Value-based decision-making of actions and tasks in human prefrontal cortex

Sven Collette

PHILOSOPHIÆ DOCTOR (PHD)
THESIS OF UNIVERSITÉ PARIS 6

ed3c - Ecole Doctorale Cerveau Cognition Comportement

submitted by

Sven COLLETTE

to obtain the grade of DOCTEUR DE L'UNIVERSITÉ PARIS 6

Value-based Decision-Making of Actions and Tasks in Human Prefrontal Cortex

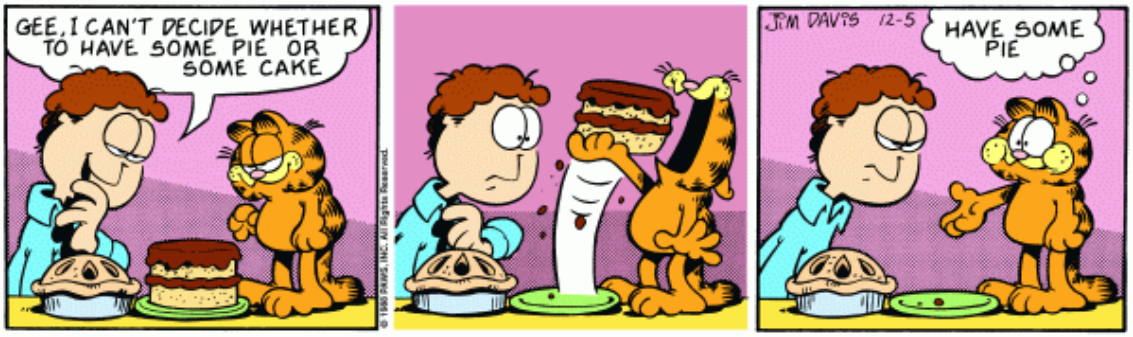
---

defended on May 30th 2012 under evaluation of a committee composed by :

Pr. Etienne KŒCHLIN	Thesis Advisor
Pr. John-Dylan Haynes	Rapporteur
Dr. Emmanuel Procyk	Rapporteur
Dr. Mathias Pessiglione	Examinateur
Pr. Alain Berthoz	Examinateur

Laboratoire de Neurosciences Cognitives
Ecole Normale Supérieure
INSERM U960
29 rue d ulm
75005 Paris

ED3C
Université Pierre et Marie Curie
4 place Jussieu
75005 Paris


GEE, I CAN'T DECIDE WHETHER TO HAVE SOME PIE OR SOME CAKE
© 1980 PAWS, INC. All Rights Reserved.
JIM DAVIS 12-5
HAVE SOME PIE

Remerciements

Premièrement, j'aimerais remercier Etienne de m'avoir accepté dans son équipe en tant que thésard, malgré le fait que la porte du métro m'avait éclaté une dent quelques minutes avant mon entretien. Il m'a appris l'autonomie, et son exigence de précision et de concision va sûrement guider mes prochains projets scientifiques. Un grand merci aussi à l Ecole de Neurosciences Paris/Ile-de-France, qui m'a permis financièrement de réaliser une thèse à Paris en tant qu'étranger, et de rencontrer des thésards de tous les coins du monde.

Salutations aux autres loulous, colocs et bouddhistes, pour des discussions scientifiques et surtout non-scientifiques: Sylvain (pour son savoir en Matlab et tant d autres discussions), Laura (pour son aide au quotidien), Nilou et sa confrérie, Fiston Morillon, Joëlle-la-Folle, el Duderino Claudio, Walter et son savoir et son expérience sur Paris, David Hasselhoff, Anne, Jan, Maël, Marion, Muriel (qui m'a enduré pendant 3 ans comme voisin de bureau) et autres coéquipiers, Jeff Lebowski, Nathan Drake, Romain(s), Tom, puis Pepe et Ronaldo pour les simulations; merci aussi aux deux Jean-Paul et à Maus, et à tous ceux que j'oublie, pour des moments agréables sur Paris.

Puis je remercie particulièrement mes parents, qui m'ont toujours soutenu de mille façons lors des longues années d'études (oui, je vous promets que là c'est fini), et ma soeur qui a su accepter des cadeaux d'anniversaire avec 10 mois de retard. Last but not least, une pensée ardente à Lisy, source de zen et tant d'autres choses: que la baia soit avec nous!

Abstract

Switching between tasks in cue-based decision-making has been studied intensively, e.g. putting on hold one task while performing another one, with a fixed reward. Here, the frontopolar cortex (FPC) is engaged when subjects are required to keep in mind a main goal while performing parallel subgoals, and the anterior cingulate cortex (ACC) associates actions to outcomes.

During my thesis, I investigated switching on different levels of action abstraction in value-based decision-making: choosing freely on the one hand between two simple actions, and on the other hand between two abstract structures, i.e. competing tasks. fMRI BOLD signals were recorded on healthy subjects during a probabilistic reversal-learning paradigm, with anti-correlated stochastic reward probabilities. I compared basic reinforcement learning and a bayesian approach to infer the subject's internal option values, which I regressed against the BOLD response.

The findings reveal that the ventromedial PFC and striatum are engaged on the level of actions, and in contrast FPC, ACC and dorsolateral PFC on a task level. FPC monitors the evidence in favor of the alternative task only, and ACC activity manifests a task effect, predicting the switching behavior between tasks, but not actions. Furthermore, I provide evidence for a specific engagement of the prefrontal cortex network in value-based decision-making of abstract behavioral structures through connectivity analyses.

Résumé

Les mécanismes de décision impliqués dans le *switching* entre tâches à la base d'un indice a été étudié intensément, p.ex. la suspension de l'exécution d'une tâche pendant la réalisation d'une autre, avec une récompense fixe. Dans ce cas, le cortex frontopolaire (CFP) est impliqué quand les sujets doivent garder en tête un but principal pendant l'exécution de buts parallèles, et le cortex cingulaire antérieur (CCA) associerait les actions à leurs résultats.

Pendant ma thèse, j'ai étudié le *switching* à différents niveaux d'abstraction de l'action dans la prise de décision fondée sur la valeur espérée: choisir librement d'un côté entre deux actions simples, de l'autre côté entre deux structures abstraites, i.e. des tâches. Les signaux BOLD ont été enregistrés en IRMf sur des sujets sains pendant une expérience d'apprentissage inversé probabiliste, avec des probabilités de récompenses stochastiques anti-corrélées. J'ai comparé des modèles d'apprentissage par renforcement et d'inférence bayésienne afin d'en déduire pour chaque sujet les valeurs des options, qui ont été régressées contre la réponse BOLD.

Les résultats montrent une implication du cortex préfrontal ventromédian et du striatum au niveau des actions, et en contraste le CFP, le CCA et le cortex préfrontal dorsolatéral au niveau des tâches. Le CFP surveille les preuves en faveur de la tâche alternative, et le CCA témoigne d'un effet tâche, prédisant le *switching* entre tâches, mais pas entre actions. En outre, j'ai montré un engagement spécifique du réseau préfrontal dans la prise de décision fondée sur la valeur espérée de structures abstraites à travers des analyses de connectivité.

Contents

CONTENTS

Preface

> A Nightingale, sitting aloft upon an oak, was seen by a Hawk, who made a swoop down, and seized him. The Nightingale earnestly besought the Hawk to let him go, saying that he was not big enough to satisfy the hunger of a Hawk, who ought to pursue the larger birds. The Hawk said: "I should indeed have lost my senses if I should let go food ready to my hand, for the sake of pursuing birds which are not yet even within sight."
>
> *Aesop*

The moral from Aesop's fable is simple: A bird in the hand is worth two in the bush. But what would the hawk say if he knew that there were three birds in the bush? Or two big lazy birds? Would he weigh the losses and potential gains, change his plan and let the bird in his hand fly? These and other problems are faced by a foraging animal, which can find itself in different decision situations. Outside of the fable world, we might consider the illustration of a squirrel. In fall, when foraging to prepare winter resources, the squirrel faces several decisions in his *collection* task: should it climb many trees to collect nuts, or should it rather focus on the ground to discover truffles? Depending on its own preferences, and the environmental dispositions, it will go for one of the two collection actions. A couple months later, the squirrel is freezing in its tree-hole, and surely collecting nuts is not an option anymore, so it stays inside and tries to eat its inventory (the *eating* task), knowing that it cannot bite a nut like it bites into a truffle - different actions are needed to fill the

belly. As illustrated through this example, we can observe several levels of action abstraction, mainly simpler actions, and tasks.

In humans, German experimental psychology around the late 19th century already discussed the use of tasks, or *sets*, as illustrated by von Kries in 1895 in fig 1.

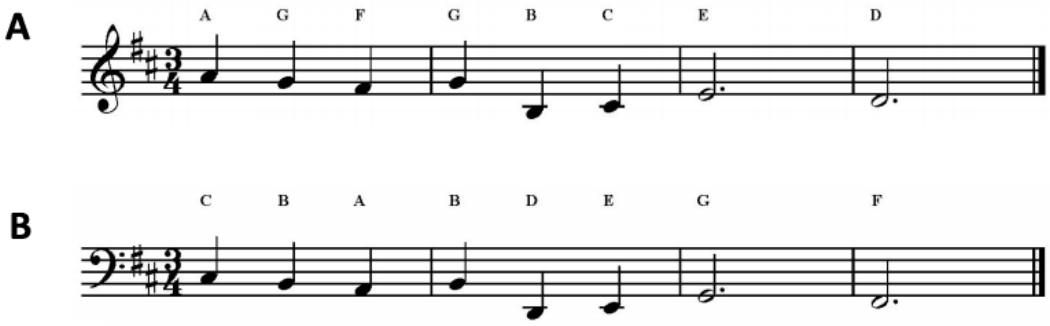


Figure 1: the clef sign changes the action performed to play a note - in A, a song is denoted in a G-clef, while in B it is written in an F-clef. Depending on the clef, a musician has to perform different actions on his instrument to play the same note position on the staff.

In the case of von Kries, the clef sign indicates the set a musician should use to play the right notes. In the context of modern psychology, this would be noted as *cue-based* decision-making, i.e. the cue indicates the set of actions to be used. In contrast, *value-based* decision-making describes situations where an agent freely chooses between tasks or actions. In the clef example, this would correspond to a songwriter who chooses the clef, depending on whether he intends to write a song for guitar or for cello.

In the brain, the specialist areas of decision-making are the frontal lobes, which underwent a considerable expansion in humans. Recently, much research has started to

investigate the behavioral principles and neural mechanisms of value-based decision-making, including experimental psychologists, ecologists, economists, neuroscientists and mathematicians. However, many questions remain uncharted, and the goal of this thesis shall be not only to clarify the literature in this domain, but specifically to distinguish variables which are often confound or not clearly stated: the level of action abstraction: in a changing environment, choosing on the one hand between simple actions and on the other hand between more abstract structures, such as tasks.

Part I

Theoretical Framework

Chapter 1

From Animal Foraging to Human Decision-Making

1.1 A strange migration

Each year between end of October and December, the world's biggest migration of mammals takes place: Up to 10 million straw-colored fruit bats (*Eidolon helvum*) fly into Kasanka National Park in Zambia (fig. 1.1). Their migratory range extends from sub-Saharan Africa down to South Africa, and their site of reunion, two small patches of mushitu forest, only covers an area of about 0.4 km^2 (Richter & Cumming, 2005). A strange migration, that could not possibly be due to coincidence. Indeed, as a recent study showed (Richter & Cumming, 2005), the colony arrives immediately before the peak of fruit production, and leaves when fruit availability is low, but before it is exhausted. During these weeks they eat up to twice their body weight per night, and many females are pregnant or recently conceived their juveniles.

Behavioral ecologists are puzzling about *E. helvum* migration and similar phenomena such as why the carnivorous Raffle's Pitcher plant serves as an overnight stay for woolly bats without devouring them (Grafe *et al.* , 2011), or why a moose decides to eat algae or grass on a particular day (Belovsky, 1984). They are trying

to explain the ultimate causes of animal (including human) behavior, and the roles of behavior to adapt to the environment. Much of their work has been devoted to link a darwinian framework with economic models to understand foraging, i.e. a quantitative analysis of the animals' decisions and a subsequent correlation of the outcomes with measures of *fitness*, the ability to survive and reproduce (Glimcher, 2002). In order to understand the advent of ecological biology and the subsequent rise of modern studies of value-based decision-making, the following section will give a brief overview on research on rewarded learning in the beginning of the 20^{th} century, which was already trying to explain animals' behavior, more specifically in heritage of the theories on *reflexes*.

Figure 1.1: Fruit bat migration in Kasanka National Park - photography of Kieran Dodds

1.2 Early Learning Theories

Ces hommes seront composés, comme nous, d'une Âme et d'un Corps. Et il faut que je vous décrive, premièrement, le corps à part, puis après l'âme aussi à part; et enfin, que je vous

montre comment ces deux natures doivent être jointes et unies, pour composer des hommes qui nous ressemblent. (Descartes, 1648)

Those are the words of Descartes, conceptualizing the beginning of physiological studies of behavior. He reasoned that all observable human behavior can eventually be classified into two categories: simple and complex behavior. Under simple behavior, Descartes understands the fact that a certain sensation always conducts to the same behavioral response, i.e. the linkage between action and sensation would be predictable and deterministic. Touching a burning hot stove with your fingers, and your hands will immediately retract. These behaviors without volitional control are part of analytic geometry, and can therefore be tested empirically. Complex behavior in contrast has a non-deterministic sensation-action linkage and underlies volition. The sensory information is forwarded through the nervous system to the siege of the nonmaterial soul (the pineal gland), which takes a decision. As the soul is outside of the physical world, no experiments can test these behaviors, and therefore the problem will stay with the philosophers.

Descartes' assumptions had a huge influence for over 250 years. One of the first to refute Descartes' dualistic system was the Russian mathematician and Nobel Prize laureate Ivan Pavlov. He argued that the deterministic reflexological system can indeed be enough to explain all behaviors, even the complex ones, which just need a more sophisticated system. While studying the dog's gastric function through measurements of the secretions of the salivary gland in reaction to conditions involving various food items, he noted that the salivary secretion of the dogs started even before they were in contact with the food. After further digging into this interesting issue, he introduced the now-called *Pavlovian* reinforcement, the acquisition (and later extinction) of conditioned reflexes. In his legacy experiment with a dog, he coupled the ringing of a bell with the presentation of food. At the beginning, the

dog salivated only at the presentation of the food, but after a short while, the ringing of the bell was sufficient to cause salivation (Pavlov, 1927). Pavlov, amongst other partisans of behaviorism, believed that such conditional mechanisms are to be found in *all* psychological processes and can account for all behaviors.

While Pavlov described this form of conditioning as an association between stimuli and automatic reflex actions, Thorndike (1898) initiated a different type of conditioning, distinguishing reflex behavior from outcome learning and subsequent changing behavior. In his experiments, cats received a reward when escaping from a cage. After the first escape, cats were much faster in finding a way out on the next trials. Skinner (1938) expanded Thorndike's *law-of-effect* and coined the term *operant conditioning*, showing through experiments with rats learning to associate pushing a lever with obtaining food, that an action-outcome association is strengthened by the presence of a positive reinforcer.

Behavioral Psychology argued that a scientific description of behaviors can be achieved without recurse to internal physiological events. However, in 1948, Tolman should lay the grounds for a conceptual change, due to the results of one of his experiments (Tolman, 1948). His Ph.D. student Ritchie trained rats as depicted in fig. 1.2. For the first days, the rats explored a T-maze, and they quickly learned in which arm to find the food, as a behavioralist would expect: Group 1 was trained to find food at F1 (e.g. right arm), Group 2 found food in F2 (left arm). After 8 days, Ritchie made a 180 degree rotation of the maze and added more paths. What happened after a few minutes of exploration was a huge surprise: The rats tended to choose not the paths which pointed in the same direction as the spots where the food has been, but rather the paths which ran perpendicularly to the corresponding sides of the room. Thus, this was evidence that rats had formed what Tolman called a spatial *cognitive map*, which can be considered as a representation of states that goes beyond simple action-associations. In other words, such a map was not appropriate for correct

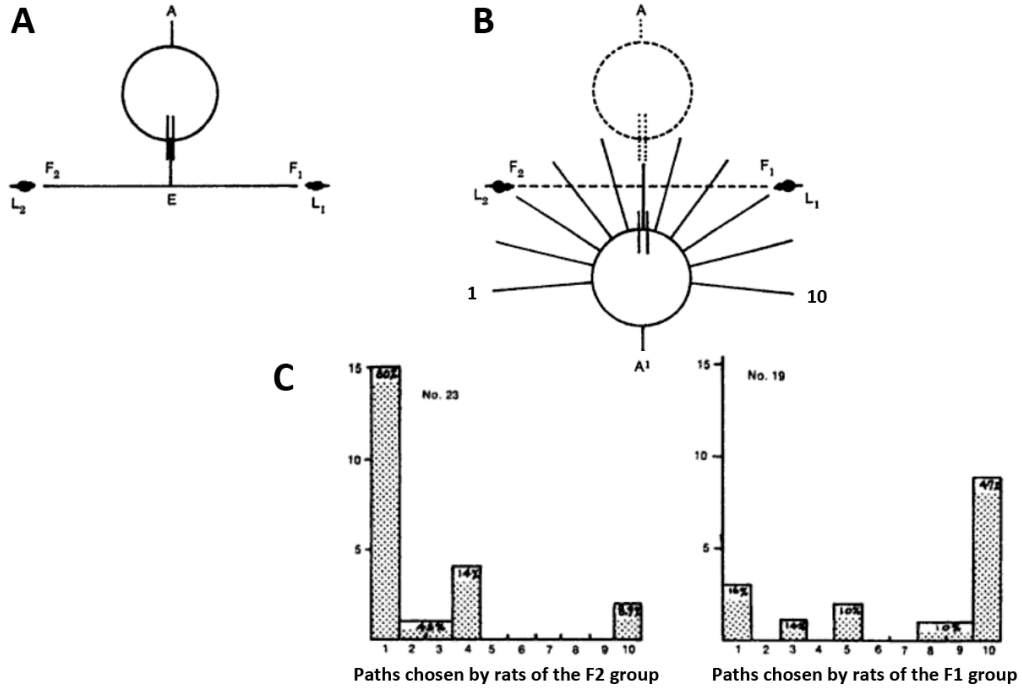


Figure 1.2: One of Tolman's experiments (1948) - **A**. Animals trained to find food in F1 or F2 ('A' represents the starting point,). Example: F1 rats turned "left" at E to get a reward. **B**. After 8 days, the table got rotated 180 degrees, and additional paths were added (dotted lines for old table configuration). **C**. percentage of rats (of each trained group) which chose each of the paths. Typical behavioralist theories from the early 20th century would predict that after rotation, the F1 rats would still turn "left" at E. However this is not what was observed. Tolman explained this observation by a *cognitive map*, an internal physiological event, which was denied by behavioralists.

goal directions, but rather in line with the correct side of the room. The evidence, that there was something like an internal variable different from a pure stimulus-response association contributed to the triggering of the *cognitive revolution*.

1.3 Animal Foraging Theory

Advent of the Prey Model

The ideas of ecological biology were rising parallel to the *cognitive revolution* of the 1950s, which placed the behaviorist school of psychology in a new context, by combining psychology, anthropology and linguistics with new fields such as artificial intelligence, computer science, and neuroscience. In the second half of the 20^{th} century, new light was shed on the learning theories of Thorndike and alikes through the groundbreaking work of researchers such as Alan Turing in artificial intelligence or Claude Shannon's *information theory*. Giving birth to the new field of behavioral ecology, Emlen (Emlen, 1966) and MacArthur (MacArthur & Pianka, 1966) published a paper in the same issue of the *American Naturalist* about the optimization problem an animal is facing whenever it has to make a decision, mainly in the context of foraging. In their papers of 1966, they wrote that this problem can be quantified and solved in economic terms, then modeled and tested experimentally.

One of the solutions proposed to the foraging problem was of particular importance to neurobiologists: the Prey Model. Imagine a monkey jumping from tree to tree in a wild jungle environment. On its search for food, it will come across a variety of patches of different types and qualities, depending on the seasons, the environment, other predators, and so on. Its panoply of food, which can consist of plants with nutritious roots, delicious insects or sweet bananas, will vary in frequency, occurring with a certain probability, varying caloric density and a different amount of time and energy investment in order to obtain it. Each patch will therefore have a specific value for our monkey, and the goal of the prey model is to give a

mathematical formulation to compute the most efficient strategy when the monkey (or any other animal) has the choice between multiple patches (Charnov & Orians, 1973; Stephens & Krebs, 1986). To put it in modeling terms, four variables need to be characterized:

- the average energy gain per prey item
- the average handling time necessary to catch and eat the item
- the energy cost of the handling process
- the frequency (time) at which the prey is detected

Once these variables are put into an equation, the Prey Model gives the optimal behavioral pattern an animal should uses in order to maximize its *fitness*, i.e. its ability to both survive and reproduce. The payoff is usually the amount of energy an animal receives per time unit. In foraging theory, animals which make decisions that maximize this payoff will persist in nature.

A classic ecological study of Decision-Making

In 1977, to empirically test the predictions of the Prey Selection Model, Krebs and Charnov designed a foraging experiment adapted to laboratory environment, making it possible to control many factors (Krebs *et al.* , 1977). They focused on five *Parus major*, a bird commonly known as the great titmouse, which favorite food is mealworms. They put the birds into a 1 m^3 cage. Near the floor of the cage was a perch through which the bird could see a black rubber conveyor belt, which moved along at about 12 cm per second (fig. 1.3). Now, the experimenters could place mealworms on the belt, which were then seen by the bird for about 0.5 seconds through the window.

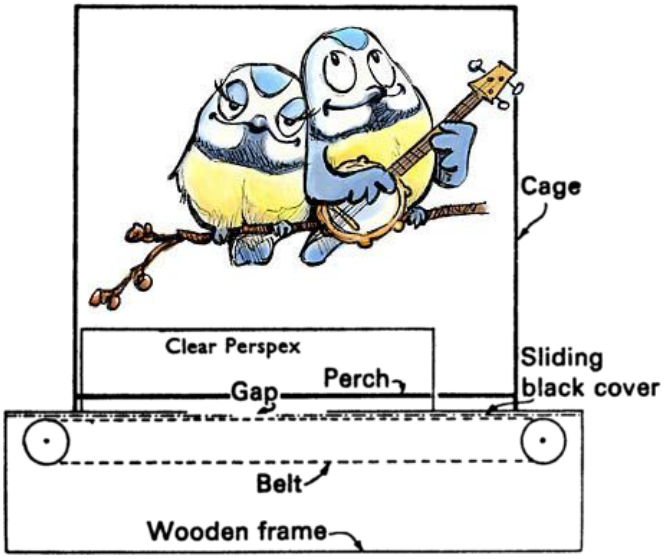


Figure 1.3: Optimal prey selection in the great tit - experimental apparatus to test predictions of the prey selection model (Krebs *et al.* , 1977)

To make the prey model of this situation quantitative, the researchers systematically manipulated three classes of variables. First, in order to control for caloric value of the worms, they were presented in two sizes: large worms and small worms (which were actually large worms cut in two). Second, the handling time was measured for each bird, and manipulated on the small worms: a piece of paper tape was attached to each small worm, which the birds had to remove before eating the prey. This made it possible to measure the *profitability* of the worms separately for each bird, since some birds striped the tape paper off much quicker than others. Third, the researchers varied the frequency at which the mealworms of each type were encountered, which made it possible to test for the minimum threshold of profitability predicted by the prey model.

Krebs and colleagues observed that in fact the prey selection model made very accurate predictions. For example, the model predicted that the birds who handle

the small worms fast should always go for the small ones. In contrast, for the slower birds, and for a frequency higher then once every 7 seconds, the birds should become selective and ignore the small worms. This was very close to how the birds actually behaved. However, the behavior also had a significant difference in comparison to the model: the *zero-one rule*, which states that if a mealworm is worth eating, go for it all the time. if not, never go for it. Interestingly, the birds did not show such an absolute preference, but continued sampling the small worms at least 10 percent of the time - a behavior we would now classify as *exploratory*.

Hierarchical Foraging

As Charnov and Orians pointed out in their 1973 paper, a first order theory of optimal foraging could focus on the level of actions during feeding interval. However, this reasoning would not reflect the integrity of the decision process. If one only considers the options still open for an animal after it started foraging, many decisions are being ignored, mainly those that were made prior to this stage. As an illustration, think of a bird with a nesting habit in a certain type of wood. Once a year this bird will have to make a decision on where to build its nest. The choice of a place for the nest will be critical for foraging and feeding the youngsters. After the nest is build, the bird can start looking for prey inside the chosen environment; some items it will like, others will be ignored. Some food patches might even change on a minute-scale, for example if the bird's favorite insect only shows in the mist of the morning, while others might be available during the whole day. In hierarchical words, the bird first selects a *nest level*, and then based on the chosen level, it chooses further actions.

Charnov classified these decisions by their place in a foraging hierarchy (fig. 1.4). However, they also state that only one level at a time can be tested experimentally, holding the other levels constant, in order to simplify the construction of such optimal foraging models.

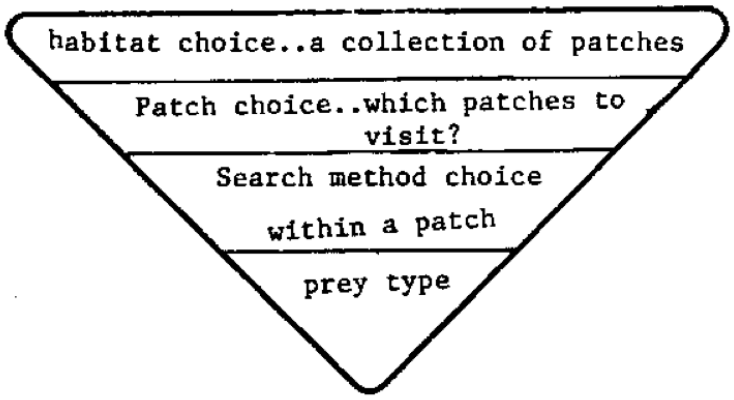


Figure 1.4: Charnov's typical hierarchical choice during foraging - (Charnov & Orians, 1973)

1.4 Utility Model and Prospect Theory

Similarly, unexpected behavior in humans called for a change of concept in behavioral economy. Until the 1970s the major theory of neoclassical decision-making was the expected utility model. Von Neumann & Morgenstern (1944) had built their model on a set of axioms (e.g. transitivity of preferences), providing criteria for the rationality of choices. As long as an individual conforms to the axioms, his behavior can be described in terms of utilities of divers outcomes. In other words, when a rational decision-maker is facing a choice, he will go for the option that offers the highest expected utility. However, people do not always behave "rational", preferences appear sometimes incompatible with expected utility theory. Consider the following questionnaires from a student survey by Tversky and Kahneman (Tversky & Kahneman, 1981) (N is denoting the number of respondents; in brackets: percentage who chose each option):

Problem 1: [N=152] Imagine that the US is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs

to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows, which of the two programs would you favor?

- If program A is adopted, 200 people will be saved *[72 percent]*
- If program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved *[28 percent]*

Problem 2: [N=155] Same cover story as before, but different formulation of the alternatives:

- If program A is adopted, 400 people will die *[22 percent]*
- If program B is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die *[78 percent]*

As one can easily see, the two problems are identical, they merely reflect a change from number of lives saved to number of lives lost. However, in the first problem, a general risk aversion can be observed, i.e. people prefer a certain, but low payoff, whereas in the second problem, the participants tend to be more risk seeking, preferring a bargain with an uncertain, but high payoff. Such a paradoxical behavior can not sufficiently be explained by Von Neumann and Morgenstern's utility concept, where an uncertain outcome is weighted by its probability. Instead in their *prospect theory*, Kahneman & Tversky (1979) propose a descriptive model based on two subject-specific functions (fig. 1.5):

A an S-shaped value function, resulting in differently perceived positive and negative outcomes: the response to losses is more extreme than the response to gains

B a weighting function, i.e. the value of an uncertain outcome is multiplied by a decision weight π, which however is not a probability: Low probabilities are overweighted, whereas high probabilities are underweighted.

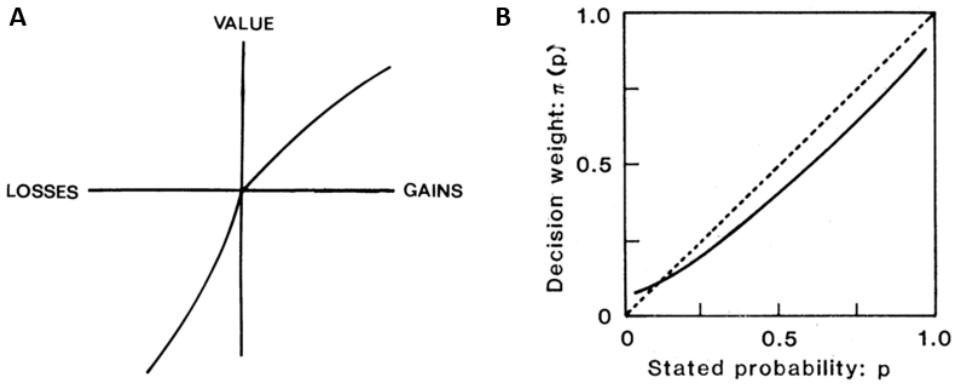


Figure 1.5: functions proposed by the prospect theory: - A. hypothetical value function and B. hypothetical weighting function (Kahneman & Tversky, 1979)

Testing hypotheses like the Prospect Theory or the Prey Selection Model is in general quite fruitful and suggests that behavior and its environment can be described pretty accurately by economic approaches. But despite the progress in this field, research still did not investigate the underlying neural architectures of the behavior. Even more - and this is the critical point where economics and biology are diverging - the models may be a good description of the behavior, but this does not necessarily mean that subjects are implementing the models; from an economical viewpoint, this is not crucial, as long as the maths fit. For a neurophysiologist, it is of great importance to discover the real implementation on the brain level. Number of laboratories have therefore started to study decision-making from a neural perspective, but the problem still persists.

1.5 The Quest for Decision Variables in the Brain

The studies of Gallistel in the 1980s and 1990s are one of the first to circumvent the experience of rewards through perceptual systems: under direct stimulation of cerebral regions (here: the medial forebrain bundle, MFB) with implanted electrodes, they observed rats making decisions when working for a reward (for an illustration, see fig. 1.6). Typically, the rat is placed in a cage and it can choose between two levers of different self-stimulation reward values, for example: likelihood of stimulation is increasing with increasing time since last lever press. By varying different factors such as the interval schedule of self-stimulation, Gallistel and colleagues were able to derive the value of a stimulation as a function of the current, the frequency and the duration of the stimulation. In other words, they showed that it might be possible to use mathematical modeling approaches to study not only behavior but also the underlying neurobiological processes. And since the value of such a stimulation can be characterized by economic variables, the question arose vice versa whether such computational approaches can be used in neurophysiological studies to characterize the observed neural activity.

Indeed, many studies since have suggested that the explicit computational models proposed by researchers from behavioral ecology, economy and artificial intelligence could be reflected on a neuro-computational level. In the following chapter, I shall first review the two learning models which recently have been extensively used to explain reward-guided behavior and to calculate parameters which are thought to be represented by neural activity in the human brain. Furthermore, to unveil the neural mechanisms behind value-based decision-making in humans, new techniques such as functional magnetic resonance imaging have proven to be of major importance. Therefore, I will devote a chapter to explain this technique, before focusing my attention on the description of specific regions in the prefrontal cortex and the striatum, as well as their respective contribution to value-based decision-making.

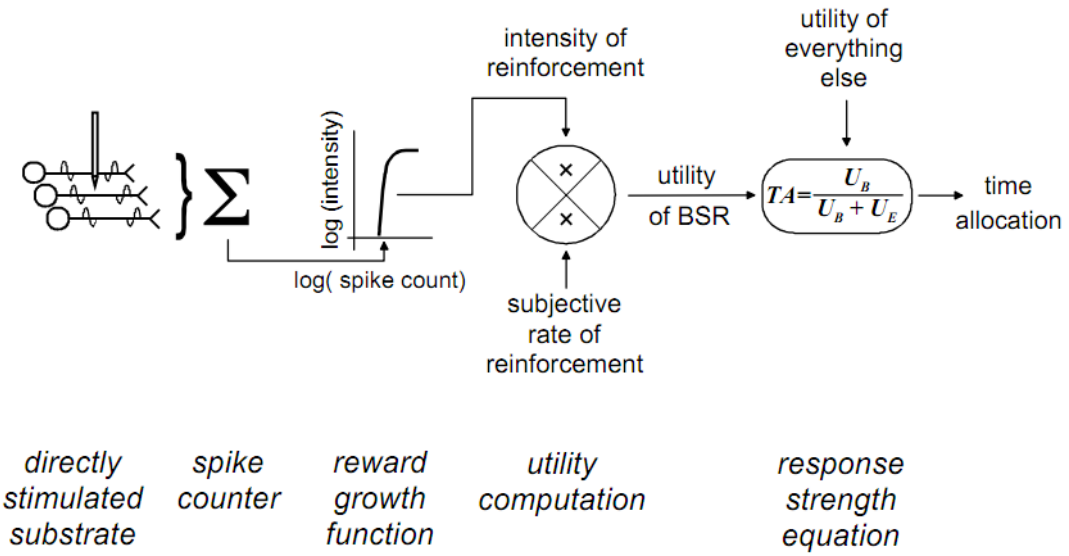


Figure 1.6: Medial forebrain bundle stimulation, utility computation and translation into behavior - schematics of Peter Shizgal (Shizgal, 1997)

Chapter 2

Modeling Behavior

2.1 Marr's three levels

In his posthumously published writings, David Marr (1982) proposed that for a complete understanding of the visual system or any other function of the cognitive system of a biological agent, we must make a distinction between three complementary levels at which information processing can be described (in parantheses: examples applied to the brain):

- computational level: description of the *goal* of the computation, and why it is appropriate (the mental representations)
- algorithmic level: description of the system's *method*, how the theory can be implemented, or how it transforms the inputs to outputs (the cognitive processes)
- implementational level: description of the system's *means*, namely how it can be realized physically (the hardware: the brain)

Marr's viewpoint was that in order to arrive at a comprehensive theory, all three levels should be taken equally seriously. Clearly, the definition of the three levels is not always very straightforward, and one sees that such a reasoning was triggered

by the rise of computers. However, the three levels can be imagined as being a much broader concept, for example in other biological disciplines like genetics: The *goal* was described by Darwin (1844) with his theory of evolution and the survival of the fittest. The *method* was described by Mendel (1865) through his experiments in plant hybridization. Finally, the *means* were discovered by Watson and Crick and the beginning of modern genetics. With Marr's three levels of explanations in mind, I will introduce in this chapter the two major current model frameworks in decision neuroscience (later used in my studies): Reinforcement Learning and the Bayesian approach.

2.2 Reinforcement Learning

Definition

Reinforcement learning (RL) in general is the algorithm used by an agent who has to learn behavior through trial-and-error interactions, through a prediction error, within a dynamic environment. Modern RL modeling is the fruit of research interaction between the fields of psychology of animal learning (Pavlovian and Instrumental conditioning), artificial intelligence and electrical engineering. In a typical RL-model, the agent is bound to its environment through perception and action. After each interaction, the agent receives an input information about the current

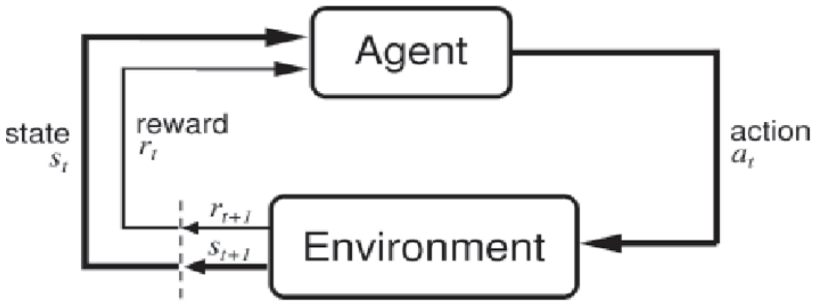


Figure 2.1: illustration of a standard RL agent in an environment -

state s of the environment. Next, the agent chooses to perform a certain action a, which changes the state of the environment. The value of this state transition is perceived by the agent as a scalar reinforcement signal r (fig. 2.1). The ultimate purpose of the agent's behavior should be to choose actions that maximize the sum of numerical rewards over time, usually formulated as discounted sum of future rewards. As an illustration, one might consider sensory-motor learning, game-playing like chess, or applications in mobile roboters like KITT.

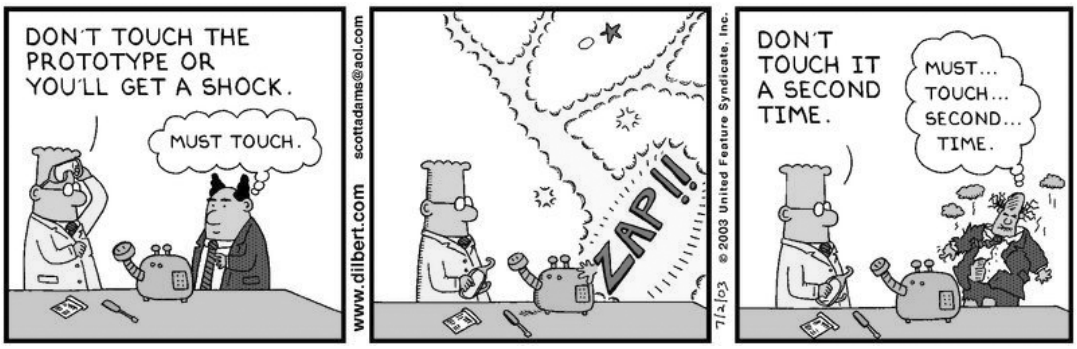


Figure 2.2: failure in reinforcement learning - The reinforcement signal can reflect the success or - like in this strip - the failure of a system after it has performed a sequence of actions - from dilbert.com

The reinforcement framework, intended to represent all essential features of the artificial intelligence problem, is characterized by three key concepts:

- a discrete set of environment states
- a discrete set of agent actions
- a set of scalar reinforcement signals (0,1, or real numbers)

Typically the decision-maker begins without knowing any rules of the environment, and he has to learn, or sample the rules, from experience. However, the RL algorithms offer much more than just a computational theory. Some RL features such

as the temporal difference learning rule, which I will discuss later, find support in neural structures such as the dopaminergic neurons.

Historical Background

During the behaviorist movement, the idea of trial-and-error learning started the field of reinforcement learning. Thorndike, whose work was briefly discussed in Chapter 1, probably was the first to describe the importance of learning to choose actions in order to achieve goals. In other words, instrumental conditioning concerns optimal choice, i.e. assigning actions to states to optimize goals. These considerations, together with the idea that learning should occur only when outcomes are surprising, led to the first efficient reinforcement learning model postulate: The Rescorla-Wagner (RW) model of pavlovian conditioning and its prediction error, which is the difference between the actual outcome and its prediction (Rescorla & Wagner, 1972). Published in 1972, the RW model was indeed able to account for a number of behavioral observations from the previous decades. Its main formulation can be written as:

$$V_X^{t+1} = V_X^t + \alpha_X(\lambda - \sum V_X^t)$$

with

- α: a learning rate, depending on the salience properties of the conditional and the unconditional stimuli being associated
- λ: the maximum conditioning possible for the unconditioned stimulus
- V_X^t: the associative strength at trial t

However, the RW model also has a few shortcomings, in particular [a] it does not allow higher-order conditioning (e.g. pairing a previously conditioned stimulus CS with a novel cue NC, in chains such as CS2 $\rightarrow$ CS1 $\rightarrow$ NC), and [b] it does not include sensitivity to potentially important temporal relationships within a learning trial.

To account for such problems, Richard Sutton proposed a temporal difference (TD) method, to handle longer action chains rather than learning simple stimulus-action associations (Sutton, 1988). As an illustration, Sutton wants us to imagine a weatherman, who wants to predict on Friday if it is going to rain on Saturday. He actually has a certain model to predict Saturday's weather, given the weather of each day in the week. With standard RL, the weatherman would wait until Saturday and then adjust all the other day models. However, by Friday the weatherman should have a good idea of the upcoming weather, and so he should be able to change Monday's predictions well before Saturday arrives (Sutton, 1988). TD learning thus uses locally obtained information and applies it to the information of a neighboring input. This allows to define the temporal difference error, a measure of the inconsistency between two predictions:

$$\delta(t) = r_t + \gamma V(s_{t+1}) - V(s_t)$$

where r_t is the reward observed at time t when in state s_t; and s_{t+1} being the next observed state of the environment. γ represents a *discount factor*, measuring the temporal devaluation of the reinforcement. The resulting learning rule can be written as follows:

$$V^{t+1}(s_t) = V^t(s_t) + \alpha\delta$$

In 1989, Watkins suggested a modified version of TD models, adapting it to control: Q-learning. Here, the agent learns the value Q(s,a) of each state-action pair, rather than only the value V(s) of each state s. The algorithm is updating these estimations as follows:

$$Q^{t+1}(s_t, a_t) = Q^t(s_t, a_t) + \alpha\delta$$

by a prediction error which is slightly different:

$$\delta(t) = r_t + \gamma \max Q(s_{t+1}, a_t) - Q(s_t, a_t)$$

This method is called *off-policy*, since the max operator used here signals that the difference is computed relative to what is believed to be the best available option. In the *on-policy* method (called SARSA), the prediction error is computed with the actual chosen action, even if this one is suboptimal. Since in Q-learning back-propagates the optimal value, independent of which action was chosen, it learns faster if a lot of exploration is performed. In a case where the reinforcement function is only dependent on the chosen action, and not of the state at t+1, i.e. when the next presented stimulus does not change according to the action, then we are only interested in γ=0, which simplifies the TD or Q-learning prediction error: δ= r - Q. This reasoning will be of a particular interest for the experimental part of this thesis.

Model-free vs Model-based Methods

Reinforcement Learning methods can be divided into two classes: model-based and model-free. Consider the *Vélib' dilemma* on a Friday morning. Based on previous experience, you have learned a mental map of bicycle roads in Paris. Model-based computation is similar to searching this map for a long-term goal. This kind of forward model has representations of the characteristics of the task, the probability of different transitions and varying immediate outcomes. Eventually after a couple of days, you constructed a favorite vélib road from home to the lab. However, on the fifth day, a friday morning, somewhere on your way near Bastille you encounter a huge protest march of joined teachers and SNCF railway workers, and the road is completely blocked. This is a momentary inconsistency inside the chain, which can be implemented with a model-free approach, an immediate negative outcome after you have had several successful bike-samples on this trajectory. Now after a few weeks, you have learned that for a given state "Friday", there is always a huge protest march, and the best action would be to take an alternative road (or the metro). In summary, the goal of model-free methods is the easier use in terms of online-decision making; it is a low cost alternative to model-based, which is always the better strategy.

Exploration vs Exploitation

After the agent chose an action in RL, he immediately knows the reward, and so the subsequent state, but he is *not* told which action would have been best in long-term. This means that a reinforcement-learner has to explore his environment explicitly to gather useful experience about all options to act optimally. Without exploration, the agent might get stuck in a local optimum. Even more, he has to find a best tradeoff between exploitation and exploration. Imagine a simple reinforcement problem: a k-armed bandit, i.e. a slot machine with k levers, each lever paying off 1 or 0 according to underlying unknown probability parameters. When the agent found a machine with sufficiently high payoff, should he choose it all the time or should he choose sometimes a different one to gather more information (cf. the great tit bird experiment by Krebs)?

Two post-hoc heuristics are widely used to solve such problems, without formal justification:

- **ϵ-greedy strategy**: reflects greedily choosing most of the time the action with the highest estimated value, reflected by the probability 1-ϵ, which leaves a probability ϵ to randomly choose another action. In other words, ϵ is the probability to choose randomly from all available actions rather than to choose another action. Such a heuristic parameter guarantees a reasonable amount of exploration. However, it treats all other possible actions as being identical, which might not always be the case. In a 3-armed bandit task, although bandit 2 seems to have a lower pay-off then bandit 1, it still might be much better than bandit 3.

- **Boltzmann exploration (softmax)**: is slightly more sophisticated than a ϵ-greedy strategy: here, the expected value Q for taking an action a is used to choose an action probabilistically according to the distribution

$$\frac{exp(\beta Q(a))}{\sum_{a'} exp(\beta Q(a')}$$

Here, β is a temperature parameter: with small β values, exploration increases. Unlike ϵ-greedy, the magnitude of the expected value influences the exploration behavior. This technique is very useful when the best action is clearly separated from the other options, however it can get a bit problematic if the values of actions are quite close.

Interestingly, neuroimaging and neurophysiology studies claim to have found the neural correlates of reinforcement algorithm signals, notably in the dopaminergic system which is thought to play a crucial role in conveying a prediction error signal. I will review some of these findings related to brain in the second part of this thesis.

2.3 Bayesian Approach

Definition

Every day, every second, the brain is facing an important task of great complexity: making sense of the external environment by integrating and glueing together pieces of information. External and internal noise sources provide uncertainty, as well as ambiguity in the stimulus itself, internal functions such as memory are distorted, speech processing is a highly challenging problem (e.g. a rich and complex acoustic waveform transformed into a phonological representation); in summary: nothing is totally certain. The goal of Bayesian approach to cognition is to model such information processing problems through the mathematical calculus of uncertain inference, namely probability theory. From a probabilistic viewpoint, beliefs are a matter of degree: each hypothesis *h* about the state of the world can be associated with a degree of belief *P(h)*. Such a probability distribution reflects any background knowledge about the chance that *h* is correct, thus they characterize *prior beliefs*. *h* might have implications for the data *D* we expect to encounter at a certain probability *P(D)* (prior beliefs of the data). This can be denoted by the likelihood *P(D|h)*: the probability of observing D given that *h* is the current state of the world. We are

interested in expressing the posterior *P(h|D)*, which represents our belief that *h* is the current state of the world after *D* has been observed. We can find this posterior from prior and likelihood by Bayes' theorem, as follows:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

respectively

$$P(h|D) = \frac{P(D|h)P(h)}{\int P(D|h)P(h)dh}$$

The latter equation makes explicit that *prior* and *likelihood* are sufficient to find the posterior. Bayes' theorem thus provides a direct method of calculating the probability of any hypothesis based on the *prior* belief that the world is a particular state, the *likelihood* function of observing data given this state, and the observed data itself.

Historical Background

Luck - or *Chance* - was long considered to be the domain of God(s), and anything divine was impossible to investigate. *Either God is or he is not. But to which view shall we be inclined?*, those are the words of Blaise Pascal around 1650. Pascal's viewpoint changed with a letter he wrote to Pierre de Fermat in summer 1654 (Devlin, 2008). Before this summer, the two famous French mathematicians were already discussing how to resolve the *problem of points*: Imagine two players playing a coin tossing game, and the best of seven wins. Unfortunately, the players are interrupted before they can finish tossing the coins. How should the prize be divided between them, if, for example, one player won three games, and the other only one? While discussing this problem, Pascal tries to give an algebraic solution to the problem in his letter, and thus gives birth to the probability theory. In 1670, he gives the answer to the question asked in 1650: *Weigh up the gain and loss involved in calling heads that God exists* or tails that he does not.

Figure 2.3: Blaise Pascal (left) and Thomas Bayes (right) -

Almost a hundred years later, Richard Price went through Reverend Thomas Bayes' papers and notes, where he discovered a theorem well worth to be published. He then arranged a posthumous publication in 1763 with the *Bayes' theorem*, later published in its discrete form by Laplace in his 1774 paper *Mémoire sur la Probabilité des Causes par les Evènements* (Laplace, 1774). Anecdotally, even though the theorem has Bayes' name, some science historians are more than ever doubting that Bayes' was the first to discover the Bayes' theorem (Stigler, 1983).

Either way, the adjective *Bayesian*, which I chose as title for this section, was not to be found in literature until the 1950s. Originally used with a pejorative connotation, it finally was embraced by a handful of statisticians to describe methods that revived inverse probability (for a deeper insight into the Bayesian connotation, see Fienberg (2006)).

After the neo-Bayesian revival in the 1950s, computational neuroscientists demonstrated a strong interest in Bayesian theories, which were used to describe neural mechanisms with probabilistic terms, such as lateral inhibition in the retina (Barlow,

1959). In contrast, human decision-making and learning had a long history of different models of conditioning, like the Rescorla-Wagner described in the previous section. Only recently, research started to reframe those models explicitly in statistical terms. Although such RL models are much easier to implement computationally in comparison to the burdensome Bayesian integrals, they only give a non-normative explanation, i.e. the *how* behind the behavior, not the *why*.

Nowadays there are strong arguments for applying a Bayesian approach to cognitive science, notably the progress in engineering sciences: Bayesian methods are successfully implemented in computer vision, speech recognition, information retrieval and so on. Although Bayesian theory provides the *goal* to the reverse-engineering of the brain, it still lacks a clear implementation process, which is important for reverse-engineering, as it is closest to what is observed.

Bayesian inference in uncertain environments

In the last decade, a growing body of work is trying to interpret conditioning with a Bayesian approach, referring to the behavior of humans and other animals as statistical deductions about the likelihood of reinforcement relative to their experience. The animal, as an intelligent problem solver, can make inferences about the structure of the environment, given some prior assumptions and experiences, and important question here is how an animal integrate the needs for change in task contingencies. A key notion in Bayesian inference is therefore *uncertainty*, which influences the estimation for volatility of the environment. In other words, unexpected events increase the internal uncertainty, and the longer an observer stays in an environment with high uncertainty levels, the higher the volatility estimate, which then, in RL terms, leads to a higher learning rate (Behrens *et al.* , 2007). Surprising events signal change, and as noted by Dayan and colleagues (Dayan *et al.* , 2000), stimuli with greater predictive uncertainty should be accorded greater attention and faster

learning. If a contingency is typically solid, then the observer only needs a few 'extinction' trials to get strong evidence in favor of a change in the state of the world.

The neural mechanisms involved in Bayesian computations are receiving increasing support. First, it was Körding & Wolpert (2004) who showed in a psychophysical experiment that humans indeed use Bayesian inferences through prior knowledge and knowledge about their sensory variability to deal with uncertainty. This concept has been incorporated by Ma *et al.* (2006) to propose a model of a neural system which uses *probabilistic population codes* to encode probability distributions over the variables of the environment. Yang & Shadlen (2007) finally provided empirical evidence for the extraction of probabilistic information from a set of stimuli in a neural system. They report a link between firing rates in the lateral intra-parietal area in a monkey and increasing evidence in favor of one outcome over another. Further, in neuroimaging, it was Behrens *et al.* (2007) who showed that volatility estimates from a Bayesian approach can be tracked down in the human anterior cingulate cortex. We are living in a highly complicated world, and since many of our decisions are made under uncertainty, understanding the basics of how the neural system treats uncertainty is essential to understanding its normal state of operation.

Hierarchy

As we have seen in the computational methodology, models can deal with simple situations, but also with more complex, hierarchical problems. In the next chapter, I will discuss the hierarchical organization of human decision-making, through behavioral implications and consequently models of hierarchy in the human brain. This will urge me to introduce the notion of *task*, a crucial factor in my thesis.

Chapter 3

Hierarchy and Task Set in Decision Making

3.1 Adapting your bolognese recipe

It is now well known that the prefrontal cortex is implicated in executive functions (also termed cognitive control), i.e. allowing flexible behavior by guiding action selection or thoughts in relation to internal goals and plans, rather then merely reflex-like acting based on the local stimulus environment. The evidence reviewed in this and the following chapters points to a hierarchical organization of cognitive control, implemented in a hierarchical, rostro-caudal axis in the prefrontal cortex.

To illustrate this theoretical framework, imagine you have just created your own special bolognese sauce for your pasta, and finally after many kitchen-hours of trying (and failing), it finally tastes heavenly. This recipe - or plan - for the whole project *making bolognese sauce for pasta* actually involves a huge sequence of sub-plans, which are again structured in sequences of actions. For example: sub-plan 1 might be: heating water for pasta, sub-plan 2 concerns *making the sauce*, which consists in serial actions (cutting onions - frying onions -frying the meat - adding red wine - etc.).

This can be seen as a hierarchical, temporal structure, since there is no point in frying the onions before cutting them! However, as you will find out very soon, your parents-in-law do not like your new special sauce, and so you adapt your recipe for the next special situation (for example you eventually figure that they don't like the amount of tabasco). You might adapt to this *episode* of your life by changing a fundamental action of your recipe (adding less tabasco), by completely changing the task (creating another dish), or by forcing the environment to adapt to you (looking for new parents-in-law). As one can easily see, these behaviors can be very concrete or very abstract (or anything in-between), temporally structured in different ways, and underlying a certain context. Let's have a closer look at those different kinds of behaviors.

3.2 Hierarchical organization of sequential behavior

Replacing the behaviorist's black box

Sequential behavior is a fundamental aspect of human activities from everyday skills to complex problem solving. Even more, it is an important part of learning in many task domains: planning, robotics, speech recognition, DNA sequencing, etc. According to the view of behaviorists such as Pavlov and Watson in the first half of the 20^{th} century, the functional mechanisms underlying this behavior is the reflex chain. This view got challenged by the experiments of Tolman (Tolman, 1948) as discussed in chapter 1, but more importantly the view changed with a classic paper from Karl Lashley, a neurophysiologist at Harvard University (Lashley, 1951), where he writes: *The input is never into a quiescent or static system, but always into a system which is already actively excited and organized* [...] *Attempts to express cerebral function in terms of the concepts of the reflex arc seem to me doomed to failure because they start with the assumption of a static nervous system*. Lashley proposes three blocs of evidence:

- Movements can occur even when sensory feedback is interrupted

- Some movement sequences occur too fast for elements of the sequences to be triggered by feedback from the preceding elements
- Errors in behavior suggest internal plans for what will be done later

To illustrate his critic of the associative chaining theory, Lashley wants us to imagine a professional pianist playing for example Rachmaninov's Fourth Piano Concerto. The pianist's finger can move at a rate of about 16/second, which is too fast for a closed-loop proprioceptive feedback control of movement. Each movement is executed before the previous step in the sequence of motor acts could have informed a central system about its valid completion. Lashley further argues that a hierarchical structure can be observed in behavior, including nested sub-behaviors. Similarly, a couple years later, Keele conceptualizes a motor program, which he characterizes as *a set of muscle commands that are structured before a movement sequence begins* (Keele, 1968). Thus, such a sequence can be executed without being influenced by peripheral feedback. In his 1960 book, Miller embedded these ideas into the cognitive concept of TOTE, Test-Operate-Test-Exit (Miller *et al.* , 1960). Imagine hammering a nail into a surface:

Test: Is the nail flush with the surface? - **No**.
Operate: continue hammering the nail
Test: Is the nail flush with the surface? - **Yes**.
Exit: Stop hammering.

TOTE was thought to replace the stimulus-response concept as basic unit of behavior. In a system with TOTE behavior, short term memory is limited, planning is a fundamental cognitive process and, importantly, behavior is hierarchically organized in chunks (or TOTE units). Specifically, scientists moved away from schemes of only low-level motor behaviors, and instead got interested in higher-level behavioral units. For example, *making a nespresso coffee* covers representations involving

a *sequence* of actions like putting the coffee pad into the machine, filling the water reservoir, pushing a button, imagining a highly stimulating discussion with George Clooney.

Sequence Planning and the Tower of London

In order to investigate planning behavior and sequences of actions, a task often used is the Tower of London (fig. 3.1) by Shallice (1982), an adapted version of the Tower of Hanoi, which is especially helpful in investigations of the executive functions in neuropsychological patients. The task was designed to generate various sequences of hypothetical states and the corresponding consequences, and thus favor the development of a structured event which then can guide movement from start to a goal state. The participant sits in front of three pegs of different sizes, through which can be put one, two or three beads respectively. The three beads of different colors are arranged in a start position, and the participant is now asked to put them into a final configuration. As one can easily see on fig. 3.1, some configurations are straightforward, whereas others are not achieved without thorough planning.

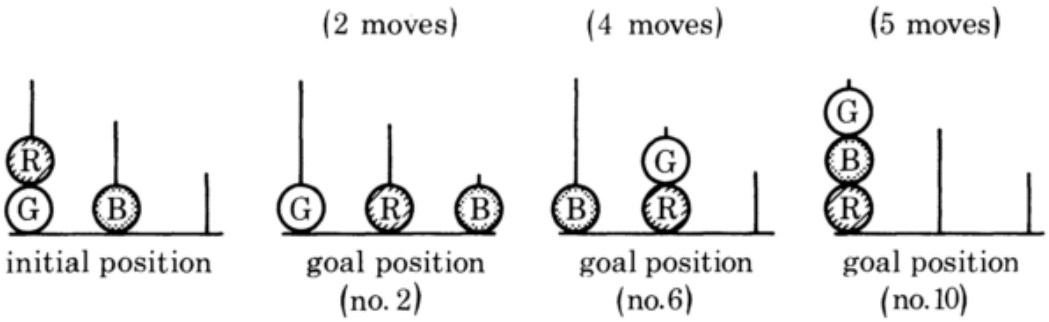


Figure 3.1: The Tower of London, as proposed by Shallice - an illustration of three different problems of the Tower of London test, with the same starting position

In 1997, Dehaene & Changeux (1997) propose a sequence planning model of three hierarchical steps, which can solve the Tower of London problem:

- **Gesture Level**: Simple sensory-motor coordination. This level is able to move a bead from one position to another.
- **Operation Level**: Sequence of basic gestures. This level codes for direct moves of several beads.
- **Plan Level**: Goal-directed exploration of a tree. This level computes indirect moves, trying out different possibilities (e.g. placing a bead on a location different from the goal, in order to find a solution for a remaining bead)

The hierarchy of this model goes from basic actions to testing different options in a tree, where the latter is the supreme level, i.e. to which the others have to *obey*. The actual planning is done on the highest level, which is made possible through a working memory system able to stock actual positions and evaluate future actions. The strength of the model comes from its efficiency in explaining the behavior of patients with prefrontal lesions, which are supposedly impaired in activating plan units.

Hierarchical organisation of action sequences

A few years later, more general hierarchical models were proposed, applicable to a whole range of data. Probably the two most discussed models are introduced on the one hand by Cooper & Shallice (2000) and on the other hand Botvinick & Plaut (2004). In the studies by Cooper, goals are decomposed into subgoals in a tree-like fashion. However, *tree* has to be understood in a much larger sense, e.g. some steps might be executed at the same time, or some behavior from a different tree might use subgroups of this tree. Now, as can be seen on fig. 3.2, the problem of this model is its inability to learn, because of the reliance on a built-in structure: goals and subgoals are rather explicit. The sequence mechanisms are inflexible, and the model would have difficulties dealing with task domains where details of subtask performance are depending on a larger task context.

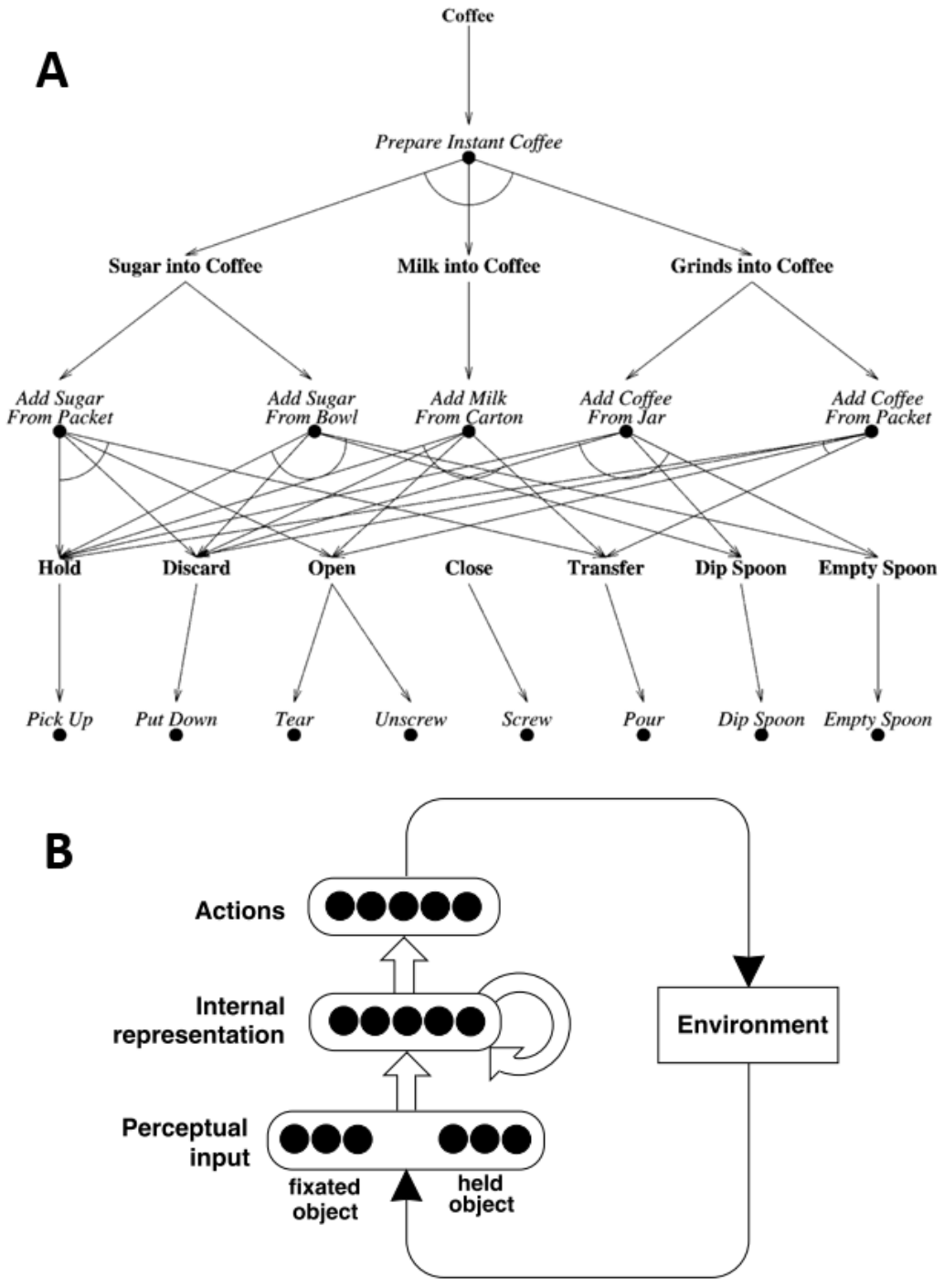


Figure 3.2: - - A. model of Cooper & Shallice (2000), illustrated by the goal "making coffee"; **B**. model of Botvinick & Plaut (2004), with emergent system properties, rather than explicit representations

Those are the critics of Botvinick & Plaut (2004), who then propose a simple recurrent network model. Here, the learned, distributed internal representations are used to map input to the output, and the recurrent connections make it possible to maintain information about a temporal context. An illustration of their implicit sequences *perceptual input - internal representation - actions* (fig. 3.2 B) might be for example: *viewed*: coffee, *held*: spoon, *action*: stir, which equals the representation "stir coffee using spoon", followed by the corresponding action. However, this is exactly the point criticized by Cooper & Shallice (2006): the total absence of explicit goal representations.

Hierarchical reinforcement learning

Before performing such sequences of behavior to obtain a goal, the structures have to be learnt. In the former chapter, I introduced the concept of reinforcement learning (RL), a computational model of learning simple behavior through a prediction error. If now sequences of such actions have to be learnt, a hierarchical model is needed (e.g. Sutton *et al.* (1999)). In such (model-based) hierarchical reinforcement learning (HRL, fig. 3.3), the goal is to expand the panoply of available actions in order to perform not only lower-level actions, but subroutines, i.e. sequences of such simple actions. Such temporally abstract actions can be referred to as options (Botvinick *et al.* , 2009), which are selected in the same way as simple actions, but they remain active until an option subgoal state is reached. In other words, HRL can be seen as learning on two levels: On the first level, the agent wants to accomplish overall task goals, and therefore he needs to select the corresponding subroutines. For example, if the goal is *going from the lab to the cinema at Les Halles*, then one of the subroutines to select might be *exit from metro 7 and walk to cinema*. Second, the agent has to learn the subroutines themselves. For our example, this would mean to find your way underground from the Metro station *Châtelet* to the exit *Les Halles*, which leaves you with many turns and options (until after a while you have learnt the quickest way).

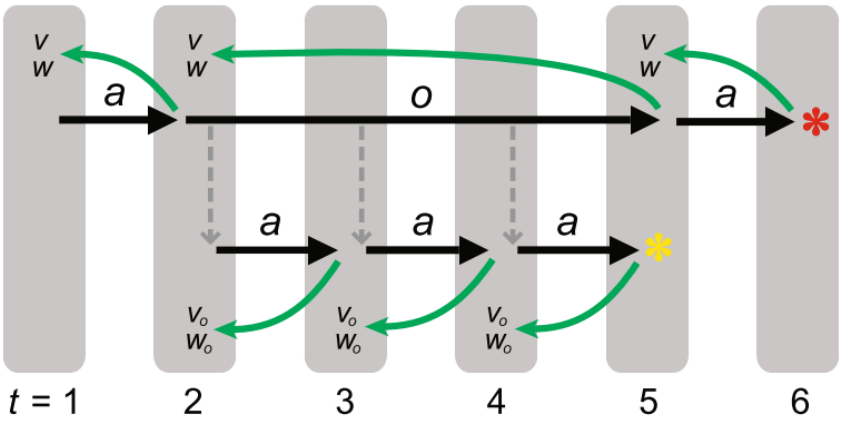


Figure 3.3: schematic illustration of Hierarchical Reinforcement Learning dynamics - At t=1, the agent performs a simple action, upon which he receives a prediction error (green arrow) at the next state, used to update value V and association strength w. At t=2, the agent selects an option O, which actually consists in a sequence of simple actions. Once the option's subgoal is reached, a prediction error of the whole option is computed (long green arrow) Botvinick *et al.* (2009)

Recently, some studies investigated the HRL with imaging techniques, to look for differences between variables such as a global prediction error and subgoal prediction error (Gläscher *et al.* , 2010; Ribas-Fernandes *et al.* , 2011). Further, Daw *et al.* (2011) tried to show that, in a multistep decision task, human choice behavior is best explained by a combination of model-based and model-free reinforcement learning, differentiating in the corresponding brain activity.

Besides the here discussed temporal hierarchy of behavior, another important point is the hierarchy in a conditional organization of behavior, be it in a learning environment or cognitive control. I will now briefly discuss the recent ideas and studies behind this topic.

3.3 Hierarchical organization in conditional action selection

The functional organisation of cognitive control has been theorized in many ways during the last decades, and recently mapped onto the brain (which I will deeper investigate in the next chapter). For example, Fuster (2001) writes of a functional hierarchy that is progressively more integrative. The highest level of this hierarchy is implemented in the prefrontal cortex and dedicated to the representation and execution of actions. On the lowest level, there are *phyletic motor* actions, on a higher level *plans*, and finally on the most complex level there is *conceptual* behavior. Badre & D'Esposito (2007) go further in the notion of abstraction and propose four levels of representational hierarchy:

- **Response**: selection of a stimulus-response mapping
- **Feature**: selection of a set of perceptual stimulus-to-response mapping
- **Dimension**: selection of a set of features, based on a perceptual cue
- **Context**: selection among sets of contextual cue-to-dimension mappings

The more abstract levels are recruiting more anterior, rostral prefrontal areas in the brain. However, in this hierarchical model the action selection depends completely on the level of abstraction, parallel treatment of information is not possible. This is taken into account in the cascade model by Koechlin (e.g. Koechlin *et al.* (2003); Koechlin & Summerfield (2007)). Indeed, four fractioned levels of control can be contrasted in parallel:

- **sensory control**: selection of an action in function of a stimulus. *the telephone rings, you pick up*
- **contextual control**: selection of a set of stimulus-action associations (i.e. *Task-Set*, cf. next section for a detailed introduction) in a given context. *the telephone rings at your friend's place, you might not pick up*

- **episodic control**: selection of a set of features, based on a perceptual cue. *the telephone rings at your friend's place, who (episodically) is having a shower and told you to pick up during that time*
- **branching control**: stand-by of a task until another is performed. *the telephone rings at your friend's place, who (episodically) is having a shower and told you to pick up during that time, but as he remembers that it might be his ex-girlfriend, he screams from the shower not to pick up this call*

Critically, the authors demonstrate that this structure is independent of the complexity of the task, but relies only on its hierarchical level, since the amount of information needed for proper execution of the task is matched. Furthermore, they confirmed that this hierarchical structure is implemented functionally in the brain in a cascade fashion, i.e. the most anterior regions of the prefrontal cortex influence the more posterior areas in a top-down manner, as can be seen in fig. 3.4. For example, on the last hierarchical level, *branching*, the task on stand-by while another is executed is encoded in the frontopolar regions.

3.4 Task Sets: Cue-based vs Value-based choice

Task Set: Definitions

When talking about different levels of abstraction in human behavior, we have to distinguish between *actions* and *task sets*. Searching for a definition of task set, one finds:

- **Rogers & Monsell (1995)**: an effective intention to perform a task which is accomplished by configuring ones mental state (e.g. attentional settings) to be in accordance with the task-specific operations which define the to be performed task when several task responses are possible
- **Sakai (2008)**: a configuration of cognitive processes that is actively maintained for subsequent task performance

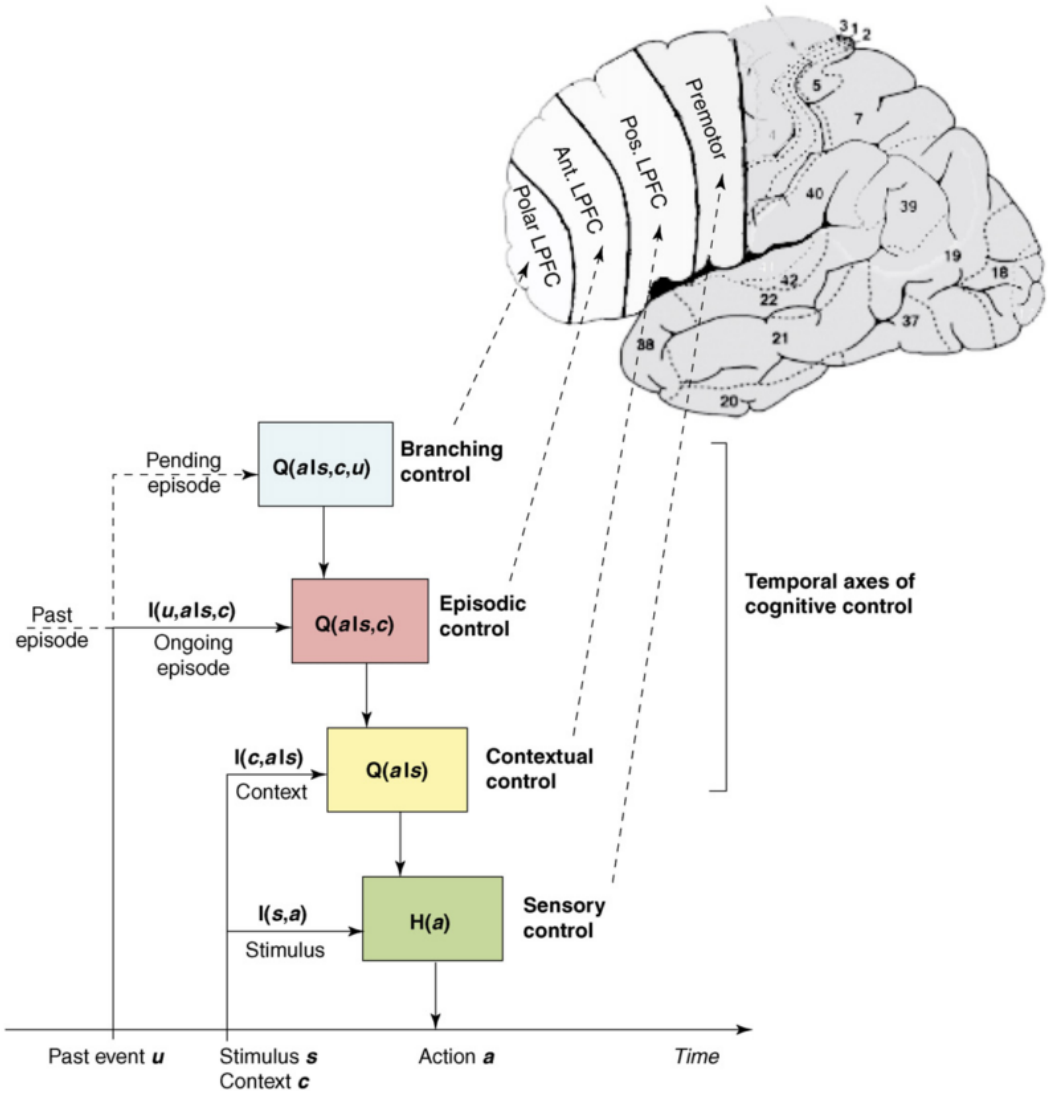


Figure 3.4: Hierarchical Cascade Model by Koechlin - taken from Koechlin & Summerfield (2007)

- **Collins & Koechlin (2012)**: active representations of behavioral strategies stored in long-term memory

These definitions lead us to the consideration of a task set as a sort of mapping of definite response decisions in the context of a specific cognitive task. A task-set has to be seen as a *whole*, with fixed configurations of actions. One important question lies in characterizing the cognitive resources implicated in alternating between two tasks, or *task-switching*. In the next section I will present different aspects of task-switching, like the cognitive cost of switching, or the learning of alternating sequences.

Cue-based vs Value-based Task switching

in general it is found that shifting back and forth from one mental set, one attitude or task to another, is a relatively ineffective mode of work
Hollingworth & Poffenberger (1919)

Back in 1927 we find one of the first studies mentioning a task switching paradigm (Jersild, 1927). It was Arthur Jersild who, in his writings *Mental set and shift*, analyzed some of the control processes that are needed to reconfigure mental resources in order to change a task. In his paradigm, he compared the duration of bloc trials in which the subject changes rapidly between tasks with blocs in which the subject performs only one task, but of course this comes with comparison confounds such as switch costs or differences in working memory load.

In general, there are two paradigms used today to study task switching: First, the *cue-based* task choice, where a cue indicates which task to choose. Different paradigms exist herein, such as the *alternating-runs paradigm*, where the task changes every N trials, or the *task-cueing paradigm*, where a task cue is presented shortly before or at the same time as the stimulus to indicate which task to choose in a trial-by-trial manner. As an illustration I might present the *Stroop Task*, in which color words such as *blue* are presented in different ink colors, and the subject usually has

to ignore the meaning of the word and name the color. Color words might be written with the ink of the written color (congruent), or in a different ink color (incongruent). In an adapted Stroop Task, the experimenter might put a cue before the appearance of the stimulus word, indicating if the subject should now name the ink color or the written color word. In such task, a typical switch effect can be observed: the *switch cost*, which is usually reflected in longer response times after a switch compared to repeated performance of the same task. Even more, as shown by Hyafil *et al.* (2009), an additional increase in reaction times are observed when a switch is occurring after two consecutive incongruent trials.

Second, the *value-based* task choice, where a subject freely chooses between two or more tasks, without any cue indication. In such paradigms, the subject is instructed to maximize the reward associated with each task, knowing that the reward contingencies can change. The *Wisconsin Card Sorting Test* represents an often used example, especially in the field of clinical neuropsychology, to detect prefrontal problems. Initially, stimulus cards are presented to the subject, varying in three dimensions: color, quantity and design. Milner (1963) aready used this task to test how patients with distinct brain lesions performed on such a problem. The subject is asked to sort cards, but he does not know by which rules (or by which dimension), he is just told whether it was correct or not. During the test, the matching rules are changed, and the subject has to learn the new rules. His mistakes are analyzed to get a clinical score. But even on a much more complex, non-laboratory level, this behavior can easily be illustrated; just think of animal foraging or the trader's behavior on a stock market.

Since this framework makes reference to a subjective value, or an internal state, the use of computational models is of major importance in order to understand such behavior or to track those variables in the brain. Typically, a gambling task is used, like a four-armed bandit in Daw *et al.* (2006), where each arm has a stochastic payoff

value. In this study however, one important point is the trial-by-trial independent diffusion of the payoff values of each arm. Therefore, the participant has to sample from all the four arms in order to get an idea of each arm's value. This reasoning permits to study periods of exploitation and exploration, i.e. switches to options that might be less rewarding than the currently chosen.

PROBE Model

However, most proposed models are limited when asked to explain human decision-making in abstract everyday environments, where one has to deal with uncertain, changing and open-ended situations. To overcome the difficulties of such an environment, Collins & Koechlin (2012) proposed a model which predicts human behavior with inter-individual differences, creates new task sets (*strategies*) and switches between learned ones: the PROBE model. The model controls creation, learning, storage, retrieval and selection of strategies, but consistent with literature on the limits of executive functions, only a limited number of concurrent task sets can be monitored. In this model, such a behavioral strategy consists of (see fig. 3.5)

- a selective mapping: stimulus-response associations
- a predictive mapping: expected action outcomes, given stimuli
- a contextual mapping: external cues predicting task set reliability

Importantly, they distinguish two distinct phases: before and after acting. Before acting, the task set *reliability* is evaluated through Bayesian inference, according to the environment's volatility and the context. Selection always happens first at the task level, then at the level of actions within a chosen task set. After acting, the stimulus-response associations (selective mapping) are adjusted through *model-free* reinforcement learning, outcome predictions are updated for predictive mappings, which so evaluate the new task set reliability. Substantially, the model successfully

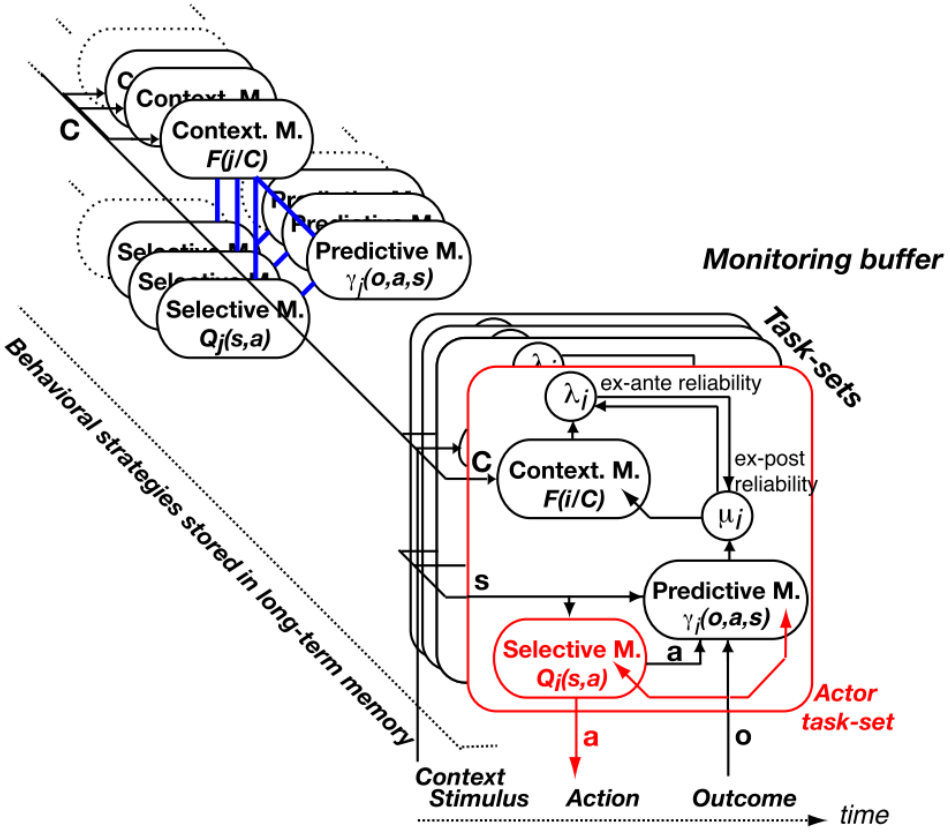


Figure 3.5: Architecture of task sets after Collins & Koechlin (2012) - In the red square, the currently used task set, preferred to the other strategies in the monitoring buffer, which can store a limited number of behavioral strategies. Each strategy comes with three mappings (M.): selective, predictive and contextual. The model monitors the reliability of the task sets at two points in time: before (*ex-ante*) and after (*ex-post*) acting. *ex-ante* is inferred from the *ex-post* of the preceding trial (according to contextual cues and the perceived volatility of the environment); *ex-post* is inferred from *ex-ante* before action and according to action outcomes. *ex-ante* of the task sets is used to choose an actor task set, which has a selective mapping determining the responses.

shows that human decisions are binary, based on the task set reliability: action selection relies on a *satisficing* criterion, i.e. if one task set appears to be reliable, it is chosen ($p > 0.5$), if no task set fulfills this criterion, a new one is created, involving *model-based* reinforcement learning based on outcome predictions.

Part II

Neural Correlates of Decision-Making

Chapter 4

Imaging the Brain

4.1 A brief history of functional imaging

Over the last two decades divers methods for imaging functional activity in living brains have given birth to many interesting insights into cognitive functioning. Magnetic Resonance Imaging (MRI) has proven to be the most important imaging progress since the discovery of X-rays by Conrad Röntgen in 1895. In 1973, the chemist P. Lauterbur was the first to describe the qualitative basics of this new technique (Lauterbur & Others, 1973), for which he would earn the Nobel Prize 30 years later. However, it still needed another twenty years until Seiji Ogawa, a researcher from the Bell Laboratories described how to take advantage of the magnetic properties of hemoglobin (Ogawa *et al.* , 1990), a technique which was then applied in 1991 by Belliveau and colleagues to show the first quantitative images of cerebral hemodynamics coming from an MRI machine (Belliveau *et al.* , 1991) (fig. 4.1). Prior to these advances researchers were limited to draw conclusions about the functions of different brain regions through patients with brain damage or from invasive studies in the primate brain. As an illustration of the impact that fMRI has on research, I performed a database query (Web of Science) with the keywords *fmri* or *functional MRI* or *functional magnetic resonance imaging*, resulting in a total of 47612 papers from 1991 until early 2012 in peer-reviewed journals (in english language). In 1992, 4 papers

were published, but this rate changed quickly and is still rising: over 6000 papers published in 2011, which is over 16 accepted papers per day.

However one has to be aware of the actual limitations of this methodology and of its interpretability. Even if the colorful blobs create a popular fascination and many speculations in the press, Nikos Logothetis reminds us that: *"fMRI is not and will never be a mind reader [...] nor is it a worthless and non-informative neophrenology"* (Logothetis, 2008). The experimental design is crucial to the interpretability of BOLD data, as is understanding the limitations, due to the circuitry and the functional organization of the brain (Logothetis, 2008). In the following section I will briefly try to give a closer look at what exactly we are measuring with fMRI.

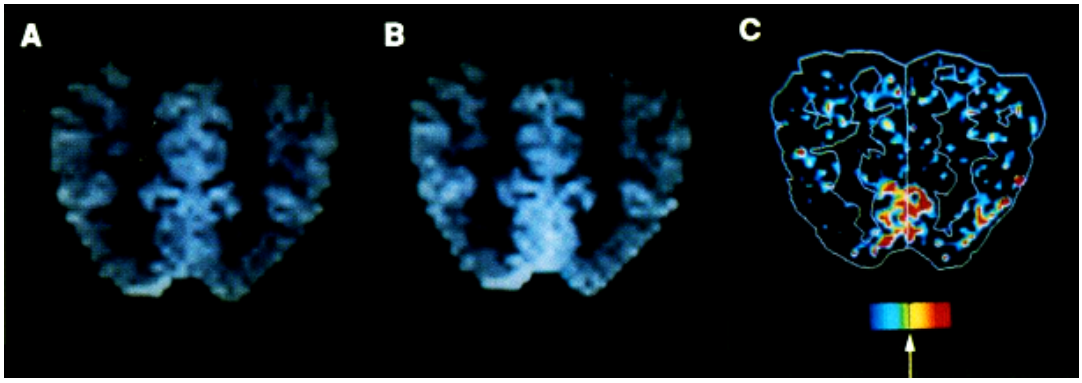


Figure 4.1: First fMRI maps of human brain activation during a passive task - A. Maps of the brain during darkness, B. during 7.8 Hz photic stimulation, C. Contrast image of changes in cerebral blood volume (B-A). taken from Belliveau et al., Science Paper from 1991 (Belliveau *et al.* , 1991)

4.2 The magnetic resonance signal

Several techniques of functional MRI scanning are used today, but the most important is based on a blood oxygen level detection (BOLD) signal, which indicates

changes in the ratio of oxyhemoglobin to deoxyhemoglobin. Deoxyhemoglobin being paramagnetic, while oxyhemoglobin is not, changes in local magnetic properties can be seen on an MRI scan, since increased blood flow to an activated brain region results in more oxygenated blood than is immediately necessary for local metabolism. In a typical MRI session like in our studies, a subject is placed in a strong vertical magnetic field (usually 1.5 or 3 Tesla; the earth's magnetic field in comparison is 0.00005 Tesla). In the brain tissue the protons, which are the nucleus of the hydrogen atom, are aligned vertically in this field. A horizontal radio frequency pulse is then applied to the tissue in order to put the protons in a synchronously rotating state in the horizontal plane, so that they are in phase. When this radio frequency is turned off, the rotating protons fall out of synchrony (dephasing time constant T_2) and return to their lower energy state, aligning with the original magnetic field ("righting" time constant T_1, which is much slower than T_2). It is exactly the rates of these two relaxation processes that are measured by an MRI. For functional MRI this means that, since deoxyhemoglobin and oxyhemoglobin have different effects on the dephasing of the protons, differential magnetic resonance signals can be measured: Deoxyhemoglobin supports efficient dephasing of the rotating protons, which results in a steep T_2 curve and so a weak magnetic resonance signal, while an a stronger signal emerges with increasing proportion of oxyhemoglobin.

For the study illustrated by fig. 4.1, this leads to the conclusion that, in simplified words, during photic stimulation, neurons in the occipital areas become active, the blood flow increases and so the proportion of deoxyhemoglobin molecules decreases. Therefore, the dephasing of the protons is slower in these regions and the measured magnetic resonance signal is stronger. However, one has to keep in mind that fMRI is only an indirect measure of neural activity. There is still a lack of clear understanding of the neurometabolic and neurovascular coupling underlying BOLD signals. Many questions are waiting to be answered: Which cell types are generating the neural activity captured by the BOLD signal? How are energy demands linked to

such activity? Recent studies are shedding light on these issues, for example the recent work by Viswanathan and Freeman (Viswanathan & Freeman, 2007): Through electrical measurements in monkeys' visual cortex, they showed a strong coupling between local field potentials and changes in tissue oxygen concentration. They concluded that the signal measured by fMRI probably reflects the input and intracortical processing in a given brain area, rather than the output firing of projection neurons. Such studies are essential in understanding the limitations and creating correct interpretations of hemodynamic signals.

4.3 Model-based fMRI

A recent development in neuroimaging research, the model-based approach, makes it possible to correlate signals derived from a computational model against fMRI data. The main argument in favor of such an approach is a progress in psychological and computational theories of brain functions, in a way that could not be achieved with conventional neuroimaging analyses or traditional behavioral studies (fig. 4.2) (O'Doherty *et al.* , 2007). The goal is to not only reveal *where* a specific process is located in the brain, but also *how* it is implemented. This is of particular interest in the field of decision-making, where different computational models of reward learning and expected value computing have been proposed. Indeed, including parametric regressors derived from such a model into a fMRI data analyses, can help to reveal regions implicated in such internal operations.

A simple example would be the Rescorla-Wagner model (seen earlier, (Rescorla & Wagner, 1972)) or related temporal difference model where learning occurs through updating expectations about an outcome through a prediction error. After fitting the model parameters (e.g. learning rate), the different model variables such as updated expectation and prediction error can be regressed against fMRI data, after convolution with the hemodynamic response function, in order to determine brain

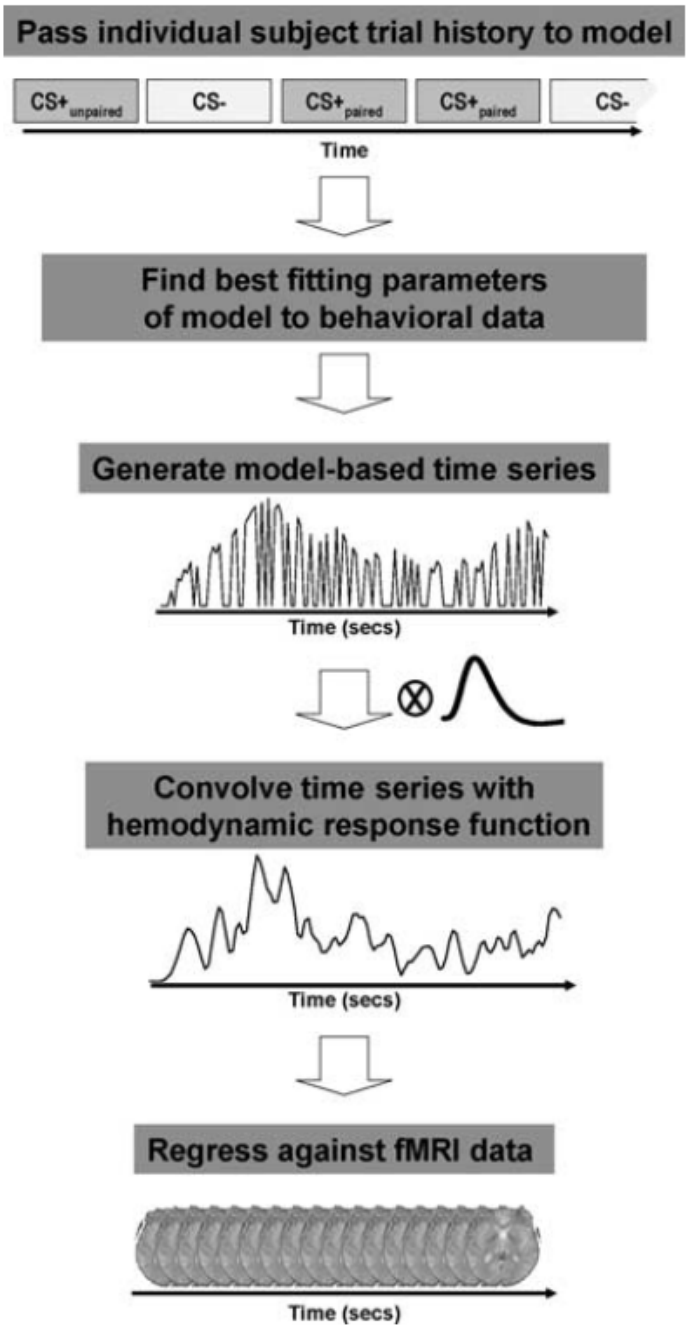


Figure 4.2: Illustration of model-based approach - taken from O'Doherty et al., 2007 (O'Doherty *et al.* , 2007)

regions for which BOLD changes correlate significantly with the model-predicted time series. For example, such prediction error signals in humans have been found to correlate with activity in the ventral striatum (O'Doherty *et al.* , 2003c).

4.4 Connectivity Analyses

Functional and Effective connectivity

It would be erroneous to conclude that activity alone is sufficient to attribute a function to a specific area. Over a hundred years ago, in August 1881 on the International Medical Congress in London, Friedrich Goltz and David Ferrier already fought and discussed the difficulties of attributing a function to an area considering the dependence of the cortical activity on the underlying connections (Phillips *et al.* , 1984). Goltz accepted the results of movements elicited by electrical stimulation of dog and monkey cortex, but he was not satisfied with the conclusions, replying that the excitation might have spread physically through the brain into distant centers, or that the origin of the movement might lie in afferent or efferent pathways. This intellectual fight resulted in ablation studies, but even then the localization of a function was tricky since sometimes restitutions of functions were observed after ablation (anecdotally, the German physiologist Hermann Munk had a rather creative explanation for this: each center with a specific function is surrounded by a *virginal cortex*, that only starts to function when the usual center is destroyed (Star, 1989)). Despite of the progress made in neuroscience, the question still remains today, and the discussion continues. In a recent review, Borchers and colleagues pointed to the fact that direct electrical stimulation (DES) of the human brain, a technique first used in 1874 and of great value for neurosurgery, does not allow us to draw unequivocal conclusions about the type of processing in a stimulated area (Borchers *et al.* , 2012). Now, can the physiological changes we observe after sensorimotor or cognitive tasks be explained by functional segregation alone or by integrated and distributed changes regulated by underlying connections (Friston & Büchel, 2003)? To

help answering such questions, new techniques, especially for fMRI, were created, which permit us to make conclusions about functional and effective connectivity.

In order to differentiate between those two types of connectivities, Friston and Büchel proposed the following definitions (Friston & Büchel, 2003): Functional connectivity is defined as *temporal correlations between spatially remote neurophysiological events*, whereas effective connectivity is *the influence one neuronal system exerts over another*. This implies that functional connectivity is simply an observed correlation, but it does not say anything about how these correlations are mediated. Importantly, as they note, functional connectivity is not necessarily due to effective connectivity (e.g. common neuromodulatory input), and if it is, effective influences might only be indirect.

In the following sections, I will briefly introduce two types of connectivities which I performed on our fMRI data: PsychoPhysiological Interactions, which might be seen more as a functional connectivity (or as very limited models of effective connectivity), and Dynamic Causal Modeling, an effective connectivity technique.

4.4.0.1 PsychoPhysiological Interactions

The general idea of PsychoPhysiological Interactions (PPI) as the name states is to capture an interaction effect between the neural activity from a cerebral region and a psychological factor. Here, it is important to notice that the interactions in the brain take place on a neural, and not hemodynamic level. Therefore to perform an adequate analysis, the neural signal as captured by the General Linear Model must be derived by deconvolving the hemodynamic response function before calculating the interaction term. Fortunately this procedure is now incorporated in softwares such as SPM. Furthermore, to prevent the interaction term to be confounded with the main effects, these variables have to be included in the new GLM as well.

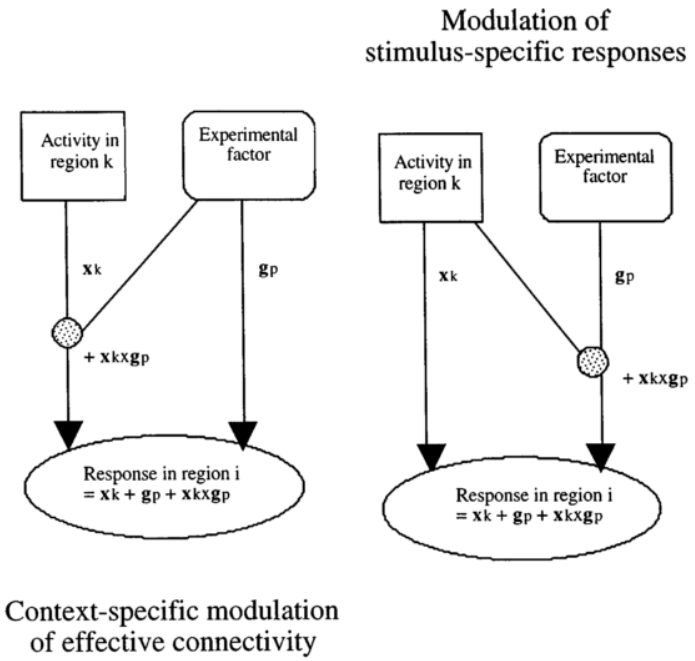


Figure 4.3: Two alternative interpretations of PPI effects - (Friston *et al.* , 1997)

In summary, a PPI analysis should include four steps: [1] performing a standard GLM analysis, [2] extracting the BOLD signal from a source region identified through the GLM analysis, [3] calculating the interaction term (source x psychological factor) and [4] performing a second GLM analysis including the interaction term, the source's extracted signal and the psychological factor (and possibly all other confounds). This leaves us with two possible interpretations of the results (fig. 4.3): [a] how the contribution of one region is altered by the psychological factor or [b] how an area's response to an experimental factor is modulated by input from another region.

4.4.0.2 Dynamic Causal Modeling

The basic idea of Dynamic Causal Modeling (DCM) is to construct a reasonably realistic neuronal model of interacting cortical regions, using a generic Bayesian framework to infer hidden neuronal states from measured brain activity.

Stephan and colleagues defined five key features for DCMs (Stephan *et al.* , 2010):

- DCMs are dynamic, using differential equations for describing neuronal dynamics
- they are causal (in the sense of control theory), i.e. they describe how one neuronal population *causes* the dynamics in another and how experimental manipulations can modulate these interactions
- the goal of DCMs is to have a valid neurophysiological interpretability
- DCMs use a biophysically motivated and parameterized forward model
- DCMs are bayesian, where each parameter is constraint by a prior distribution.

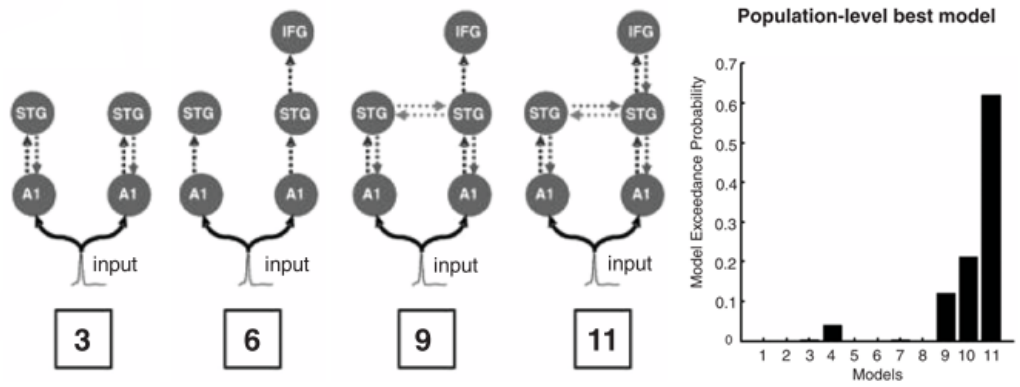


Figure 4.4: an example of DCM - Illustration of 4 models (out of 11 tested models) with different connectivities. Right: Through Bayesian Model Selection, the *winning* model, and therefore explaining best the observed data, is model 11 (Boly *et al.* , 2011)

The least to say is that DCM is relatively complex compared to conventional analysis techniques. Therefore it is important to know a bit about its theoretical foundations. For example, there is no point in using DCM without a more or less clear hypothesis. Performing DCMs can be quite an adventure in the beginning, since to be in line with one's hypothesis and the experimental design, it is crucial to choose an appropriate method for group-level inference on model structure and parameters, leading to correct interpretations. Once this has been integrated, DCM can be a powerful method for inferring causal mechanisms in systems, whose dynamics are observed indirectly (fig. 4.4).

Chapter 5

The Human Brain in Decision-Making

5.1 The Human Brain: General Approach

Figure 5.1: First written account of the brain - as found on the Edwin Smith Surgical Papyrus, written around 1700 BC in Ancient Egypt

Dated at approximately 1700 BC, the Edwin Smith Papyrus (fig. 5.1) is believed to be the oldest known surviving scientific treatise on surgery, specifically on head traumas. It is even thought that the anonymous Egyptian author was indeed making a copy of a much older document, which origins lie around 2700 BC (Feldman & Goodrich, 1999). The Edwin Smith Papyrus (named after its discoverer) is of great interest and value to neurosurgeons; through its descriptions of brain and spinal

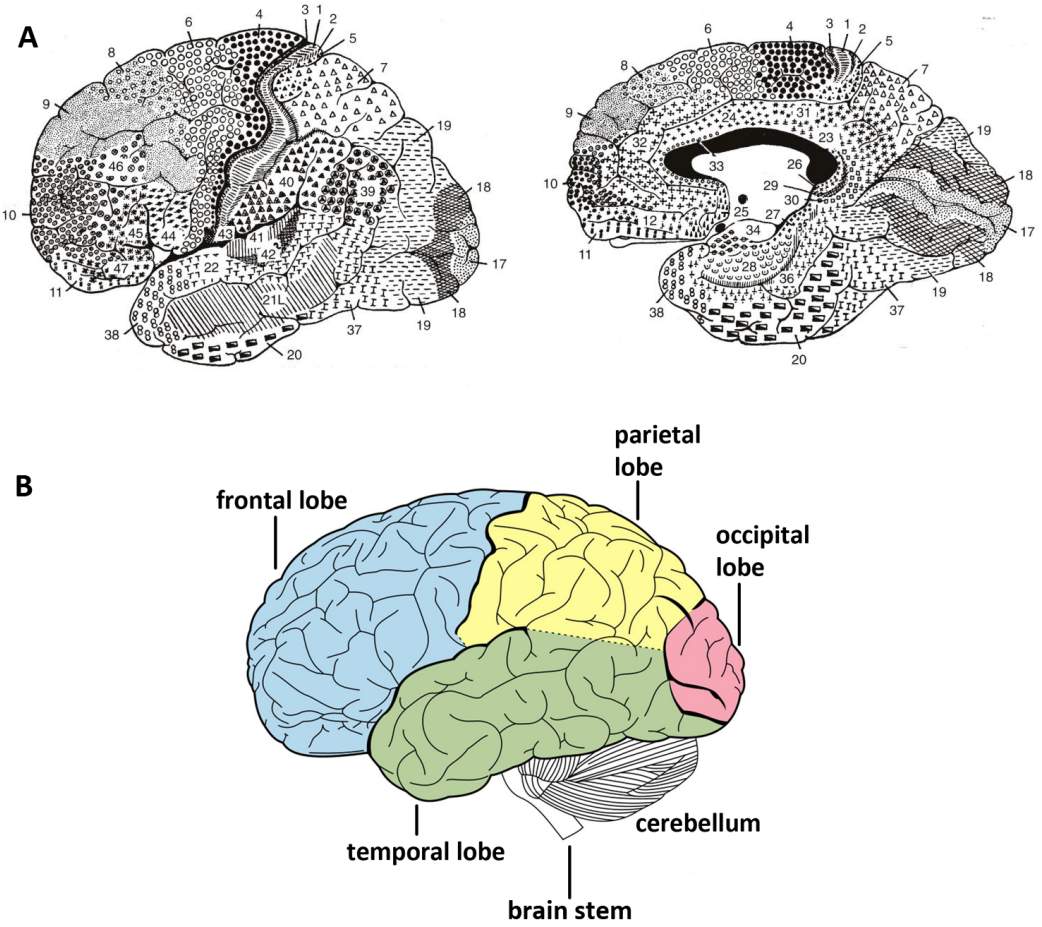


Figure 5.2: Views of the brain - **A**. lateral and medial view of the cortex, labelled and numbered after Brodmann (1909); **B**. The neocortex consists of four lobes: frontal, parietal, temporal, and occipital. It can be further subdivided into structurally and functionally distinct areas mediating our most important mental abilities. The surface of the human neocortex is about ten times larger as that of a macaque monkey, and around a thousand times larger as that of a mouse.

cord trauma and infections, neurosurgeons can look through a window in time and wonder at the insights and challenges faced by their ancestors at the dawn of a scientific medical profession.

But in the early days of medicine and philosophy, not much attention was given to the brain; the egyptians even removed it while mummifying their dead. Distancing from his teacher Plato, Aristotle believed the brain to be a less important organ, mainly serving as a cooling agent for the heart - the seat of intelligence, motion, and sensation. Over two millennia later, it is now clear that the brain is the center of the nervous system in nearly all animals (all vertebrates and most invertebrates).

The human brain has three main structures: the cerebrum, the cerebellum and the brainstem (fig. 5.2); and it is divided into two symmetrical hemispheres, interconnected by the inter-hemispheric commissures. The cerebral cortex, that is the 2 to 4 mm thick top-layer of the cerebrum, is usually referred to as *grey matter*, made up of a dense assembly of neuronal cell bodies, consisting of about 20 billion neurons (Koch, 1999). The other part of the cerebrum, the *white matter*, consists of glial cells and myelinated axons. They enable communication between neurons from one region of the cerebrum to another, and between the cerebrum and lower brain structures. The cerebral cortex is divided into four lobes (fig. 5.2): the frontal lobe (including the prefrontal cortex), separated from the parietal lobe by the Rolando fissure. The temporal lobe, separated from the parietal and the frontal lobes by the Sylvius fissure., and finally the occipital lobe, which is mainly implicated in vision functions. Furthermore, the cerebrum has additional internal structures, such as: the insular cortex, which is a part of the cerebral cortex folded upon itself between the temporal and the frontal lobe. Or the basal ganglia which consists of different nuclei such as the putamen, englobed by the white matter.

Furthermore, the microcellular structure of the cortex shows multiple local variations. It was Brodmann (1909) who first underwent the colossal work of classifying brain regions based on the local cytoarchitectonics, or structure and organization of cells. Even though his number-labeling of regions is a hundred years old, and underwent only a few minor modifications, it still is commonly used in scientific publications for referencing, for example: Brodmann Area 16 and 17 in the visual lobe, which are primary visual areas.

5.1.1 The Prefrontal Cortex

The studies of my PhD Thesis will mainly focus on the prefrontal cortex, which is the most anterior part of the frontal lobes of the brain, the posterior part forming the motor and premotor areas. Specifically elaborated in primates and humans, it is the cortical region that underwent the greatest expansion during evolution, and the peak of its maturity arrives only at the end of adolescence (Paus, 1999). This late maturation is mostly due to a late myelination of the axonal connections. Kennedy *et al.* (1998) reported that in the human adult, 41 percent of the total cerebral cortex volume is to be found in the frontal lobes, nearly one-third of the volume in the prefrontal cortex.

Patients with frontal lesions can present a panoply of cognitive dysfunctioning, sometimes even paradoxical: Levy & Dubois (2006) report patients with an inability to choose between several options (or even no action at all if they get an order); Bechara *et al.* (2000) noticed a perseveration of their patients on actions, being guided by immediate outcomes, while being insensitive to positive or negative future consequences. Another neuropsychological illustration comes from Lhermitte *et al.* (1986): While waiting for his frontal patients behind his office in Paris, he put a hammer, a nail and a painting on a desk near the entrance to his office. When the patient arrived and saw the desk, he impulsively took the painting and nailed it to the wall. Such deficits made clear that the *frontal function*, or *executive function*, can be

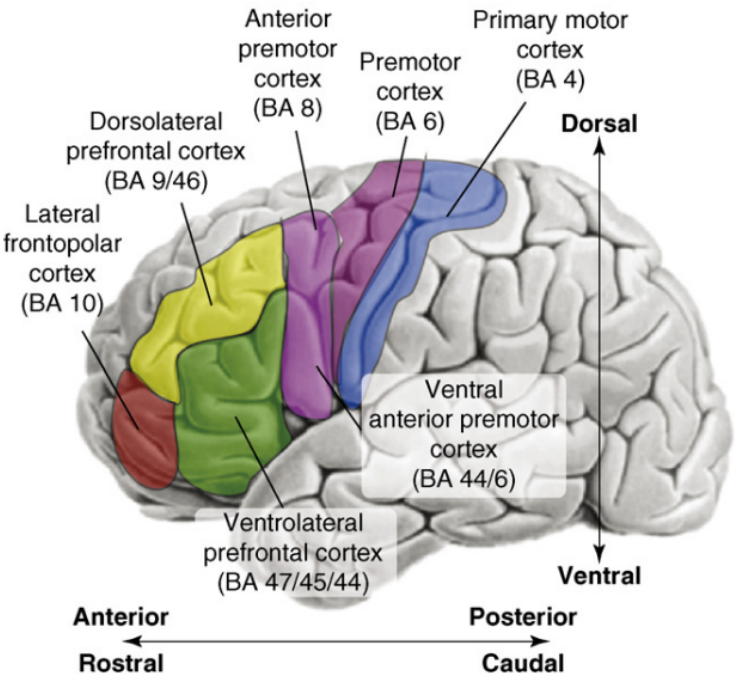


Figure 5.3: Major anatomical sub-divisions of the frontal lobe - lateral view - from Badre (2008)

seen as a whole of cognitive functions, ultimately organizing and controlling goal-directed behavior by evaluation and planification to adapt behavior to a context.

5.2 Cue-Based Decision-Making

The fascinating mystery of human cognition, that is is our capacity for cognitive control, is still under hot debate: how are these functions achieved by the prefrontal cortex? Here, I differentiate between cue-based and value-based paradigms, as introduced in chapter 3. I shall start with the former, where a cue, a rule or an instruction indicates the selection. Herein, a number of theories of PFC functions have been proposed, the most important will be presented in this section, focusing first on the (dorso-) lateral PFC, then on the function of the anterior cingulate cortex, and finally a short introduction to the neural face of the previously presented cascade model, a unifying theory of the prefrontal cortex.

5.2.1 Lateral Prefrontal Theories

Working memory and temporal organization models

First coined by Miller *et al.* (1960), cognitive psychologists use the term *working memory* to characterize an active maintenance of information relevant only for a short duration. Baddeley (1986) argued that working memory is a common resource with limited capacity that can be used by a variety of cognitive tasks. He described two subsystems (the *phonological loop* and the *visuospatial sketchpad*), storing different types of perceptive information, governed by a central executive. Goldman-Rakic (1987) describes working memory as the essential function of the prefrontal cortex. Most of her research is based on neuropsychological and electrophysiological studies in nonhuman primates, showing for example that many cells in the dorsolateral PFC fire when they must maintain spatial information in memory for a short delay (Funahashi *et al.* , 1989). However, since then some experimental studies showed

that working memory is not always necessary or sufficient for an activation of prefrontal regions (D'Esposito *et al.* , 1998).

In a larger approach, Fuster proposes that the PFC is implicated in a temporal organization of behavior in terms of short-time memory, motor attention and inhibitory control of interference through attentional selection (e.g. Fuster (2001)). In his theory, PFC subserves mechanisms assisting a link between perception and a corresponding, temporally separated action, ensuring a correct performance of behavioral sequences. Such a temporal integration is computed by the PFC neurons and by parieto-frontal interactions. Furthermore, Fuster advances a representational hierarchy within the PFC, where abstract or complex action representations are to be found in frontopolar cortex, and the simplest habitual actions stored in premotor areas (and the basal ganglia).

Attentional control models

Norman & Shallice (1986) propose two mechanisms involved in behavior monitoring: First, the *contention scheduler*, which is a subsystem with a posterior neural localization that schedules and orchestrates automated actions. Second, in a novel, conflict or threat situation, the supervisory attentional system (SAS) imposes top-down goals to coordinate multiple action schemata or to integrate larger behavioral goals. The SAS is necessary for more abstract behavior rather than simple action responses to stimuli, and it is located in the PFC. Lesion studies argued in favor of their model, since damages to the SAS regions show an impairment in planning abilities (Shallice, 1982), and Baddeley considers the SAS a valuable candidate for his central executive. Shallice eventually also proposed a hierarchical model, which has been discussed in chapter 3 (Cooper & Shallice, 2000, 2006).

Passingham proposes an alternative theory regarding top-down attentional control, also within the prefrontal cortex (Sakai & Passingham, 2003; Passingham &

Sakai, 2004). They suggest that the prefrontal regions transcribe sensory information of different representations, e.g. sensory cue or task rules, into an appropriate action response. Importantly, the representations can be of various abstraction levels, and attentional top-down modulation is exerted from anterior to posterior regions of the frontal lobes or to parietal areas. In this sense, Passingham's theory of the PFC does not rely on the maintenance of sensory information, but rather in how to prospectively use such information in behavior, e.g. while learning a task set (Sakai & Passingham, 2003).

Finally, Miller & Cohen (2001) introduced an attention-modulation theory of the PFC, implicated in guiding activations of posterior goal-related information, especially in complex tasks. Depending on the context, the representations in PFC can serve as task-specific rules, attentional templates or goals, and so bias other regions through top-down signals. PFC thus has an essential modulatory role to choose (or to learn to choose) a correct action in relation to a context.

Although these theories use a large array of terms and divers backgrounds to depict cognitive control, some key concepts clearly overlap, such as the existence of a top-down control system in the dorsolateral PFC, outlining an action representation hierarchy. Anatomically this hierarchical expansion stretches from premotor areas to anterior frontopolar cortex, increasing in abstraction rules or systems in order to select a correct action for a given context.

5.2.2 Medial Prefrontal Theories

The answer to the *where* of cognitive control seems to root in the lateral parts of the PFC, but another important issue was first discarded: how does the brain know *when* it should implement cognitive control? Or, as Botvinick *et al.* (2001) puts it: *in addition to the regulative dimension of control, by which its top-down influence is exerted, there must also exist an evaluative component that monitors information processing, making*

an assessment of current demands. Along this line of reasoning, the medial PFC saw a sudden increase in interest around the year 2000, as the theories on the lateral PFC began to refine. At first, two different theoretical philosophies arose, supported by empirical evidence, rivaling each other: conflict monitoring and error detection. Both theories are involved in performance monitoring, and subsequent signaling the urge for a behavioral adjustment.

Error detection

The anterior cingulate cortex theory of error detection relies first of all on well-known behavioral results on error commission (and correction), which is followed by less errors and slower reaction times on subsequent trials (Rabbitt, 1966; Laming, 1979). Such error signals might point to a need in increase of cognitive control, and indeed such signals were found in monkeys' ACC, on individual neurons (Niki & Watanabe, 1979) and local field potentials (Gemba *et al.* , 1986). A few years later, first traces of these signals were also found in humans, referred to as *error-related negativity* (ERN), i.e. a peak around 80 to 100 milliseconds in an event-related potential occurring after an error commission, for example Gehring *et al.* (1990) with EEG, and Ullsperger & von Cramon (2001) with fMRI and EEG.

While investigating the origin of error detection and the associated signals of medial PFC, Scheffers & Coles (2000) proposes that the ERN might be due to the comparison between representations of an appropriate and an actual response. As reported by Schultz *et al.* (1997), such reinforcement learning signals are identified in dopaminergic neurons in the midbrain. Holroyd & Coles (2002) then forward a model of two neural systems, linking the ERN to the prediction error: the dopamine system, involved in computing prediction errors, might be conveying a difference signal to an error-processing system within the ACC.

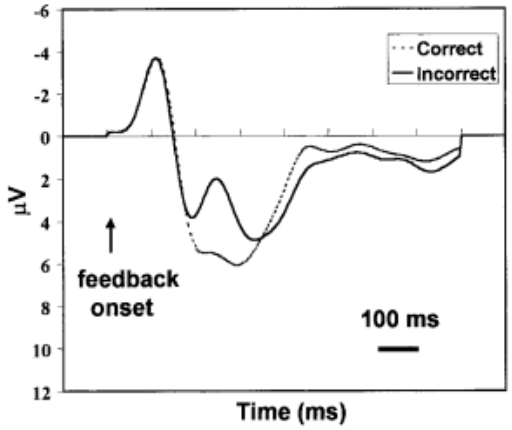


Figure 5.4: Error-related negativity component - appearing around 100 ms after the feedback- Holroyd & Coles (2002)

Conflict monitoring

In parallel to the error detection theories, Botvinick *et al.* (1999, 2001) explained the error effect in medial PFC as a signal of *response conflict* between correct and incorrect response processes. They state that a response conflict appears whenever an incorrect response representation is more active than a parallel activated correct response representation. Similarly, MacDonald (2000) argued in favor of the conflict monitoring theory by showing that the ACC activation is higher in response to incongruent than to congruent stimuli. Most of these studies investigate such a response conflict with the Stroop Task (illustrated in chapter 3.4), which reliably induces a response conflict when a participant is asked to name the color word when written in incongruent color ink.

Importantly however, in situation of a conflict, the ACC activity can be greater even if *no* error occurred. As Botvinick (2007) states later in a review, this conflict effect and subsequent medial PFC activity correlations can be found in many tasks of different quality: attentional tasks, reversal learning, stimulus detection, all explain-

able within a conflict monitoring scheme. Kerns *et al.* (2004) even go further, showing evidence that ACC conflict-related activity can predict subsequent neural and behavioral adjustments in lateral prefrontal cognitive control, revealing a medial-to-lateral interaction.

Action-Outcome association

However, the work of Gehring & Willoughby (2002) and others start reporting that the ERN does not necessarily reflect an error detection, but the medial PFC might participate in computations related to reward outcomes. Similarly, many neurophysiological studies, for example Ito *et al.* (2003), showed ACC responses even in tasks with no cognitive conflict, but rather neurons coding the consequences of actions. In a sequence learning task by Procyk *et al.* (2000), a change in activity of a population of ACC neurons was observed when their monkeys were learning which action sequence was followed by reward. Matsumoto *et al.* (2003) observed that the neuronal activity of the ACC depended on various factors such as the expectation of reward or non-reward, the intention to do an action or not, or a combination of both. Taken together, these results are the groundwork for action-outcome theory forwarded by Rushworth *et al.* (2004). They propose a selection of action function in the pre-SMA, and a crucial role of the ACC in relating actions to consequences, be it positive outcomes or errors.

Importantly, as indicated here, the induction of a viewpoint shift from cue-based to value-based decision-making is crucial to further clarify the functional role specifically of the medial PFC, which I will discuss in the next section.

5.2.3 Multitasking in Frontopolar Cortex

The frontopolar cortex, or Brodmann Area 10 (see fig. 5.3), was much a mystery region until 10 years ago. Not much was known about these cortices, mainly because its location underneath a bony air sinus, making any electrophysiological

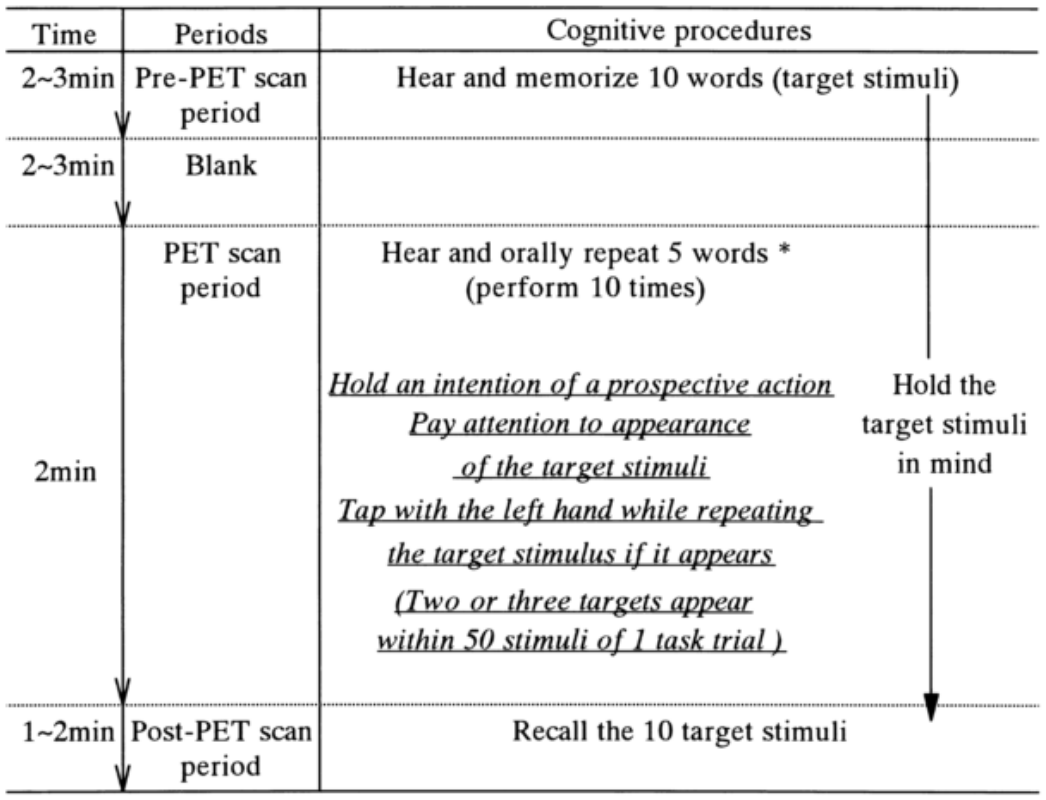


Figure 5.5: simple prospective memory task or Multitasking? - paradigm designed by Okuda *et al.* (1998). They explain the observed frontopolar activations in their task by a prospective memory involvement, contrary to the results of a prospective memory task by Coull *et al.* (1996). At a closer look, it appears that Okuda's participants were multitasking (and branching): most of the time, the participants would hear non-target stimuli and repeat them until the next trial begins. Simultaneously, they hold 10 target stimuli on stand-by (learned in a prior session), and on appearance of such a target stimulus, participants should start tapping with the left hand.

recordings quite difficult. This changed with the advent of modern imaging techniques, and so within a few years numbers of theories regarding BA 10 were mushrooming from different labs. Ramnani & Owen (2004) or Koechlin & Summerfield (2007) reviewed in some detail the different hypothetical functions, amongst others for example mind wandering and introspective evaluation of thoughts (Christoff & Gabrieli, 2000). Tulving (2002) attributed an essential role in episodic memory to the frontopolar cortex, the *retrieval mode*, based on his own neuropsychological patients and on studies like Düzel *et al.* (1999) implicating BA10 in memory retrieval.

More paradoxical results were reported about a *prospective memory* hypothesis, i.e. remembering to perform intended actions after a delay. While Burgess *et al.* (2001) identified frontopolar regions to be more active in prospective memory conditions, Coull *et al.* (1996) reported that frontopolar activity did not vary with prospective memory load. Interestingly, Okuda *et al.* (1998) designed a task to put prospective memory to the test, and they effectively found a higher activation in the frontopolar cortices in comparison to a control task. However, as illustrated in fig. 5.5, the effect they observed might be described as *branching* in a multitask environment, i.e. the maintenance of primary task goals while simultaneously allocating attentional resources to secondary goals.

Koechlin *et al.* (1999) first proposed the concept of cognitive branching for the frontopolar region. In their experiment, they carefully control for other processes that are supposed to be integrated in prefrontal regions, in order to isolate the function of BA 10. On account of this, they could successfully demonstrate that frontopolar activation does not raise from working memory tasks, dual task performance or episodic information processing, but rather from alternating task sequences which require a temporary hold of a goal while performing another task, only to resume this goal at a later stage. These results have been reliably replicated several times with slightly different tasks, e.g. semantic processing (Braver & Bongiolatti, 2002),

and similarly Dreher *et al.* (2008) showed an impairment in multitasking behavior in frontopolar-lesioned patients. In the same line, Charron & Koechlin (2010) added to this conception amongst others that human prefrontal function is limited to the simultaneous chase of two concurrent goals. Importantly, in an earlier study, Koechlin *et al.* (2002) demonstrated that te frontopolar region is specifically recruited when learning a sequence of tasks, but not of motor responses. Although these results illustrate well the frontopolar function in environments where an external cue indicates the return to a task on stand-by, it does not treat the question of holding a primary task goal while performing another one based on its current higher profit. We will discuss this more in the next section, on value-based decision-making.

5.2.4 Neural Integration of the Cascade Hierarchy

To come to an end of the section discussing cue-based decision-making, I shall mention the neural correlates of the cascade model (described in chapter 3 and on fig. 3.4 (Koechlin *et al.* , 2003; Kouneiher *et al.* , 2009)), which relies on this large theoretical background, resulting in a consolidation of the executive function theory in human PFC. The cascade model is of major importance, since there are no other attempts of model theorizations to be found in literature trying to include all these regions and results into a common framework. Their results suggest a parallel functional architecture in medial and lateral prefrontal PFC, for motivating resp. selecting behavior.

On the one hand, the temporal hierarchy in cognitive control is organized on the lateral prefrontal cortex: contextual control in posterior lateral PFC, episodic control slightly more anterior in mid-lateral PFC, and branching in the frontopolar region. On the other hand, motivational control is hierarchically organized along the medial PFC: contextual motivation, which represents transient incentives, in pre-SMA, and episodic motivation, i.e. a sustained (high or low) reward environment, in ACC. Interestingly, the medial PFC regions are insensitive to cognitive control factors, but

they influence the lateral areas by their motivational control. Importantly, Kouneiher *et al.* (2009) observed enhanced connectivity by episodic motivation between dACC and mid-lateral PFC, and mid-lateral to posterior lateral PFC, as well as an enhancement by contextual motivation between pre-SMA and posterior lateral PFC.

As mentioned before, it is important to notice that participants were asked to make cue-based choices, i.e. previously learned cue associations or instructions before a bloc indicated the task to perform, reward was not a choice factor. In the following section we will have a closer look at value-based decision-making, i.e. when the choice is based on a subjective value.

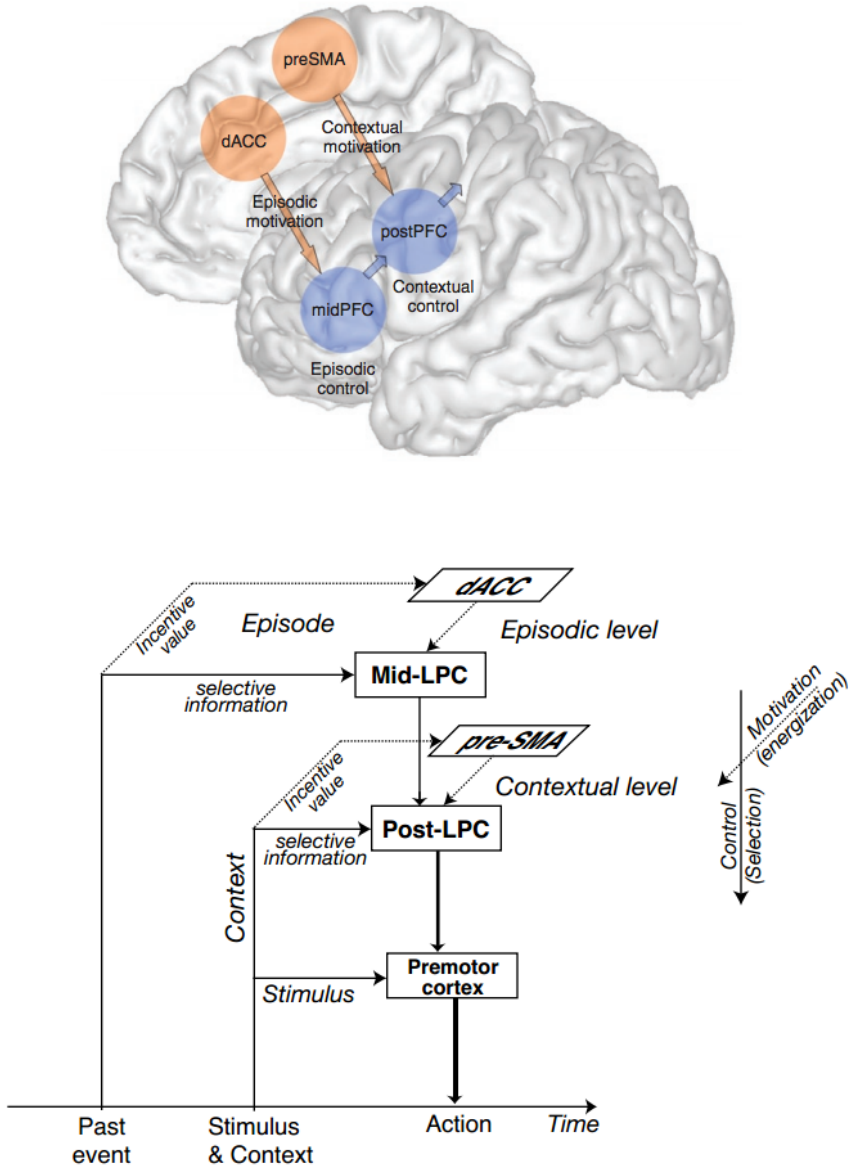


Figure 5.6: the cascade model: parallel architecture in medial and lateral prefrontal cortex -

5.3 Value-Based Decision-Making

Most of the experiments described in the previous section use paradigms with static and learned or overtrained contingencies. In contrast to the categorically correct and incorrect trials in laboratory set-ups, such black-and-white situations rarely occur in the real, uncertain world. When an animal is foraging, it has to make decisions based on a history of information coming from both more positive and more negative outcomes, and hence creating a subjective value of an action or a task. Most of the more recent research on decision-making is therefore focusing now on value-based decision-making, showing an implication of a network of prefrontal and striatal regions, which I will discuss now.

5.3.1 Basal ganglia and Neurotransmitters

Parkinson's Disease (PD) is surely the most prominent of the basal ganglia disorders, characterized by a progressive loss of the ascending dopaminergic projections. PD can tell us much about the function and dysfunction of the midbrain, but it is not the only source of information. Some neuropsychiatric disorders have their origin here and addictions disrupt basal ganglia function. Basal ganglia consists of several nuclei, each might have different clinical implications: amongst others striatum (consisting of putamen and caudate nucleus), pallidum and substantia nigra (see fig. 5.7). Due to the location, this subcortical structure is an excellent candidate for interactions with the executive functions; the striatum in particular receives massive input from the neocortex.

In contrast to the cerebral cortex which has excitatory, glutamatergic projection neurons, the basal ganglia network comprises inhibitory, GABA-ergic projection neurons. The spiny projection neurons are inhibitory cells found in the striatum (input), and the aspiny neurons are the main component of the pallidum (output) (Swanson, 2000).

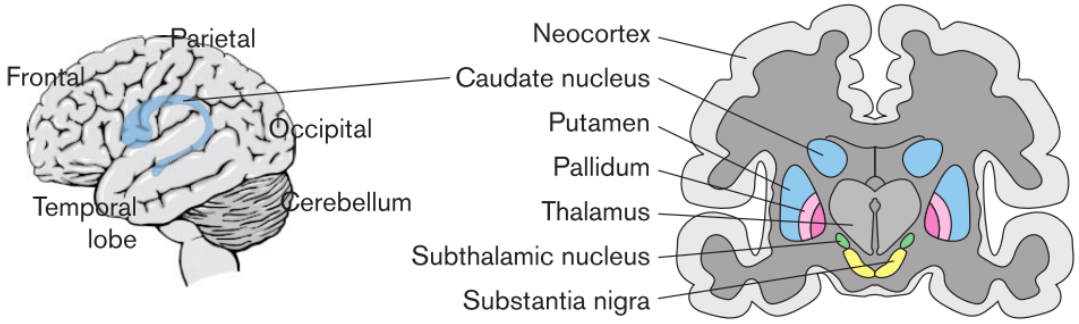


Figure 5.7: basal ganglia - medial view of left hemisphere - from Graybiel (2000)

Konorski (1967) was among the first to analyze lesions of the basal ganglia. Although fine movements were relatively intact, the animals showed a decrease in action acquisition needed to get a specific reward. Since then, many studies proved that neurons in several brain structures are sensitive to rewards, especially dopaminergic neurons (i.e. molecular compound released by nerve cells to communicate with other cells) projecting from the substantia nigra, the ventral tegmental area and the hypothalamus, which show short, phasic activation after presentation of rewards or after stimuli that predict reward (e.g. Schultz (1986); Schultz & Romo (1990)). At a closer look at the properties of the phasic dopamine response, an activity increase is only observed for surprising rewards, not for accurately predicted ones (Mirenowicz & Schultz, 1994). Hollerman & Schultz (1998) also report such a reward prediction error during learning.

As discussed in the chapter about modeling behavior, prediction errors play a crucial role in reinforcement learning, and thus such dopaminergic activity seems to be a central factor in learning and adaptive optimization (Schultz *et al.* , 1997). Indeed, artificial networks using such a teaching signal can learn for example to play high-level backgammon (Tesauro, 1994). Dopamine-containing neurons also seem to encode the uncertainty of the prediction in maintained firing levels (Fiorillo *et al.* , 2003), and they might be coding a temporal aspect of reward delivery

as shown by Fiorillo *et al.* (2008): the neural activity was highly sensitive to the duration of a stimulus-reward interval, although only weakly sensitive to the exact timing of reward delivery after a well conditioned stimulus. Satoh *et al.* (2003) further demonstrate that midbrain dopamine reflect both reward magnitude and trial-specific reward probability. Interestingly, while recording from single dopamine neurons, Bayer & Glimcher (2005) reported a correlation prediction error (or more precisely with the actual reward and a weighted average of previous reward), but only for positive outcomes; no activity changes were found to correlate with negative prediction errors. Additionnally, these midbrain dopamine neurons are even sensitive to a reward prediction error in a context-dependent manner (Nakahara *et al.* , 2004).

In humans, O'Doherty *et al.* (2003c) was one of the first to investigate pavlovian conditioning with fMRI in humans. They showed a BOLD correlation of the reward prediction error in the ventral striatum, a target area of dopamine neurons. Furthermore, they observed a signed prediction error, meaning that a negative prediction error led to a negative BOLD signal. It should be noted that interestingly even when the learning process is unconscious, prediction error signals can be observed in the striatum, as well as a behavioral effect (Pessiglione *et al.* , 2008). Further, O'Doherty *et al.* (2004) attributed separate functions to the ventral and dorsal striatum, in the neurocomputational context of *actor* and *critic* components of reinforcement learning. The neural correlates of the critic, i.e. the structure learning to predict future outcomes, were found in the ventral striatum, whereas the actor-related BOLD activity, i.e. involved in maintaining information about the rewarding action outcomes, was situated in the dorsal striatum. By showing an interaction of pavlovian and instrumental learning through conditioned reinforcement in these same regions, they were able to counter much criticism of the actor/critic model, e.g. from Dayan (2002) arguing that the actor/critic model can not account for the two parallel pavlovian and instrumental learning systems of conditioning.

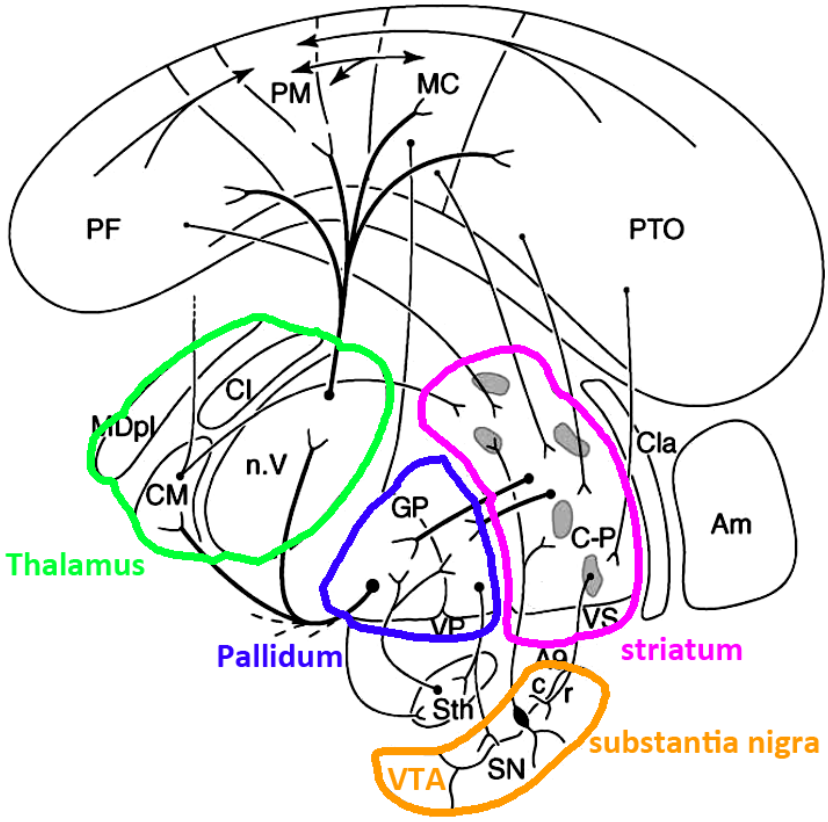


Figure 5.8: cortico-basal ganglia circuits (omitting a.o. descending connections) - adapted from Graybiel (2005). Abbreviations: A9, cell group A9; Am, amygdala; c, pars compacta of substantia nigra; Cl, nucleus centralis lateralis of thalamus; Cla, claustrum; CM, centre median nucleus of thalamus; C-P caudate nucleus-putamen; Glu, glutamate; GP, globus pallidus; MC, motor cortex; MDpl, pars lateralis of thalamic mediodorsal nucleus; n. V, ventral nuclear complex of thalamus; PF, prefrontal cortex; PM, premotor cortex; PTO, parieto-temporo-occipital cortex; r, pars compacta of substantia nigra; SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; Sth, subthalamic nucleus; VP, ventral pallidum; VS, ventral striatum; VTA, ventral tegmental area

However, even though there is heavy convincing evidence of the striatum entanglement in learning-related functions, it would be misleading to reason that this is the only structure implicated in learning. Importantly, while dopamine neurons are reported to code the expected value of an action that soon will be taken, their activity does not contribute directly to the selection of action itself (Morris *et al.* , 2006). In the same line, the presence of prediction error signals are not necessarily followed by adjustments in action selection (Bayer & Glimcher, 2005). Indeed, the striatum harmonizes with other regions, especially with the prefrontal cortex, to form a much more extended network defined by distinct anatomical areas with specific relations to one another.

5.3.2 Frontal Pole I: Ventromedial PFC and the Chosen

The frontal pole cortex, or BA 10, is the prefrontal area that underwent the largest evolutionary shift from apes to humans, mostly through modifications in local circuitry and a better spatial organization (Semendeferi *et al.* , 2001, 2010). In humans, studies usually distinguish between lateral and medial parts, namely the ventromedial PFC (vmPFC) and the lateral frontopolar cortex (although the nomenclature can vary between papers).

Unfortunately for neurophysiologists, as noted earlier in this chapter, the frontal pole's anatomical location makes it very difficult to record from electrodes in monkeys. Therefore, apart from its anatomy, most insights are coming from fMRI studies in humans. Recently however, Tsujimoto *et al.* (2010) managed to get data of neuronal activity with a direct approach through the air sinus. They found that neurons in the frontopolar cortex encoded the monkey's decision at feedback, monitoring action outcomes. Interestingly, as pointed out by Koechlin (2011) and in line with anatomical studies (Carmichael & Price, 1996; Ongür *et al.* , 2003), this supports the view that the frontopolar cortex in monkeys corresponds only to the vmPFC in humans (see fig. 5.11).

Patients with vmPFC lesions are reported to be impaired in adaptive decision-making, regarding their insensitivity to more distant outcomes and consequences and their focus on immediate prospects (Bechara *et al.* , 1994, 2000; Camille *et al.* , 2011). Camille *et al.* (2004) also reported that vmPFC patients did not experience regret, i.e. they were unable to anticipate the negative consequences of their actions. Although Tsujimoto *et al.* (2011) argues that vmPFC does have a rather simple encoding function, the multifunctional aspect of vmPFC is pointed out in a recent review by Roy *et al.* (2012), but mainly its implication in economic valuation and prospection. It has major connections to subcortical structures such as amygdala and ventral striatum (Carmichael & Price, 1995), as well as to other prefrontal regions (mainly the dorsolateral prefrontal cortex), to the anterior superior temporal lobe and the posterior cingulate cortex (Carmichael & Price, 1996).

Response to stimulus value and action outcome

It has been known for a while that the frontal pole in monkeys is implicated in processing information about rewards and punishments. Kim *et al.* (2006) showed such a pattern in human subjects, but more interestingly they reported an increase in activity after successful avoidance of punishment, besides the typical increase with rewarding outcome (see fig. 5.9). Indeed, avoidance of a negative reward can be seen as an intrinsic reward correlation, serving to reinforce avoidance. Further, Plassmann *et al.* (2007) revealed correlations of the *willingness-to-pay* in the vmPFC, i.e. the maximum amount of money subjects are willing to pay in exchange for a presented food item., or in other terms: the value of a goal. However, the vmPFC does not respond only to money, it's rather a global currency integrator: pleasant odor correlations can be found (Rolls *et al.* , 2003), as well as for taste values (Small *et al.* , 2003) or face attractiveness (O'Doherty *et al.* , 2003a).

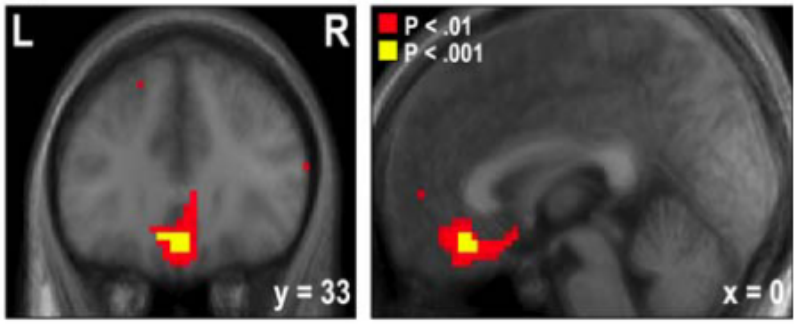


Figure 5.9: typical activation pattern for ventromedial PFC - responses to avoidance of aversive outcomes as well as to obtaining reward, from Kim *et al.* (2006)

Prediction of outcomes

Alternatively instead of solely monitoring object values and subsequent outcomes , subjects might use these outcomes subsequently to make decisions (O'Doherty *et al.* , 2003b), for example by choosing between different possible options over many trials in order to identify the option with the largest benefit. To this purpose, computational algorithms such as reinforcement learning can be used to estimate the expected value of each stimulus based on past experience; e.g. if a choice gets a larger output than expected, then the value is corrected upwards.

Usually values derived from such models are positively correlated with activity in vmPFC (Tanaka *et al.* , 2004; Knutson *et al.* , 2005; Daw *et al.* , 2006; Gläscher *et al.* , 2009), even for expected values of rewards delayed in time (Kable & Glimcher, 2007). In the outcome phase, Hampton *et al.* (2006) additionally reported an activity pattern in the feedback phase which was highest when a reward (25 cts) was delivered. Besides these typical results, some studies recently noted that vmPFC does not only signal expectations of monetary gain, but it also reflects expectations of monetary losses (Tom *et al.* , 2007; Basten *et al.* , 2010). Boorman *et al.* (2009) demonstrated that vmPFC correlated positively with the chosen option and negatively with the rejected one, signaling a comparison between the different options. In the same

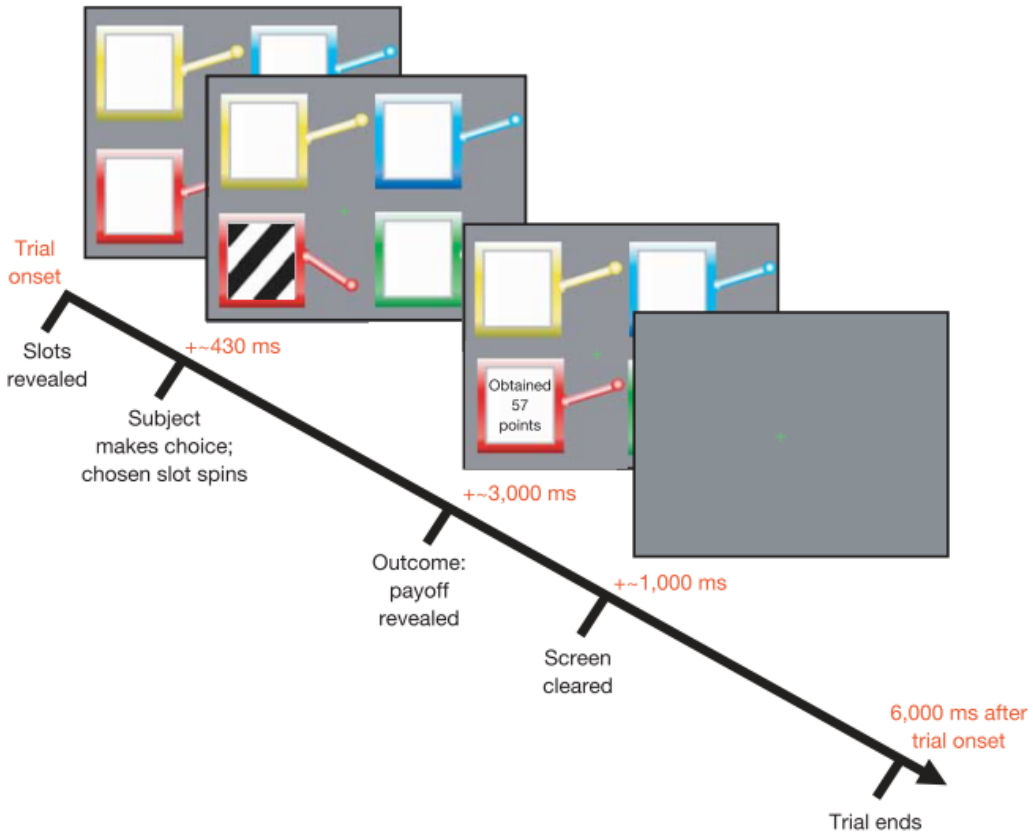


Figure 5.10: example of task design in value-based decision-making - in this 4-armed bandit task, the participant can freely choose between 4 options which are differently rewarded and change over time. After the choice, the outcome is revealed, and a new trial starts (Daw *et al.* , 2006)

context, using multivariate pattern recognition Kahnt *et al.* (2010) showed that, although vmPFC encodes the value of anticipated outcomes independent of sensory properties, each fMRI voxel in this region is selective for either high or low values.

These studies are in agreement with electrophysiological studies revealing a more heterogeneous relationship between value and neural activity, like Padoa-Schioppa & Assad (2006). In their 2006 paper they show that some cells in vmPFC are correlating positively with the chosen value, whereas others correlate negatively and further that the neurons encode the two economic options independently of motor response values. In a second paper, it becomes even clearer that the neurons in this region have a very stable heterogeneous firing (Padoa-Schioppa & Assad, 2008), e.g. the firing correlating with the value of one of the proposed juices does not depend on what other juices there are. Seen in contrast with the vmpfc results of Boorman *et al.* (2009) or Kahnt *et al.* (2010), that different cells in the vmpfc might encode different values, but overall through fMRI we can observe the global correlation.

In summary, vmPFC is a critical region for adaptive decision-making. It encodes the global value of stimuli as they are revealed, it maintains representations of expected reward, neuronal firing can be very heterogeneous, it receives prediction error signals from the striatum and it guides predictions of future rewards, possibly even computing the decision.

5.3.3 Frontal Pole II: Lateral Frontal Pole and the Unchosen

The functions of lateral frontopolar cortex (lFPC) and vmPFC seem to be dissociable, since results in human neuroimaging studies conclude that lFPC and vmPFC are hardly ever co-activated. They are found in addition to lateral prefrontal and ACC activations (Charron & Koechlin, 2010), and further it is reported that activity in lPFC and vmPFC can be of opposite sign, e.g. O'Doherty *et al.* (2001). Here,

voxels in vmPFC correlated positively with reward, whereas lFPC correlated positively with punishment. In the same line, using reinforcement learning algorithms Daw *et al.* (2006) reported that lPFC activity is observed during exploration periods, while vmPFC is active during exploitation; similar results are reported by Yoshida & Ishii (2006) with a bayesian learner approach. Yoshida and colleagues designed a virtual maze task and asked their participants to navigate their way out it while lying in an MRI machine. lPFC correlated with the uncertainty about their position in the maze, i.e. when they were exploring the maze, and considered several possible alternatives.

However in general, literature about lFPC in value-based decision-making is rather sparse; not much is known about this area. Two recent studies from the same team investigated this question more deeply (Boorman *et al.* , 2009, 2011). First, in the case of a binary choice, participants could freely choose between a left or a right gamble option and subsequent button press on every trial. Then, the potential reward for each of the options was shown, varying randomly from trial to trial, but the probability of getting a reward could be estimated from experience. Boorman and colleagues provide clear evidence for an enrollment of lFPC in monitoring an alternative outcome relative to a current outcome and the subsequent adjustment of behavior. In other terms, they show the ability of lFPC to track increasing evidence in favor of an alternative option, transmitting information to the inferior parietal sulcus if enough evidence favors a switch. This is broadly consistent with theories of multitasking, like branching (Koechlin & Summerfield, 2007); in such a scenario FPC might encode the alternative option rather than simply the evidence for it.

The second study, from 2011, considers multiple uncertain options, not only a binary environment. They measured BOLD signals while participants performed a three-option gambling task, in which each option has a distinct magnitude of reward and a certain probability of getting that reward. After choosing one option, they got

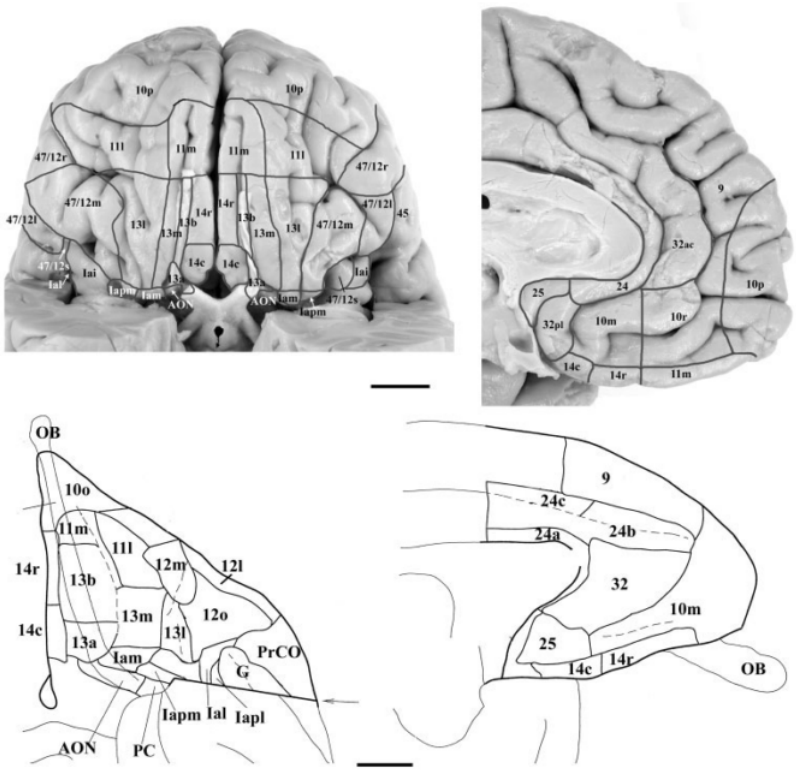


Figure 5.11: orbital and medial surface of human and monkey brain - taken from Ongür *et al.* (2003); the difference in area 10 between human and monkey can clearly be seen on this figure. For the human brain, the region corresponding to 10m in monkeys was renamed into 10m and 10r, and area 10o is now labelled 10p (polar)

feedback on whether each option would have paid out. What the authors found, was that the brain activity in lFPC did not reflect the estimated value of multiple counterfactual choices, but the accumulated evidence in favor of the second-best option. Instead of comparing the best two options, it appears that lFPC represents the value of the next-best alternative.

However, Boorman et al. found similar results in other regions, amongst the ACC which we will discuss now. They present a network of regions involved in monitoring unchosen options, but they are not able to discriminate their unique functional roles. Future work should therefore look into the specific contribution of each of those regions to learning from the outcomes of unchosen options.

5.3.4 Anterior Cingulate Cortex

In the previous section about the enrollment of the ACC, I concluded that the activity in this region could not solely be explained by error detection or response conflict theories. In an fMRI experiment on humans, Walton *et al.* (2004) related activity of ACC neurons to an interaction of choice decision and consequences of a freely made choice. Under this viewpoint ACC might use both rewards and errors to construct a history of choice-outcome associations upon which future decisions will be taken. Further evidence in this line comes from single-unit recordings in monkeys by Kennerley *et al.* (2006). The authors trained monkeys on a reinforcement-guided reversal task to choose between one of two joystick movements. In the first experiment, only one of the movements was rewarded during a bloc of trials, before a switch occurred, whereas in the second experiment the responses were rewarded with different probabilities. Kennerley and colleagues could demonstrate that monkeys with ACC lesions were not impaired on behavior that followed errors

(i.e. when contingencies switched), but they could not integrate values over time to adapt responses in a changing environment.

From a distant viewpoint it might seem like ACC neurons have an activity pattern similar to the dopaminergic neurons, using reward and prediction to code a prediction error. However, there are important differences. The same dopamine cell codes a positive and a negative prediction error by a phasic increase in their firing rate. In contrast, as shown by Matsumoto *et al.* (2007), cells inside the ACC have specific roles; some encode positive feedback, others negative feedback or rewards that were not received (Hayden *et al.* , 2009) and even a few encode both directions. This might be some evidence that the ACC is implicated in adjustment for future behavior by explicitly specifying the direction and amount of the errors. Similarly, Quilodran *et al.* (2008) were able to link adaption of action value and subsequent adjustment of behavior, and a study by Amiez *et al.* (2006) showed that monkeys with temporary lesions were not able anymore to choose the best stimulus, since they were impaired in integrating task value and received rewards. It seems like a general agreement that the ACC is crucial to link behavioral regulation and monitoring action valuation.

However in general, positive feedback-related activity is seen less often, and an important criteria to see or not to see ACC activity is the informative value of the actual feedback; the subject should *know* the value of information. This knowledge is generally referred to as the *volatility* of the environment, towards which the subject adapts his learning rate. In an fMRI experiment with healthy humans, Behrens *et al.* (2007) asked participants to perform a gambling task with a changing reward rate, from stable phases to phases with high volatility. The authors demonstrated that the tracking of the volatility estimate (or the related uncertainty) can be traced back to the ACC.

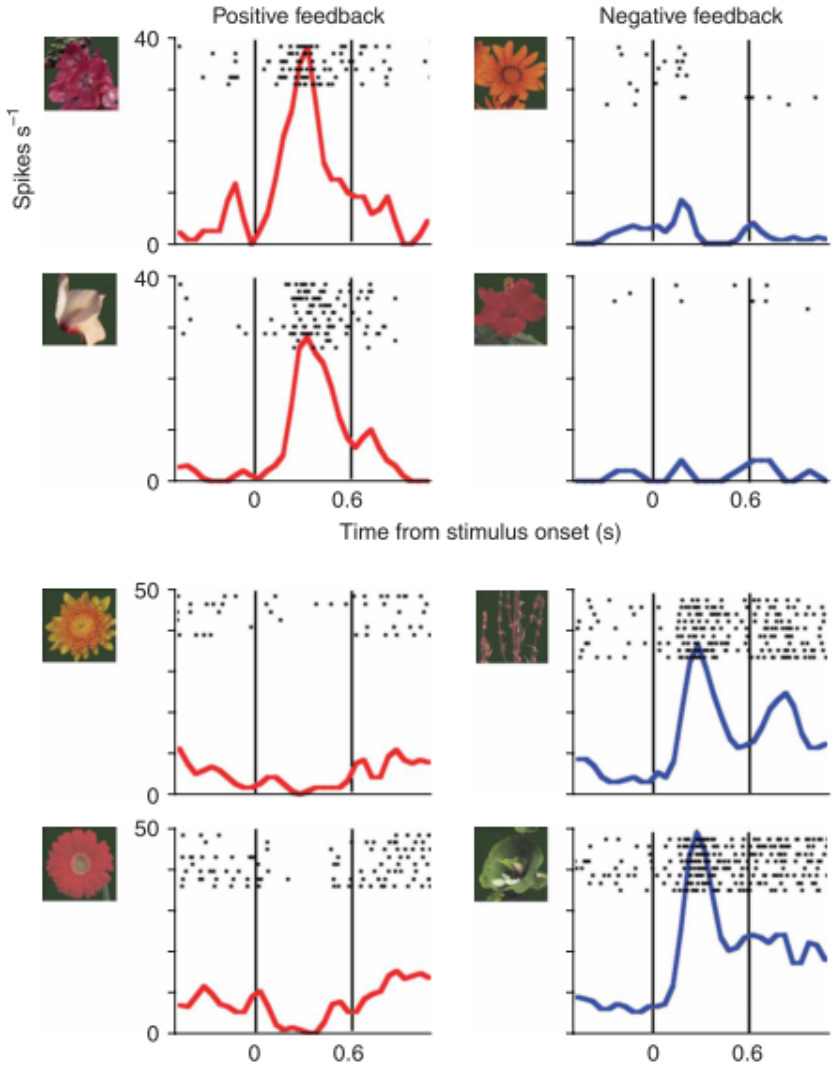


Figure 5.12: ACC cells responding to positive and negative feedback respectively - taken from Matsumoto *et al.* (2007); in red: a cell preferring positive feedback, in blue a cell preferring negative feedback

As one might have noticed, most of what we know about the ACC is coming from monkey studies. Of course, humans are not monkeys, like some comparison studies underline, but nevertheless we have common ancestors from which the brain evolved. As I tried to argue throughout these introduction chapters, there are many common interests between behavioral ecologists studying foraging in monkeys and other animals, and cognitive neuroscientists studying decision making in humans. A common language over multiple research domains has evolved, but there are still many gaps that need to be bridged (or not!) to further understand the specificity of the human brain. I tried to contribute to this knowledge with the studies of my PhD, which goals I will now specify.

Chapter 6

Towards a specific Thesis question

So far we have shown that studies of decision-making involving rewards have a long history of constantly evolving ideas, mostly due to advances in methodological techniques. An integration of results from behavioral ecology, economy, computational modeling, cognitive psychology and neuroimaging contribute to a better understanding of the brain and human behavior in decision-making. We showed that decision-making and cognitive control are tightly linked, involving especially the basal ganglia and the prefrontal cortex. The ability of the human brain to coordinate behavior of different levels of abstraction relative to a specific context or episode is associated with its hierarchical organization. We discussed task sets, representing the first hierarchical brick of cognitive control, which are on a different level of action abstraction than motor responses, and recruit higher level brain areas. We mentioned the flexibility of the human brain to switch between tasks in order to adapt behavior according to the requirements of the environment. We showed that two classes of models are used to theorize value-based decision-making, that is reinforcement learning et bayesian inference models.

The goal of this thesis is the investigation of the neural basics of decision-making based on values of expected outcomes of actions, at different levels of action abstraction: choosing between simple actions and between tasks, i.e. more abstract

structures. Directly comparing actions and tasks in a situation of free choice in the absence of cues is a novel undergoing.

Indeed, most of the insights of recent years are coming from research on monkeys, specifically through neurophysiological experiments. While they help us to understand the role of single cells within a brain region, they rarely can conclude about the global function of an area in comparison to other areas. In the same line, as discussed before, although humans have the same ancestors as monkeys, comparing both brains to make conclusions might not always be an accurate approach, especially regarding higher cognitive functions. On the other side, neuroimaging studies in humans in value-based decision-making rarely discriminate between the level of action abstraction, or report similar activation patterns in different regions, making it difficult to attribute functional roles to specific regions. We wanted to clarify these issues and therefore we proposed an fMRI paradigm, which will be presented in the next chapter.

Part III

Experimental Work

Chapter 7

Disentangling action and task value-based decision-making

7.1 Introduction

Adapting behavior according to dynamic values of chosen and unchosen options has been of a particular interest in recent studies of value-based decision-making. Besides the well-studied dopaminergic projections in the striatum encoding a positive prediciton error (Bayer & Glimcher, 2005), a network of four prefrontal regions have been identified in humans and other primates, each with potentially different functional roles in value-guided behavior: ventromedial prefrontal cortex (vmPFC), frontopolar cortex (FPC) and adjacent lateral orbitofrontal cortex (lOFC), dorsolateral prefrontal cortex (dlPFC), and dorsal anterior cingulate cortex (dACC) (Rushworth *et al.* , 2011).

Typically, studies investigate value-based decision-making in humans using gambling tasks, in which participants are instructed to freely choose between two or more simple options, mostly in the form of different visual stimuli or motor responses. The subjects' internal strategies and subjacent fitted states like choice value

or choice probability are inferred through computational models such as Reinforcement Learning (RL) (Daw *et al.* , 2006) or Bayesian Learners (Behrens *et al.* , 2007). In a second step, functional magnetic resonance imaging (fMRI) is used to find the underlying neural correlations of the model's internal signals (O'Doherty *et al.* , 2007).

In such studies, the ventral striatum is well known to correlate with a temporal difference prediction error (O'Doherty *et al.* , 2003c). The human vmPFC has been found to increase activity with increasing expected value of a chosen option (Hampton et al., 2006), whereas the FPC signals values of choices that have not been taken (Boorman *et al.* , 2009). dACC and dlPFC are thought to play a major role in action valuation and outcome-contingent changes in behavior (Quilodran *et al.* , 2008; Kennerley *et al.* , 2006), although most of the knowledge is coming from monkey studies. The exact nature of the ACC activity in contrast to the ventral striatum and the FPC has still to be clarified: Matsumoto *et al.* (2007) reports signed prediction error signals in ACC cells, Hayden *et al.* (2009) concludes that these cells respond to reward outcomes of unchosen options, and Behrens *et al.* (2007) finds a correlation with the tracking of environmental volatility.

While much is known about tracking the values of simple actions, it is unclear whether the same neural dynamics apply within the conceptual framework of task sets, i.e. active representations of an ensemble of actions. On the one hand, switching between tasks in a rule-guided, non-volatile environment has been studied intensively, e.g. participants are instructed to put one task on hold while performing another one, with a fixed reward. Findings indicate that in such experimental settings the FPC is engaged when subjects are required to delay the execution of one task in order to perform another first (Koechlin *et al.* , 1999). Moreover, patients with frontopolar lesions are impaired in such multitasking behavior (Dreher *et al.* , 2008). However, how the brain is implementing switches and tracking values

between competing tasks in a volatile environment, i.e. where the participant can freely choose between two complex behavioral structures, still remains uncharted.

To address this issue, we conducted a probabilistic reversal-learning problem in fMRI. On a trial-by-trial basis, subjects were instructed to voluntary select either between two actions or between two tasks with anti-correlated stochastic reward probabilities in a block-wise manner, followed by a feedback displaying the reward of the chosen option. Importantly, subjects were not explicitly cued about switches in reward contingencies. We compared basic RL and bayesian models to infer the individual subject's internal signals of the options at each trial, which we then used as parameters to regress against the blood oxygen level-dependent (BOLD) response of the whole brain separately for the decision phase and the feedback phase. We reveal that vmPFC and striatum are engaged on the level of actions, and in contrast FPC, ACC and dlPFC on a task level. FPC monitors the evidence in favor of the alternative task only, and ACC activity manifests a task effect, predicting the switching behavior between tasks, but not actions. Furthermore, we provide evidence for a specific engagement of the prefrontal cortex network in value-based decision-making of abstract behavioral structures.

7.2 Methods

7.2.1 Participants.

Twenty-four right-handed volunteers with normal or corrected-to-normal vision were tested on the experimental paradigm (ages 20-28 years). All participants were recruited on campus at the Université Pierre et Marie Curie in Paris, France, and underwent a prescreening for neurological or psychological disorders, use of medications and contraindications for MRI. Informed consent from every subject was obtained from our on-site physician.

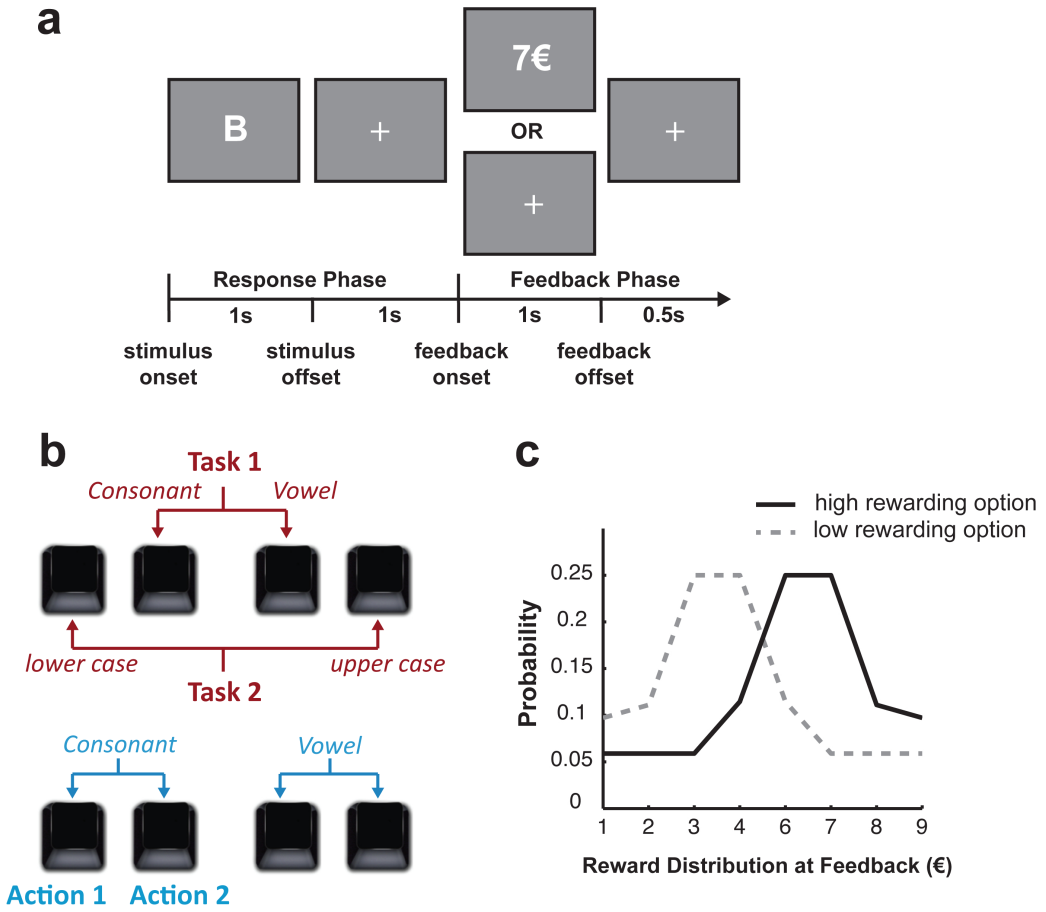


Figure 7.1: Experimental design - **A** At the beginning of each trial a letter was presented, followed by a fixation cross. Participants made a choice according to the past outcomes and bloc type, and pushed a button at any time during the Response Phase. Subsequently a reward based on their choice was revealed or a null event was displayed (fixation cross), followed by an inter-stimulus interval (fixation cross) before the next trial started. **B**. Example of button mapping according to bloc type (Task or Action Condition). **C**. Reward probability distribution for the two correct, but differently rewarded options.

7.2.2 Experimental task.

Each block began with an information display to reveal the ongoing condition to the participants: task or action condition (fig. 7.1). Each trial began with the display of a single letter stimulus (1 s), followed by a blank screen (1s). During this 2s-decision phase, subjects were instructed to make a free choice, depending on the condition: In tasks, they could choose freely to perform one task out of two possible tasks (e.g. task 1: lower/upper case or task 2: vowel/consonant). In actions, they could choose freely one out of two actions to respond (pushing button with index finger or middle finger). Subjects were informed that the two tasks respectively the two actions were differently and variably rewarded in a stochastic manner (between 1 and 9 euros), one task resp. one action yielding a higher reward in average across trials of an episode (trials in between contingency changes). These changes were not cued, the only source of information for the subject was the feedback. The decision-phase was followed either by a feedback revealing a reward (50 percent of trials), or a null event (blank screen). After the 1s feedback-phase, an inter-trial interval was fixed at 0.5s (blank screen), before the next trial started.

Subjects participated in two MRI sessions, a total of 8 pseudo-randomized blocs of tasks and 4 blocs of actions, each consisting of 96 trials and 4 episodes, one episode containing randomly 8, 12, 16, 24 or 32 trials before a contingency switch. Before those runs in the MRI, subjects had the opportunity to practice the experimental task outside the MRI scanner. The first training session was one day before the MRI and lasted one hour to familiarize with the paradigm. The second training session just before the MRI session lasted 15 minutes. Participants were informed that after the experiment, one bloc would be picked randomly to define their reward.

All stimuli were created using the Psychophysics Toolbox for MATLAB (Mathworks, inc.), and visual stimuli projection onto a screen were viewed through a mir-

ror attached to the head coil. Subjects responded with the index and middle fingers of each hand via a four-button MR-compatible response box.

7.2.3 Computational Modeling.

We implemented two learning models, which hypothesize different methods by which participants might use experience with rewards to learn choice values, and compared them through LogLikelihood (LLH) computation and Bayesian Model Comparison (BMC).

a. Reinforcement Learner.

A reward prediction error is derived using model-free Q-learning, a variant of classic RL (Sutton & Barto, 1998) (see also Fig. 7.2). The agent intends to estimate a Q-value for each chosen and unchosen task and action. These values are initialized to 0.5 at the start of each bloc of the experiment, and updated after each trial if a reward is experienced. The updating algorithm is as follows:

$$Q_{t+1}(a) = Q_t(a) + \alpha\delta$$

where *Q(s.a)* represents the value of the response choice *a* for the foregone trial, α is a free parameter for the estimated learning rate, and δ represents the prediction error, *r* being the experienced reward:

$$\delta(t) = r_t - Q_t(a)$$

Here, we compare two approaches inside the RL scheme, differing in the updating procedure of the unchosen option. In the *normalized RL* variant, the agent updates the chosen value through the prediction error, from which he also deduces the value of the unchosen option *a'* (since the two options are dependent):

$$Q_{t+1}(a') = 1 - (Q_t(a) - \alpha\delta)$$

Second, in the *standard RL* variant, the agent updates only the chosen option upon reward, leaving the unchosen value constant until re-chosen:

$$Q_{t+1}(a') = Q_t(a')$$

b. Bayesian Learner.

Inferences about the probabilities for selecting the most rewarded option are made through prior assumptions about the reward distribution (see Fig. 7.1).

$z_t \in \{1, 2\}$ denotes the most rewarded option in each trial, with a probability τ of changes between two consecutive trials, thus capturing the decay. In each trial the agent performs a choice $a_t \in \{1, 2\}$. If the choice is the currently high rewarded, then a_t=z_t. The reward is being sampled from a noisy distribution, denoted as β_{hi} for the high rewarded, and β_{lo} for the low rewarded (fig. 7.1 c). If for example the agent chooses the high rewarded option, he obtains a reward r_t with the corresponding probability from this distribution β_{hi,r_t}. If a feedback is received, r_t is then distributed as

$$r_t|a_t, z_t \sim \beta_{hi}^{f(a_t=z_t)} \beta_{lo}^{f(a_t \neq z_t)}$$

where f is a function returning f(k)=1, if k is true, and 0 otherwise.

Since the goal is to infer the most rewarded option on the next trial z_{t+1}, the agent has to infer z_t, after choosing a_t and obtaining r_t, knowing τ and β. If $\gamma_t = p(z_t = 1|a_{1:t}, r_{1:t})$ is the confidence in the most rewarded choice selection being 1, given all the information including trial t, $z_{t+1} = 1|a_{1:t}, r_{1:t}$ is computed from γ_t:

$$p(z_{t+1} = 1|a_{1:t}, r_{1:t}) = \tau(1 - \gamma_t) + (1 - \tau)\gamma_t$$

Then we can find γ_t incrementally:

$$\begin{aligned} \gamma_t &= p(z_t = 1|a_{1:t}, r_{1:t}) \\ &= \frac{p(r_t|z_t = 1, a_t)p(z_t = 1|a_{1:t-1}, r_{1:t-1})}{p(r_t|a_{1:t-1}, r_{1:t-1})} \end{aligned}$$

$$= \frac{\beta_{hi,r_t}^{a_t}\beta_{lo,r_t}^{1-a_t}\Sigma_{z-1}p(z_t=1|z_{t-1})p(z_{t-1}|a_{1:t-1},r_{1:t-1})}{\beta_{hi,r_t}^{a_t}\beta_{lo,r_t}^{1-a_t}(\tau(1-\gamma_{t-1})+(1-\tau)\gamma_{t-1})+\beta_{lo,r_t}^{a_t}\beta_{hi,r_t}^{1-a_t}(1-\tau(1-\gamma_{t-1})+\tau\gamma_{t-1})}$$

$$= \frac{\beta_{hi,r_t}^{a_t}\beta_{lo,r_t}^{1-a_t}(\tau(1-\gamma_{t-1})+(1-\tau)\gamma_{t-1})}{\beta_{hi,r_t}^{a_t}\beta_{lo,r_t}^{1-a_t}(\tau(1-\gamma_{t-1})+(1-\tau)\gamma_{t-1})+\beta_{lo,r_t}^{a_t}\beta_{hi,r_t}^{1-a_t}(1-\tau(1-\gamma_{t-1})+\tau\gamma_{t-1})}$$

Note that at the beginning of a bloc, γ is set to 0.5, random choice.

c. Choice selection

Additionally we introduced a softmax rule for each model, controlling for stochastic choice selection (denoted Q for simplicity) according to probabilities from the distribution:

$$p(a_{t+1}) = \frac{exp(\beta Q_t(a))}{exp(\beta(Q_t(a) - Q_t(a')))}$$

where β is the *inverse temperature*, controlling the agent's tendency to exploit the option with the highest value. Through the softmax rule, an agent's behavior towards exploration of suboptimal choices are determined probabilistically on the grounds of the values of available options. Previous research proposed evidence for the softmax rule over other methods, such as ε-greedy (Daw *et al.* , 2006).

7.2.4 fMRI methods.

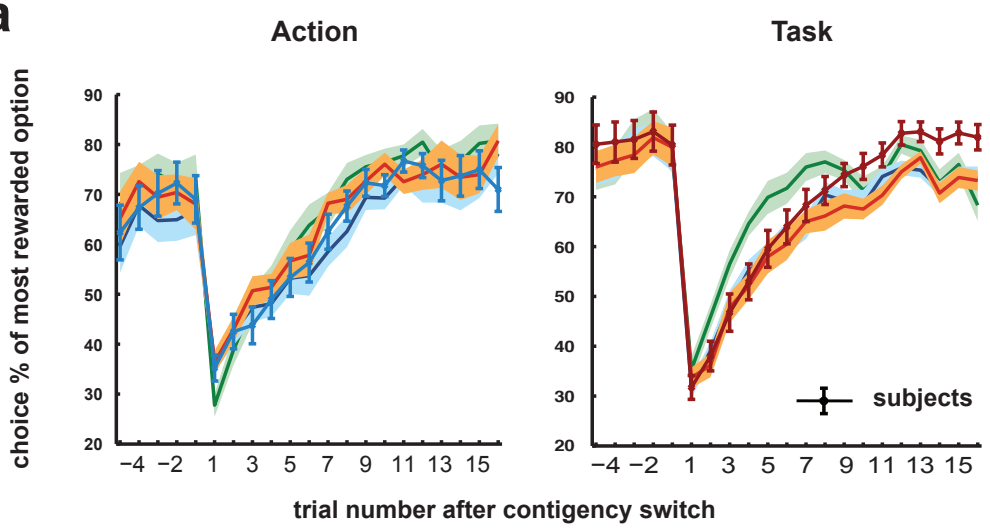
fMRI Data Acquisition.

Magnetic resonance images were acquired with a 3T Siemens (Erlangen, Germany) Trio whole-body and RF coil scanner (TR 2000 ms, TE 30 ms, flip angle 90 deg, echo spacing 0.53 ms, FOV 192 x 192 mm2, acquisition matrix 64x64, thickness 3 mm, 37 interleaved and joint slices, voxel size 3 x 3 x 3 mm3). High-resolution T1-weighted (MP-RAGE) anatomical images were collected for anatomical visualization.

fMRI Preprocessing and Analyses.

Imaging data were analyzed with SPM8 (Wellcome Dept. of Cognitive Neurology, London, available at http://www.fil.ion.ucl.ac.uk/spm). All volumes from all sessions were realigned to the first volume, nonlinearly normalized to the stereotaxic Talairach atlas (Montreal Neurological Institute EPI-template, images resampled at 4 x 4 x 4 mm3) and spatial smoothing (isotropic 3D Gaussian kernel, 8 mm). Statistical models were constructed under the assumptions of the general linear model. Statistical parametric maps were obtained from local fMRI signals using a linear multiple regression model with event-related regressors, six realignment and three session parameters as covariates. In single-subject analyses, regressors (Dirac functions convolved with the canonical hemodynamic response function) were modeled separately: 1) the average BOLD response for each condition (task Set/action) at stimulus and feedback onset, 2) parametric regressors for model-derived value signals for each condition at stimulus and feedback onset, 3) parametric regressors for reward value at feedback onset for each condition, 4) nuisance partition including instructions and error trials. In a second GLM, we split the task and action regressors from decision phase into separate regressors for switch and stay trials. Parametric signals were entered unorthogonalized into the first-level design matrices. Then, parameter estimates were entered in a between-subject, random-effect analysis for obtaining statistical parametric maps. We identified brain activations exhibiting significant contrasts of parameter estimates with a voxel-wise (T= 3.78, $P < 0.0005$, uncorrected) and cluster-wise ($P < 0.05$, corrected for multiple comparisons over search volumes) significance thresholds. To check for supplementary activations, we lowered the threshold to P=0.005 uncorrected, where we reported an additional cluster in the orbitofrontal cortex. Region-of-Interest analyses were conducted on significant clusters of activation identified from the first whole-brain voxel-wise analysis.

a



b

Model	Action		Task	
	LLH	BMC	LLH	BMC
normalized RL	-166.59	-74.31	-295.07	-130.54
Bayesian Learner	-167.11	-75.61	-296.83	-131.79
standard RL	-203.4	-90.12	-303.52	-134.55

Figure 7.2: Behavioral Model Comparison - **A**. Probability of choosing the currently most rewarded option, plotted separately for Task and Action blocs, in function of trial number before and after a contingency switch. The learning curves depict averaged subject behavior and the estimated model fitting for three tested models: Normalized RL, classic RL and Bayesian Learner. Error bars (participants) and shading (model) represent standard errors. **B**. Behavioral model fits. Table of LogLikelihoods (LLH) and Bayesian Model Comparison (BMC) for each of the three models.

PPI Analyses.

Seed voxels for dACC, vmPFC, dlPFC and FPC were defined as a 5mm sphere around the group peak voxel sensitive to the appropriate contrast (defined from the same maps as previously at $p < 0.0005$, uncorrected, at the group level). Using standard analysis techniques, the "physiological" time series extracted at these voxels were corrected for variance associated with parameters of no interest, deconvolved with the hemodynamic responses, multiplied by a parameter encoding the relevant "psychological" contrast (e.g., ΔQ), and reconvolved to form a "psychophysiological interaction" (PPI) regressor. This regressor was entered into a design matrix alongside the physiological and psychological regressors and all parameters of the main GLM independently, encoding all the main effects not included in contrast and time series. Results are reported at $p<0.01$ within the previously defined ROIs.

Supplementary DCM Analyses.

Voxels of interest were extracted as for the PPI analyses. Twenty models were constructed, based on the PPI results, each with different connectivity patterns between the four regions (backward, forward, reciprocal), modulated or not by the regressor of interest (ΔQ). Bayesian Model Comparison was used to compare the different models and select the best before performing quantitative analyses. In such a case, the best model, given the data, is the one with the highest log-evidence, $p(y|m)$. We then performed a random-effects Bayesian Model Selection across all subjects to determine the winning model at the group level. This latter model was used for a final between-subject analysis of the estimates of connectivity, to test for the conditional differences in strength of connections.

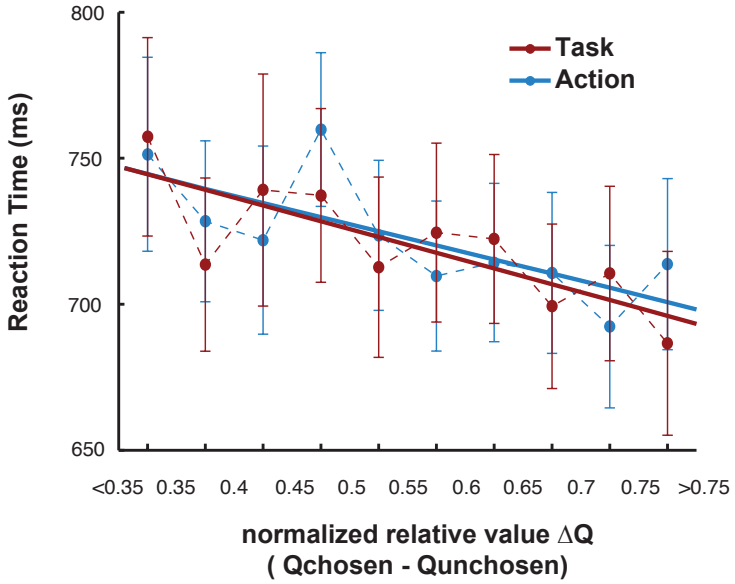


Figure 7.3: Reaction Time - Reaction Time plotted against normalized relative value difference between the two options (ΔQ), and the corresponding linear regressions, separately for Task (red, R=0.63, $p<0.05$) and Action (blue, R=0.53, $p<0.05$). Error bars are standard errors across participants.

7.3 Results Study 1: Decision Phase

7.3.1 Computational Model Comparison and Behavior

To investigate the behavior of subjects quantitatively, we compared two (RL) strategies differing in their computation of the values of the unchosen option, and a bayesian learner. In RL theories, the model's prediction error is used to update the value of the currently chosen option according to the associated reward of the subject's choice. As the choices in our study are symmetrically dependent, we might hypothesize a concurrent tracking of both the value of the chosen and the unchosen option. In this first strategy (normalized RL) both chosen and unchosen values are constantly updated after each feedback, whereas in the second strategy (standard RL) only the estimated value of the chosen option is revised after feedback, while the value of the unchosen option stays invariant. The choice selection difference in the two proposed strategies results from the 'softmax' rule, applied to account for the probability of exploration, a tendency to choose a suboptimal action. According to this rule, the probability of exploration increases gradually as the difference between the chosen and unchosen value functions approaches zero. Computational details of the three strategies can be found in the Methods section.

We compared the three implicit strategies by using the LogLikelihoods (LLH) of the subjects' choices and Bayesian Model Comparison (BMC), optimized over two free parameters (learning rate for RL theories, decay for Bayesian Learner, exploration/exploitation trade-off for both). We found that the normalized RL and bayesian model provided a better explanation of subjects' behavior than did the standard RL for tasks and actions ($p=0.039$, fig. 7.2b). As expected, normalized RL and our bayesian model fitted similarly with no difference between conditions ($p=0.844$, fig. 7.2b); the normalized RL had slightly better LLHs, albeit not statistically significant (fig. 7.2). Additionally we tested for conditional differences in reaction time, but no effect between tasks and actions was found ($p>0.8$, fig. 7.3). As the

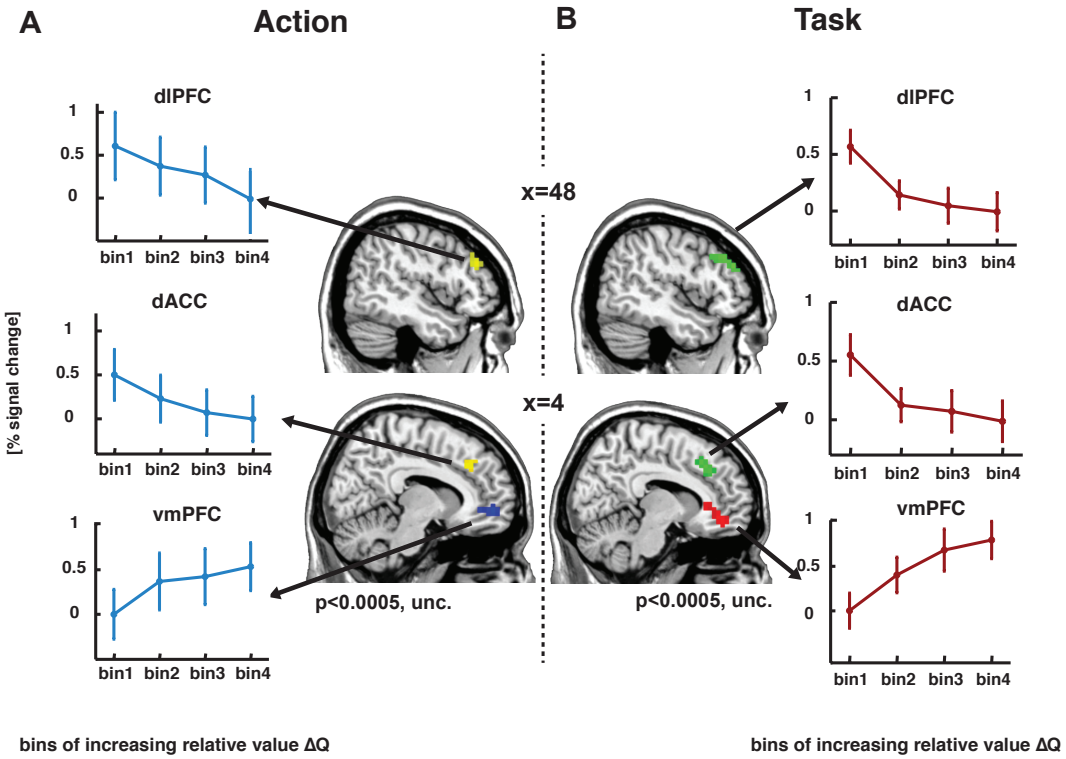


Figure 7.4: Common regions - Sagittal slices through SPMs for mean BOLD correlation with relative value ΔQ and plots of percentage BOLD signal change in bins of increasing relative value ΔQ for **A**. Action condition (blue plots) and **B**. Task condition (red plots). ROIs from upper to lower: dorsolateral prefrontal cortex (dlPFC), dorsal anterior cingulate cortex (dACC), ventromedial prefrontal cortex (vmPFC). Maps correspond to $p<0.0005$ uncorrected.

experiment was not designed to find differences between models, we do not make claims about which of the two models are used by the participants to flexibly adapt in a changing environment as provided by our design. We suppose that participants used the feedback provided for their most recent choice to continuously update both choice values. Since the LLHs and BMC of the normalized RL are numerically more advantageous, we will use the term Q-value in the results presented here, although identical imaging results are observed with the internal variables from the bayesian approach.

7.3.2 Common neural network of action and task Values

After computational characterization of the subjects' behavior, we used the best-fitting model, the normalized RL, to generate regressors containing task and action value predictions on each trial for each subject. Importantly, since the value for the chosen and the unchosen option are anti-correlated in this model, one regressor gives information about positive and negative correlations with ΔQ, the *relative chosen value*. To search for a neural network with such a signature, we tested throughout the whole brain for regions correlating with the model's internal signals. Our first analysis focused on common neural structures implicated in value computing of actions as well as tasks. The statistical parametric maps yielded three common regions: on the one hand the ventromedial prefrontal cortex (vmPFC; p_{unc}= 0.0005; action: MNI x=4, y=52, z=-4; task: MNI x=-4, y=52, z=-4) (fig. 7.4) correlated with the relative chosen value ΔQ. On the other hand the dorsal anterior cingulate cortex (dACC, p_{unc}= 0.0005; action: MNI x=8, y=32, z=36; task: MNI x=4, y=24, z=44) (fig. 7.4) and the dorsolateral prefrontal cortex (dlPFC, p_{unc}= 0.0005; action: MNI x=52, y=36, z=28; task: MNI x=48, y=32, z=36) (fig. 7.4) correlated with the relative unchosen value -ΔQ during the response phase.

Additionally to the reported clusters, the right orbitofrontal cortex (OFC) exhibited a correlation with the relative unchosen value at a lower statistical threshold

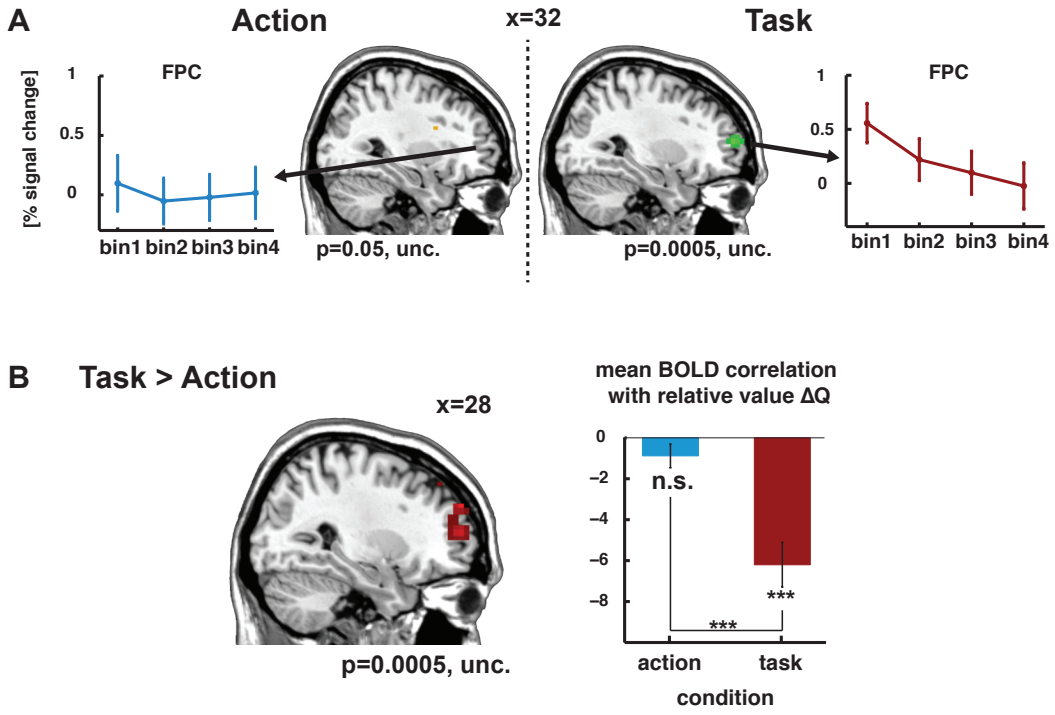


Figure 7.5: Unique Task correlates - **A**. Sagittal slices through SPMs for mean BOLD correlation with relative value ΔQ for Actions (p=0.05 unc.) and Tasks (p=0.0005, unc.). **B**. BOLD activity extracted from ROI defined by the contrast ΔQ Task – ΔQ Action (p=0.0005 unc.).

(OFC, p_{unc}= 0.005, task: MNI x=28, y=56, z=-12; action: MNI x=28, y=52, z=-4, see Supplementary Material fig. 7.10).

7.3.3 Unique Frontopolar Correlations with relative unchosen task Value

After having identified the common regions, we looked for differences in ΔQ correlations between tasks and actions. A cluster in the right frontopolar cortex (FPC) correlated with the relative unchosen task value (FPC, p_{unc}= 0.0005, MNI x=32, y=56, z=12, fig. 7.5A), but no correlation in the lateral regions of the FPC in the action condition was to be found, even at a very liberal threshold (p_{unc} = 0.05, Fig. 7.5A). In order to use correct statistical procedures to report significant differences in the two conditions, as pointed out by Nieuwenhuis *et al.* (2011), we checked thoroughly for differences between action and task by computing a difference contrast (ΔQ task - ΔQ action). Our analysis revealed one single significant cluster in the prefrontal cortex, namely in the FPC, specifically correlating with the relative unchosen task value (FPC, p_{unc}= 0.0005, MNI x=32, y=56, z=12) (Fig. 7.5B). Beta extraction of the cluster from this contrast confirmed a significant difference between task and actions (paired t-test: $p<0.001$), as well as a main effect for tasks (t-test: $p<0.001$) but not for action (t-test: $p>0.05$) (Fig. 7.5B).

The correlations described here are not affected by the current behavioral choices or restrictions, since the clusters survived the inclusion of confound regressors in the GLM, namely switch versus stay trials, task 1 versus task 2 (in tasks) and left versus right hand (in actions). Notably, as stated before, using the variables provided by the bayesian model instead of the normalized RL values revealed identical correlations throughout the brain. Consequently we do not claim to reveal the exact nature of the value function represented in these regions, but rather that the observed imaging results are independent of the computational model in the case of our paradigm.

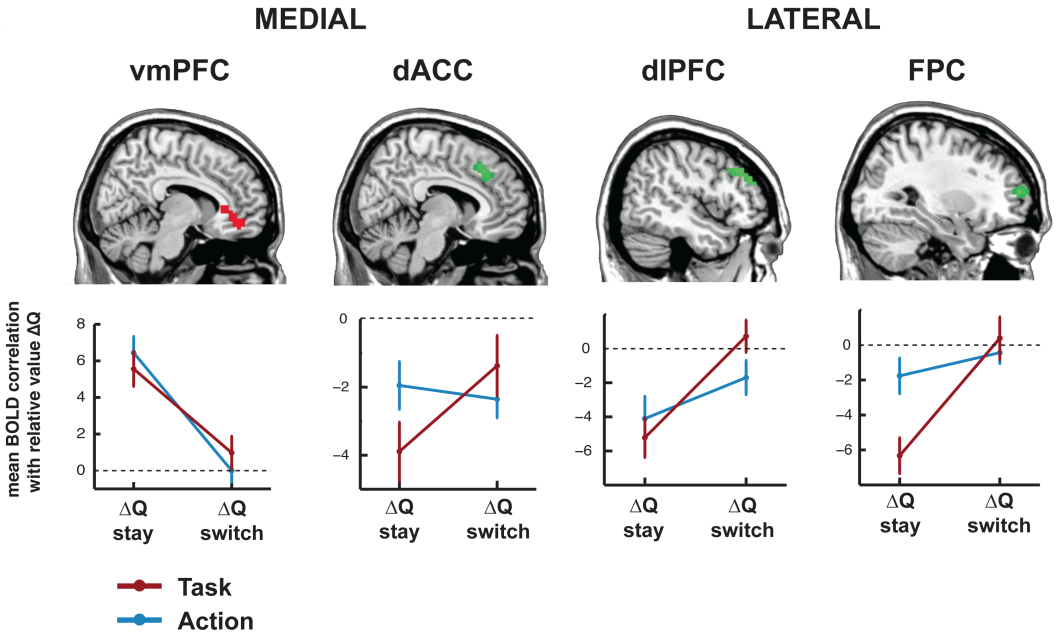


Figure 7.6: Stay versus Switch trials - BOLD activity extracted from the four prefrontal ROIs for Bloc (Task/Action) x Trial type (ΔQ stay/switch) interaction. The vmPFC reveals the same pattern of activity for both conditions, in contrast to the three regions with ΔQ anti-correlations (dACC, dlPFC and FPC) where a significant bloc x trialtype interaction is observed (p=0.02).

7.3.4 Distinct correlations of switch/stay trials

Since the participants are performing only one task in the action condition, but switching between two tasks in the task condition, we cogitated that areas exhibiting a specific task-value effect should not differentiate between switch and stay trials in actions (since subjects are not switching between tasks, but between actions), but such regions should show a differential activation pattern for each trial type in tasks. Vice versa, in regions involved in tracking choice values of actions, we should observe a differential activation per trial-type in actions. To test these predictions, we constructed a second GLM with regressors separated by switch/stay trial type, each with the same parametric modulations as before. To be coherent in our cluster analysis, we used the same ROIs as defined by the results above to extract the beta values. None of the regions correlating with the relative unchosen value ΔQ, i.e. dACC, dlPFC and FPC, differentiated between switch and stay trials in the action condition (all $p>0.20$), contrary to the task condition, in which a significant trial type effect is reported in each of the clusters (all $p<0.01$). Even more, on trials where the subject switches from one task to the alternative task, the correlation not only diminished, but vanished. Importantly we notice a significant trial-type-by-condition interaction in these three regions ($F(1,22)=5.893$, $p=0.02$) with no region effect on this interaction ($F(2,44)=0.652$, $p=0.53$), exposing a substantial task-effect in dACC, dlPFC and FPC. In contrast, correlations with the relative chosen value ΔQ within the vmPFC showed statistically identical results for tasks and for actions, namely a strong effect on stay trials, but which is reset when the participants perform a switch.

7.3.5 Network Connectivity

To test for differences between actions and tasks in the underlying connectivity of this four-region prefrontal network, we performed Psychophysiological Interaction (PPI) analyses. We reasoned that, if engaged in monitoring, FPC should not show any significant outgoing connections to other prefrontal regions. PPI analyses for tasks revealed significant interactions between dACC and all three ROIs (all

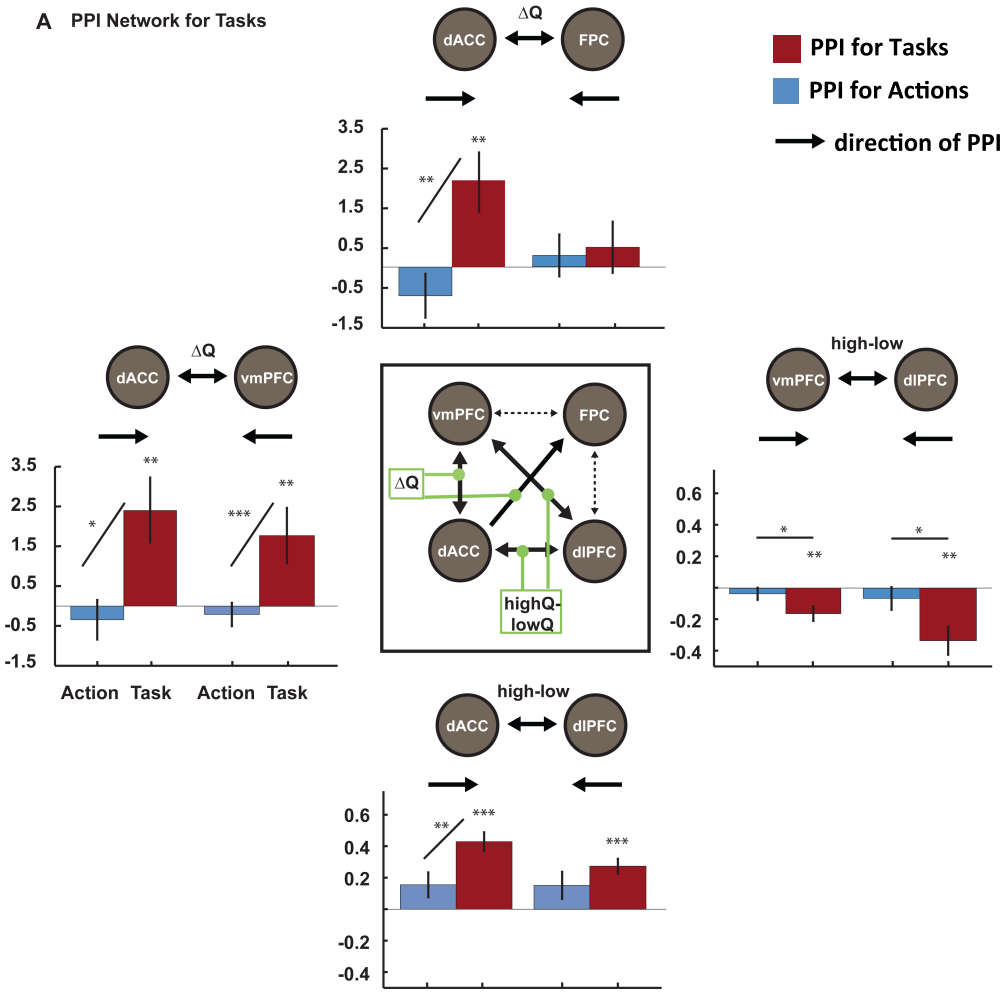


Fig 7. Connectivity

Figure 7.7: PsychoPhysiological Interactions - Psychophysiological Interaction between the prefrontal ROIs for both conditions. Full lines show a significant increase in connectivity between regions with increasing value difference ΔQ, respectively increased connectivity for high Q values compared to low Q values. Dotted lines show non-significant PPIs. Strikingly, no significant interactions are observed in actions.

$p < 0.01$), and no outgoing connections from the FPC to other prefrontal regions. Importantly, for actions we did not find any evidence for interregional connectivity (fig. 7.7), which leads us to the assumption that the psychophysiological interactions between connections in this prefrontal network specifically emerge in the task condition.

Additionally, to confirm this four-region prefrontal network and to further specify the direction of the connections, we used Dynamic Causal Modeling (DCM) to compare several models of effective connectivity differing in backward and forward connections. Bayesian Model Selection unveiled a best model, confirming a unidirectional connectivity from dACC to FPC, modulated by the relative chosen value, and no significant outgoing connection from FPC to other prefrontal regions (Supplementary Material).

7.4 Results Study 2: Feedback Phase

7.4.1 Disentangling Value and Reward from Prediction Error

In reinforcement learning, at each trial the agent is supposed to use the value functions to predict the rewards that a choice will yield. The reward obtained during the feedback phase after choice selection is then compared to the prediction. The discrepancy between the choice value and the reward value is referred to as prediction error, which plays a critical role in reinforcement learning theories. To dismantle the neural contributions to the encoding of a prediction error, we separated the choice value and reward value into distinct regressors in our GLM. We reasoned that a region implicated in prediction error signals should fulfill two correlation criteria. First, the BOLD activity should significantly correlate with choice value and reward value. Second, these two regressors should show an opposite relationship. In case a cluster correlates with the prediction error signal, but does not satisfy these criteria, it can not be said to be implicated in an prediction error encoding.

We ran a whole brain analysis with these two parametric modulations for each of the conditions (task/action) at the onset of the feedback phase, and computed one prediction error contrast per condition at the group level (Reward - Q). We report several regions correlating with the global prediction error from this contrast, but after beta extraction for choice and reward value only the bilateral ventral striatum suffices the two criteria (p_{unc} = 0.0005, fig. 7.8), for both task and action. Interestingly although correlating positively with PE, the vmPFC shows a strong increase in activity with increasing reward value, but no significant correlation with choice value in the feedback phase, and no difference between task and actions (fig. 7.8).

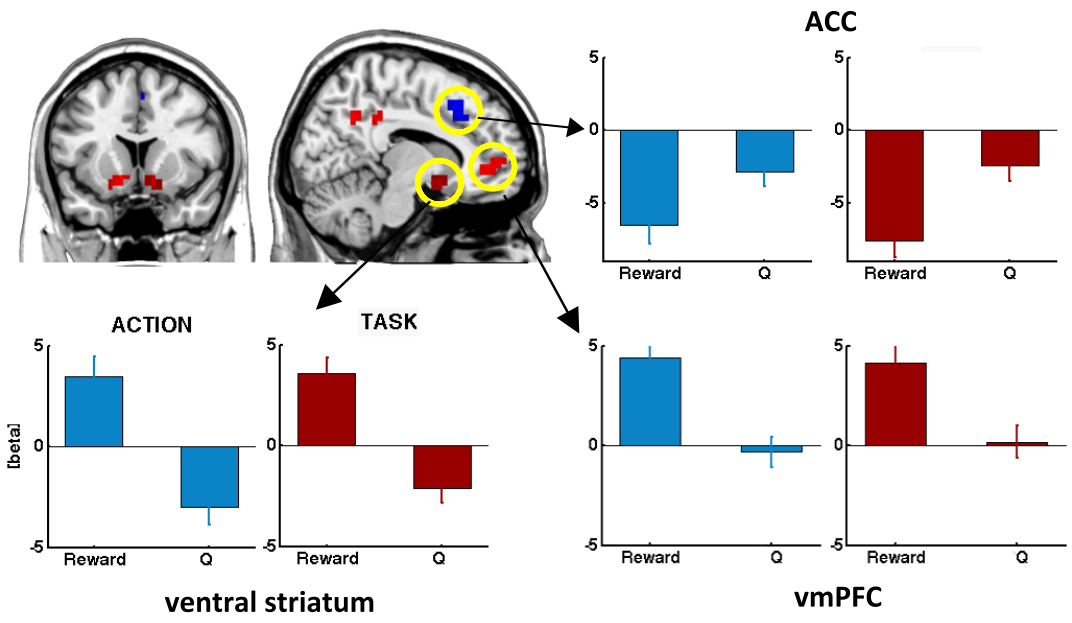


Figure 7.8: Prediction error during feedback phase - The ventral striatum is the only region satisfying the two prediction error criteria. vmPFC exhibits strong correlation only with reward value, whereas anti-correlates with both choice and reward value ($p<0.0005$ uncorrected)

7.4.2 ACC predicts individual task switching behavior

Furthermore, although the ACC correlates negatively with PE (p_{unc} = 0.0005, fig. 7.8), after beta extraction we show that this activity actually reflects negative relationship with both choice and reward value, for task and actions (fig. 7.8). To further investigate these activations, we conducted a 2-by-2 factorial analysis, separating the parametric regressors into 4 bins: high/low values for choice and reward, depicted in fig. 7.9A. Again, no effect of task/action is observed (F=0.122, p=0.730), however we report a significant interaction $Q_{hi,lo}$ x $R_{hi,lo}$ (F=5.341, p=0.031). The activity within the ACC cluster peaks on trials of low reward and low choice value, i.e. on trials where the the participant chose the less rewarded option upon which he got a low reward. This activity was significantly higher than on trials with an actual negative prediction error (high choice value and low reward value).

Since the ACC seems to be implicated in behavior adjustment, we wanted to test whether this is true for switching between tasks or only between actions. Matsuzaka *et al.* (2012) showed that in monkeys, the posterior medial PFC, a region adjacent to the preSMA, exhibits a selective action preference, a discrimination which is not found in the ACC on the action level. Additionally, in the decision phase we reported that the ACC manifests a task effect. Therefore we have reason to hypothesize that ACC provides a signal to switch between tasks, not actions. First we applied the same factorial design on the switch probabilities (fig. 7.9B), where we observe a pattern similar to the ACC activity in tasks, but not action. In the task condition, group results show that subjects are more prone to switches when a low task value is followed by a low reward in comparison to high task values followed by a low reward. In contrast, in actions no value effect on the switching behavior is observed, only a highly significant reward effect. Second, we tested for a correlation between inter-individual switching behavior and ACC activity, regressing the ACC activity against switching probabilities (fig. 7.9C). We observe a highly significant correlation between task switching probability and ACC activity (r=0.53, p=0.007), but no

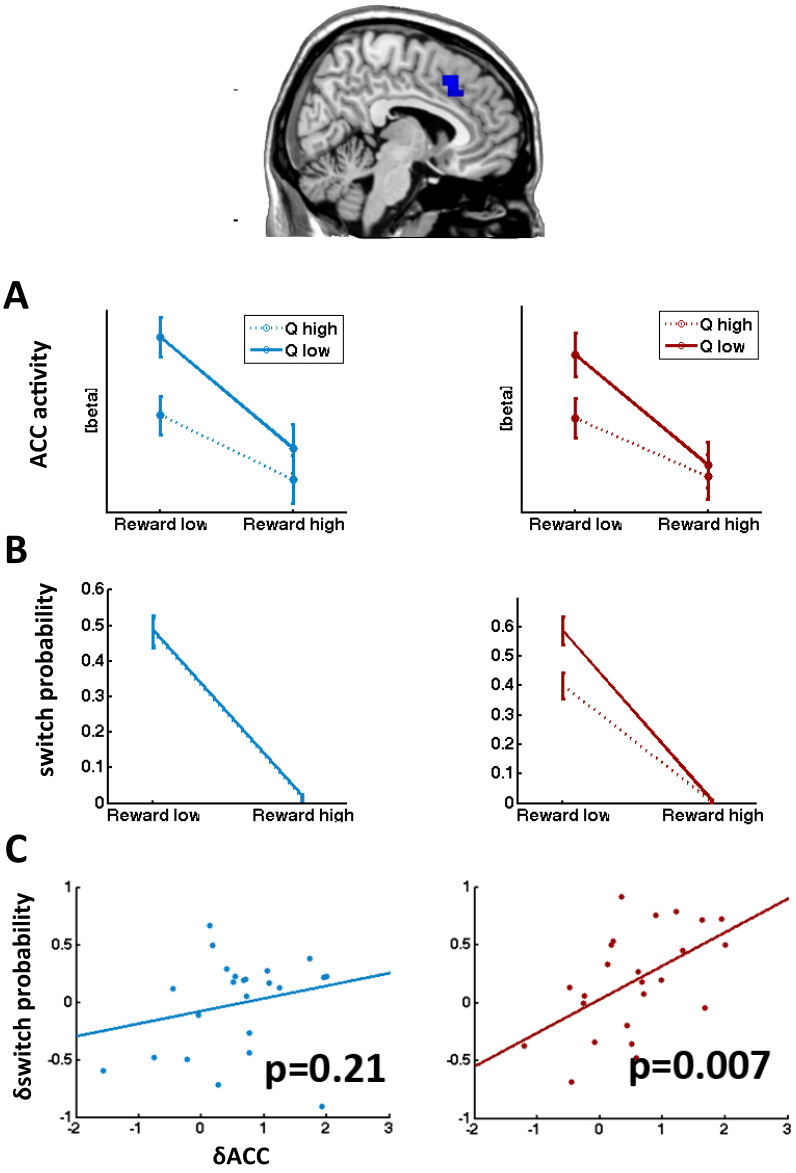


Figure 7.9: ACC activity correlates with subsequent task switching behavior - **A** ACC region of interest activity extracted from 2x2 GLM for high Q values (dashed lines) and low Q values (solid lines) on low reward and high reward trials. identical activation pattern for task and action, with highest activity for low Q on low reward trials. **B** switch probability after a Q x reward trials. No Q effect observed for actions, task switching behavior with similar pattern to ACC activity **C** regression for switching behavior against ACC activity yields a significant correlation in tasks

correlations in actions (r=0.27, p=0.2), providing further evidence for a task effect in the ACC.

7.5 Discussion

Our data reveal the neural contributions of specific prefrontal cortices and striatal regions to value-based decision-making between concrete actions and more abstract structures, such as tasks. We discriminate between regions exhibiting an action effect (vmPFC and ventral striatum), and a task effect (ACC, dlPFC, FPC). vmPFC integrates the value of an action and the ventral striatum signals an reward prediction error of actions, irrespective of the condition, i.e. including actions within a task. We also detected an encoding of the non-evidence of the chosen task in dACC, dlPFC and FPC (Fig 4.). Importantly, activity in FPC corresponds to monitoring the evidence in favor of alternative task only (Fig. 5). During the feedback phase, ACC activity correlates with the individual subject's task switching behavior. Finally through psychophysiological interactions, we demonstrate that in the decision phase the network of four prefrontal regions changed their functional connectivity in tasks, but not in actions (Fig. 7). These results emphasize the critical role of the prefrontal cortices in deciding between more abstract behavioral structures rather than concrete actions to adjust behavior in a dynamic environment.

Several studies have shown that FPC is engaged in dual task conditions, i.e. monitoring the task that soon has to be performed, where an explicit cue tells them to switch (Koechlin *et al.* , 1999; Charron & Koechlin, 2010). We provide evidence here that FPC is also tracking the alternative task in environments where the subjects can freely choose between two tasks. The results presented here are in agreement with the recent findings by Boorman and colleagues, who showed a tracking of a counterfactual choice in the frontal pole (Boorman *et al.* , 2009). We demonstrate that this is valid for more complex behavior, but not for simple action value tracking. It should

be noted however that the voxel coordinates they indicated seem to be more ventral and closer to the orbitofrontal cluster we identified with a lower threshold, which is involved in actions as well as tasks. Furthermore, previous studies reported an activity increase in the frontopolar cortices when subjects leave optimal choices to explore alternative options (Daw *et al.* , 2006). We argue that the exploration-related activity in FPC can be understood as a correlation with the value of an unchosen structure, which increases when subjects perform suboptimal choices. A recent study showed that switching behavior in patients with frontopolar lesions is severely impaired (Rowe *et al.* , 2007). No evidence for an FPC implication in switching behavior was found in our analyses. We forward the hypothesis that FPC intervenes in case of rivalry between multiple unchosen options, and hence might not be engaged in switching in our design, since there is only one alternative.

Our results further demonstrate that vmPFC integrates the advantage in favor of the current action choice during the decision phase, independent of the condition. Previous studies show a general unison about the enrollment of vmPFC in monitoring expected outcomes of an ongoing course of action. The vmPFC BOLD signal correlates with the value of a chosen stimulus (Wunderlich *et al.* , 2010), or with the difference between a chosen and an unchosen option (Boorman *et al.* , 2009). In further agreement with our results, Smith and colleagues found an increasing vmPFC activity across different reward categories (Smith *et al.* , 2010), demonstrating a correlation with both increasing monetary and social rewards. Importantly our results point to a vmPFC implication in monitoring expected and actual outcomes, rather than making a decision per se.

The design of our study provides the means to conclude about the functional role of the ACC, and to clearly isolate its distinct contribution to behavior compared to other regions embraced in value-based decision-making. A huge body of literature implicates the ACC in learning and predicting the value of actions (Walton

et al. , 2004), signaling signed prediction errors, i.e. discrepancies between actual and predicted outcomes (Matsumoto *et al.* , 2007) and correlating with the environmental uncertainty (Behrens *et al.* , 2007). The findings presented here contribute to an understanding of the panoply of recent ACC activity interpretations. We were able to demonstrate an implication of ACC in tracking the non-evidence specific to the current task during the decision-phase, and an interaction between the non-evidence and the reward during feedback. Concretely, in the action condition where the subject performs choices within one task, but switches between actions, the ACC tracks continuously the non-evidence of the current task on switch and stay decision phases, whereas in the task condition, we observe a reset of the ACC correlation on trials where the subject performs a switch to the alternative task. Additionally a correlation between ACC activity during the feedback phase and subsequent task, and not action switching behavior is observed. As forwarded by the functional connectivity, the ACC serves as a hub specifically in the task condition, sending relevant information about the urge to adapt behavior to all other prefrontal clusters. In the framework of the parallel medial-lateral hierarchy in the cascade model, Kouneiher *et al.* (2009) showed that the medial prefrontal cortex regulates the need for cognitive control in the lateral prefrontal cortex according to the rewards at stake. In our study, motivational information might be sent from ACC to the dlPFC in the task condition before switching, but the ACC is not performing the switch.

Previous studies have investigated the anatomical connectivities in the prefrontal cortices (Medalla and Barbas, 2010; Beckmann et al., 2009). However, it still remains elusive how exactly the distinct prefrontal regions functionally engage in a network and share information to make decisions between several options. First, we found interactions between the four regions only for the task condition. This confirms our assumption that the human prefrontal cortex has evolved to go beyond deciding between simple actions. Furthermore we report that the FPC has no outgoing functional connections, which points to a role of monitoring the alternative task. This

finding might be specific to our paradigm; there is no need for subjects to track multiple options, since they only switch between two dependent options. Hence when the subject notices a value decrease of the chosen option, he might prefer the now unchosen option on the next trial without having to decide between multiple alternatives.

In most investigations on value-based decision-making the level of abstraction is not clearly defined. Our findings reveal a crucial difference between value-based decision-making in tasks and actions through the distinct nature of BOLD activity correlations and connectivity interactions between four prefrontal regions and the ventral striatum. We contribute to the sparse literature on the frontopolar cortices in humans, proposing a role in monitoring the value of alternative complex behavioral structures.

7.6 Supplementary Material

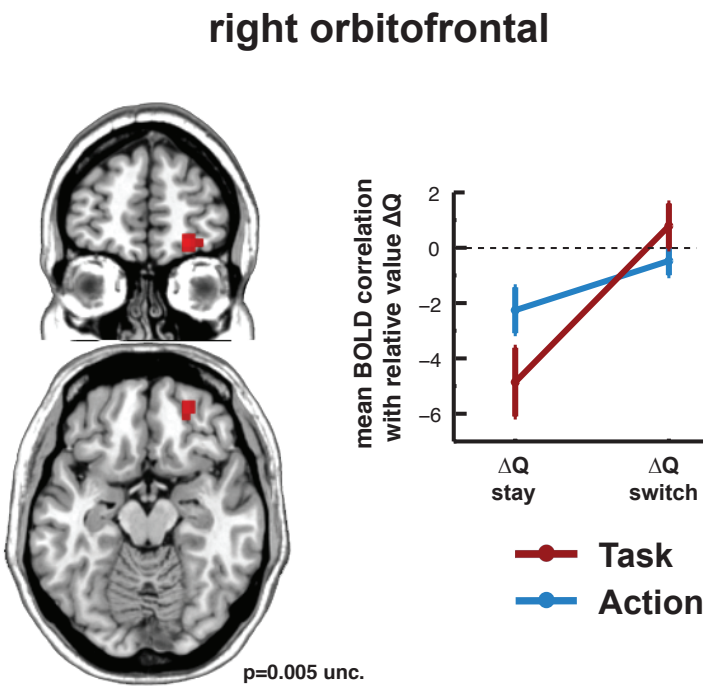


Figure 7.10: Supplementary 1: Orbitofrontal cluster - Activations in the orbitofrontal cortex were observed for both task and action conditions after lowering the threshold to p=0.005 uncorrected

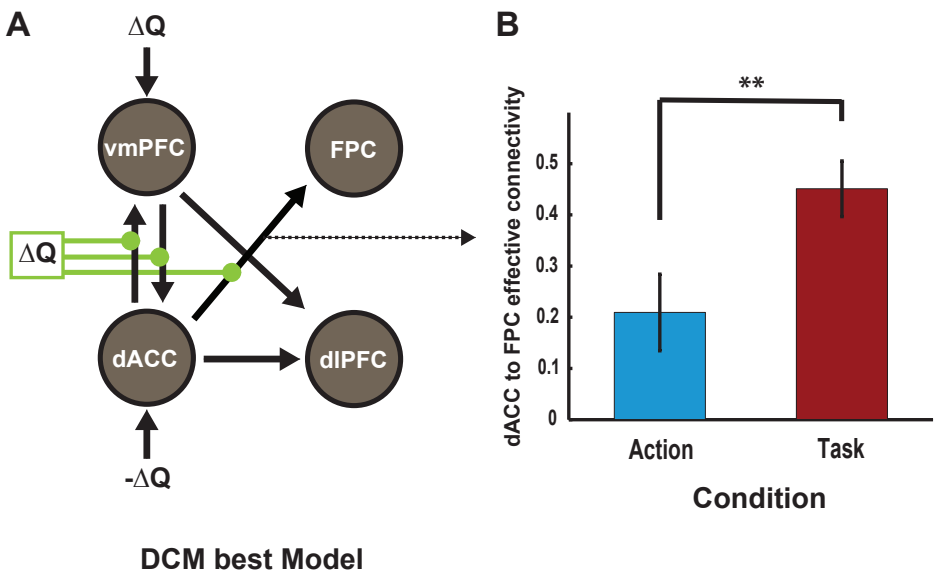


Figure 7.11: Supplementary 2: DCM - A. best model through bayesian model selection against other models with different connectivity patterns, confirming the PPIs. **B** Random effects analysis of connection parameter estimates shows a significant difference between task and action between the ACC and the FPC.

Part IV

Discussion

Chapter 8

Discussion

8.1 Implications of findings

The last decade has been marked by a rise of interest in grasping the neural correlates of value-based decision-making. This is partially due to the advances of behavioral ecology, describing the evolutionary pressure on foraging (Glimcher, 2002), but also on the reflections of decision theory in economics and progress in computational modeling of the neural structures (Doya, 2008). The multidisciplinary results try to characterize value computation in the brain, which can be tracked down in humans with the use of modern neuroimaging methods and recent analyses techniques.

As a review recently pointed out (Rushworth *et al.* , 2011), at least four prefrontal regions are implicated in value-based decision-making: ventromedial PFC, anterior cingulate cortex, dorsolateral PFC and the frontopolar cortices (and adjacent orbitofrontal areas). Much research has focused on one of these regions, mostly on monkeys, and only a handful have tried to differentiate the different functional roles in the same experiment (e.g. Kennerley *et al.* (2011)). Some controversies can be found between studies, and therefore there is still a huge amount of material for a forthcoming debate.

Even in studies where results are reported for several regions, same activity patterns emerge, making it difficult to declare a specific function for each region. For example, Boorman *et al.* (2011) found similar results in the frontopolar cortex, the dorsal ACC and the posterior cingulate cortex (fig. 8.1). Although their findings clearly suggest a three region-network for monitoring the value of unchosen options, respectively the second best option, one can raise the question about their distinct contributions. The conclusion they were able to draw out is that the use of the encoded information guides changes in behavior, all together. Regarding such problems, our studies give clearer interpretations of the different roles: we differentiate between ACC activity, predicting the switching to the alternative task on the subsequent trial, and the frontopolar cortex, which specifically encodes the evidence in favor of the alternative task, but does not seem to be engaged in switching, at least not in our design with only two dependent options.

As briefly discussed in a previous chapter, it is not always obvious to compare human neuroimaging results with neuron cell recordings from animals, especially because of their macro resp. micro resolution. Whereas studies like Matsumoto *et al.* (2007) report positive and negative prediction error signals in the same cell, neuroimaging in humans report global results, such as a environmental volatility correlation (Behrens *et al.* , 2007). More than merely describing what correlations can be observed in ACC, we find differences in ACC activity between conditions, and more interestingly an impact on behavior. Our results confirm that the ACC activity might look like a negative prediction error at some trials - but not on all trials; other correlations are hiding behind this pattern. The peak of the activity occurs on trials where subjects chose the less rewarded option and subsequently got a low reward. There is no prediction error in this case, since the prediction agrees with the reward outcome. It seems more like a confirmation that they truly are performing the worst option, and thus they collect more evidence in favor of a change in behavior.

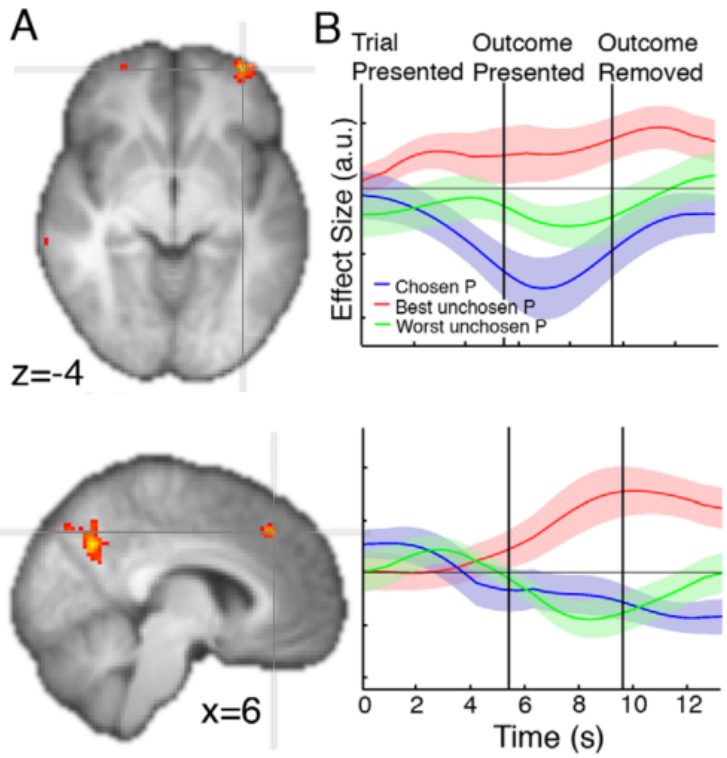


Figure 8.1: similar signals in ACC and FPC - from Boorman *et al.* (2011), making it difficult to conclude about the distinct functional roles - **A**. statistic maps representing the effect of reward probability of the best unchosen option. **B**. time course of the effects of the reward probability for the chosen option (blue), the best unchosen option (red) and the worst unchosen option (green) across decision and feedback phase.

Anatomical connection studies (Barbas & Pandya, 1989; Cavada *et al.* , 2000; Petrides & Pandya, 2007) have shown that the area around the ventromedial PFC is innerving both lateral and medial prefrontal areas. They are also heavily projecting to the striatum, more specifically the ventral regions of the putamen and caudate nucleus, over a large longitudinal dimension including head, body and tail of the caudate nucleus. It is still not well understood if these neurons also project to nucleus accumbens or not.

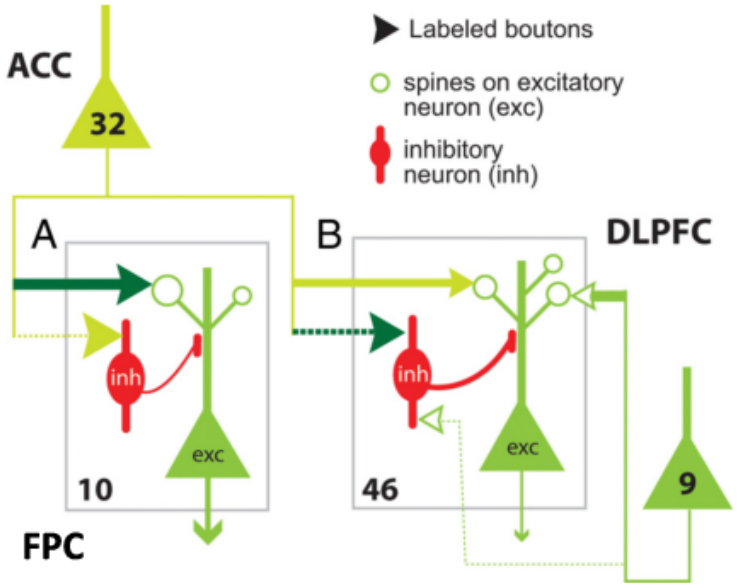


Figure 8.2: different ACC pathways to FPC and dlPFC - summary of the findings of Medalla & Barbas (2010) showing common (light green) and unique (dark green) synaptic connections to dlPFC and FPC. Line thickness represents prevalence of pathways, and size of arrowhead represents size of boutons. The pathway to FPC revealed more synapses on spines of putative excitatory neurons, and in contrast more synapses on the pathway to putative dlPFC inhibitory neurons

Medalla & Barbas (2010) investigated the projections from the ACC, specifically the innervations to dlPFC and FPC by injecting tracers into the ACC and analyzing

the properties of the axon terminals to the lateral prefrontal cortex. They found several differences between the dlPFC and FPC projections (fig. 8.2): First, synapses into FPC excitatory neurons have a larger size than those to other regions, even compared to the inhibitory neurons. Second, the projections to the dlPFC connect to more inhibitory neurons than in other areas. In an earlier study, Medalla & Barbas (2009) also reported that the medial-to-lateral projections conclude in more prevalent and larger synapses with inhibitory neurons than lateral-to-lateral projections and in larger innervations to calbindin inhibitory neurons. In the context of cognitive control, they explain that ACC neurons thus might be implicated in reducing the noise by inhibiting excitatory neurons in situations of high cognitive demands.

Although these findings might not be easily extrapolated from monkeys to humans, for the reasons discussed before, the synaptic interactions shown by these studies can help to a better understanding of our results. In tasks, ACC and vmPFC are interacting in both directions, pointing to an information exchange about the value and the non-evidence of the currently chosen task. Note however that the value of the unchosen task and the non-evidence of the chosen task are correlated in our study, since the two tasks are dependent; low task value means automatically that the other task has a high value. Future studies will have to investigate different scenarios, as discussed in the next section. The ACC who tracks the urge to an adaption of behavior, might send related information to the dlPFC to increase cognitive control. The dlPFC might thus be implicated in implementing the switch. The ACC also might send information related to the non-evidence of the current task to the FPC, which updates the value of the unchosen option. However since low value of one task implicates high value of the alternative, the FPC is not recruited for switching purposes in our design.

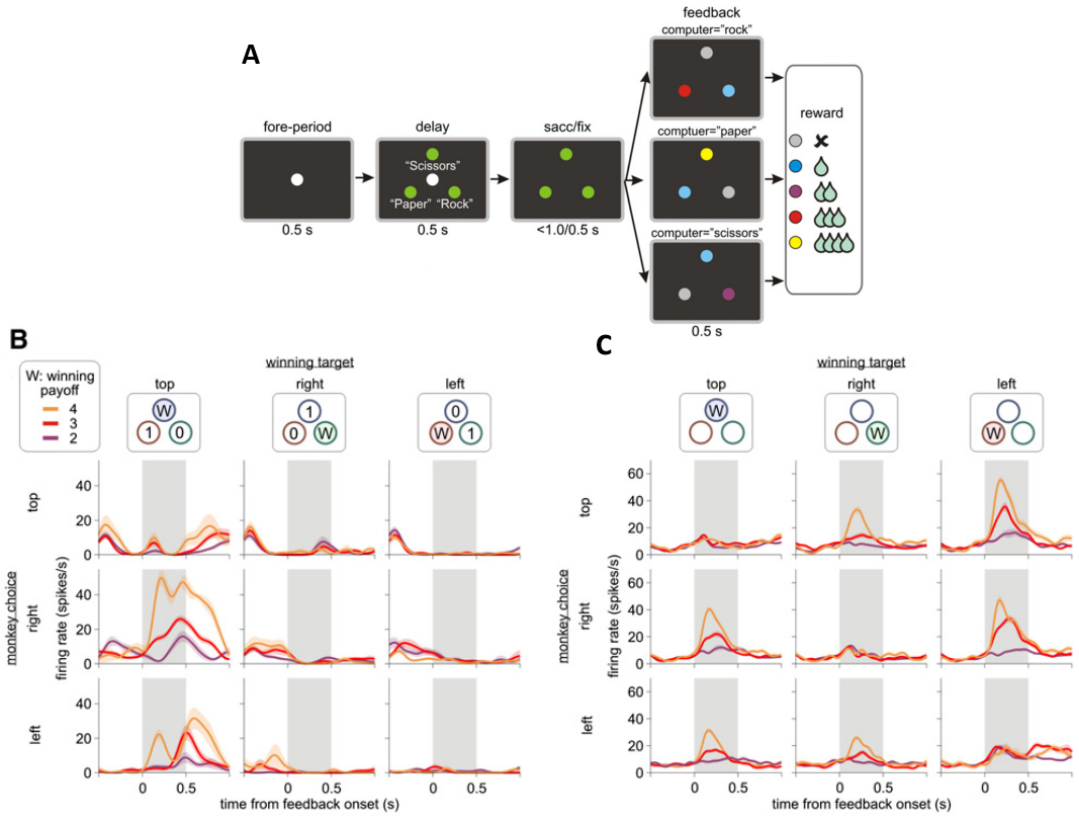


Figure 8.3: Unchosen action outcome signals in dlPFC - **A** a rock-paper-scissor game on monkeys (Abe & Lee, 2011) - **B** Spike density functions (SDF) of dlPFC neurons according to position and payoff of the winning target, and the choice of the monkey. **C** SPF of orbitofrontal neurons. - Both cell regions encode outcomes they did not receive in a similar manner, the only difference seems to be a position preference)

8.2 Limitations and perspectives

The paradigm and findings presented in this thesis are the first to directly compare actions and tasks in human value-based decision-making. In order to investigate our fundamental hypotheses, the design was chosen carefully: two options with dependent values, and a fixed reward distribution. However after this first work other questions are already waiting to be answered. In our daily life we often have to face more than two possible options: which recipe should I choose to cook for the parents-in-law? From which author should I buy a book for my vacation? The most obvious questions in the line of the results can be tested by a slightly adjusted paradigm: what if there are more than two tasks competing for choice? And at the same time, what if the values are not dependent, i.e. values of options evolving through time without being coupled to each other?

A recent paper by Abe & Lee (2011) investigated a similar issue on monkeys. Monkeys had to play a rock-paper-scissor against a computerized opponent, thus a three-choice task (fig. 8.3). The monkeys adapted their behavior according to rewards they did not receive, and cells in the dlPFC and orbitofrontal cortex correlated with these *fictive* reward signals. However, the authors were not able to clearly discriminate the specific contribution of each region. Boorman *et al.* (2011) addressed similar questions to humans, as discussed before, however again without differentiating on the level of action abstraction; the participants had to track continually the expected values of image categories with randomized screen locations, so that the values can not be reduced to simple motor responses. The BOLD signal in FPC reflected the accumulated evidence of the second-best option: when the value of the best option increases, activity in FPC decreases, and in contrast when the value of the third best option decreases, FPC activity increases, since now the second best is more valuable. In this case, the FPC is not only interested in comparing the best two options, but it represents the relative value of the second option in comparison to other choices.

Further, one should note that in our study presented in this thesis, the subjects do not learn task or action configurations, rather the configurations are supposed to be integrated during the excessive training session. The more complex modeling question involving learning, changing and even creating task sets has previously been addressed by our team (Collins & Koechlin, 2012). Here, participants start the experiment without previous knowledge of task set configurations of the underlying actions, as discussed in chapter 3. Two young, motivated team members are currently scrutinizing the neural correlations of this model's subject-specific variables such as the reliability of the task sets at offer.

Another interesting issue is the integration of reward. In a real-world scenario we do not always update our beliefs based on one categorical reward value, but rather through multiple information sources, just like a trader on a stock market trying to collect information from as many sources as possible before deciding to buy or not to buy. Until now, neuroimaging studies and computational models only focused on the integration of a single reward per trial, respectively a learning process implicating the rewards of unchosen options. Such hypotheses might be tested best in a social interaction environment, where confederates give their opinion about which action to take before the actual participant makes a decision. An exciting point might be to control the probability of being correct of the confederates, so that the participants should track beliefs about the reliability of their opponents to maximize the gains. On a neural level, beliefs in reward information is still waiting to be dissociated from evidence in favor of a choice.

Finally, this leads us to one more intriguing issue on behalf of the reward integration and its relation to the probabilistic reward distribution. In our experimental set-up, due to the specific reward distribution (cf. fig 7.1C), the reward values of 6€ and 7€ are highly informative about the probability of performing the most rewarded action/task, whereas 8€ and 9€ are less informative, du to the decreased

difference between the high and low rewarded options. Thus, one might expect that, if subjects perform optimally, reinforcement learning algorithms would perform less good due to the linear correlation of the prediction error with the reward magnitude, and a purely bayesian subject would perform better. However we show that, at least in our design, this is not the case. A future study might therefore be optimized to distinguish first between situations where an RL resp. a bayesian strategy would have a different impact on subjects' behavior, second between the reward magnitude and the actual information a reward value contains, and third reveal neural networks that are engaged specifically for RL or for bayesian computations.

Bibliography

Abe, Hiroshi, & Lee, Daeyeol. 2011. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron*, **70**(4), 731–41. 146, 147

Amiez, C, Joseph, J P, & Procyk, E. 2006. Reward encoding in the monkey anterior cingulate cortex. *Cerebral cortex (New York, N.Y. : 1991)*, **16**(7), 1040–55. 101

Baddeley, A. 1986. *Working Memory*. New York: Oxford University Press. 78

Badre, David. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in cognitive sciences*, **12**(5), 193–200. 77

Badre, David, & D'Esposito, Mark. 2007. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of cognitive neuroscience*, **19**(12), 2082–99. 53

Barbas, H, & Pandya, D N. 1989. Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, **286**(3), 353–375. 144

Barlow, H.B. 1959. Sensory mechanisms, the reduction of redundancy, and intelligence. *The mechanisation of thought processes*, 535–539. 42

Basten, Ulrike, Biele, Guido, Heekeren, Hauke R, & Fiebach, Christian J. 2010. How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(50), 21767–72. 95

Bayer, Hannah M, & Glimcher, Paul W. 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, **47**(1), 129–41. 91, 93, 109

Bechara, A, Damasio, a R, Damasio, H, & Anderson, S W. 1994. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, **50**(1-3), 7–15. 94

Bechara, A, Tranel, D, & Damasio, H. 2000. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, **123 (Pt 1**(Nov.), 2189–202. 76, 94

Behrens, Timothy E J, Woolrich, Mark W, Walton, Mark E, & Rushworth, Matthew F S. 2007. Learning the value of information in an uncertain world. *Nature neuroscience*, **10**(9), 1214–21. 43, 44, 101, 110, 135, 142

Belliveau, JW, Kennedy, DN, McKinstry, RC, Buchbinder, BR, Weisskoff, RM, Cohen, MS, Vevea, JM, Brady, TJ, & Rosen, BR. 1991. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, **254**(5032), 716. 63, 64

Belovsky, Gary E. 1984. Herbivore Optimal Foraging : A Comparative Test of Three Models. *The American Naturalist*, **124**, 97–115. 19

Boly, Melanie, Garrido, Marta Isabel, Gosseries, Olivia, Bruno, Marie-Aurélie, Boveroux, Pierre, Schnakers, Caroline, Massimini, Marcello, Litvak, Vladimir, Laureys, Steven, & Friston, Karl. 2011. Preserved feedforward but impaired top-down processes in the vegetative state. *Science*, **332**(6031), 858–62. 72

Boorman, Erie D, Behrens, Timothy E J, Woolrich, Mark W, & Rushworth, Matthew F S. 2009. How green is the grass on the other side? Frontopolar cortex and the

evidence in favor of alternative courses of action. *Neuron*, **62**(5), 733–43. 95, 97, 98, 110, 133, 134

Boorman, Erie D, Behrens, Timothy E., & Rushworth, Matthew F. 2011. Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex. *PLoS Biology*, **9**(6), e1001093. 98, 142, 143, 147

Borchers, Svenja, Himmelbach, Marc, Logothetis, Nikos, & Karnath, Hans-Otto. 2012. Direct electrical stimulation of human cortex - the gold standard for mapping brain functions? *Nature reviews. Neuroscience*, **13**(1), 63–70. 68

Botvinick, Mattew M. 2007. Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, affective & behavioral neuroscience*, **7**(4), 356–66. 82

Botvinick, Matthew M, & Plaut, David C. 2004. Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. *Psychological review*, **111**(2), 395–429. 49, 50, 51

Botvinick, Matthew M, Nystrom, L E, Fissell, K, Carter, C S, & Cohen, J D. 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, **402**(6758), 179–81. 82

Botvinick, Matthew M, Braver, Todd S, Barch, Deanna M, Carter, Cameron S, & Cohen, Jonathan D. 2001. Conflict Monitoring and Cognitive Control. *Psychological Review*, **108**(3), 624–652. 80, 82

Botvinick, Matthew M, Niv, Yael, & Barto, Andrew C. 2009. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition*, **113**(3), 262–80. 51, 52

Braver, Todd S, & Bongiolatti, Susan R. 2002. The role of frontopolar cortex in subgoal processing during working memory. *NeuroImage*, **15**(3), 523–36. 85

Brodmann, K. 1909. *Vergleichende Lokalisationslehre der Grosshirnrinde*. Barth Verlag, Leipzig. 74, 76

Burgess, P W, Quayle, a, & Frith, C D. 2001. Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia*, **39**(6), 545–55. 85

Camille, Nathalie, Coricelli, Giorgio, Sallet, Jerome, Pradat-Diehl, Pascale, Duhamel, Jean-Rene, & Sirigu, Angela. 2004. The Involvement of the Orbitofrontal Cortex in the Experience of Regret. *Science*, **304**(May), 1167–1170. 94

Camille, Nathalie, Griffiths, Cathryn a, Vo, Khoi, Fellows, Lesley K, & Kable, Joseph W. 2011. Ventromedial frontal lobe damage disrupts value maximization in humans. *The Journal of Neuroscience*, **31**(20), 7527–32. 94

Carmichael, S T, & Price, J L. 1995. Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, **363**(4), 615–641. 94

Carmichael, S T, & Price, J L. 1996. Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology*, **371**(2), 179–207. 93, 94

Cavada, C, Compañy, T, Tejedor, J, Cruz-Rizzolo, R J, & Reinoso-Suárez, F. 2000. The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cerebral cortex (New York, N.Y. : 1991)*, **10**(3), 220–42. 144

Charnov, Eric L, & Orians, Gordon H. 1973. *Optimal Foraging: Some theoretical explorations*. 25, 28

Charron, Sylvain, & Koechlin, Etienne. 2010. Divided representation of concurrent goals in the human frontal lobes. *Science*, **328**(5976), 360–3. 86, 97, 133

Christoff, Kalina, & Gabrieli, John D E. 2000. The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organisation within the human prefrontal cortex. *Psychobiology*, **28**(2), 168–186. 85

Collins, Anne, & Koechlin, Etienne. 2012. Reasoning , Learning and Creativity : Frontal Lobes Function and Human Decision-Making. *Plos Biology*, **in press.** 56, 58, 59, 148

Cooper, Richard P, & Shallice, Tim. 2000. Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, **17**(4), 297–338. 49, 50, 79

Cooper, Richard P, & Shallice, Tim. 2006. Hierarchical schemas and goals in the control of sequential behavior. *Psychological review*, **113**(4), 887–916; discussion 917–31. 51, 79

Coull, J T, Frith, C D, Frackowiak, R S, & Grasby, P M. 1996. A fronto-parietal network for rapid visual information processing: a PET study of sustained attention and working memory. *Neuropsychologia*, **34**(11), 1085–95. 84, 85

Daw, Nathaniel D, O'Doherty, John P, Dayan, Peter, Seymour, Ben, & Dolan, Raymond J. 2006. Cortical substrates for exploratory decisions in humans. *Nature*, **441**(7095), 876–9. 57, 95, 96, 98, 110, 116, 134

Daw, Nathaniel D, Gershman, Samuel J, Seymour, Ben, Dayan, Peter, & Dolan, Raymond J. 2011. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, **69**(6), 1204–15. 52

Dayan, P, Kakade, S, & Montague, P R. 2000. Learning and selective attention. *Nature neuroscience*, **3 Suppl**(november), 1218–23. 43

Dayan, Peter. 2002. Reward, motivation, and reinforcement learning. *Neuron*, **36**, 285–298. 91

Dehaene, S, & Changeux, J P. 1997. A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Sciences of the United States of America*, **94**(24), 13293–8. 48

Descartes, René. 1648. *Traité de l'Homme*. 21

D'Esposito, M, Ballard, D, Aguirre, G K, & Zarahn, E. 1998. Human prefrontal cortex is not specific for working memory: a functional MRI study. *NeuroImage*, **8**(3), 274–82. 79

Devlin, K. 2008. *The Unfinished Game: Pascal, Fermat, and the Seventeenth-Century Letter That Made the World Modern*. 1st editio edn. Basic Books. 41

Doya, Kenji. 2008. Modulators of decision making. *Nature neuroscience*, **11**(4), 410–6. 141

Dreher, Jean-Claude, Koechlin, Etienne, Tierney, Michael, & Grafman, Jordan. 2008. Damage to the fronto-polar cortex is associated with impaired multitasking. *PloS one*, **3**(9), e3227. 86, 110

Düzel, E, Cabeza, R, Picton, T W, Yonelinas, a P, Scheich, H, Heinze, H J, & Tulving, E. 1999. Task-related and item-related brain processes of memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(4), 1794–9. 85

Emlen, JM. 1966. The role of time and energy in food preference. *The American Naturalist*, **100**(916), 611–617. 24

Feldman, R P, & Goodrich, J T. 1999. The Edwin Smith Surgical Papyrus. *Child's nervous system*, **15**(6-7), 281–4. 73

Fienberg, Stephen E. 2006. When Did Bayesian Inference Become 'Bayesian'? *Bayesian Analysis*, **1**(1), 1–40. 42

Fiorillo, Christopher D, Tobler, Philippe N, & Schultz, Wolfram. 2003. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science (New York, N.Y.)*, **299**(5614), 1898–902. 90

Fiorillo, Christopher D, Newsome, William T, & Schultz, Wolfram. 2008. The temporal precision of reward prediction in dopamine neurons. *Nature neuroscience*, **11**(8), 966–973. 91

Friston, K, & Büchel, C. 2003. Functional connectivity: eigenimages and multivariate analyses. *In: Human Brain Function*, 2nd editio edn. Academic Press. 68, 69

Friston, K J, Buechel, C, Fink, G R, Morris, J, Rolls, E, & Dolan, R J. 1997. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, **6**(3), 218–29. 70

Funahashi, S, Bruce, CJ, & Goldman-Rakic, P. 1989. Mnemonic Coding of Visual Space in the Monkey s Dorsolateral Prefrontal Cortex. *Journal of neurophysiology*, **61**(2), 331–349. 78

Fuster, J.M. 2001. The Prefrontal Cortex - An Update : Time is of the essence. *Neuron*, **30**(2), 319–333. 53, 79

Gehring, W J, Coles, M G H, Meyer, D E, & Donchin, E. 1990. The error-related negativity: an event-related brain potential accompanying errors. *Psychophysiology*, **27**, s34. 81

Gehring, William J, & Willoughby, Adrian R. 2002. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, **295**(5563), 2279–82. 83

Gemba, H, Sasaki, K, & Brooks, V B. 1986. 'Error' potentials in limbic cortex (anterior cingulate area 24) of monkeys during motor learning. *Neuroscience Letters*, **70**(2), 223–227. 81

Gläscher, Jan, Hampton, Alan N, & O'Doherty, John P. 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral cortex*, **19**(2), 483–95. 95

Gläscher, Jan, Daw, Nathaniel D, Dayan, Peter, & O'Doherty, John P. 2010. States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. *Neuron*, **66**(4), 585–595. 52

Glimcher, Paul. 2002. Decisions, decisions, decisions: choosing a biological science of choice. *Neuron*, **36**(2), 323–32. 20, 141

Goldman-Rakic, P. 1987. Circuitry of Primate Prefrontal Cortex and Regulation of Behavior by Representational Memory. *Pages 373–417 of: Handbook of physiology*. Bethesda: American Physiological Society. 78

Grafe, T Ulmar, Schöner, Caroline R, Kerth, Gerald, Junaidi, Anissa, & Schöner, Michael G. 2011. A novel resource-service mutualism between bats and pitcher plants. *Biology letters*, **7**(3), 436–9. 19

Graybiel, A M. 2000. The basal ganglia. *Current Biology*, **10**(14), R509–11. 90

Graybiel, A M. 2005. The basal ganglia: learning new tricks and loving it. *Current opinion in neurobiology*, **15**(6), 638–44. 92

Hampton, Alan N, Bossaerts, Peter, & O'Doherty, John P. 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *The Journal of Neuroscience*, **26**(32), 8360–7. 95

Hayden, B.Y., Pearson, J.M., & Platt, M.L. 2009. Fictive reward signals in the anterior cingulate cortex. *science*, **324**(5929), 948. 101, 110

Hollerman, J R, & Schultz, W. 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature neuroscience*, **1**(4), 304–9. 90

Hollingworth, Harry L, & Poffenberger, AT. 1919. *Applied Psychology*. D. Appleton and Company. 56

Holroyd, Clay B., & Coles, Michael G.H. 2002. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, **109**(4), 679–709. 81, 82

Hyafil, Alexandre, Summerfield, Christopher, & Koechlin, Etienne. 2009. Two mechanisms for task switching in the prefrontal cortex. *The Journal of Neuroscience*, **29**(16), 5135–42. 57

Ito, Shigehiko, Stuphorn, Veit, Brown, Joshua W, & Schall, Jeffrey D. 2003. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*, **302**(5642), 120–2. 83

Jersild, AT. 1927. Mental set and shift. *Archives of psychology*. 56

Kable, Joseph W, & Glimcher, Paul W. 2007. The neural correlates of subjective value during intertemporal choice. *Nature neuroscience*, **10**(12), 1625–33. 95

Kahneman, D, & Tversky, A. 1979. Prospect Theory: An analysis of decision under risk. *Econometrica*, **47**(2), 263–291. 29, 30

Kahnt, Thorsten, Heinzle, Jakob, Park, S.Q., & Haynes, J.D. 2010. The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, **107**(13), 6010. 97

Keele, SW. 1968. Movement control in skilled motor performance. *Psychological bulletin*, **70**(6), 387–403. 47

Kennedy, D N, Lange, N, Makris, N, Bates, J, Meyer, J, & Caviness, V S. 1998. Gyri of the human neocortex: an MRI-based analysis of volume and variance. *Cerebral Cortex*, **8**(4), 372–84. 76

Kennerley, Steven W, Walton, Mark E, Behrens, Timothy E J, Buckley, Mark J, & Rushworth, Matthew F S. 2006. Optimal decision making and the anterior cingulate cortex. *Nature neuroscience*, **9**(7), 940–7. 100, 110

Kennerley, Steven W, Behrens, Timothy E J, & Wallis, Jonathan D. 2011. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience*, **14**(12), 1581–1589. 141

Kerns, John G, Cohen, Jonathan D, MacDonald, Angus W, Cho, Raymond Y, Stenger, V Andrew, & Carter, Cameron S. 2004. Anterior cingulate conflict monitoring and adjustments in control. *Science*, **303**(5660), 1023–6. 83

Kim, Hackjin, Shimojo, Shinsuke, & O'Doherty, John P. 2006. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS biology*, **4**(8), e233. 94, 95

Knutson, Brian, Taylor, Jonathan, Kaufman, Matthew, Peterson, Richard, & Glover, Gary. 2005. Distributed neural representation of expected value. *The Journal of Neuroscience*, **25**(19), 4806–12. 95

Koch, Christof. 1999. *Biophysics of Computation. Information Processing in Single Neurons*. New York: Oxford University Press. 75

Koechlin, Etienne. 2011. Frontal pole function: what is specifically human? *Trends in cognitive sciences*, **15**(6), 241. 93

Koechlin, Etienne, & Summerfield, Christopher. 2007. An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences*, **11**(6), 229–35. 53, 55, 85, 98

Koechlin, Etienne, Basso, G, Pietrini, P, Panzer, S, & Grafman, J. 1999. The role of the anterior prefrontal cortex in human cognition. *Nature*, **399**(6732), 148–51. 85, 110, 133

Koechlin, Etienne, Danek, Adrian, Burnod, Yves, & Grafman, Jordan. 2002. Medial prefrontal and subcortical mechanisms underlying the acquisition of motor and cognitive action sequences in humans. *Neuron*, **35**(2), 371–81. 86

Koechlin, Etienne, Ody, Chrystèle, & Kouneiher, Frédérique. 2003. The architecture of cognitive control in the human prefrontal cortex. *Science*, **302**(5648), 1181–5. 53, 86

Konorski, J. 1967. *Integrative Activity of the Brain*. Vol. 40. Chicago: University of Chicago Press. 90

Körding, Konrad P, & Wolpert, Daniel M. 2004. Bayesian integration in sensorimotor learning. *Nature*, **427**(6971), 244–7. 44

Kouneiher, Frédérique, Charron, Sylvain, & Koechlin, Etienne. 2009. Motivation and cognitive control in the human prefrontal cortex. *Nature neuroscience*, **12**(7), 939–45. 86, 87, 135

Krebs, JR, JT, Erichsen, Webber, MI, & Charnov, Eric L. 1977. Optimal prey selection in the great tit. *Animal Behavior*, **25**, 30–38. 25, 26

Laming, D. 1979. Choice reaction performance following an error. *Acta Psychologica*, **43**(3), 199–224. 81

Laplace, P S. 1774. Mémoires sur la probabilité des causes par les évènements. *Mémoires de mathématique et des physiques presenteés à lAcadémie royale des sciences par divers savans lûs dans ses assemblées*, **6**, 621–656. 42

Lashley, K. 1951. The problem of serial order in behavior. *Cerebral Mechanisms in Behavior*, **2**(7), 112–147. 46

Lauterbur, P.C., & Others. 1973. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, **242**(5394), 190–191. 63

Levy, Richard, & Dubois, Bruno. 2006. Apathy and the functional anatomy of the prefrontal cortex-basal ganglia circuits. *Cerebral cortex*, **16**(7), 916–28. 76

Lhermitte, F, Pillon, B, & Serdaru, M. 1986. Human autonomy and the frontal lobes. Part I: Imitation and utilization behavior: a neuropsychological study of 75 patients. *Annals of neurology*, **19**(4), 326–34. 76

Logothetis, Nikos K. 2008. What we can do and what we cannot do with fMRI. *Nature*, **453**(7197), 869–78. 64

Ma, Wei Ji, Beck, Jeffrey M, Latham, Peter E, & Pouget, Alexandre. 2006. Bayesian inference with probabilistic population codes. *Nature neuroscience*, **9**(11), 1432–8. 44

MacArthur, RH, & Pianka, ER. 1966. On optimal use of a patchy environment. *The American Naturalist*, **100**(916), 603–609. 24

MacDonald, a. W. 2000. Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science*, **288**(5472), 1835–1838. 82

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Vol. 50. W.H. Freeman. 33

Matsumoto, Kenji, Suzuki, Wataru, & Tanaka, Keiji. 2003. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science*, **301**(5630), 229–32. 83

Matsumoto, Madoka, Matsumoto, Kenji, Abe, Hiroshi, & Tanaka, Keiji. 2007. Medial prefrontal cell activity signaling prediction errors of action values. *Nature neuroscience*, **10**(5), 647–56. 101, 102, 110, 135, 142

Matsuzaka, Yoshiya, Akiyama, Tetsuya, Tanji, Jun, & Mushiake, Hajime. 2012. Neuronal activity in the primate dorsomedial prefrontal cortex contributes to strategic selection of response tactics. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(12), 4633–4638. 131

Medalla, Maria, & Barbas, Helen. 2009. Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. *Neuron*, **61**(4), 609–20. 145

Medalla, Maria, & Barbas, Helen. 2010. Anterior cingulate synapses in prefrontal areas 10 and 46 suggest differential influence in cognitive control. *The Journal of Neuroscience*, **30**(48), 16068–81. 144

Miller, E K, & Cohen, J D. 2001. An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, **24**(Jan.), 167–202. 80

Miller, G.A., Galanter, E., & Pribram, K.H. 1960. *Plans and the structure of behavior.* Holt, Rinehart & Winston. 47, 78

Milner, B. 1963. Effects of different brain lesions on card sorting. *Archives of neurology*, **9**(July), 100–110. 57

Mirenowicz, J, & Schultz, W. 1994. Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology*, **72**(2), 1024–1027. 90

Morris, Genela, Nevet, Alon, Arkadir, David, Vaadia, Eilon, & Bergman, Hagai. 2006. Midbrain dopamine neurons encode decisions for future action. *Nature neuroscience*, **9**(8), 1057–63. 93

Nakahara, Hiroyuki, Itoh, Hideaki, Kawagoe, Reiko, Takikawa, Yoriko, & Hikosaka, Okihide. 2004. Dopamine neurons can represent context-dependent prediction error. *Neuron*, **41**(2), 269–80. 91

Nieuwenhuis, Sander, Forstmann, Birte U, & Wagenmakers, Eric-Jan. 2011. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, **14**(9), 1105–1107. 125

Niki, H, & Watanabe, M. 1979. Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research*, **171**(2), 213–224. 81

Norman, DA, & Shallice, Tim. 1986. Attention io Action: Willed and Auiomatic Control of Behavior. *Pages 1–18 of:* Davidson, RJ, Schwartz, GE, & Shapiro, D (eds), *Consciousness and selfregulation*. New York: Plenum Press. 79

O'Doherty, John P, Kringelbach, M L, Rolls, E T, Hornak, J, & Andrews, C. 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature neuroscience*, **4**(1), 95–102. 97

O'Doherty, John P, Winston, J., Critchley, H., Perrett, D., Burt, DM, & Dolan, RJ. 2003a. Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, **41**(2), 147–155. 94

O'Doherty, John P, Critchley, Hugo, Deichmann, Ralf, & Dolan, Raymond J. 2003b. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **23**(21), 7931–9. 95

O'Doherty, John P, Dayan, Peter, Friston, Karl, Critchley, Hugo, & Dolan, Raymond J. 2003c. Temporal difference models and reward-related learning in the human brain. *Neuron*, **38**(2), 329–37. 68, 91, 110

O'Doherty, John P, Dayan, Peter, Schultz, Johannes, Deichmann, Ralf, Friston, Karl, & Dolan, Raymond J. 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, **304**(5669), 452–4. 91

O'Doherty, John P, Hampton, Alan N, & Kim, Hackjin. 2007. Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, **1104**(May), 35–53. 66, 67, 110

Ogawa, S, Lee, T M, Kay, a R, & Tank, D W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, **87**(24), 9868–72. 63

Okuda, J, Fujii, T, Yamadori, a, Kawashima, R, Tsukiura, T, Fukatsu, R, Suzuki, K, Ito, M, & Fukuda, H. 1998. Participation of the prefrontal cortices in prospective memory: evidence from a PET study in humans. *Neuroscience letters*, **253**(2), 127–30. 84, 85

Ongür, Dost, Ferry, Amon T, & Price, Joseph L. 2003. Architectonic subdivision of the human orbital and medial prefrontal cortex. *The Journal of comparative neurology*, **460**(3), 425–49. 93, 99

Padoa-Schioppa, Camillo, & Assad, John a. 2006. Neurons in the orbitofrontal cortex encode economic value. *Nature*, **441**(7090), 223–6. 97

Padoa-Schioppa, Camillo, & Assad, John a. 2008. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature neuroscience*, **11**(1), 95–102. 97

Passingham, Dick, & Sakai, Katsuyuki. 2004. The prefrontal cortex and working memory: physiology and brain imaging. *Current opinion in neurobiology*, **14**(2), 163–8. 79

Paus, T. S. a. a. 1999. Structural Maturation of Neural Pathways in Children and Adolescents: In Vivo Study. *Science*, **283**(5409), 1908–1911. 76

Pavlov, I P. 1927. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford University Press. 22

Pessiglione, Mathias, Petrovic, Predrag, Daunizeau, Jean, Palminteri, Stefano, Dolan, Raymond J, & Frith, Chris D. 2008. Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*, **59**(4), 561–7. 91

Petrides, Michael, & Pandya, Deepak N. 2007. Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *The Journal of Neuroscience*, **27**(43), 11573–86. 144

Phillips, GC, Zeki, SM, & Barlow, HB. 1984. Localisation of function in the cerebral cortex. Past, present and future. *Brain*, **107**, 327–361. 68

Plassmann, Hilke, O'Doherty, John P, & Rangel, Antonio. 2007. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **27**(37), 9984–8. 94

Procyk, E, Tanaka, Y L, & Joseph, J P. 2000. Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nature neuroscience*, **3**(5), 502–8. 83

Quilodran, R, Rothe, M, & Procyk, E. 2008. Behavioral Shifts and Action Valuation in the Anterior Cingulate Cortex. *Neuron*, **57**(2), 314–325. 101, 110

Rabbitt, P M. 1966. Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, **71**(2), 264–272. 81

Ramnani, Narender, & Owen, Adrian M. 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature reviews. Neuroscience*, **5**(3), 184–94. 85

Rescorla, RA, & Wagner, AR. 1972. a theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. *Pages 64–99 of: Classical Conditioning II: Current Research and Theory*. Appleton Crofts, New York. 36, 66

Ribas-Fernandes, José J F, Solway, Alec, Diuk, Carlos, McGuire, Joseph T, Barto, Andrew G, Niv, Yael, & Botvinick, Matthew M. 2011. A neural signature of hierarchical reinforcement learning. *Neuron*, **71**(2), 370–9. 52

Richter, H. V., & Cumming, G. S. 2005. Food availability and annual migration of the straw-colored fruit bat (Eidolon helvum). *Journal of Zoology*, **268**(1), 35–44. 19

Rogers, RD, & Monsell, S. 1995. The cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology*, **124**(2), 207–231. 54

Rolls, Edmund T., Kringelbach, Morten L., & de Araujo, Ivan E. T. 2003. Different representations of pleasant and unpleasant odours in the human brain. *European Journal of Neuroscience*, **18**(3), 695–703. 94

Rowe, James B, Sakai, Katsuyuki, Lund, Torben E, Ramsø y, Thomas, Christensen, Mark Schram, Baare, William F C, Paulson, Olaf B, & Passingham, RE. 2007. Is the prefrontal cortex necessary for establishing cognitive sets? *The Journal of Neuroscience*, **27**(48), 13303–10. 134

Roy, Mathieu, Shohamy, Daphna, & Wager, Tor D. 2012. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in cognitive sciences*, Feb. 94

Rushworth, M F S, Walton, M E, Kennerley, S W, & Bannerman, D M. 2004. Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, **8**(9). 83

Rushworth, Matthew F S, Noonan, Maryann P, Boorman, Erie D, Walton, Mark E, & Behrens, Timothy E. 2011. Frontal cortex and reward-guided learning and decision-making. *Neuron*, **70**(6), 1054–69. 109, 141

Sakai, Katsuyuki. 2008. Task set and prefrontal cortex. *Annual review of neuroscience*, **31**(Jan.), 219–45. 54

Sakai, Katsuyuki, & Passingham, Richard E. 2003. Prefrontal interactions reflect future task operations. *Nature neuroscience*, **6**(1), 75–81. 79, 80

Satoh, Takemasa, Nakai, Sadamu, Sato, Tatsuo, & Kimura, Minoru. 2003. Correlated coding of motivation and outcome of decision by dopamine neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **23**(30), 9913–23. 91

Scheffers, Marten K., & Coles, Michael G. H. 2000. Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, **26**(1), 141–151. 81

Schultz, W. 1986. Responses of Midbrain Dopamine Neurons to Behavioral Stimuli in the Monkey. *Journal of Neurophsyiology*, **56**(5), 1439–1461. 90

Schultz, W, & Romo, R. 1990. Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *Journal of Neurophysiology*, **63**(3), 607–624. 90

Schultz, W, Dayan, P, & Montague, P R. 1997. A neural substrate of prediction and reward. *Science*, **275**(5306), 1593–9. 81, 90

Semendeferi, K, Armstrong, E, Schleicher, a, Zilles, K, & Van Hoesen, G W. 2001. Prefrontal cortex in humans and apes: a comparative study of area 10. *American journal of physical anthropology*, **114**(3), 224–41. 93

Semendeferi, Katerina, Teffer, Kate, Buxhoeveden, Dan P, Park, Min S, Bludau, Sebastian, Amunts, Katrin, Travis, Katie, & Buckwalter, Joseph. 2010. Spatial Organization of Neurons in the Frontal Pole Sets Humans Apart from Great Apes. *Cerebral cortex*, Nov., 1485–1497. 93

Shallice, Tim. 1982. Specific impairments of planning. *Transactions of the Royal Society of London*, **298**, 199–209. 48, 79

Shizgal, P. 1997. Neural basis of utility estimation. *Current opinion in neurobiology*, **7**(2), 198–208. 32

Skinner, B F. 1938. *The Behavior of Organisms: An Experimental Analysis*. Century psychology series. Appleton-Century. 22

Small, Dana M, Gregory, Michael D, Mak, Y Erica, Gitelman, Darren, Mesulam, M Marsel, & Parrish, Todd. 2003. Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron*, **39**(4), 701–11. 94

Smith, David V, Hayden, Benjamin Y, Truong, Trong-Kha, Song, Allen W, Platt, Michael L, & Huettel, Scott a. 2010. Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *The Journal of Neuroscience*, **30**(7), 2490–5. 134

Star, Susan Leigh. 1989. Goltz. *Chap. Appendix A, page 202 of: Regions of the mind*. Stanford University Press. 68

Stephan, K E, Penny, W D, Moran, R J, den Ouden, H E M, Daunizeau, J, & Friston, K J. 2010. Ten simple rules for dynamic causal modeling. *NeuroImage*, **49**(4), 3099–109. 71

Stephens, DW, & Krebs, JR. 1986. *Foraging Theory*. Princeton University Press. 25

Stigler, S.M. 1983. Who discovered Bayes's theorem? *The American Statistician*, **37**(4), 290–296. 42

Sutton, Richard S., Precup, Doina, & Singh, Satinder. 1999. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, **112**(1-2), 181–211. 51

Sutton, RS. 1988. Learning to predict by the methods of temporal differences. *Machine Learning*, **3**, 9–44. 37

Sutton, R.S., & Barto, A.G. 1998. *Reinforcement Learning: An Introduction*. Vol. 9. Cambridge, MA: MIT Press. 114

Swanson, L W. 2000. Cerebral hemisphere regulation of motivated behavior. *Brain research*, **886**(1-2), 113–164. 89

Tanaka, Saori C, Doya, Kenji, Okada, Go, Ueda, Kazutaka, Okamoto, Yasumasa, & Yamawaki, Shigeto. 2004. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature neuroscience*, **7**(8), 887–93. 95

Tesauro, Gerald. 1994. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural computation*, **6**(2), 215–219. 90

Thorndike, E L. 1898. Animal Intelligence: An experimental Study of the Associate Process in Animals. *Psychological Review Monograph Supplement*, **2**(4), 1–8. 22

Tolman, Edward C. 1948. Cognitive Maps in Rats and Men. *Psychological Review*, **55**(4), 189–208. 22, 46

Tom, Sabrina M, Fox, Craig R, Trepel, Christopher, & Poldrack, Russell a. 2007. The neural basis of loss aversion in decision-making under risk. *Science*, **315**(5811), 515–8. 95

Tsujimoto, Satoshi, Genovesio, Aldo, & Wise, Steven P. 2010. Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nature neuroscience*, **13**(1), 120–6. 93

Tsujimoto, Satoshi, Genovesio, Aldo, & Wise, Steven P. 2011. Frontal pole cortex: encoding ends at the end of the endbrain. *Trends in cognitive sciences*, **15**(4), 169–76. 94

Tulving, Endel. 2002. episodic memory: From Mind to Brain. *Annual review of psychology*, **53**, 1–25. 85

Tversky, A., & Kahneman, D. 1981. The framing of decisions and the psychology of choice. *Science*, **211**(4481), 453. 28

Ullsperger, M, & von Cramon, D Y. 2001. Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *NeuroImage*, **14**(6), 1387–401. 81

Viswanathan, Ahalya, & Freeman, Ralph D. 2007. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nature neuroscience*, **10**(10), 1308–12. 66

Von Neumann, J, & Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Vol. 2. Princeton University Press. 28

Walton, Mark E, Devlin, Joseph T, & Rushworth, Matthew F S. 2004. Interactions between decision making and performance monitoring within prefrontal cortex. *Nature neuroscience*, **7**(11), 1259–65. 100, 134

Wunderlich, Klaus, Rangel, Antonio, & O'Doherty, John P. 2010. Economic choices can be made using only stimulus values. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(34), 15005–10. 134

Yang, Tianming, & Shadlen, Michael N. 2007. Probabilistic reasoning by neurons. *Nature*, **447**(7148), 1075–80. 44

Yoshida, Wako, & Ishii, Shin. 2006. Resolution of uncertainty in prefrontal cortex. *Neuron*, **50**(5), 781–9. 98