledged that the criticism has some validity in the case of the first-generation experiments in the behavioral-economics-in-the-scanner program. The results of these experiments merely corroborated in terms of biological markers the explanations of behavioral economics in terms of mental states and processes in certain contexts. As noted by Vernon Smith a few years ago: "*initially new tools tend naturally to be applied to the old questions*" (Smith, 2008, p. 313). Hence, prospect theory, as a main alternative to expected utility theory in behavioral economics (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), became the subject of numerous neuroeconomic experiments that confirmed some of its central hypotheses such as loss aversion and non-linearity of the probability weighting function, but reference-dependence in decision making and framing are currently unclear (see Louie & De Martino, 2014; Fox & Poldrack 2014)[24]. Likewise, many explanations of a psychological nature put forward in behavioral economics

---

[24] Kahneman was himself 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, the experiment was intended to relate to reward learning theory; yet it is clear that prospect theory provides, there also, the

so as to understand the so-called “anomalies” or “irrationalities” observed in situations of social interaction (voluntary participation in collective action, help given to perfect strangers, costly punishment for those who infringe social norms…) were confirmed by numerous neuroeconomic experiments. Social neuroeconomics helps further define the emotional content of “other-regarding” preferences as an explanatory factor. In so doing, it clarifies the neural bases of the motivations that behavioral economics generally attributes to participants to explain their cooperative behavior such that trust, altruism, reciprocity, empathy, generosity, guilt or concern for equity (overviews can be found in Sanfey & Rilling, 2011; Fehr & Krajbich, 2014; Declerck & Boone, 2016; Serra, 2016; Dreher & Tremblay, 2017).

That being said, this analysis must be nuanced. It may be objected that not only do these hypotheses gain a stronger *objective* justification but also some of these experiments are beginning to shed new light on interesting issues for economists in many areas. By revealing the neural bases of specific choice elements in areas as varied as desire, regret, anticipation, and risk, ambiguity or time-attitude, the behavioral-economics-in-the-scanner program contributes to improving understanding of the motivations specific to individual decisions.

Some studies, for example, explore the difference between disappointment and regret. Regret generates more intense physiological responses than disappointment. More precisely, the two emotions appear to be cumulative: one may be both saddened by the obtained outcome and sad to have made the wrong choice (Camille et al., 2004; Coricelli et al., 2005). Other experiments provide more detailed explanations than behavioral economics on loss aversion or ambiguity aversion, two common emotions frequently observed in financial decisions in particular. Loss aversion is explained by the interaction between the neural structures involved in the anticipation, encoding and computation of the hedonic effect of a risky decision. We express loss aversion because we anticipate the negative emotional impact of the loss. Neuroeconomics also casts an original light on ambiguity aversion. In many experimental settings, participants show a strong preference for outcomes that are risky (i.e., with known probabilities) rather than ambiguous (i.e., with unknown probabilities). Although decision theory treats ambiguity simply as a specific case of risky decision-making, several neuroeconomic studies have shown that neural-mechanism supports for risky and ambiguous decisions are actually distinct. We express ambiguity aversion because we anticipate a

---

experimental paradigm. Neural data is consecutively considered to provide parameters of the Kahneman-Tversky model (*exogenous* variables that must be evaluated in order to “calibrate” the model) and 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”, i.e., the disparities between “ideal” economic rationality and observed behavior, into the theoretical framework of reward learning. In neurobiology, irrational behavior is assessed by reference to the learning dynamic (Fox & Poldrack, 2014).

more intense negative emotion when faced with an ambiguous situation rather than a risky one (Hsu et al., 2005; Huettel et al., 2006; Taya, 2012).

Generally speaking, neuroeconomic experiments deal with a lot of various economic issues including day-to-day purchasing decisions, financial choices (where attitudes to risk and uncertainty play a significant role), intertemporal choices (where psychological discounting of the future and impulsivity come into play), and interpersonal choices and social decisions (where social cognition and the forces of social/moral/cultural norms affect relationships with others). Despite the very specific features of each of these decision classes, some qualitative regularities are already detectable in the experimental results. In the fields of both interactive and individual decisions, it is now accepted that emotions and affect are fully involved in economic decision-making, linking closely with more sophisticated deliberation and evaluation cognitive processes. This lesson is unanimously accepted: highlighting the role of emotions in economic decision-making is a common denominator of many neuroeconomic studies in the behavioral-economics-in-the-scanner program.

This broad lesson however may appear in different ways: sometimes emotions are presented as primordial to "rational" decision-making (in line with the findings of Antonio Damasio and colleagues in the 1990s: Bechara, Damasio et al*.,* 1994, 1997), sometimes they are considered to intervene in close cooperation with cognitive control and evaluation processes (which explains in particular some apparently irrational behavior, in agreement with certain behavioral economic models), and sometimes emotions appear to compete with the cognitive process of decision-making as the brain arbitrates, so to speak, between reason and emotion (particularly in intertemporal decisions)[25]. Furthermore, experiments on interpersonal choices, in particular those designed to explain the prosocial nature of some observed behavior, confirm that social cognition systems are indeed present in cerebral mechanisms engaged especially in mentalizing and in the mirror neuron system as components of empathy[26] (see in particular Rizzolatti & Siniglia, 2006; Singer & Tusche, 2014; Serra, 2016; Alos-Ferrer, 2018).

---

[25] Results from-the behavioral-economics-in-the scanner program are generally interpreted, even now, in a single "dualistic perspective" opposing "instinctive, emotional" choices to "reasoned, cognitive" choices. The less radical interpretation that we adopt here appears more in keeping with the diversity of results obtained by these studies.

[26] In a somewhat reductive manner, we can acknowledge that social cognition in human brain is governed by 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) (see e.g., Singer & Tusche, 2014). But it is possible to make more explicit relationships between empathy, mentalizing, and the mirror system, suggesting to separate empathy into two distinct dimensions: i) "cognitive" empathy (*understanding* of other people's feelings) specifically in relation to the mentalizing system, and ii) "affective" empathy (*sharing* of other people's feelings) specifically related to the mirror system. Yet, during an empathic process, generally, the two systems are likely to maintain cooperative interactions (see e.g, Oliver et al. 2018).

It is also the case that some neuroeconomic experiments demonstrate the constraints that particular brain activity imposes on people's strategic reasoning. Hence, it was shown that the human brain is unable to exceed two or three levels of iteration in a common reasoning process: brain function itself restricts the "depth" of strategic reasoning. The experiment of Coricelli & Nagel (2009) using the "beauty contest" as a behavioral paradigm fully illustrates this result[27]. Further information on strategic reasoning in neuroeconomic experiments are provided by Camerer & Hare (2014).

4.2.2. Neural economics

Since the first cognitive science studies in the early 1950s, the brain has popularly been compared to a computer. This comparison, although appropriate in many ways, disregards one critical functional constraint of the brain – it is a computer specially designed for providing the motivation required for the appropriate behavior to preserve both the individual and the species (search for food, reproduction, hazard avoidance…) (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.

Valuation and learning

Contemporary neuroscience research showed that the cerebral reward 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 world, survival and reproduction determines 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).

It should be pointed that this conception of the reward system reflects a radical paradigm change in neuroscience in the 1990s: actions are no longer considered to be conditioned only by reflexes (Sherrington's studies dating from the beginning of the 20th century) but also and more specifically by motivations and intentions (Glimcher, 2003).

---

[27] This neuroeconomic experiment is in line with a number of older cognitive psychology experiments which testify to the brain's cognitive limits. For example, in Miller (1956), it is shown that the human brain can, at best, only deal with seven data points at once. The "mental accounting" concept introduced in behavioral economics to convey the mental simplifications operated by the brain to deal with complex choices (see in particular Thaler,1999) also finds an objective justification in this limited information-processing capacity of the prefrontal cortex.

To date, at least 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.

For several years, most neurobiologists suspect that dopamine (one of the main neurotransmitters that send chemical signals from one neuron to the next) is playing a decisive role in the reward system. The most frequent hypothesis was that dopaminergic neurons were the "pleasure center of the brain". In the 1990s, it was challenged by a set of experiments revealing the role that expectations were playing in modulation of dopamine activity too. Dopaminergic neurons 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). It is that "reward prediction error" (RPE) – difference between expected and perceived rewards – which is used to continuously update value that the brain is attaching to stimuli (Montague et al., 1996; Schultz, Dayan, & Montague, 1997) (see Daw & Tobler, 2014, for a more formal presentation).

These pioneer experiments revealing the role of dopamine in reward system were carried out in non-human primates. Human studies have long been rare because it is difficult for technical reason to scan by fMRI the target regions for dopamine in the midbrain (see however McClure, Berns, & Montague, 2003, and O'Doherty et al., 2003). Today we have a lot of studies proving the role of dopamine in the encoding of positive RPEs in humans after the pioneering works of Pessiglione et al. (2006) (the first pharmacological manipulation study proving a causal role for dopamine), d'Ardenne et al. (2008) (an experiment which adopts an original protocol combining EGC data with fMRI images), and Zaghloul et al. (2009) (the first electrophysiology study on human during deep brain stimulation therapy in patients with Parkinson's disease).

Remark that the first studies on the role of dopamine in reward learning in non-human primates date from the 1990s, i.e., before the paper written by Platt & Glimcher (1999) that generally is accepted as the first neuroeconomic study. So, we may consider them as a kind of "theoretical prehistory" of neural economics. The research of Platt and Glimcher indeed is the first experiment 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. Thank to this new experimental tool of electrophysiology, the neural activity of conscious animals can be recorded before they made their choice. As the observation precedes the decision, it becomes possible to understand the mechanism of the choice, i.e., to understand how the monkey, conditioned to make eye movements, constructs and gives a subjective value to each option. And overall, even if some cytoarchitectonic differences can be observed between the monkey and human brains (e.g., Mackey & Petrides, 2014; Wallis, 2012), it turns out that *brain processes* appear to be largely similar.

Broadly, the neural economics program is associated at least with three main ideas: (i) the brain "valuates" multiple stimuli in terms of rewards/punishments independently of sensory or motor systems (Platt & Glimcher, 1999; Padoa-Schioppa & Assad, 2006; Plassman, O'Doherty, & Rangel, 2007); (ii) there is a "common currency" that allows the brain to compare on a single scale the subjective values of these multiple stimuli that are *a priori* not comparable; the release of dopamine appears to be one of the main components of a "centralized" neural mechanism that resolves comparability problems (Montague & Berns, 2002; Landreh & Bickle, 2008; Schultz, 2009 ; Balleine, Daw, & O'Doherty, 2009; Rangel, 2009; Chib et al*.,* 2009; Levy & Glimcher, 2012); (iii) uncertainty is fundamental to an understanding of brain functioning (Glimcher, 2003; Glimcher, Dorris, & Bayer, 2005; Faisal, Selen, & Wolpert, 2008; Rolls & Deco, 2010; Glimcher, 2011; Fehr & Rangel, 2011; Rolls, 2014); faced with uncertainty, the brain undergoes continuous "updates" through various learning mechanisms that are beginning to be identified, including Pavlovian, habitual and goal-directed systems (Bouton, 2007; Niv & Montague, 2009; Balleine, Daw, & O'Doherty, 2009; Balleine & O'Doherty, 2010, 2014; Rolls, 2014; Daw & Tobler, 2014). The central lesson of neural economics is that these three interdependent properties are essential in understanding the neural mechanisms that govern decision-making.

Ultimately, rather than borrowing models from behavioral economics in order to improve them, neural economic studies aim to develop a genuine neural theory of behavioral learning. In so doing, neuroeconomics is distancing itself from behavioral economics. "Anomalies" or "biases" in particular no longer have the same meaning: certain behavior is no longer considered "irrational" with reference to an "ideal" economic rationality but instead by reference to an "optimal" neural learning process. At the same time, yet, we can argue that the nature essentially brain-centric of these works appears rather unsatisfactory in an economic perspective. Findings of neural economics – at least initially – could appear not very useful for economists. In fact, neural economics incorporates ideas, concepts and method from economic theory without genuine reciprocation (Gul & Pesendorfer, 2008; Harrison, 2008a). We can pretend however that the recent direction taken by neuroeconomic research – which could be seen as the emergence of a "neural-and-behavioral-economics" research program – should validate the specificity of neuroeconomics in its capacity to propose original ideas to economists (Vromen, 2011; Serra, 2016).

4.2.3. Neural and behavioral economics

The common ambition of many neuroeconomists is now to develop a computational model of choice that would integrate the constraints applied to the economic decision-making process by the brain function. Driven by the theory of revealed preference, economics traditionally had not been interested in the neural process underlying human choice. As a result, standard economic models are "as if" as opposed to "as is" descriptions of decision making. In contrast, neuroeconomists are interested in the actual computational and neurobiological processes behind human behavior. Neuroeconomics aims for "structural" models of decision making (Glimcher, 2011; Fehr & Rangel, 2011).

The distinguishing feature of the model emerging today in neuroeconomic literature – that may be designed as "standard" in that it covers the various versions proposed– is that it integrates the numerous experimental results of behavioral economics in the scanner over the last fifteen years into the theoretical framework developed by neural economics.

Let's briefly outline the key components of this model. Broadly, two mechanisms are assumed to intervene sequentially in the brain during the decision-making process: (i) an valuation mechanism which is supposed to learn, store and retrieves the "values" given to the various options attached to the choice; (ii) a selection mechanism which uses the results of the valuation mechanism to generate by comparison the actual choice of one of the options. Moreover, both computation and comparison of decision values are modulated by attention (i.e., the brain's capacity to vary the computational resources that are used in different circumstances insofar as these resources are scares and costly in terms of consuming energy).

The valuation mechanism

The concept of the subjective value of a decision (in short, "decision value"), thus, is at the heart of the system and is connected with the notion of "experienced utility". Whilst the value of a decision coincides with a signal computed at the time of choice, experienced utility depends on a signal computed at the time of the actual experience of the consequences of the option selected (e.g., Kable & Glimcher, 2009; Rangel & Hare, 2010; Fehr & Rangel, 2011; Padoa-Schioppa, 2011; Rangel & Clithero, 2014; Berridge & O'Doherty, 2014; Glimcher, 2011; Glimcher, 2014). Of course, we cannot fail to draw a parallel between this distinction and the one often put forward in behavioral economics between "decision utility" (i.e., the weight of potential output in decisions) which corresponds to the concept of utility in standard economic theory, and "experienced utility" with its hedonic quality, symbolizing subjective well-being (e.g., Kahneman, Wakker, & Sarin, 1997; Kahneman, 2003; Kahneman, 2011; Brocas & Carillo, 2014)[28].

The existence of decision value signals at the time of choice was the single most frequently tested hypothesis, as well as the most systematically replicated finding, in neuroeconomic literature. Multiple human and nonhuman studies have shown that neural activity in areas of the ventromedial prefrontal cortex increases with behavioral measures of the decision values assigned to options at the time of choice[29]. If decision value corresponds to a signal computed at the time of choice (whether the option is chosen or not), experienced utility is supposed to represent a signal computed at the time of "consumption" that reflects the actual consequences for the organism on consuming the chosen option. Consequences of decisions have necessarily to leave a trace in brain. Otherwise, it would be unable to learn how to make a choice in the future. One of the key factors in the learning and storage process is the computation of this experienced utility signal. How these signals are used by the brain to update future decision values is still an exploratory issue. However, authors agree that the reward system and reinforcement learning processes mentioned above are likely taking a central part by favoring the building of a stimuli valuation common scale and by feeding the memorization system that keeps track relevant environmental information collected and continuously updated. Where in the brain are the experienced utility signals computed? Evidence here is more preliminary than for the case of decision values. Nonetheless, several studies have found that such signals are present in various parts of the brain, including the orbitofrontal cortex (OFC) and nucleus accumbens (in the ventral striatum) at the time of "consuming" a variety of primary rewards (foods, liquids, music, works of art…) and secondary rewards (monetary gains) during short time courses, and in the dorsolateral prefrontal cortex, which provides a complementary function to the OFC in value building, for the valuation over generally

---

[28] Others have highlighted the importance of the utility related to anticipating a positive or negative outcome (e.g., Loewenstein, 1987, referred to as "anticipation utility"). Other distinctions are developed in the literature: for instance, Bossaerts, Preuschoff, & Hsu (2009) mention "true" preferences (what individuals want) and "revealed" preferences (what individuals do). 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" (p. 242).

[29] Several researchers add the ventral striatum to the ventromedial prefrontal cortex as contributing to value decision (Knutson et al., 2009; Glimcher, 2011; Levy & Glimcher, 2012). But this is still a controversial hypothesis, at least for humans (Padoa-Schioppa, 2011; Rangel & Clithero, 2014).

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 (for further detail see Glimcher, 2011, 2014; Lempert & Phelps, 2014; Serra, 2016, 2019). Some properties of the experienced utility also are found: for instance, it is shown that experience utility depends not only on what is consumed but also on the extent to which that consumption was expected (positive surprises increase experienced utility while negative surprises decrease it).

The selection mechanism

The second phase taking place in decision making corresponds to the selection of one particular option. Once the value of the different options has been computed, the "best" one has to be selected by comparing them. The topic of which neural structures are taking place in this process of choice implementation still remains unclear. Some neuroeconomists suggest however that a frontoparietal network, including the dorsomedial prefrontal cortex and parietal cortex, is involved (Kable & Glimcher, 2009; Hare et al., 2011; Glimcher, 2011, 2014).

Concerning the mechanism of choice implementation, an increasing consensus in the neuroeconomic literature favors one of the two basic models commonly developed in theoretical neuroscience, the "drift-diffusion" model. The "winner-take-all" algorithm, the neural equivalent of the "arg max" function in the standard static economic model (i.e., maximization of utility) seems to be supplanted by the more direct "drift-diffusion" algorithm: a decision is made once the activation rate of the relevant neurons exceed a critical threshold (Glimcher, 2011, 2014; Fehr & Rangel, 2011; Wang, 2014). How is this threshold determined? It should be function of the more or less complex nature of decisions. The brain should engage in a trade-off between speed and accuracy: a lower threshold value entails a much simpler choice but with the risk of not choosing the best option whereas with a high value, the choice becomes more difficult and takes longer but the risk of error tends to be lower (Roitman & Shadlen, 2002; Palmer, Huk, & Shadlen, 2005)[30]. The components of the drift-diffusion model have empirically been tested using both behavioral and neural data but authors recognize that additional research is necessary for precisely delimiting the field in which the model can be useful (Krajbich, Oud, & Fehr, 2014). One of the key lessons of the computational model of economic choice is that mistakes are fundamentally part and parcel of economic choices: an optimal decision is to choose the option associated with the largest experienced utility signal at the moment of consumption; if not, the individual makes a decision error.

---

[30] The literature on drift-diffusion model is extensive. See the survey of Ratcliff et al. (2016).

Of course, this model remains largely exploratory. It applies mainly to goal-directed decisions and only concerns “simple” choices (i.e., with a restricted number of common goods, without information asymmetry, strategic considerations, self-control problems, financial, temporal or social dimensions). The authors hope that this model might constitute a satisfactory testing ground for the systematic examination of economic issues and that some of its main computational and neurobiological properties might be preserved in more complex choices. The first available results on risky choices (Levy et al*.,* 2010) and on intertemporal choices (Kable & Glimcher, 2009; Fehr & Rangel, 2011; Kable, 2014) suggest that this hypothesis may be reasonable. As regards social decisions involving a neural representation of oneself and others, things seem to be more complicated (Hare et al*.,* 2010). However, recent studies suggest the existence of a likely unified mechanism for motivational control of behavior that may incorporate social and non-social factors even though some aspects of the results tend to indicate that there are differences between social and monetary neural values (Ruff & Fehr, 2014; Hutcherson, Bushong, & Rangel., 2015; Krajbich et al., 2015; Wake & Izuma, 2017).

5. CONCLUSION

Neuroeconomics is a very recent discipline; two decades at the most. As with any emerging discipline – consider for example mathematical economy, econometrics or experimental economics at different times – it is rather controversial with economists. It has attracted strong criticism from several authors but quarrels between experts tend to be somewhat opaque for the non-initiated. We tried in this paper to clarify the topic. Neuroeconomics raises many issues that the paper has organized around three major questions about the results produced so far: Are they reliable? Are they scientifically-legitimate? Are they useful for economists? Without claiming to be exhaustive, this analytical framework enables understanding of the thrust of the major criticisms of neuroeconomics and at the same time the nature of the likely responses.

As concern the reliability of results and criticisms focusing primarily on technical and limits of neuroimaging – the main tool used so far in this new field–, we agree that there is clearly a need to encourage the use of a wider range of neuroscientific tools and methods in experiments. Some neuroeconomists have already begun to move in this direction with studies employing multivariate machine learning technique for decoding brain images, using brain stimulation and/or pharmacological interventions in experiments. With regard to criticisms that have argued that neuroscience and economics are scientifically incompatible, we can purport that such a charge ultimately fades away once the doctrine of revealed preference is no longer accepted as a dogma.

There remains the question of the usefulness of this new scientific field for economists. Following criticism that neuroeconomics was of no real use to economics given the contributions of behavioral economics (such criticism focused essentially on the behavioral-economics-in-the-scanner program), it was then argued by some economists that neuroeconomics is essentially a brain-centric enterprise that incorporates ideas from economics without real reciprocation (such criticism targeted the neural economics program). All this is true, no doubt, but excessive. First, even though early work in the behavioral-economics-in-the-scanner program merely corroborated the explanations of behavioural economics, results have acquired a stronger objective justification in this way. Then, and above all, the recent evolution of neuroeconomics towards a “neural and behavioral economics”, which incorporates the numerous experimental results of behavioral economics in the scanner into the theoretical framework developed by neural economics, shows that neuroscience should bring benefits to economics similar to those it has received from economics. By opening the “black box” behind economic behavior, neuroeconomics provides foundations for more comprehensive and powerful economic models and should allow economists to answer fundamental questions that are unable to address so far.

REFERENCES

Allos-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.

Aydinonat, N.M. (2010). Neuroeconomics: More than inspiration, less than revolution. *Journal of Economic Methodology*, *17*(2), 159-169.

Aydogan, G., Furtner, N.C., Kern, B., Jobst, A., Müller, N., and Kocher, M.G. (2017). Oxytocin promotes altruistic punishment. *Social Cognitive and Affective Neuroscience, 12 (*11*),* 1540-1747.

Balleine, B.W, Daw, N.D., and O'Doherty, J.P. (2009). Multiple forms of value learning and the function of dopamine. In P.W. Glimcher, C.F. Camerer, E. Fehr, & R.A. Poldrack (Eds.) (2009), *Neuroeconomics. 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. *Neuropsychopharmacolog*y, *35*, 48-69.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., & Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton: Princeton University Press.

Barraza, J.A., McCullogh, M.E., Ahmadi, S., and Zak, P.J. (2011). Oxytocin infusion increases charitable donations regardless of monetary resources. *Hormones and Behavior*, *60* (2), 148-151.

Barraza, J.A. & Zak, P.J. (2009). Empathy toward stranger triggers oxytocin release and subsequent generosity. *Annals of the New York Academy of Science*, *1167*, 182-189.

Baumgartner, T., Gianotti, L.R.R., & Knoch, D. (2013). Who is honest and why: Baseline activation in anterior insula predicts inter-individual differences in deceptive behavior. *Biological Psychology*, *94*, 192-197.

Baumgartner; T., Heinrichs, T., 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, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, *14* (11), 1468-1474.

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

Becker, G. (1976). Altruism, egoism, and genetic fitness: economics and sociobiology. *Journal of Economic Literature*, *14*(3), 817-826.

Bernheim, B.D. (2009). On the potential of neuroeconomics: A critical (but hopeful) appraisal. *American Economic Journal: Microeconomics*, *1*(2), 1-41.

Berridge, K.C. and 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.

Bickart, K.C., Wright, C.I., Dautoff, R.J., Dickerson, B.C., & Barrett, L.F. (2012). Amygdala volume and social network size in humans. *Nature Neuroscience*, *14*, 163-164).

Bird, A. (1998). *Philosophy of science*. Montreal & Kingston: McGill-Queen's University Press.

Bjork, J.M., Momenan, R., & Hommer, D.W. (2009). Delay discounting correlates with proportional lateral frontal cortex volumes. *Biological Psychiatry*, *65* (8), 710-713.

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.

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.

Brocas, I. & Carillo, J.D. (2014). Dual-process theories of decision-making: A selective survey. *Journal of Economic Psychology, 41*, 45-54.

Bzdok, D. & Yeo, B.T.T. (2017). Inference in the age of big data: Future perspectives on neuroscience, *NeuroImage*, *155*, 549-564.

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

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

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., 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.

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.

Camus, M., Shimojo, S., Camerer, C.F., O'Doherty, J.P., & Rangel, A. (2009). rTMS of right dorsolateral prefrontal cortex disrupts the computation of goal-directed values. *European Journal of Neuroscience*, *30*, 1980-1988.

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. & Glimcher, P. (2014). Basic methods from neoclassical economics. In P.W. Glimcher & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 3-17). Amsterdam: Elsevier.

Caplin, A. & Schotter, A (Eds.) (2008). *The foundation of positive and normative economics: A handbook*. New York: Oxford University Press.

Chalmers, A.F. (1999). *What is this thing called science?* Maidenhead and New York: Open University Press.

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.

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

Coates, J.M. & Herbert, J. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proceedings of the National Academy of Sciences U.S.A, 105*, 6167-6172.

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. & 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.

Corrado, G.S. & Doya, K. (2007). Understanding neural coding through the model-based analysis of decision making. *Journal of Neuroscience*, *27*, 8178-8180.

Corrado, G.S., Sugrue, L.P., Brown, J.R., & Newsome, W.T. (2009). The trouble with choice: Studying decision variables in the brain. In P.W. Glimcher, C.F. Camerer, E. Fehr, & R.A. Poldrack (Eds.), *Neuroeconomics. Decision making and the brain* (pp. 463-480). Amsterdam: Elsevier.

Cozic, M. (2012). Economie 'sans esprit' et données cognitives. *Revue de philosophie économique*, *13* (1), 127-153.

Craig, A.D. (2009). How do you feel now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*: 59-70.

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

Crockett, M.J., Clark, L., Tabibnia, G., Liebermann, M.D., & Robbins, T.W. (2008). Serotonin modulates behavioural reactions to unfairness. *Science, 320*: 1739.

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.

Damasio, A.R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: A. Grosset/Putnam 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.

Daw; N.D. & O'Doherty, J.P. (2014). Multiple systems for value learning. In P. 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, E. & Cohen, L.G. (2011). Neuroplasticity subserving motor skill learning. *Neuron*, 72, 443-454.

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. 162-203). Oxford, UK.:Oxford University Press.

Declerck, C. & Boone, C. (2016). *Neuroeconomics of prosocial behaviour. The compassionate egoist*. London: Elsevier.

De Dreu, K. (2012a). Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion. *Psychoneuroendocrinology*, *37*, 871-880.

De Dreu, K. (2012b). Oxytocin modulates cooperation within and competition between groups: An integrative review and research agenda. *Hormones and Behavior*, *61*, 419-428.

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

D'Esposito, M, Ballard, D. Aguirre, G.K., & Zarahn, E. (1998). Human prefrontal cortex is not specific for working memory: A functional MRI study. *NeuroImage*, *8*, 274-282.

Diamond, P. & Vartiainen, H. (2012). *Behavioral economics and applications*. Princeton University Press.

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

Economics and Philosophy (2008) *24* (3), Special issue on Neuroeconomics.

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

Eisenegger, C., Haushofer, J., & Fehr, E. (2011). The role of testosterone in social interaction. *Trends in Cognitive Sciences*, *15* (6), 263-271.

Evans, S.L., Dal Monte, O., Noble, P., & Averbeck, B.B. (2013). Intranasal oxytocin effects on social cognition: A critique. *Brain Research*, Science Direct (e-pub).

Faisal, A., Selen, L., & Wolpert, D. (2008). Noise in the nervous system. *Nature Reviews. Neuroscience*, *9*, 292-303.

Fann, K.T. (1970). *Peirce's theory of abduction*. The Hague: Martnus Nijhoff.

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). Neuroeconomics foundations of economic choices – Recent advances. *Journal of Economic Perspectives*, *25* (4), 3-30.

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.

Fioretti, M. & Marden, S. (2015). Suboptimal dishonesty: Rationality in the absence of strategic behaviour in honesty experiments. *Journal of Neuroscience*, *35* (5), 1817-1818.

Forest, D. (2014). *Neuroscepticisme*. Paris: Ithaque.

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.

Fudenberg, D. (2006). Advancing beyond advances in behavioural economics. *Journal of Economic Literature, 44*, 694-711.

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.

Gilaie-Dotan, S., Tymula, A., Cooper, N., Kable, J.W., Glimcher, P.W., Levy, I. (2014). Neuroanatomy predicts individual risk attitudes. Journal of Neuroscience, *34* (37), 12394-12401.

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

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

Glimcher, P.W. (2014). 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., Camerer, C.F., Fehr, E., & Poldrack, R.A. (Eds.) (2009). *Neuroeconomics. Decision making and the brain.* 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.) (2014). *Neuroeconomics. Decision making and the brain* (2nd ed.). Amsterdam: Elsevier.

Goldman, A.I. (1988). *Epistemology and cognition*. Harvard University Press.

Gospic, K., Mohlin, E., Fransson, P., Petrovic, P., Johannesson, M., & Ingvar, M. (2011). Limbic justice – Amygdala involvement in immediate rejection in the ultimatum game. *PLoS Biology*, *9* (5), 1-8

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

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., 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.

Hausman, D.M. (1992). *The inexact and separate science of economics*. Cambridge University Press.

Hausman, D.M. (2008). The mindless or mindful economics: A methodological evaluation. In A. Caplin & A. Schotter (Eds.), *The foundation of positive and normative economics: A handbook* (pp. 125-151). New York: Oxford University Press.

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.

Heinrichs, M., von Dawans, B., & Domes, G. (2009). Oxytocin, vasopressin, and human social behavior. *Frontiers in Neuroendocrinology*, *30*, 548-557.

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologists. *Quarterly Journal of Experimental Psychology*, *A 58*, 193-233.

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.

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.

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.

Hutzler, F. (2013). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage*, *94,* 1061-1069.

Jordan, M.I. & Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*. *349* (6245), 255-260.

Journal of Economic Methodology (2010), *17* (2). Special issue. Neuroeconomics: Hype or Hope?

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

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. (2009).The neurobiology of decision: Consensus and controversy. *Neuron*, *63*, 733-745.

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

Kahneman, D. (2011), *Thinking, fast and slow*. London: FSG.

Kahneman, D. & Tvserky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*, 313-327.

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

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

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

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 diminishes the human ability to build a good reputation. *Proceedings of the National Academy of Science USA*, *106*(49), 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.

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 Behavioral Neurosciences*, *11* (December), Art 251, 1-11.

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

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

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

Lambert, B., Declerck, C.H., & Boone, C. (2014). Oxytocin does not make a face appear more trustworthy but improves the accuracy of trustworthiness judgments. *Psychoendocrinology, 40*, 60-68.

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

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. 173-192). Amsterdam: Elsevier.

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.

Lo, A.W & Repin, D.V. (2002). The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience*, *14*, 323-339.

Lo, A.W, Repin, D.V., & Steenbarger, B.N. (2005). Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review*, *95*(2), 352-359.

Lorentz, 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 Science USA*, *104*, 9494-9498.

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.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *Economic Journal*, *97*, 666-684.

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.

Mäki, U. (2003). 'The methodology of positive economics' (1953) does not give us *the* methodology of positive economics. *Journal of Economic Methodology*, *10*, 495-506.

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

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

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

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.

Miller, G.A (1956).The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

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., Berns, G.S., Cohen, J.D., McClure, S.M., & Pagnoni, G. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage*, *16*, 1159-1164.

Montague, P.R., Dayan, P., & Sejnowski, T.J. (1996). A framework from mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of neuroscience*, *16*, 1836-1947.

Morishima, Y., Schunk, D., Brunhin, A., Ruff, C.C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*, *75*, 73-79.

Münte, T.F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473-478.

Nagel, R., Brovelli, A., Heinemann, F., & Coricelli, G. (2018). Neural mechanisms mediating degrees of strategic uncertainty. *Social Cognitive and Affective Neuroscience*, *13*(1), 52-62.

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, *Trends in Cognitive Science*, *10*(9), 424-430.

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., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*, 35-53.

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, E.A., Collins, P.F., Hooper, C.J., Muetzel, R., Lim, K.O., & Luciana, M. (2009). White matter integrity delay discounting behaviour in 9-to 23- year-old: A diffusion tensor imaging study. *Journal of Cognitive Neuroscience*, *21*(7), 1406-1421.

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. & 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.

Peirce, C.S. (1878). Deduction, induction, and hypothesis. *Popular Science Monthly*, *13*: 470-482.

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.

Pfeiffer, U.J. (2013). Oxytocin not always a moral molecule. *Frontiers in Human Neuroscience*, *7*(10), 1-2.

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.

Plassman, H., O'Doherty, J., & 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.

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

Poldrack, RA. (2011). Inferring mental states from neuroimaging data: From reverse inference to largescale decoding. *Neuron*; 72(5), 692-697.

Poldrack, R.A. (2018). *The new mind reader. What neuroimaging can and cannot reveal about our thoughts*, 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 reproductible neuroimaging resource. *Nature Review Neuroscience*, *18*(2), 115-126.

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

Price, C.J. & Friston, K.J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*, 262-275.

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

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

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

Reuter, M. & Montag, C. (Eds.) (2016). *Neuroeconomics* (3rd ed.). Berlin: Springer-Verlag.

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, DeMarco, A.C., Hackett, P.D., Thompson, R., Ditzen, B., Patel, R., & Pagnoni, G. (2012). Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology*, *37*, 447-461.

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

Rizzolatti, G. & Sinigaglia, C. (2006). *So quel che fai. Il cervelloche agisce e i neuroni specchio*. Milano: Raffaello Cortina Editore (French translation: *Les neurones miroirs*. Paris : Odile Jacob, 2011).

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, UK: Oxford University Press.

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

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

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

Ross, D. (2011). Neuroeconomics and economic methodology. In J.B. Davis & D.W. Hands (Eds.), *Handbook of economic methodology* (pp. 61-93). Cheltenham: Edward Elgar.

Rubinstein, A. (2008), Comments on neuroeconomics. *Economics and Philosophy, 24*, 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 P. & E. Fehr (Eds.), *Neuroeconomics. Decision making and the brain* (2nd ed.) (pp. 77-108). Amsterdam: Elsevier.

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

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). New York: Psychology Press.

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

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., 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.

Serra, D. (2012). Principes méthodologiques et pratiques de l'économie expérimentale : une vue d'ensemble. *Revue de Philosophie Economique*, *13*(1), 21-78.

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

Serra, D. (2019). Neuroeconomics and modern neuroscience. CEE-M Working Paper 2019-12.

Shamay-Tsoory, S.G., Abu-Akel, A., Palgi, S. et al. (2013). Giving peace a chance: Oxytocin increases empathy to pain in the context of the Israeli-Palestinian conflict. *Psychoneuroendocrinology*, *38*(12), 3139-44.

Siegel, J.Z. & Crockett, M.J. (2013). How serotonin shapes moral behavior. *Annals of the New York Academy of Sciences, 1299*(1), 42-51.

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* (pp. 513-532). Amsterdam: Elsevier.

Sinnott-Armstrong, W., Roskies, A., Brown, T., & Murphy, E. (2008). Brain images as legal evidence. *Episteme: A Journal of Social Philosophy* (special issue: Evidence and Law), *5*(3), 359-373.

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

Smith V.L. (2010). Theory and experiment: What are the questions? *Journal of Economic Behavior and Organization*, Special issue on the methodology, *73*, 3-15.

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

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

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

Tirole, J. (2007). Interview. *La Recherche*, n° 414 (décembre).

Tverski, A. & Kakneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323.

Uttal, W. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge: MIT Press.

Van Rooij, M. & Van Orden, G. (2011). It's about space, it's about time, neuroeconomics and the brain sublime. *Journal of Economic Perspectives*, *25*(4), 31-56.

Vercoe, M. & Zak, P.J. (2010). Inductive modelling using causal studies in neuroeconomics: Brains and drugs. *Journal of Economic Methodology*, *17*(2), 133-146.

Vromen, J. (2007). Neuroeconomics as a natural extension of bioeconomics: The shifting scope as 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.

Vul, E., Harris, C., Winkielman, P., & Pshler H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Sciences*, *4*(3), 274-290.

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.

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

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

Walliser, B (2012). From psycho-economics to neuro-economics. *Recherches économiques de Louvain*, *78*(3), 13-28.

Wibral, M., Dohnen, T., Kingmüller, D., Weber, B., & Falk, A (2012). Testosterone administration reduces lying in men. *PLos ONE*, *7*(10), e46774.

Wilkinson, N. (2008). *An introduction to behavioral economics*, Palgrave Macmillan (3rd ed by Wilkinson N. & Klaes, M., 2018)

Wilson, E.O. (1975). *Sociobiology, the new synthesis*, Harvard University Press.

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 Science*, *359*(1451), 1737-1748.

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

Zak, P.J. (2012). *The moral molecule. The source of love and properity*. New York, NY: Dutton Press.

Zak, P.J., Borja, K., Matzner, W.T., & Kurzban, R. (2005a). The neuroeconomics of distrust: Sex differences in behavior and physiology. *American Economic Review*, *95*, 360-363.

Zak, P.J., Kurzban, R., & Matzner, W.T. (2004). The neurobiology of trust. *Annals of the New York Academy of Sciences*, *1032*, 224-227.

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

Zak, P.J., Kurzban, R., Ahmadi, S., Swerdloff, R.S., Park, J., et al. (2009). Testosterone administration decreases generosity in the ultimatum game. *PLoS ONE*, *4*, e8330.

Zak, P.J., Stanton, A., & Amadi, S. (2007). Oxytocin increases generosity in humans. *PLoS ONE*, *2*, e1128.

CEE-M Working Papers[1] - 2020

WP 2020 - 01	Francisco Cabo & **Mabel Tidball** « Cooperation in a dynamic setting with asymmetric environmental valuation and responsibility »
WP 2020 - 02	**Mathias Berthod** « Commitment and efficiency-inducing tax and subsidy scheme in the development of a clean technology »
WP 2020 - 03	Manon Authelet, **Julie Subervie**, Patrick Meyfroidt, Nigel Asquith & Driss Ezzine-de-Blas « Economic, Pro-social and Pro-environmental Factors Influencing Participation in an Incentive-based Conservation Program in Bolivia »
WP 2020 - 04	Lesly Cassin Paolo Melindi Ghidi & **Fabien Prieur** « Confronting climate change: Adaptation vs. migration strategies in Small Island Developing States »
WP 2020 - 05	Victor Champonnois, & **Katrin Erdlenbruch** « Willingness of households to reduce flood risk in southern France »
WP 2020 - 06	Sophie Clot, **Gilles Grolleau** & **Lisette Ibanez** « The Reference Point Bias in Judging Cheaters »
WP 2020 - 07	Jean-Marc Blazy, **Julie Subervie**, Jacky Paul, François Causeret, Loic Guindé, Sarah Moulla, Alban Thomas & Jorge Sierra « Ex ante assessment of the cost-effectiveness of Agri-Environmental Schemes promoting compost use to sequester carbon in soils in Guadeloupe»
WP 2020 - 08	Sophie Clot, Gilles grolleau & **Lisette Ibanez** « Projection bias in environmental attitudes and behavioral intentions »
WP 2020-09	**Annaïs Lamour** & **Julie Subervie** « Achieving Mitigation and Adaptation to Climate Change through Coffee Agroforestry: A Choice Experiment Study in Costa Rica »
WP 2020-10	**Douadia Bougherara, Margaux Lapierre, Raphaele Preget & Alexandre Sauquet** « Do farmers prefer increasing, decreasing, or stable payments in Agri-Environmental Schemes? »

---

[1] **CEE-M Working Papers / Contact** : laurent.garnier@inra.fr
- RePEc https://ideas.repec.org/s/hal/wpceem.html
- HAL https://halshs.archives-ouvertes.fr/CEE-M-WP/

WP 2020-11	**Gabriela Demarchi, Julie Subervie**, Thibault Catry & Isabelle Tritsch « Using publicly available remote sensing products to evaluate REDD+ projects in Brazil »
WP 2020-12	Ines Macho-Stadler, David Perez-Castrillo & **Nicolas Quérou** « Goal-oriented agents in a market »
WP 2020-13	Daniel Serra « Neuroeconomics: reliable, scientifically legitimate and useful knowledge for economists?»