"The role of cognitive control and emotions in decision-making: a neuroeconomics perspective"

Gabriele Chierchia[1,2] and Giorgio Coricelli[1,3]

[1]Center for Mind/Brain Science, University of Trento, [2]Max Planck Institute for Human Cognitive

and Brain Sciences, [3]Economics Department, University of Southern California

Abstract
One of the problems with treating cognitive control in economic decision-making is that there is a resilient idea that control makes behavior rational and that emotions make it irrational. The relatively dualistic or unitary nature of decision processes is critically compared from a neuroeconomics perspective in different well known behavioral settings, such as loss aversion, temporal discounting, ambiguity aversion, framing effects and strategic interactions. We claim that neither dualistic models – that stress the relative independence of sub-systems that can *independently* generate a decision - nor unitary models, capture some important aspects of how the brain processes decisions. Our interpretation of the neural data will be in line with a "weak" dual model approach. Such a model constitutes, arguably, an example of how decision theory might go beyond descriptive adequacy thanks to the balanced use of traditional modeling approaches and neuroscientific ones.

1. Introduction

Let us consider the following behaviors and try to understand what, if anything, differentiates them: i. Giving a small kick when one hits our knee (the osteo-patellar reflex); ii. A snail ceasing to retract itself to a touch when, through repetition, the touch has proved to be harmless (procedural learning in animals); iii. Scratching a mosquito bite; (less stereotyped reactions); iv. Someone with a compulsive disorder washing his hands (pathological behavior); v. A smoker lighting a cigarette (addictive behavior); vi. One taking his/her own somewhat "idiosyncratic" position when seating or sleeping (well-rooted routines); vii. Going to the door when it rings (more complex routines); viii. Quickly assessing how to circumnavigate moving cars and people in a crowded parking lot (visuo-motor decision-making); ix. Thinking of whether to have an easy microwave dinner or to a take time to cook a healthier meal (relatively simple decision-making); x. Reflecting over when to get married or have children (more complex decision-making); xi. Making market investments (deciding about the decisions of others).

Reconsidering such behaviors, we feel that the former (i.e. the osteo-patellar reflex or the snail's tail) require less (if any) decision-making or control, and that latter do. The problem is that it is very hard to say exactly when and how this switch occurs.

The same problem exists even *within* the same behaviors. For instance, as we write we make up our minds on which general ideas to use and this requires attention and effort. However, virtually in the same moment, the choice of specific words or phrasing, that is, the sub-processes of writing, require less deliberation; and the sub-processes of those sub-processes, i.e. finding the individual letters on the keyboard, emerge in an even more automatic fashion. Similar reasoning holds for a same action in different moments of time. For example, if we're learning a new language or sport, the demand of control will decrease as our proficiency increases.

It is commonly held that rational choice theory was born when economics was separated from psychology (Bruni&Sugden, 2007) The new definition of economics promoted by Vilfredo Pareto at the beginning of the 20th century was: "the science of logical action". Where by "logical" Pareto intended those actions that were "means to an end". Thought to be an anticipator of revealed preference theory, Pareto, wrote, "I'm not interested in the reason why man is indifferent between [one thing and another]: I notice the pure and naked fact". His was not to be a theory of sensation but of choice (Bruni&Sugden, 2007).

Nonetheless, Pareto believed that all not actions are logical. Behaviors such as "a man removing his hat whenever he enters a drawing room or a Catholic who regularly attends mass" were *not* to the object of economic analysis. It follows that the distinction between behaviors that are the result of routine or norms, from those that are properly logical, directed towards an end, constituted thus one of the cornerstones of rational choice theory. We are convinced that there is such thing as logical actions; however, our list of behaviors above aimed to show how gradually they emerge from a background of norms, routines, instincts and reflexes, thus how difficult that distinction actually is.

The fact that cognitive control (henceforth, CC) and decision processes can theoretically come to encompass, through a *continuum* of varying levels, virtually all behaviors (as Alex Kacelnik put it, "decision making is an approach to *behavior*, rather than a sub-category of behavior") (personal communication) makes revealed preference methodology inapt to circumscribing Pareto's aspired domain. The fact that even the same actions or behaviors may rely on different levels of deliberation makes even behavioral economics insufficient to break up the continuum.

In cognitive neuroscience cognitive control refers to our capacity to go beyond relatively "reflexive

reactions to salient stimuli in accordance to internal often extended and far-removed goals" (Miller, 2000) and one of cognitive neuroscientists' most challenging and yet open-ended goals is to see whether the brain shows evidence of the proposed distinctions, thus characterizing how decision making and CC emerge and are articulated throughout our actions.

We believe the economist should follow this progress with curiosity, as some believe it may help resolve some of the lingering problems of economics (Camerer et al., 2005).

We will proceed as follows. In section 1 we review some important models of CC in a part of the brain that has received particular attention in this regards, the prefrontal cortex (henceforth, PFC). Such models are not primarily inspired by economic paradigms rather by a psychophysiological tradition and should thus afford an unbiased general vision of how simple CC functions are implemented in the brain. In section 2 we attempt to draw the borders between broader opposing tendencies in the neuro-cognitive explanations of cognitive control, focusing on a dual vs. unitary framework opposition (Rustichini, 2008). In section 3 we pass to the neuroeconomic data and will compare the relatively dualistic or unitary nature of decision processes in different behavioral settings, such as loss aversion, temporal discounting, ambiguity aversion, framing effects and economic games. In light of the neuroeconomic data we will argue that neither the dual or unitary models capture some important aspects of how the brain processes economic decisions, however we suggest that there is space for a "weak" dual model that is born out of the realization that emotional/intuitive processes are *not* the cause of economic "irrationality".

2. The neuroscience of cognitive control

There are plausibly different ways to think of and subdivide the environmental or cognitive *factors* that recruit cognitive control. Norman&Shallice (2000) propose that there are five general classes of them, among which novelty, complexity of the environments/tasks, performance error, uncertainty and conflict. Such factors are clearly connected to those known to mediate economic utility. For instance, risk is related to the "uncertainty" criteria above and ambiguity to complexity/novelty.

By manipulating such factors, neuroscience studies on CC converged largely on the prefrontal cortex (PFC) (Miller, 2000) - though there has also been much attention to posterior cortices (Platt&Glimcher, 1999). The following neurocognitive models of CC attempt to "break" the apparent "CC continuum", by investigating what differences in tasks/environments yield differentiable activation patterns in the brain.

Following the intuitive definition of CC given in the introduction, it is a system that enables to adapt behavior to internal goals and override reflexive reaction to salient stimuli. At a very basic level, such a system should be able to i) monitor performance and, subsequently, ii) adjust it. Ridderinkhof et al (2004) reviews a growing corpus of evidence that attributes these two functions respectively to the posterior medial prefrontal cortex (pmPFC) and the lateral PFC (LPFC). In what follows we briefly illustrate them separately.

i) CC in the pmPFC
Single cell recordings on non-human primates and imaging studies on humans converge on the pmPFC's role as general *error tracking* system.

Following Ridderinkhof (2004), a brain region that does this should be sensitive to at least 4 "declinations of error": (i.1) valence of outcomes, (i.2) response errors, (i.3) pre-response conflict and (i.4) decision uncertainty, we will explain them as we go. To these we suggest to add two further factors that seem to play an important role in CC, (i.5.) motivation and (i.6.) decision costs. In addition to the "error" type categorizations, in (i.7) we report of several studies supporting the

idea of "gradients" running along different pmPFC axes (i.e. aligned sub-regions increasingly sensitive to specific characteristics of information).

(i.1) *Outcome valence*
A neural error-processing area should be sensitive to the *value* of outcomes - it should be able to tell us what outcomes are "good or bad". Indeed, single neuron recordings in non-human primates show that neuronal firing rates in the pmPFC track reward expectancy, delivery and omission (Shidara&Richmond, 2002; Stuphorn et al., 2000; Ito et al., 2003).

In line with this function, at outcome, electroencephalogram (EEG) recordings in humans show a negative polarity voltage deflection, or error-related negativity (feedback ERN), that tracks the degree of error, with respect to expected reward mean and variance, as early as 250-300 ms after outcome delivery. Moreover, this signal is greater for monetary losses than for gains (Holroyd&Coles, 2002), constituting a very early neural antecedent of loss-aversion.

Tracking the valence of outcomes is necessary to orient our behavior, particularly in any novel situation in which we might be exploring a scenario to understand which actions have positive and negative outcomes. Indeed, "novelty" is one of the five factors held to trigger CC (Norman&Shallice, 2000).

(i.2) *Performance errors*
In many contexts however we may *already* know "what's good and bad" and nonetheless simply mistake. For example, if we have to respond quickly, or if a behavior is difficult or not well practiced (also related to Norman&Shallice's novelty/complexity criteria).

In areas overlapping with those above, both monkey and human neural activation patterns differentially track correct and incorrect responses, peaking as 100 ms after the first muscle contraction (independently of the effector, i.e. hand, foot, voice etc.). Both types of feedback error and response error signals are associated with phasic signals from midbrain dopamine neurons (Ridderinkhof, 2004), which, in turn, are known to fire in accordance to reinforcement learning principles and have been strongly associated with reward (Schultz et al., 2006). For example, large performance-error ERNs are associated with midbrain phasic decreases when outcomes are worse than expected (negative prediction error), and phasic increases when outcomes are better the expected (positive prediction error). This ties the pmPFC activity to well known learning mechanisms.

(i.3) *Response conflict and (i.4) decision uncertainty*
The error types described above are "a posteriori", that is, they are detected once they've already occurred. Other types of error-related processes are a priori, in that they should serve the purpose of *anticipating* errors. These type of factors are close to Norman&Shallice's CC-inducing "conflict" and "uncertainty".

This form of CC is recruited when a stimulus activates more than one response. In which case, two scenarios may be given: (i.3) the incorrect responses must be inhibited, where we speak of pre-response conflict; or (i.4) there is more than one response that appears correct, where we speak of decision uncertainty. In contrast with performance-errors described above (i.2), we now perform the "desired" responses, but this demands CC, usually translating into slowed response times. The link between these two error-related functions is that they both signal the reduced *probability* of obtaining a reward, in contrast with performance-error and negative outcomes, which were closer to

4

tracking *magnitude* of unattained rewards.

A large number of studies show a priori error-related responses in the pmPFC (Ridderinkhof et al. (2004); Beckmann et al., 2009).

(i.5) and (i.6) *Motivation and decision costs*
Cognitive control is certainly recruited by the "declinations of error". These however might not be sufficient conditions to induce cognitive control. In particular, motivation and decision costs are important factors.

Regarding motivation, subjects are not equally motivated by rewards (or "demotivated" by their errors). A recent study by Locke and Braver (2008) reported that pmPFC (and LPFC) contributions to a task were modulated by the presence of task incentives (i.e. paying subjects). The same regions were shown to be sensitive to decision costs in another recent experiment (Botvinick et al., 2009). Thus, the effect of error-related information on cognitive control is modulated by how sensitive subjects are to rewards and punishments (their motivation) and what the costs of adjusting behavior are. These computations, though not explicitly considered by Norman&Schallice, also recruit the LPFC-PMPFC network, thus, arguably CC.

Together, all these error-related processes (i.1-6) in the pmPFC can signal when an adjustment of behavior is required. According to a meta-analysis it appears that some areas (i.e. transition zone between cingulate and paracingulate cortex) are activated by *all* kinds of monitoring events (i.1-4), thus acting as a unified monitoring system; other areas appear however activated preferentially by conflict/uncertainty-related processes (i.3-4) or by response/feedback-error related stimuli, thus yielding first dissociable articulations of CC.

(i.7) Organization of CC along the posterior-anterior axis of the mPFC
A particularly interesting fMRI study by Venkatraman et al. (2009) further qualified anatomo-functional specialization of CC within the mPFC. In their design, subjects were required to take part in three tasks, which appeared to tap on three different types of increasingly complex forms of CC. The three specifications will also serve as exemplifications of the forms of CC mentioned above, they are: a) response control, b) decision control and c) strategy control. Venkatraman et al. found such functions to recruit regions disposed on a posterior-anterior axis of the MPFC. In what follows we illustrate the distinctions separately.

a) response control
The first, plausibly simpler form of CC was elicited by manipulating pre-response conflict (i.3 above). Subjects had to count the number of subsequently presented words in 2 conditions, a congruent condition, in which the words were names of animals, and an incongruent condition, in which the words were numbers. The incongruent block clearly required more control the congruent one, as counting objects is harder when those objects are unaligned numbers. This is also a clear case of pre-response conflict, in that the same stimulus (the current number) is likely to activate conflicting responses, one of which has to be suppressed.

b) decision control
In this task subjects chose between one of 3 stocks, each with two either balanced or unbalanced rankings (i.e. one stock could have been ranked to be worth 50 and 50 – balanced rankings - by separate agencies, while the other could have been ranked 40 and 60 - unbalanced). Within this task, decision uncertainty (i.4 above) was manipulated by making the overall expected value of rankings either more similar (thus requiring more control to decide) or more dissimilar (an easier decision).

c) strategy control

In this condition subjects kept choosing between 3 stocks. Strategic control was expressed as the limit in which they chose against their usual biases. For instance, it takes a risk-averse subject more control than a risk-prone subject to accept a risk, thus suppressing his/her usual strategy. This type of manipulation would appear linked to the decision-cost factor mentioned above (i.6).

Contrasting between such increasingly complex levels of cognitive control, the authors found that they separately recruited regions disposed along an anterior-posterior axis of the mPFC: more posterior regions were recruited by response control. As this form of control is relatively more superficial, it makes sense for it to recruit regions with a majority of connections targeting lateral motor and premotor areas. Increasingly elaborate forms of CC were associated with more anterior regions, which in contrast, establish connections with more rostral aspects of the LPFC. As we will see in the next section on the LPFC, these results well fit existing models of CC in the LPFC, where a similar posterior-anterior axis is proposed to support increasingly complex forms of control.

In synthesis, the distinction proposed by Ridderinkhof between performance monitoring vs. adjustment may parallel that between a more "afferent" and a more "efferent" component of cognitive control: the pmPFC codes situations that require adjustment; the LPFC receives this information and adjusts behavior accordingly. This division is not clear cut, the proponents of the model acknowledge the many direct connections between PMPFC and motor cortices are likely to subserve a role of the PMPFC directly on execution. Nonetheless, to date this dissociation seems to have been supported by much evidence and it is a useful schematization.

ii) CC in the LPFC

Above, we distinguished between two important articulations of CC, one having to do with monitoring, plausibly based along the posterior medial wall of the PFC, the other with performance adjustment.

Adjustments can occur in two ways: 1) as a shift towards accuracy, in a theoretical speed/accuracy tradeoff, experimentally observable through post-error or post-conflict slowing of response times (Botvinick et al., 2001) (i.e. in the coin flip gamble example we gave above this coincides with the slowing of decision times following an increase in decision uncertainty); or 2) as an increase in the efficiency of control, which can translate into a *de*crease in reaction time differences between congruent vs. incongruent stimuli (Ridderinkhof et al., 2002) (i.e. after mistaking, subjects are less distracted by interfering stimuli). In line with this, a long tradition of studies attributes a function of response *inhibition* to the LPFC (cfr. Ridderinkhof et al., 2004), as well as a number of performance execution-related functions (for an extensive review see Tanji&Hoshi, 2008).

This notion is consistent with a functional subdivision within the LPFC, proposed by Koechlin and colleagues (2003). The authors take an "information theory" approach (Berlyne, 1957) in quantifying cognitive control as the amount of information needed to perform an action *"a"* among alternative actions. The account given proposes similar distinctions as those suggested by Venkatraman.

In synthesis, Koechlin et al. define cognitive control as Q (a | s) in:

(1) $Q(a \mid s) = H(a) - I(s, a)$

Where H(a), the demand of CC, is the total information needed to perform action "a", "s" is a given stimulus and I(s,a) is the mutual information between "a" and "s", as in :

(2) $I(s,a) = \log_2[p(s,a)/p(a)p(s)]$

If "s" largely determines "a", as in a very automatic response to a stimulus (i.e. osteo-patellar reflex or, though less so, scratching a mosquito bite), then $p(s \mid a) = p(s,a)/p(s) = 1$, $I(s,a)$ becomes equal to $H(a)$ and $Q(a \mid s)$ (CC) becomes equal to 0. In other words, cognitive control (in (2)) is defined in opposition to automatic responses, where an automatic response is defined as a response that is completely determined by stimulus. In this sense Koechlin's notion of cognitive control is close to that of response inhibition, which can thus be taken as an important interpretative tool for understanding CC. However, Koechlin's more innovative contribution is that of a further subdivision of cognitive control. They use information theory to describe how executive function can be divided into hierarchically ordered control processes. In what follows we illustrate.

The main idea is that CC may be exerted with different degrees of complexity. Such varying degrees of complexity may occur either within or between different, potentially hierarchically nested levels. Koechlin et al. propose that such levels correspond to (a) sensory, (b) contextual, (c) episodic and (d) branching forms of control; and, on the basis of experimental evidence, they suggest that such sub-functions would be implemented respectively in the premotor cortex, the caudal LPFC, the rostral PFC and the frontopolar cortex (FPC) respectively. In what follows we informally illustrate Koechlin's classification (the same classifications are formalized in Koechlin et al., 2003, 2007):

(a) Sensory-motor control
At the most superficial level of cognitive control correct, stimulus-response associations are given. We imagine such couplings to be particularly simple, constant or well-practiced. Such that, if other factors at higher levels of control do not interfere, all that is necessary is to attend to the stimulus, for the associated response to be activated and delivered.

For instance, lets say I want an apple and I know there are some in the fridge. While I browse the shelves, what is likely keep online is the correct stimulus-response association, i.e. "apple-take". So that when the right stimulus will be perceived, the corresponding action will be executed (when I see the apple, I'll take it).

Clearly, even within this "superficial" level of CC, stimulus-response associations can be added one to another, thus increasing CC demand. Lets say that the apple was in the middle of unwanted pears in a fridge drawer. In my new search task I'll have to keep two stimulus-response associations, "apple-take" and "pear-shift". This increases the information I must retain in order for the appropriate actions to be executed. This kind of control, which directly links stimuli to responses, is called sensory-motor control by the authors. They provide evidence that its parametric manipulation (the progressive increase in the number of associations) activates the premotor cortex.

(b) Contextual control
In the preceding paragraph, we gave an example of how stimulus-response associations can be chained together in a serial manner, as in sensori-motor contro (i.e. "take if apple", n=1; "take if apple, shift if prune", n=2 etc.). However, stimulus response associations can also be chained together hierarchically. That is, we can link *sets* of stimulus-response associations to stimuli. In this case, the set-setting stimulus plays a role similar to that of a context. Let us provide a few examples.

Changing our "fruit" example slightly, one could reserve different stimulus-response associations according to somewhat "contextual" aspects that relate to the fruit. For instance, it could be that I prepare to take an apple or a banana, though if I find them in an advanced state of maturation I

might prefer to blend them. Thus, the banana could trigger the set [(eat iff yellow)& (blend iff black)], while the apple will trigger [(blend iff mushy)&(eat iff red)].

In the given example the two stimuli, "apple" and "banana", don't directly cue a response, as they did in sensori-motor control. Instead, they trigger *other* stimulus-response associations, or action sets, hierarchically embedded in a superordinate clause.

In synthesis, CC can operate adding elements serially, as in sensorimotor control, i.e. (take iff apple)&(shift iff pear), or it can operate in depth, as in contextual control, which incorporates sensory-motor control. For example: {[(peel iff yellow)& (blend iff black)] iff banana}&{[(fry iff black)&(wash iff red)] iff apple}.

Koechlin et al. reported that control of this form elicited activity in posterior LPFC regions.

(c) Episodic control
Such an in depth hierarchical form of control may by triggered by contextual cues, that are presented more or less simultaneously to the response cues, or by different temporal episodes, that is by stimuli or contexts that temporally precede the response cue.

In one experiment conducted by Koechlin et al. (2003), either of 2 colors, lets say blue and green, would appear. After the color disappeared, either of 2 stimuli, say A or B appeared. If "green" appeared, then subjects were to respond with action 1 to stimulus A, and with action 2 to stimulus B. If "blue appeared", they would have to do the reverse. In this context, the colors are considered cues that trigger "episodic control", the letters as "response cues".

This manipulation elicited neural activity in the rostral LPFC.

(d) Branching
In many economic decision making scenarios, one of the big problems a cognitive agent faces is that of analyzing and confronting a number of possible plans of actions. However, the LPFC is functionally limited, and only serially represented plans can be processed, as in a bottleneck (Koechlin 2007). The authors suggest that "branching" is the function that counters the bottleneck problem in the LPFC.

Branching is attributed to the frontopolar cortex (FPC) and enables the exploration/execution of a target task, while maintaining a previously selected task in a pending state for subsequent automatic retrieval and execution. In other words, branching enables agents to go in and out of related goal/plan-related branches and is proposed to be the "core" of a number of other functions previously ascribed to FPC in other brain imaging studies, such as inferential reasoning, multitasking and explorative learning.

It is fairly straightforward to understand how such a function could subserve decision making. For instance, say we have to go to the movies and are evaluating whether to go by car or bike. Keeping a branch-like terminology, the main "tree" (the overall task) (1) would consist of us reaching the cinema. From there, two branches depart, that of going by bike (1.1) and that of going by car (1.2). However, given the bottleneck-informational restraints, only one of these branches can be processed at the time.

The branching processes subserved by the FPC would however maintain the "car" branch (1.2) pending, while the "bike" branch (1.1) is explored, giving way to yet lower level branches. For instance, to go by bike we would "have to go the cellar to get the bike" (1.1.1), but "who has the

cellar keys?" (1.1.1.1) etc. Lets say the bike alternative loses its attractiveness for a moment, and that I may want to explore the car option. While thinking of the bike our agent hasn't forgot what he was doing: the alternative "car branch"(1.2.), has remained pendant - thanks to proposed branching processes – and can thus be quickly reactivated.

Though the above example is clearly anecdotal, it should resemble a valid extension of what Koechlin et al. (2007) demonstrates in a neurocomputational FPC model of branching.

Synthesis of CC in the PFC
We gave an overview of the type of functional specializations that neuroscience proposes to differentiate CC in the PFC. A distinction between more monitoring-related processes in the pmPFC and adjustment related ones in the LPFC is supported by several lines of evidence (Ridderinkhof et al., 2004).

Within the pmPFC there is empirical support for specialization with regards to a number of factors: i) whether errors are detected or anticipated (cases i.1-2 vs. i.3-4); ii) whether cues trigger more or less complex forms of CC (Venkatraman et al., 2009). This latter finding well fits with proposed models of CC in the LPFC (Koechlin et al., 2003, 2007), which suggest that specific, hierarchically organized levels of CC recruit PFC regions along a posterior to anterior axis.

Such a neurocognitive model is a simplification of how CC is implemented in the PFC. Routinely debates in neuroscience simultaneously work in two directions, one more empirical, the other more interpretative; these translate respectively in 1) finding progressively finer-grained distinctions at the functional and anatomical level and 2) obtaining broad models that explain what such distinctions have in common and how they may be integrated. In what follows we focus on the latter debates.

3. Cognitive control and emotions in economic decision-making: dual or unitary system?

Let us think by extremes: perhaps the largest doubt one can have about cognitive control is whether it *exists at all,* as a dissociable anatomical and functional system. At the opposite extreme, cognitive control could be completely integrated with other control unrelated structures/functions, perhaps, functionally emerging from a more distributed network. Were this the case Pareto's aspiration of circumscribing "logical actions" would probably be harder to reach. This schematization lends itself to a very broad and yet open-ended debate in cognitive neuroscience regarding the relatively dualistic or unitary nature of decision processes.

Dual models stress the relative "independence" of "decision sub-systems", that is, systems that can *independently* generate a decision, 'as if' we had different 'selves' competing for different options (Rustichini, 2008). From a theoretical point of view it is telling that models within the dual framework have adopted *game theory* constructs to analyze individual decision behavior, thus modeling behavior it as the outcome of an interaction between separate "self-interested selves". The unitary approach however also predicts the involvement of a number of "sub-systems", however, none of these can generate an independent decision. From the neuroscientific viewpoint, dual models well support the prediction that there are two dissociable neuroanatomical networks, one subserving CC, the other subserving emotion-related processes. In a unitary framework on the other hand there is no need for a functionally or anatomically distinguishable CC system.

The dual system distinction often runs parallel to the one between emotional and deliberative processes (Ochsner&Gross, 2005) (or between variously labeled fast and frugal, automatic/effortless, intuitive, experiential, hot processes, on one hand, vs. effortful, analytic, rule-

based, verbal, cool or rational processes, on the other (Mukherjee, 2010)). In support of this distinction, what may be seen as an "evolutionary dualism" within the anatomical and cytoarchitectonic structure of the brain has often been stressed. It is in fact possible to distinguish between a phylogenetically older emotion-related limbic system, which we largely share with our primates ancestors, and the prefrontal cortex (PFC), which is relatively more developed in humans (for neurophysiological properties that are unique to the PFC see Huey et al., 2006). Supporters of the dual system view frequently stress cases in which emotional responses are associated to sub-optimal or inconsistent decision-making.

Broadly speaking, both models have apparent strong and weak points. For instance, it is nearly a truism to say that the unitary approach is simpler, as it explains decision phenomena with one rather than two systems. The dual system on the other hand appears particularly appealing to explain "inconsistencies" observed in decision behavior. In what follows we explain why this is so, reviewing the neuroeconomic literature that has focused on a number of such behavioral inconsistencies. We will show that, in many cases, unitary frameworks can also accommodate the data. Throughout, we will argue that neither of the models captures some important aspects of how the brain processes decisions.

Loss-aversion
Loss-aversion (Kahneman&Tversky, 1979) is one of those clear cases of "inconsistency" of decision-making.
It seems reasonable in fact to assume that a consistent value-tracking device would weigh gains and losses equally, though oppositely. In other words, we should be indifferent when asked to accept or reject a fair coin flip offering the possibility to win 100 dollars if heads, and to lose 100 dollars if tails. However, in a number of experimental settings, people (as well as young children (Harbaugh et al., 2001) and nonhuman primates (Chen et al., 2006)) tend to refuse similarly structured bets. Normally, they require that potential gains nearly double potential losses to take the risk.

As we mentioned, it is tempting to explain decision inconsistencies with dual models. In the case of loss aversion for instance, there is an intuitive appeal in hypothesizing that the different impacts that gains and losses have on behavior could be explained by different underlying neurocognitive systems. In particular, it would be consistent with a classic dual system approach to predict that that losses might loom larger as a result of their being processed in more emotion/somato-sensory-related cortical regions. An alternative explanation however, more consistent with the unitary approach, would be that a same neural network is *differentially* recruited by the processing of both gains and losses.

Earlier neuroeconomic studies in on healthy decision makers appeared to support the dual systems hypothesis, as the anticipation and experience of economic losses has been repeatedly connected with activity in structures associated with affective and autonomic processing, such as the amygdala and the anterior insula (Knutson&Bossaerts, 2007). With some exceptions (Smith et al., 2009), the same regions were not sensitive to gains, which have instead been shown to recruit regions of the mesolimbic dopaminergic system, a system centered on the midbrain and the striatum, and branching to various regions of the PFC (Schultz, 2006).

In support of the opposite "unitary" view, one study showed by Tom et al. (2007) that increasing potential losses and gains recruited a same network, which was activated by gains and deactivated by losses. The deactivation for losses was steeper than the activation for gains, hence the behavioral loss-aversion. The difference with previous results could have been explained by differences in Tom et al.'s task, which elicited processing of decision utility rather than experienced or anticipated utility.

However, a study by Canessa et al. (in preparation) that included Tom and colleagues' task, as a control, was unable to replicate their results. In contrast, Canessa et al. found loss-related activity in regions that were not recruited by gains, such as the amygdala and somatosensory cortex. These are part of Damasio's (1994) *somatic marker* network, suggesting that sensorial and emotion-related activity is relatively specific to experiencing losses. Furthermore, Canessa et al. found that individual differences in loss-aversion were predicted by interindividual differences in amygdala volume, specifically in the centromedial nuclei. In line with this finding, De Martino et al. (2010) used the same task in patients with circumscribed damage to the amygdala. These patients' decision behavior clearly dissociated from their matched controls, as they apparently did not exhibit loss-aversion. Though some studies employing different tasks do show that the amygdala may also be recruited for positively valenced cues (Haman&Mao, 2002), those focusing specifically on loss-aversion seem to tilt in favor of a dual view.

Risk
Imagine being offered to choose between (a) $100 for sure, or (b) $200 if heads comes up on a fair coin flip. If you consistently tend to have a preference for the sure thing (a) or consistently prefer (b) in similarly structured bets you are susceptible to risk - in contrast to loss-aversion, here no losses are involved. One definition of risk is variance of outcomes. Indeed, the two gambles above have the same expected value, but different variance. "Rational" people should be completely indifferent between the above options, that is, they should be risk neutral. On average however, people are risk averse (Binswager, 1980).

In a dual view, biased decision behaviors, such as risk attitudes, could be mediated by emotions (and emotion related cortices) and unbiased risk-neutrality could be the result of deliberative processes overriding them. At first sight there is some evidence that appears to corroborate this idea.

For instance, we already mentioned Venkatraman's results (2009), which showed that overriding one's own risk tendencies involves a specific form of CC, which the authors called strategy control, which recruited highly CC-related areas in the rostral LPFC. In line with this idea that CC attenuates decision biases, patients with lesions in the OFC (Hsu et al., 2006), as healthy subjects with higher IQ (Rustichini, 2008) exhibit behavior that is close to risk neutrality. Benjamin et al. also showed that submitting subjects to cognitive loads (i.e. memorizing a large number of names) exacerbated risk-attitudes (usually making them more risk averse) (Harbaugh et al., 2009). Consistently with a dual view, it was suggested that cognitive loads would reduce the available CC to counteract our impulsive decisions. Moreover, a recent imaging study suggests that the LPFC (ventrolateral PFC) has an important role in mediating aversion to risk; where we earlier mentioned that this area is associated with CC (Ridderinkhof et al., 2004) (Tobler et al., 2009). These findings are consistent with TMS studies showing the causal regulatory link between other CC-related regions, such the right dorsolateral PFC (DLPFC) and risky behavior. In particular, one study disrupted DLPFC activity via repetitive TMS and found subjects to become risk-seeking (Knoch et al., 2006). The authors of the study propose that, when facing choices between options with different levels of risk the risky option is more salient and attractive, as it usually features a higher outcome. For instance, imagine having to choose between (a) 100 for sure or (b) 2000 dollars with 3% probability. Disrupting CC-related processes would make us more susceptible to the superficial salience of stimuli. In the above case, this might translate into subjects' preference for option (b), because of the immediacy of the "100 to 1000" comparison. Such automatic attraction towards higher outcomes would require the intervention of control processes, which would support a cooler, more analytical, assessment of the options (i.e. enabling to weigh the higher paying option by its probability, making it overall *less* attractive).

Now, though the above data and interpretations appear to support a dual system approach to risk processing, Rustichini (2008) argues that many of the same results are also compatible with a unitary view. For instance, if the DLPFC serves general information processing, *regardless* the emotional or deliberative nature of the stimulus, its disruption would nonetheless lead to a reduced capability of integrating different informational sources such as values and probabilities, which, in Knoch and collegues' task (2006), could easily translate in risk-taking behavior, as we would drift towards the higher payoff. The same argument could explain the effect of higher IQ on risky behavior.

With regards to "risk" the available brain data, though interesting, appears compatible with both dual and unitary views.

Temporal discounting

Humans discount the value of goods as the time to their receipt increases (Samuelson, 1937), a behavioral pattern that is resistant to both ontogenetic (Green et al., 1996a) and phylogenetic differences (as macaques (Hayden&Platt, 2007) and pigeons (Ainslie, 1974) also discount the value of delayed goods). Much of this behavior can be accounted for by exponential temporal discounting, which decreases the value of goods *constantly* for different time horizons.

However, compare the following two choices: the first between options (A) 50$ now and (B) 100$ in 1 month; and the second between (C) 50$ in 12 months and (D) 100$ in 13 months. Though the quantities and time lag between each option of the two choices is the same, subjects tend to switch their preference, from (A) in the first choice to (D) in the second, apparently becoming more patient when both payoffs are delayed (Green et al., 1994b). If we were to discount goods in a constant fashion (i.e. exponentially) such preference reversals shouldn't occur.

Theoretical models to account for temporal discounting and preference reversals have been proposed both within (i) dual and (ii) unitary models.

(i) *Dual models of temporal discounting*

Dual type explanations hinge on the idea that competition for guiding choice in temporal discounting occurs between an "impulsive" and a "patient" device. To represent this, a 2 parameter temporal discounting model was proposed. One parameter (delta) discounts evenly across different time points, and is consistent with the normal exponential discounting; the other (beta) gives the function a steep curvature for immediate rewards, and allows for a kink in the overall function (Phelps&Pollak, 1968; Laibson, 1997). This model was shown to fit experimental data better than models that adopted a single parameter, which were thus more consistent with the unitary approach (Frederick et al., 2002).

In line with this, a game theoretical approach has also been applied to temporal discounting (Fudenberg&Levine, 2006). It describes preference reversals as the result of competing selves: a short run self (SRS) and a long run self (LRS). Only the SRS (i.e. the beta system, in the previous model) chooses what to do in every moment, and its utility depends only on the rewards it can consume now. The LRS on the other hand can operate on choice only indirectly, by influencing the SRS in any direction through cognitive control, which is costly. The payoff function of the emerging game is the net utility of the SRS's consumption minus the LRS's control cost. In other words, the higher the immediate reward, the higher temptation is for the SRS, the more cognitive control has to be spent by the LRS, and only if the future reward is high enough the LRS might decide to invest in cognitive control. Thus, when immediate rewards are at hand, behavior will normally be biased towards the impatient choices of the SRS. If however both payoffs are in the future the SRS will not be tempted (as it is only interested in immediate rewards), and the LRS will

not have to spend in cognitive control. Consequently, when both rewards are in the future, the difference between the higher but delayed reward and lower but nearer one won't have to be as large as when one of them is immediate. Similar interpretations that view preference reversals as the result of competition between CC and impulsive systems accommodate the empirical observation that higher IQ correlates with more patient temporal discounting (Mischel et al., 1989) and that cognitive loads increase it (Hinson&Whitney, 2003). Higher IQ could for instance enter the above model in terms of *efficiency,* such that a "skilled" cognitive control would incur fewer costs than an "unskilled" one in order to prevent a same SRS to resist a temptation.

(ii) Unitary models of temporal discounting
Though these parallelisms between formal dual models and brain activity appear very appealing, there are models and neural evidence pointing in favor of a unitary view.

From a theoretical point of view, much of the dual model's appeal resides in its ability to explain "inconsistencies" in decision behavior (i.e. preference reversals). However it isn't clear whether this is something that only dual systems can do. For instance, unitary views have been defended by highlighting parallelisms with other better-known psychophysical systems (rather than developing elaborate parallels with game theoretic constructs, as in the SRS vs. LRS description).

For instance, following Rustichini (2008), a perceptual system's output could be considered as the result of two factors: i) its "utility", which would consist in the particular stimulus that triggers it; and ii) its discriminatory power. Rustichini argues that the combination of utility and poor discriminatory power would be enough to yield *semi*-ordered preferences in decision-making, which, in turn, can accommodate inconsistent behavior such as preference reversals.

To have a gist of what this means consider the following example of preference reversal (by Armstrong, 1939): you like sugar in your coffee but are indifferent between a cup with one grain (1g) and a cup with 2 (2g). You are also indifferent between 2g and 3g, between 3g and 4g and so on, until a threshold of *n* grains of sugar is reached and preference emerges. Now, in the notion of order that we are probably familiar with, that of strict orders, indifference is transitive (i.e. if A=B and B=C, A = C). However, if this order determined our sugar preferences we would be indifferent between 1 grain and n grains, and this clearly isn't the case: when the *n* grains reach a certain threshold (i.e. 2 spoons full) you will prefer *n* grains to 1 grain, thus violating indifference transitivity. From the point of view of our taste perception it is clear why this happens: we discriminate well between large differences (i.e. between 1g and 2 spoons) but we don't discriminate between the smaller, grain to grain, amounts that bridge between them. That is, we have a low discriminative power. Orders that violate indifference transitivity are called semi-orders (Luce, 1956).

Rustichini illustrates how the notion of semi-order (which indeed is perhaps more "psychophysically friendly" than game theoretic accounts) allows unitary views to account for preference reversals in temporal discounting. Let us explain why. When facing complicated decisions subjects tend to deconstruct the options in order to confront their specific dimensions, one by one. Take the example we gave at the beginning of this paragraph, in which we are to decide between 100$ now and 200$ in 1 month. We might prefer the first option. Indeed, as we saw in the sugar grain case, we are sensitive to large rather than small differences, and the difference between now and 30 days appears larger than the one between 100$ and 200$. However, this reasoning is reversed if we are offered to choose between 100$ in 12 months and 200$ in 13 months. In the latter case, the difference between 100$ and 200$ might have a greater impact than the difference between 12 months and 13 months, and this might push us to choose the patient option. In synthesis, semi-orders, together with the strategies we adopt to compensate low discriminatory

power, can accommodate preference reversals in temporal discounting, which cease to be a monopoly of dual explanations.

Temporal discounting in the brain
The models illustrated had originally no explicit *explanatory* aspirations, they were primarily theoretical constructs meant to better describe behavior. However, within a dual model approach, a first study by McClure investigated the possibility to dissociate between beta and delta-pliant neural systems. More specifically, in line with the emotion vs. CC distinction that is implicated in dual models, the group predicted that meso-limbic regions would have been primarily activated by immediate rewards and that LPFC regions would have been associated with all decisions, independent of delay.

To investigate this McClure et al. had subjects make a series of decisions between payoffs that were nearer in time and payoffs that were farther in time. The decisions were of two types: in one, the nearer payoff was immediate; in the other it was in the future, so both payoffs were delayed. When the authors contrasted the brain activity elicited by these two conditions they found that decisions containing an immediate outcome preferentially recruited a network comprehending the middle orbitofrontal cortex (mOFC), an area implicated in the integration of emotion and cognition (Damasio, 1994). Other areas of this network were the medial PFC (mPFC), the posterior cingulate cortex (PCC) and the ventral striatum. Conversely, areas previously associated with CC, such as the DLPFC and ventro lateral PFC (VLPFC), were recruited for all choices, apparently tracking reward independently of delay. Importantly, activation of these latter regions was enhanced when subjects chose the delayed reward, suggesting that CC is necessary to delay rewards. This pattern of results was replicated with primary rewards (McClure et al., 2007).

Kable&Glimcher (2007) adopted another approach, closer to the psychophysics tradition mentioned above within the description of unitary models. The results of their fMRI study contrasted rather clearly from those of McClure et al. However there were important differences in the tasks employed by the two groups and the conflicting results could be due to this. In particular, in Kable&Glimcher, subjects chose repeatedly only between an immediate reward (kept constant) and a delayed one (which varied in magnitude and delay), while, as we mentioned, McClure's study also employed choices between two delayed payoffs. Such differences considered, Kable&Glimcher showed that activity in mPFC, PCC and ventral striatum tracked subjective utility, as inferred from pre-scanning behavioral sessions. This correlation was independent of the immediate payoff, which was constant throughout the experiment. This was taken to indicate that regions such as the ventral striatum and mPFC are not exclusively related to processing immediate rewards, as previously suggested by McClure's study, but delayed ones as well. Furthermore, the neural responses in these regions did not fit alpha and beta weights of the decomposed subjective discount rates, indicating that quasi-hyperbolic discounting functions described neural responses worse than models employing single parameter.

One recent study (Ballard&Knutson, 2009) reported results partially consistent with both viewpoints. This study adopted an orthogonalized parametric task that temporally displaced reward magnitude and delay, so as to separately assess the neural responses to the two temporal discounting components, independently of one another. Consistently with the "dual system" view (McClure et al., 2004), Ballard&Knutson observed a dissociation between regions implicated in CC (such as the DLPFC and VLPFC) and limbic regions (i.e. nucleus accumbens and mPFC) - which are strongly related to emotions. However, inconsistently with McClure's results the two systems respectively tracked reward delay and magnitude, thus suggesting that no systems is uniquely sensitive to immediate rewards. Consistently with Kable&Glimcher's results, the same reward-tracking regions also coded delay, but only in the most "impulsive" subjects.

Overall, while there appear to be some "dualisms" in the brain, they don't align well to those of a typical dual model. For instance, dual models predict that a neural system would be preferentially activated by immediate as opposed to future rewards; however, Ballard and Knutson's study suggests that a key dissociation might be between reward magnitude and reward delay, which is compatible with Kable&Glimcher's results. In general, the most consistent result appears that of an LPFC involvement in the processing of the delay of rewards, as this is confirmed by two of the preceding studies, an electrophysiological study on monkeys,[32] and several patient and imaging studies in different but related tasks (Knoch&Fehr, 2007). Moreover, this idea is not in conflict with Kable&Glimcher's findings, as their study could not differentiate well between reward magnitude and delay.[1] The LPFC's involvement for processing rewards that are delayed in time is consistent with the notion that this region is needed to override prepotent responses such as those that could derive from the temptation to accept immediate payoffs.

In conclusion, the LPFC remains a good candidate for representing CC in the PFC; a TMS study could possibly confirm the causal role of this region in regulating myopic decisions. However, the modality with which this control is exerted remains largely underdetermined.

Decisions under Ambiguity
Imagine you were offered two extraction-type lotteries, presented as two boxes. For either box, you win $50 if a red ball is extracted. In box 1, there is one red ball and one blue ball. In box 2, there are initially two red balls and two blue balls, then two balls are extracted but you are not shown their colors. Thus, in box 2 there could be either two balls of the same color (either red or blue) or one ball of each color. Which box do you prefer to bet on? If you choose box 1 you are susceptible to ambiguity. Indeed, the two boxes offer the same chances of winning (they have the same expected value). The simplest definition of ambiguity is that outcome probabilities are unknown to the subject (Ellsberg, 1961).

As in the previous cases of decision inconsistencies, it could be tempting to explain ambiguity aversion within a dual framework. Not knowing the contingencies of our decision environments could easily "frighten" us, perhaps so quickly and automatically that we don't give ourselves the time to consider the possible situations and make a more analytical choice. The first neuroimaging research by Huettel et al. (2006) to directly confront neural responses to risk vs. ambiguity showed that subjects that chose the ambiguous lotteries more often exhibited enhanced inferior frontal gyrus (IFG) activity. Such activity was interpreted to be a signature of CC, which would override the impulsive decision of automatically avoiding ambiguity and plausibly mobilize cognitive resources to explore the ambiguous scenario.

A second study, by Hsu et al. (2005) was particularly consistent with dual models, as it showed that emotion-related cortices, among which, the amygdala and the OFC, responded preferentially to ambiguity and that striatal responses were more sensitive to risk. Interestingly, the two types of responses also differed in timing, as the amygdala was activated seconds earlier than the striatum. Furthermore, the causal role of the OFC in ambiguity processing was demonstrated by the observation that patients with lesions in this area were less sensitive, and even became prone to both ambiguity and risk, relative to their matched controls. Together, the functional magnetic resonance imaging (fMRI) and lesion data leaded the authors to speak of an amygdala-OFC centered *vigilance-evaluation* system (requiring CC, via the dorsomedial PFC, or dmPFC) that quickly tracks salient aspects of the stimuli that carry uncertainty-related information (i.e., signaling that information is missing).

Though Hsu and colleagues' results seem to support the idea that risk and ambiguity are processed

by distinct mechanisms in the brain, their neuropsychological results also suggested that ambiguity and risk tendencies are connected, as they seemed to correlate in both the control and patient samples (Bossaerts et al., 2010). In line with this, and closer to a unitary perspective, a study by Levy et al. (2010) found that the activity in the set of regions, including the medial PFC, striatum, amygdala, and posterior cingulate cortex (PCC) covaried with subjective value in both risky and ambiguous decisions. There was however evidence for differential activation patterns (rather than segregation), as connectivity analysis suggested that connection "weights" are stronger between the amygdala and the striatum under ambiguous than risky choices.

In conclusion, there does seem to be fast responding, emotion related activity preferentially induced by the processing of ambiguity, which can be regulated by CC-related cortices. The available neural evidence thus appears to support a dual view.

Framing Effects
Imagine being offered 100 euros to make two separate choices, 50 prior to each.
In choice 1, you are offered to decide between (A) keeping 20 of your 50 euros and (B) betting everything on a "wheel of fortune" type lottery with a 65% chance to keep all and a 35% chance to lose all. In choice 2, you are to choose between (C) *losing* 30 of your 50 euros and (D) betting everything on the same lottery above. If you chose A in choice 1 and D in choice 2, you are in line with the majority of subjects; alternatively, you might have realized that the two decisions are equivalent. In fact, B = D, but also A = C, since in one case you keep 20, in the other you lose 30 from the originally endowed 50 euros. Indeed, it all boils down to preferring a half empty glass or a half full one: the two glasses refer to the same object, that is, they are extensionally equivalent, as are the previous prospects; however, subjects tend to reverse their choices according to how the options are framed (Kahneman&Tversky, 1979).

Consistently with a dual systems approach, it has been proposed that emotional processes may underlie subjects' susceptibility to the way in which choices are framed, for instance as losses or gains. Such a model would predict that frame-driven behavior would correlate with activity in emotion-related regions and that behavioral consistency across frames (the "rational" behavior) would elicit activity in areas associated with CC. In line with this, a study by De Martino and colleagues (2006) showed that amygdala activity correlated with risk-averse behavior in "gain frames" and risk-seeking behavior in games framed negatively, which is consistent with the idea that this limbic structure amplifies risk-related biases by processing contextual cues. In contrast, when subjects "resisted" frames, the anterior cingulate cortex (ACC) was preferentially recruited in a subregion associated with strategic control (Venkatraman et al., 2009). Moreover, the authors obtained individual "rationality" indexes from behavior (a measure of their subjects' degree of susceptibility to frames) that correlated with medial OFC (mOFC) activity. The OFC is considered to integrate emotional valence and goal-oriented behavior (Damasio, 1994). As such, the authors suggested that subjects who chose more "rationally" had richer representations of their own emotional biases, enabling them to better modify their behavior.

The role of the amygdala in mediating framing effects was further supported in a recent genetic study by Roiser et al. (2009). Genetic variation at the serotonin transporter-linked polymorphic region is associated with altered amygdala activity and lack of CC on part of the PFC. The decisions on a framing effect task of subjects with the short allele variant (ss) were confronted with those with the long variant. The ss group exhibited enhanced susceptibility to frames, and their amygdalas were more active during frame-pliant choices. However, a study on patients with selective amygdala damage study by Talmi et al. (2010) found no effects of the lesion on frame-related behaviors, but did on other decisions. Thus, it remains possible that the amygdala is involved but does not cause frame-related biases.

Strategic interactions

Deciding under uncertainty means that the outcomes of our decisions depend on the realization of certain states of the world. In all the above cases (perhaps with the exception of temporal discounting) these states are the result of relatively *mechanistic* processes, such as lotteries or coin flips. In social decision contexts, on the other hand, such states are the result of *mental* processes. This distinction runs parallel to previous ones between exogenous uncertainty and endogenous (Knight, 1921) or strategic uncertainty (Van Huyck et al., 1990), and raises interesting questions revolving around the neural correlates of CC under strategic uncertainty. A number of studies recently began to investigate the issue. In what follows we review relevant results.

Prisoner's dilemma (PD)

In one shot PDs, subjects have been shown to cooperate only 10% of the times. In repeated versions the percentage goes up but drops back towards 10% as the repeated game reaches its end (Rapaport&Dale, 1967). An fMRI study Rilling et al. (2002) reported that enhanced striatal activity was associated to C-C outcomes, and was deactivated by C-D outcomes, consistently with neural models of reinforcement learning under exogenous uncertainty (Schultz et al., 2006). Moreover, when subjects chose C, after their counterparts had previously chosen C, they exhibited enhanced ACC activity. This is consistent with a role of the ACC in mediating strategic as well as exogenous uncertainty: we mentioned that subjects gradually stop cooperating as repetitions go on, this might reflect the fact that their uncertainty regarding mutual willingness to cooperate increases when prior outcome was already a C-C. Interestingly, in a follow up study (Rilling et al., 2007) decreased amygdala responses to C-D outcomes correlated with individual psychopathy scores in male players that tended to defect. This results supports the idea that amygdala activity mediates our ability to cooperate in PD-like situations. Serotonin has a documented role in amygdala activity and emotional responses in social contexts (for a review see, Canli&Lesch, 2007). The role of serotonin in cooperative behavior has been demonstrated by several neuropharmacological studies that were able to increase and decrease cooperation in PD games, respectively, by increasing and decreasing extracellular levels of serotonin (Tse&Bond, 2002; Wood et al., 2006). Overall this data supports the view that emotional processes are more strongly implicated players' decisions to cooperate, rather than to defect, however this could have to do with a number of factors, such as risk perception or different game analysis strategies (Kuo et al., 2009; Polonio et al., in preparation).

Ultimatum game and dictator game (UG&DG)

The ultimatum game is one of the most experimented games in the literature (Camerer, 2003). It consists of a 2 stage sequential game between 2 players, a proposer and responder. The proposer is endowed with a sum X, which he may split however he pleases between himself and the responder. The responder will then decide whether to accept or reject. If he accepts, both players receive the split offered by the proposer, if he rejects both players receive 0. Though game theory prescribes responders to offer the minimum possible and for responders to accept, a large number of people offer 50/50 splits, and offers of less than 20% have a half chance of being rejected (even when stakes are high). The dictator game is like the UG, with the exception that responders cannot respond, but normally only accept whatever the proposer proposes. This variation on the UG was done to examine the extent in which positive offers in UG were due to fear of punishment, rather than authentic altruism.

As we mentioned for the case of risk one of the most difficult problems of dual models is that it isn't always clear what predictions they make, and this also make it harder to confront them with unitary models. In the case of risk the ambiguity was suggested to regard neural circuitry. In the ultimatum game, it regards behavior. Specifically, it appears less clear which of the choices is to label as the "tempting" one – i.e. thus favored by an emotional system - and which as rational – i.e.

putatively mediated by CC.

For instance, a game theoretic dual model could predict that rejecting unfair offers (larger than 0) is the irrational option, and that CC is needed to override it. On the other hand, however, it could also be that our instinctive reaction to an unfair offer is actually to accept it, as this is a self-interested behavior. In this latter case, CC might intervene to "remind us" of social moral standards that should push us to take a stance against such proposals, rejecting them. Both of these interpretations are dual model ones, however they make opposite predictions. Indeed, both selfish and fairness impulses could have emotional components (Knoch et al., 2007). This is frequent problem of evaluating dual model predictions as they appear to predict A and not A.

The neuroeconomic experiments on the UG tell an interesting story in this regard.
The first neuroimaging study employing an UG (Sanfey et al., 2003) supported the 1ˢᵗ view. It showed that increased anterior insula activity positively correlated with probability to reject an unfair option, while DLPFC activity was reduced when subjects rejected. A "re-balancing" of insula vs. DLPFC activity occurred when subjects accepted more unfair offers. It was proposed that the anterior insula, an area associated with a number of autonomic processes (such as disgust), could subserve a bodily, repulsive reaction to the unfair offers. When this activation dominated over DLPFC responses, which were hypothesized to mediate self-interest, "irrational" rejections were more frequent. When however cognitive control, via the DLPFC, appeared to compete more efficiently, behavioral reactions appeared somewhat cooler/analytical and subjects were more likely to accept. Perhaps consistent with the "emotional" interpretation of rejecting unfair offers, a behavioral follow-up on Sanfey and colleagues' experiment (Xiao&Houser, 2005) showed that enabling subjects to express their negative feelings to their unfair counterparts leaded them to decrease rejections.

Though this overall scenario is compelling successive data suggest a more complex view. Unpredicted results came both from the CC and the emotion "sides" of the hypothetical dual system. Regarding the former, a TMS study (Knoch et al., 2006) delivering low-frequency pulses showed that acceptance rates were increased relative to a right DLPFC (vs. left DLPFC) vs. sham contrast. The authors suggested that the DLPFC doesn't have a crucial role in overriding impulses to reject unfair offers, rather, that it may mediates an impulse towards fairness, thus its disruption favored self-interested behavior. Regarding the effect of emotions on rejections, studies focused on the role of vmPFC lesions. According to the somatic marker hypothesis (Damasio, 1994) the vmPFC and OFC, mediate the attachment of emotional "markers" on to behaviors, events or objects. This was proposed following the observation that patients with OFC lesions are subject to a decreased emotional impact of punishers (Damasio, 1994) something that appeared to make it harder for them to learn from their mistakes. Together with Sanfey and colleagues' results, this data favored the prediction that patients with lesions in the vmPFC wouldn't have felt the "sting" of iniquity and that, being their DLPFC preserved, this would have tilted their behavioral balance towards hyper-rational decisions (i.e. rejecting less). This would have also fitted well with the "moral decisions" of these patients, which have been shown to be utilitarian (i.e. they are more willing than controls to endorse harmful acts for the greater good) (Koenigs et al., 2007), decisions that were interpreted to support a dual process model (Greene, 2007). In contrast to all this, vmPFC patients were shown to actually hyper-reject unfair offers (Koenigs, 2007).

A suggestion to make a piece of these conflicting results was proposed by Moll&Oliveira-Souza. In their view, a dissociation between "pro-social" sentiments (i.e. guilt, compassion, interpersonal attachment) in the vmPFC and more anger and indignation-related sentiments in the vlPFC would account for both utilitarian judgments in moral decision-making and excessive rejections in the ultimatum game (Moll&de Oliveira-Souza, 2007). Focus on the vlPFC was supported also by a

study by Tabibnia et al. (2008), which showed that vlPFC activity related to accepting offers, thus implicating this region in the regulation of resentment. However, another explanation, also consistent with the lesion and TMS data, seems to support the idea that inclination towards fairness requires cognitive control, while emotional processing would underlie a relatively automatic response towards self-interested behavior.

This latter interpretation seems also supported by literature on the DG. DGs were introduced (Forsythe et al., 1994) to assess whether "irrationally" fair proposals were driven by genuine "sympathy" or just by fear of punishment. The first experimental DG showed that only 21% of dictators kept all their money, thus favoring the "sympathy" view. A following experiment employed a double-blind design (Hoffman et al., 1996), in that dictators were assured that not even the experimenter was able to trace their offers back to them. In this design, more than 60% of dictators kept all their money. This still isn't a clear-cut result, as a large percent of participants gave a part of their endowment and could thus still be interpreted to be "sympathetic". However, it appears sufficient to suggest that the "experimenters' eyes" may prime a form of self-control in subjects and that this control is relieved in double-blind experiments, leaving subjects free to gravitate towards self-interest. Indeed, the risk of a loss in reputation might weigh more than a monetary one.

The impact of punishments on the brain and the effect of this on respect of social norms (i.e. fairness) was the center of fMRI experiment by Spitzer et al. (2007). They scanned subjects under 2 conditions. One condition was a classical DG, in the other however receivers were endowed additional money that they could invest to punish a selfish dictator (a variation of the UG). 1 unit of such endowment would cost the dictator 5, so the threat was credible, and dictators were motivated to change their average offers when passing between the conditions. Spitzer et al. found that, when dictators made their offers, DLPFC and lateral OFC activity was preferentially activated in the punishment condition than in the non-punishment one (the DG). Moreover, DLPFC activity positively tracked the difference of dictators' average offers between conditions. This was held to reflect the fact that the risk of incurring a punishment triggered CC, which, in turn, regulated the temptation to keep all the money.

Trust Game (TD)
A trust game is usually a 2 player 2 stage game. The 2 players are endowed with a sum X. Player 1 moves first by giving whatever part of his endowment, Y, (between 0 and X, included) to player 2. Whatever he offers is multiplied by some factor k (>1) by the experimenters. Then player 2 can give any part of his new endowment (X+kY) to player 1. Then the game ends.

Though it often comes easy to us, it is theoretically hard to trust others. In fact, we know by now that a number of deeply rooted responses would push against it: 1) while we attend reciprocation, there is a moment in which we give away money, which goes against the most basic "rule" of utility, that is, that "we prefer more to less"; 2) we are not sure that we will be reciprocated, so we are taking a risk, where, as we saw, we tend to be risk-averse in the gain domain, as is this one; 3) even if reciprocation occurs we will only be compensated later on, meaning that, our tendency to discount future rewards might lead us to devalue the future compensation. Thus, at least 3 potentially strong responses would have to be overridden, in order to make a trusting decision.

One line of research has focused on the neural mechanisms underlying the development of reputation in social decision-making (King-Casas et al., 2005; Delgado et al., 2008). Relevantly to our focus on dual models, a study by Krueger and colleagues (2008) suggested that two different systems could contribute to the development of trust. The authors distinguish between an "unconditional" and a "conditional" system. The former is characterized by earlier activation of the

anterior rostral mPFC, an area strongly linked to inferring the mental states of others (Frith&Frith, 2006), followed by activity in the septal area. Considered a limbic region, this latter area regulates the release of oxytocin via the hypothalamus; where oxytocin has been shown to support social bonds of various nature: from milk-let down (thus supporting mother-kin bonding) to a number of more complex social environments (cfr. review by Kreuger et al., 2008). A second "conditional" trust system appeared to initially rely less on mentalizing networks and more on reward-related networks in the ventral tegmental area, implicated in reinforcement learning (Schultz et al., 2006). Counterproof of the causal effect of oxytocin on trust came from a study by Kosfeld et al. (2005), where intranasal administrations of oxytocin were associated with increased trusty decisions. Interestingly, a subsequent fMRI study showed that subjects treated with oxytocin had decreased amygdala responses when the trustees betrayed their trust. Such results support the idea that two distinct neural systems underlie the processing of trust and of reputation.

Between games
The above studies were interested in dissociating alternative responses within games at the neural level. Importantly however, one of the reasons to differentiate between games is that often they greatly differ in the strategies that are required to resolve them. One of the clearest and most interesting differences is that between dominance-solvable games and coordination games (Gibbons, 2003), we illustrate. If strategy A is dominated by strategy B, for player 1, it means that selecting B yield the best payoff to player 1 no matter what player 2 chooses. A game is dominance solvable if iteratively eliminating all dominated strategies for both players yields a unique pair of responses. An example is the following by Kuo et al. (2009): both players have to pick a number between 0 and 3, player 1 wins if he picks player "2's number +1", player 2 wins if he/she picks player "1's number". For player 1, Choice "0" is dominated by choice "1", because whatever player 2 chooses, "0" cannot be "player 2's number +1", which is player 1's target. This eliminates option "0" for player 2 as well, given he/she knows that player 1 will never choose it. The process is iterated until only option "3" remains for both players. This game is thus dominance solvable. To turn this into a coordination game it is sufficient to change player 1's target number to "player 2's number", everything else being equal. Indeed, in this game there are no dominated strategies for either of the players. Any matched response (1-1,2-2,etc.) is a NE, but there is no mathematical reasoning that can indicate to the players which one is correct. This doesn't mean that they choose by chance. For instance, following Kuo et al.'s discussion, when subjects were asked to name any year only 7% answered the current year, however, when they were paid depending on whether they matched the response of others this percentage rose to 60%. The current year became a "focal point", that is, an object with "connotative and symbolic characteristics". It is clear that the strategies that are necessary to solve these games are rather different: recognizing focal points in coordination games requires fast and effortless, possibly emotional, recognition of salient characteristics of complex stimuli, something typical of intuitive processes. Iterative elimination of dominated strategies on the other hand is a relatively slow and effortful, typical signatures of cognitive control. These premises leaded Kuo et al. (2009) to predict that strategizing in coordination vs. dominance solvable games would have recruited distinct neural networks related respectively with deliberative reasoning and affective processes. Indeed, they found that regions in frontal and parietal cortex, principally middle frontal gyrus and inferior parietal lobule, were preferentially active in the dominance solvable condition, consistently with previous studies implicating such regions in working memory, attention, and logical reasoning (cfr. Kuo et al., 2009). The coordination game on the other hand recruited insular cortices and ACC, regions consistently implicated in coding internal body states, environmental salience and uncertainty. Activity in the two networks moreover correlated respectively with the number of steps necessary to solve the dominance solvable games (a measure of their difficulty) and with a score measuring how evident focal points were in the coordination game. Coordination games constitute a case in which intuitive processes are crucial in reaching optimal decisions, as they cannot be resolved by

deliberative strategies.

4. Discussion and Conclusions

Neuroscience enables us to go beyond choice/behavioral data to investigate the complex articulations of CC by observing what differences in environments and tasks yield dissociable signals in the brain. This enables us to divide that apparent continuum of behaviors connecting reflexes to deliberate actions, and to begin to discuss about its building blocks. These deliberate actions are arguably, Pareto's logical actions, which he considered the true object of economics. Understanding which actions are controlled and which are not, might then enable to understand in which domains, behaviors or contexts, rational decision theory has higher chances of working, and where it needs to be integrated. We reviewed how CC can be conceptualized in an information theory framework (Koechlin). In a near future it might be possible, with the aid of such biologically relevant models, to understand what procedures underlie the analysis and resolution of decision problems.

One broad model that predates neuroscience is that of the relative dual or unitary nature of our decisions. For instance, one very resilient idea is that deliberation makes behavior rational and that emotions make it irrational. A line of literature coming from neuropsychological observations supports this idea: patients with lesions in the amygdala do not exhibit loss aversion (De Martino et al., 2010), patients with lesions in the OFC/vmPFC are less risk and ambiguity averse, and are close to neutrality in both domains (Hsu et al., 2005) they are also utilitarian in moral decision making (Greene, 2007) and are less influenced by regret in economic decisions (Camille et al., 2004); similarly, subjects with autistic syndromes are less susceptible to framing effects (De Martino et al., 2007). All these pathologies are thus associated with *increased* "economic rationality" in a number of contexts.

This interpretation, however, ignores the most prominent and consequential behavioral feature of these patients; that is, they are also severely impaired in everyday decision-making. In experimental tasks, this is suggested by vmPFC/OFC patients' inability to learn from negative decision outcomes (Damasio, 1994) their impairments in reversal learning (Fellows&Farah, 2005), their violations of preference transitivity (i.e., they are more likely to exhibit inconsistent preferences of the type A > B, B > C, but A < C) (Fellows&Farah, 2007) and abnormal decision making in a number of interactive choice contexts (Van den Bos&Guroglu, 2009). Thus, overall, emotions take part in inconsistent *and* consistent/adaptive decisions.

The same holds for the neuroeconomic data reviewed above. On one side there are many studies indicating that i) CC-related cortices in the PFC suppress myopic behavior in temporal discounting, reduce our susceptibility to choice-irrelevant frames, attenuate our natural fear of risk and ambiguity and make us more strategic in interactions such as the UGs, DGs and dominance solvable games; while ii) emotion-related cortices induce us to sub-optimal decisions in nearly all these contexts. On the other hand, nearly as many studies stress the beneficial role for emotions in decisions, such as occurs in fictive learning (i.e. regret), obtaining higher coordination rates in games with multiple equilibria, trusting others and cooperating in PDs.

All this has implications for the dual versus unitary discussion. We suggest that there is a "strong" interpretation of dual models and a "weak" one. The weak version makes only the first of the following two claims, the strong one makes both: (1) that there are two relatively distinct broad systems in the brain, one that preferentially takes part in fast, effortless, emotional, and context-related processes, another that is preferentially activated in situations requiring control and deliberation; and (2) that these two systems make separate contributions to, respectively, "rational"

and "irrational" economic decision making. The stronger version appears at odds with current neuroscientific evidence. The weak one on the other hand accommodates those findings indicating that 2 large subsystems do appear in some form of competition for guiding decisions. It also seems like a strong request for a dual theory to claim that its subsystems are capable of completely independent decision, when most of its subsystems, especially in the PFC are so intricately connected. However, in many cases, limbic subsystems, do seem to unilaterally drive us towards one decision or another, as appears to be the case of loss-aversion for instance, where a number of regions seem to *specifically* be involved in the mediation of aversion or fear. In other cases, they seem to act in concert with CC-related regions, as shown by connectivity analysis in ambiguous decision-making or as highlighted in the paradigmatic case of regret. We briefly illustrate the latter case.

Results from a neuroimaging study (Coricelli et al. 2005) showed that increasing regret was correlated with enhanced activity in the medial orbitofrontal region, the dorsal ACC and anterior hippocampus. The hippocampal activity is consistent with the idea that a cognitive-based declarative process of regret is engaged by the experimental task. This supports a modulation of declarative (consciously accessible) memory such that after a bad outcome the lesson to be learned is: "In the future pay more attention to the potential consequences of your choice." Moreover, and of particular interest for our current discussion, affective consequences of choice can induce specific mechanisms of cognitive control (Yarkoni et al., 2005). Coricelli et al. (2005) observed enhanced responses in right DLPFC, right lateral OFC, and inferior parietal lobule during a choice phase after the experience of regret, where subsequent choice processes induced reinforcement, or avoidance of, the experienced behavior (Clark et al., 2004). Corroborating results from Simon-Thomas et al. (2005) show that negative emotions can recruit "cognitive" right hemisphere responses. Thus, negative affective consequences (regret) induce specific mechanisms of cognitive control on subsequent choices. These data suggest a mechanism through which comparing choice outcome with its alternatives (fictive error), and the associated feeling of regret, promotes behavioral flexibility and exploratory strategies in dynamic environments so as to minimize the likelihood of emotionally negative outcomes. These studies stress a more interdependent nature of controlled and controlling processes in the brain.

The debate over the unitary or dual nature of decision-making plausibly goes as far back as Plato's *Fedro*, who described the soul (*psyche*) as a chariot pulled by a black "bodily" impulsive horse, driven towards the contingent and immediate satisfaction of the senses, and a white "spiritual" one, passing through Descarte's dualism to Freud's subdivision of the self.

We are still far from a solution to these "hard" problems, though we've learned much and some of the recent developments in neuroscience seem very promising.

References

1. Ainslie GW. 1974. Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior*, 21, 485–489.
2. Armstrong WE. (1939). The determinateness of the utility function. *Economic Journal*, 49: 453-467.
3. Ballard K, Knutson B. 2009. Dissociable neural representations of future reward magnitude and delay during temporal discounting. *Neuroimage* 45: 143–150.
4. Benjamin DJ, Brown SA, Shapiro JM. (2006). "Who is behavioral?" Cognitive ability and anomalous preferences. Mimeo: Harvard University.
5. Berlyne DE. 1957. Uncertainty and conflict: A point of contact between information-theory and behavior-theory concepts. Psychological Review, 64: 329-339.
6. Binswager HP. 1980. Attitudes toward risk: experimental measurement in rural India. *Am J Agric Econ* 62: 395–407.
7. Bossaerts P, Ghirardato P, Guarnaschelli S, Zame WR. 2010. Ambiguity in asset markets: theory and experiment. *Rev Financ Stud* 23: 1325–1359.

8. Botvinick MM, Braver T, Barch DM, Carter, CS, Coher JD. 2001. Conflict monitoring and cognitive control. *Pscychological Review,* 108, 624-652.

9. Botvinick MM, Huffstetler S, McGuire JT. 2009. Effort discounting in human nucleus accumbens. *Affective & Behavioral Neuroscience*, 9: 16-27.

10. Bruni L, Sugden R. 2007. The road not taken: how psychology was removed from economics, and how it might be brought back. *Econ J* 117: 146– 173.

11. Camerer C. 2003. Behavioral Game Theory: Experiments in Strategic Interactions. Princeton Press.

12. Camerer CF, Loewenstein G, Prelec D. 2005. Neuroeconomics: how neuroscience can inform economics. *J Econ Lit* 34: 9–64.

13. Canessa N, Chierchia G, Motterlini M, Baud-Bovy G, Tettamanti M, Cappa S. (in preparation). Distinct neural correlates for the processing of magnitude, probability and uncertainty of potential monetary wins and losses.

14. Canli T, Lesch KP. 2007. Long story short: the serotonin transporter in emotion regulation and social cognition. *Nature Neuroscience,* 10.

15. Chen MK, Lakshminaryanan V, Santos LR. 2006. The evolution of our preferences: evidence from capuchin monkey trading behavior. *J Polit Econ* 114: 517–537.

16. Coricelli G, Critchley HD, Joffily M, O'Doherty JP, Sirigu A, Dolan RJ. 2005. Regret and its avoidance: a neuroimaging study of choice behavior. *Nature Neuroscience,* 8: 1255–1262.

17. Damasio AR. 1994. Decartes' Error: Emotion, Reason, and the Human Brain. New York: Putnam Publishing.

18. De Martino B, Camerer CF, Adolphs R. 2010. Amygdala damage eliminates monetary loss aversion. *Proc Natl Acad Sci USA* 107: 3788–3792.

19. De Martino B, Harrison NA, Knafo S, Bird G, Dolan RJ. 2007. Explaining enhanced logical consistency during decision making in autism. *J Neurosci* 28: 10746–10750.

20. De Martino B, Kumaran D, Seymour B, Dolan RJ. 2006. Frames, biases, and rational decision-making in the human brain. *Science* 313: 684–687.

21. Delgado MR, Frank RH, Phelps EA. 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, 8: 1611–1618.

22. Ellsberg D. 1961. Risk, ambiguity, and the savage axioms. *Quaterly Journal of Economics,* 75: 643–699.

23. Fellows LK, Farah MJ. 2005. Different underlying impairments in decision- making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb Cortex* 15: 58–63.

24. Fellows LK, Farah MJ. 2007. The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cereb Cortex* 17: 2669–2674.

25. Forsythe R, Horowitz JL, Savin NE, Sefton M. 1994. Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6: 347-369.

26. Frederick S, Loewenstein G, O'Donoghue T. 2002. Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature.*

27. Frith CF, Frith U. 2006. The neural basis of mentalizing. *Neuron*, 50: 531-534.

28. Fudenberg D, Levine DK. 2006. A dual-self model of impulse control. *American Economic Review*, 96: 1449-1476.

29. Gibbons R. 1992. Game Theory for Applied Economists. Princeton, New Jersey: Princeton University Press

30. Green L, Fristoe N, Myerson. 1994b. J. Temporal discounting and preference reversals in choice between delayed outcomes**.** *Psychonomic Bulletin & Review*, 1: 383-389.

31. Green L., Fry, AF, Myerson J. 1994a. Discounting of delayed rewards: A life-span comparison. *Psychological Science*, 5, 33– 36.

32. Greene JD. 2007. Why are VMPFC patients more utilitarian? A dual-process theory of moral judgments explains. *Trends in Cognitive Sciences*, 11:322-323.

33. Hamann S, Mao H. 2002. Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport* 13: 15–19.

34. Harbaugh WT, Krause K, Vesterlund L. 2001. Are adults better behaved than children? Age, experience, and the endowment effect. *Econ Lett* 70: 175– 181.

35. Harbaugh WT, Krause K, Vesterlund L. 2009. The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal*, 5-17.

36. Hayden BY, Platt ML. 2007. Temporal Discounting Predicts Risk Sensitivity in Rhesus Macaques. *Current Biology*, 17: 49-53.

37. Hinson JM, Whitney P. 2003. Impulsive decision making and working memory. *Journal of experimental Psychology: Learning, Memory, and Cognition*, 20: 298-306.

38. Hoffman E, McCabe K, Smith VL. 1996. Social Distance and Other-Regarding Behavior in Dictator Games. *The American Economic Review*, 86.

39. Holroyd CB, Coles MGH. 2002. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679-709.

40. Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF. 2005. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310: 1680–1683.

41. Huettel SA, Stowe CJ, Gordon EM, Warner BT, Platt ML. 2006. Neural signatures of economic preferences

for risk and ambiguity. *Neuron* 49: 765– 775.

42. Huey ED, Krueger, Grafman J. 2006. Representations in the Human Prefrontal Cortex. *Current Directions in Psychological Science*, 15: 167-171.

43. Ito S, Stuphorn V, Brown JW, Schall JD. 2003. Performance Monitoring by the Anterior Cingulate Cortex During Saccade Countermanding.

44. Kable JW, Glimcher PW. 2007. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience.*

45. Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263– 291.

46. King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz, SR, Montague PR. 2005. Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308: 78 –83.

47. Knight FH. 1921. Risk, Uncertainty, and Profit. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

48. Knoch D, Fehr E. 2007. Resisting the power of temptation. *Annals of the New York Academy of Sciences*, 1104:123-124.

49. Knoch D, Gianotti LR, Pascual-Leone A, Treyer V, Regard M, Hohmann M, Brugger P. 2006. Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior. *J Neurosci* 26: 6469–6472.

50. Knutson B, Bossaerts P. 2007. Neural antecedents of financial decisions. *J Neurosci* 27: 8174–8177.

51. Koechlin E, Hyafil A. 2007. Anterior Prefrontal Function and the Limits of Human Decision-Making. *Science*, 318: 594-598.

52. Koechlin E, Ody C, Kouneiher F. 2003. The architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, 302: 1181.

53. Koenigs M, Tranel D. 2007. Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game. *The Journal of Neuroscience*, 24: 951-956.

54. Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. 2005. Oxytocin increases trust in humans. *Nature*, 435: 673–676.

55. Krueger F, Grafman J, McCabe K. 2008. Neural correlates of economic game playing. *Philosophical Transaction of the Royal Society*, 363: 3859-3874.

56. Krueger F, McCabe K, Moll J, Kriegeskorte N, Zahn R, Strenziok M, Heinecke A, Grafman J. 2007. Neural correlates of trust. *PNAS:* 20084–20089.

57. Kuo WJ, Sjöström T, Chen YP, Wang YH, Huang CY. 2009. Intuition and Deliberation: Two Systems for Strategizing in the Brain. *Science,* 324: 519-522.

58. Laibson D. 1997. "Golden Eggs and Hyperbolic Discounting". *Quarterly Journal of Economics*, 112: 443-477.

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

60. Locke H, Braver T. 2008. Motivational influences on cognitive control: Behavior, brain activation, and individual differences. Cognitive, *Affective & Behavioral Neuroscience,* 8: 99-112.

61. Luce RD. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24: 178-191.

62. McClure SM, Laibson DI, Loewenstein G, Cohen JD. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503– 507.

63. Michael Falkenstein M, Hoormann J, Christ J, Hohnsbein J. 2000. ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology,* 51: 87-107.

64. Miller EK. 2000. The prefrontal cortex and cognitive control. *Nat Rev Neurosci* 1: 59–65.

65. Mischel W, Shoda Y, Rodriguez MI. 1989. Delay of gratification in children. *Science* 244: 933–938.

66. Moll J, de Oliveira-Souza, R. (2007). Moral judgements, emotions and the utilitaria brain. *Trends in Cognitive Sciences*, 11: 319-321.

67. Moll J, Oliveira-Souza R. 2007. Response to Greene: Moral sentiments and reason: friends or foes? *Trends in Cognitive Sciences,* 11: 322-323.

68. Mukherjee A. 2010. Dual system model of preferences under risk. *Psychol Rev* 117: 243–255.

69. Norman DA, Shallice T. 2000. Attention to action: willed and automatic control of behaviour. In: CHIP Report 99. San Diego: University of California.

70. Ochsner KN, Gross JJ. 2005. The cognitive control of emotion. *Trends Cogn Sci* 9: 242–249.

71. Phelps ES, Pollak RA. 1968. On second-best national saving and game- equilibrium growth. *Rev Econ Stud* 35: 185–199.

72. Platt ML, Glimcher PW, 1999. Neural correlates of decision variabes in parietal cortex. *Nature* 400: 233-238.

73. Polonio L, Bonini G, Coricelli G. In preparation. Disentangling mental processes in interactive game play.

74. Rapoport A, Dale P. 1967. The "end" and "start" effects in iterated Prisonder's Dilemma. *Journal of Conflict Resolution*, 10: 363-366.

75. Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. 2004. The role of the medial frontal cortex in cognitive control. *Science,* 306: 443–447.

76. Ridderinkhof RK, de Vlugt Y, Bramlage A, Spaan M, Elton M, Snel J, Band, GPH. 2002. Alcohol

Consumption Impairs Detection of Performance Errors in Mediofrontal Cortex. *Science*, 298: 2209-2211.

77. Rilling JK, Glenn AL, Jalram MR, Pagnoni G, Goldsmith DR, Elfenbein HA, Scott OL. 2007. Neural Correlates of Social Cooperation and Non-Cooperation as a Function of Psychopathy. *Biological Psychiatry*, 61: 1260-1271.

78. Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, Clinton DK. 2002. A neural Basis for Social Cooperation. *Neuron*, 35: 395-405.

79. Roiser JP, de Martino B, Geoffrey CYT, Kumaran D, Seymour B, Wood N, Dolan R. 2009. A genetically Mediated Bias in Decision Making Driven by Failure of Amygdala Control. *The Journal of Neuroscience*, 29: 5985-5991.

80. Rustichini A. 2008. Dual or unitary system? Two alternative models of decision making. *Cogn Affect Behav Neurosci* 8: 355–362.

81. Samuelson PA. 1937. A note on measurement of utility. *Rev Econ Stud* 4: 155–161.

82. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. 2003. The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science*, 300: 1755-1758.

83. Schultz W. 2006. Behavioral theories and the neurophysiology of reward. *Annu Rev Psychol* 57: 87–115.

84. Schultz W. 2006. Behavioral theories and the neurophysiology of reward. *Annu Rev Psychol* 57: 87–115.

85. Shidara M, Richmond B. 2002. Anterior Cingulate: Single Neuronal Signals Related to Degree of Reward Expectancy. *Science,* 296 (5573): 1709-1711.

86. Simon-Thomas ER, Role KO, Knight RT. 2005. Behavioral and electrophysiological evidence of a right hemisphere bias for the influence of negative emotion on higher cognition. *J Cogn Neurosci* 17: 518–529

87. Smith BW, Mitchell DG, Hardin MG, Jazbec S, Fridberg D, Blair RJ, Ernst M. 2009. Neural substrates of reward magnitude, probability, and risk during a wheel of fortune decision-making task. *Neuroimage* 44: 600–609.

88. Spitzer M, Fischbacher U, Hernberger B, Gron G, Fehr E. 2007. The Neural Signature of Social Norm Compliance. *Neuron*, 56: 185-196.

89. Stuphorn V, Taylor TL, Schall JD. 2000. Performance monitoring by the supplementary eye field. *Nature*, 408: 857-860.

90. Tabibnia G, Satpute A, Lieberman M. 2008. The sunny side of fairness. *Psychological Science*, 19: 339-347.

91. Talmi D, Hurlemann R, Patin A, Dolan R. 2010. Framing effect following bilateral amygdala lesion. *Neuropsychologia*, 48: 1823-1827.

92. Tanji J, Hoshi E. 2009. Role of the Lateral Prefrontal Cortex in Executive Behavioral Control. *Physiological reviews*, 88: 37-57.

93. Tobler PN, Christopoulos GI, O'Doherty JP, Dolan RJ, Schultz W. 2009. Risk-dependent reward value signal in human prefrontal cortex. *Proc Natl Acad Sci USA* 106: 7185–7190.

94. Tom SM, Fox CR, Trepel C, Poldrack RA. 2007. The neural basis of loss aversion in decision-making under risk. *Science* 315: 515–518.

95. Tse WS, Bond AJ. 2002. Serotoninergic intervention affects both social dominance and affiliative behavior. Psychofarmacology, 161: 324-330.

96. Van den Bos W, Guroglu B. 2009. The role of the ventral medial prefrontal cortex in social decision making. *Journal od Neuroscience*, 29: 7631-7632.

97. Van Huyck JB., Battalio R, O. Beil R. 1990. Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure, *American Economic Review* 80, 234- 248.

98. Venkatraman V, Alexandra GR, Taran AA, Huettel SA. 2009. Resolving response, decision, and strategic control: evidence for a functional topography in dorsomedial prefrontal cortex. *Journal of Neurosciene* 29: 13158-13164.

99. Wood R, Rilling J, Sanfey A, Bhagwager Z, Rogers D. 2006. Effects of Tryptophan Depletion on the Performance of an Iterated Prisoner's Dilemma Game in Healthy Adults. *Neuropsychopharmacology*, 31: 1075-1084.

100. Xiao E, Houser D. 2005. Emotion expression in human punishment behavior. *PNAS*, 102: 7398-7401.

101. Yarkoni T, Gray JR, Chrastil ER, Brach DM, Green L Braver TS. 2005. Sustained neural activity associated with cognitive control during temporally extended decision making. Brain Res Cogn Brain Res 23: 71-84.