Alma Mater Studiorum – Università di Bologna

&

University College of London

DOTTORATO DI RICERCA IN

COGNITIVE NEUROSCIENCE

Ciclo XXII

Settore scientifico-disciplinare di afferenza
M-PSI/02

COGNITIVE AND AFFECTIVE PROCESSES IN SOCIAL ACTIONS AND DECISIONS

Presentata da: **Giovanna Moretto**

Coordinatore Dottorato

Prof.ssa Elisabetta Ladavas

Relatori

Prof. Giuseppe di Pellegrino

Prof. Patrick Haggard

Esame finale
anno 2010

CONTENTS

PART II

NEW EVIDENCES: EXPERIMENTAL STUDIES

Study I

Prefrontal damage reduces betrayal aversion in economic exchanges

Study II

Affective modulation of economic decision-making

Study III

Moral Judgment after Ventromedial Prefrontal Damage

Study VI
Experience of agency and sense of responsibility

INTRODUCTION

Our lives consist of an ongoing stream of decisions, choices and actions. The question of how we make, and how we should make judgments and decisions, has occupied thinkers for many centuries. Even if research on decision making has made enormous progress in the last decade, decision making still remains one of the most complex and least understood animal behaviours (Lee, 2008).

The study of decision making attempts to explain our fundamental ability to process multiple alternatives and to choose an 'optimal' course of action. This ability has been studied by various disciplines with different theoretical assumptions and measurement techniques although with relatively little integration of findings. Recently, cognitive neuroscience offers the possibility to readdress the lack of integration between disciplines. This integration is particular evident in the new interdisciplinary field popularly known as Neuroeconomics (Sanfey 2007). Neuroeconomics investigates the psychological and neuronal correlates of individual and social decision-making using tasks derived from experimental economics and Game Theory. These tasks create a simplified and controlled social interaction between subjects, often requiring sophisticated reasoning.

The classical approach to decision making typically suggests that in order to take a rational decision we should produce a list of costs and benefits for all alternatives in each

single decision, and choose on the basis of a rational calculation (Hansson, 1994). In economics, rational calculation translates into finding the alternative with the highest benefit and lowest cost. In economic terms. people make decisions that maximize their utility. However, producing a complete list of cost/benefit for each option, and compare each possible alternative could constitute a very long process, as well as cognitively be very demanding. In real life, we do not always have the time to produce this long list and make appropriate comparisons, as sometimes the variables involved in the decision are different and numerous, and the circumstances often require a fast decision. We can be rational within the limits of our cognitive capacities (Simon 1955, 1956), so the rational process to take a decision has to take in consideration of the limits of our cognition (Khaneman, 2003). Cohen (2005) suggests that people are simply incapable of optimally maximizing their utility for a variety of reasons, including limited access to information (about the past, present or future), limited ability to learn, and limited ability to focus and control behaviour. Even in circumstances where it could be easy to produce the list of alternatives and compare them following economic criteria, subjects may prefer an intuitive solution based more on a generic impact of the circumstances than on a rational process (Greene et al. 2001). This is especially true when they take a decision in a situation where opposing values (for example, moral and economic) compete with each other.

One possible way to provide evidence for an individual's preference for the intuitive strategy can be the involvement of emotions in the decision making process. Emotions have evolved because of their adaptive functions for genotypic and phenotypic survival (Ketelaar, 2004). Emotions may assist in prioritizing certain goals, allowing the individual to mobilize energy and give direction to behaviour (Bagozzi, et al. 2000; Frijda, 2006). Decision making is indeed often based on emotional processes (Damasio, 2004). Before deciding, we consider the possible practical and emotional effects of our choice, especially when the consequences of our decisions impact other people. Emotions are also present subsequent to the decision-making process. After having made a choice and before the outcomes are known we are often in state between hope and fear. When the outcomes materialize, they may again be a source of emotion, such as elation, happiness, surprise, regret and disappointment (Mellers, 2000). I think there are good reasons for emotions to be so pervasive in all phases of decision making. In contrast to the commonly held view that emotion blurs the decision-making process, cognitive neuroscience has shown that

emotions may help us in making the right, and perhaps the more advantageous and 'rational', decision (Bechara et al, 1999).

The rational and emotive approaches to decision making could represent two separate processes in our mind. Minsky defined mind (1986) as a "society of minds" to explain the idea that the human mind can be understood as an aggregate of separate processes, each with its own goals and operating according to its own principles. Psychologists have long recognized the distinction between efficient but highly specialized "automatic" processes and less efficient but more general mechanisms involved in "controlled" processing (Cohen, Dunbar and McClelland, 1990; Kahneman and Treisman, 1984). This idea has also found its way into the decision-making and economic literatures, where a distinction has been made between intuitive-system and deliberative-system mechanisms (Kahneman, 2003; Camerer, Loewenstein and Prelec, 2005). The intuitive/emotive system corresponds closely to automatic processing; it quickly proposes intuitive answers to problems as they arise. The rational/deliberative system corresponds closely to controlled processes; it monitors the quality of answers provided by the intuitive system and, in some situations, corrects or overrides these judgments.

Recent investigations have tried to locate in the brain the two systems (Montague et al 2004; Craig, 2002; O'Doherty et al., 2001; Phelps et al., 2001; Rolls, 2000; Whalen et al., 2001, Greene, 2007). The intuitive system is seen to involve subcortical structures responding directly to rewarding events, and is involved in fundamental forms of reinforcement learning (Montague et al., 2004), such as brainstem that release the neurotransmitter dopamine and the striatum that are influenced by the release of dopamine. These, and other subcortical structures responsive to valenced events (that is, events associated with positive or negative utility), make direct connections between several structures within the frontal lobes and temporal lobes (Craig, 2002; O'Doherty et al., 2001; Phelps et al., 2001; Rolls, 2000; Whalen et al., 2001). These cortical areas include medial and orbital regions of frontal cortex, the amygdala, and insular. These cortical structures are classically referred to as the limbic system of the brain, and are thought to be critical to emotional processing (Dalgleish, 2004). In contrast the system including deliberative thought, abstract reasoning, problem solving, planning seem located in the anterior and dorsolateral regions of prefrontal cortex, lying along the upper and front most surfaces of

the frontal lobes (Duncan, 1986; Koechlin et al., 1999; Miller and Cohen, 2001; Shallice and Burgess, 1991)

Emotions have been considered important variables affecting decisions (Cohen, 2005). Emotions, however, did not make it into decision research because they were seen as a subjective variable, experimentally difficult to measure and control, intrinsically unstable and unpredictable. There are surprisingly few studies on social and economic decision making that have explicitly manipulated or measured emotion. Even in the last years, the influence of emotions is inferred rather than effectively measured. For this reason, the aim of the experimental part of this research work is to increase our knowledge of the role of emotions in decision making. I propose several tasks where I attempt to manipulate and measure emotions in social situations.

In this research two different kinds of decisions are investigated. The first kind of decision regards the choice to trust a stranger in order to have an economic advantage (study I and II). The second kind of decision involves moral dilemmas (study III and IV). These two kinds of choice are particular useful in investigating the role of emotions in decision making process because both decisions induce a clear conflict between the rational/economic solution and the more emotive/intuitive solution (Greene et al., 2001; Greene, 2007; McCabe, 2001; Berg et al., 1995). In the case of moral decisions, a smaller number of victims (economic criteria) is opposed to a greater sense of guilt and regret. In trust decisions, the fear of being betrayed and humiliated is opposed to the possibility of securing greater economic income.

For both type of decisions: i) a formal model of rationality proposed by theorists of decisions or by game theory was identified; ii) I tested whether subjects follow normative standards and how; iii) I tested patients with focal lesion (lesion method) in brain areas specifically involved in emotion and decision making; iv) and I explained why subject follow/fail to follow normative standards.

The overall aim of this research work concerns the hypothesis that rational and deliberative process may work in conjunction with, and even helped by, emotions. Emotions could restrict the size of the consideration set and focus the decision maker on certain, relevant aspects of the options (Hanoch, 2001). Emotions serve to assign value to

objects, aid learning of how to obtain those objects, and could provide the motivation for doing so (Gifford, 2002). In accordance with Loewestein & Lerner (2003), I believe that emotions can influence the decision in several different ways: i) by predicting the emotional consequences of decision outcomes (expected emotions) before a decision is actually made, ii) by the actual emotional reaction related to the decision itself, and iii) by incidental emotions present before and/or during the process of decision. These three means of emotional involvement are not independent of each other; there is a mutual influence between them. However, they could have specific attributes. In the experimental part studies are presented that focus on the different means of emotional involvement.

Specifically in study III, I investigated how anticipated emotions can shape moral decisions. The study provides psychophysiological evidence (skin conductance activity) of the shaping role of emotions on decision making. In this study, decisions are proposed in which it is possible to foresee the sure effect of our action (non-ambiguous situations). I hypothesized that subjects should choose the option with the least negative emotional impact. This implies the capacity to consider the emotional content of effects (affective forecasting) in relation to the values (for example, economic or moral) used in taking a decision. Moreover the investigation of moral choice in patients with medial prefrontal cortex lesion (specifically the ventro-medial prefrontal cortex, vmPFC) suggests that this area could have a specific role in forecasting the expected emotions.

In study I, I investigated the role that emotions play in the decision to trust a stranger. In case of a decision under uncertainty, I hypothesis that emotions could have a more intense influence on the decision making process because with expected emotions there are also emotions, especially fear, that are experienced at the time of decision making in a situation with a very high level of uncertainty and risk. An example of this situation is the 'Trust Game' (Berg, 1995). This game involves real monetary exchanges between two anonymous individuals, the investor and the trustee, who receive a sum of money from the experimenter. The investor can keep all the money or decide to invest some amount, which is tripled by the experimenter and sent to the trustee. Next, the trustee decides how much of the tripled amount to return. Money sent by the investor is used to measure her trust, while money returned by the trustee is used to measure her reciprocity. This game generates a conflict between the possibilities of a trustful relation with a greater income for both counterparts and the possibility of being betrayed, humiliated and losing one's own

money. Healthy subjects usually develop strategies of choice in order to balance the positive and negative instances of the game. In study I it is observed how patients with vmPFC lesion fail in this balance.

The indirect influence of emotion on decision making was investigated in study II, where emotional stimuli are introduced incidentally during a decision making process in the trust game task. Indirect effects of incidental emotion are those that are mediated by changes in expected emotions or changes in the quality and/or quantity of information processing. In order to understand how incidental emotion can affect decisions we measured the level of trust in a group of healthy subjects playing the trust game in which incidental emotional expression are presented.

Study IV investigates the impact of emotion on the actions of choice. The last component of a decision is the selection of an action representing our choice. The selection of action is not a merely instrumental realization of our choice but is actually the real, and probably the unique, evidence of the choice itself and the possible linkage between decision and outcome. For this reason we hypothesise that the emotive content of the action could affect also the subjects' experience of linkage of actions to their effects (sense of agency).

This work starts with three introductory chapters outlining current knowledge about the decision making process. In chapter I, the definition of decision making is introduced, starting from the Decision Theory point of view and Economic definitions, and arriving at the definition proposed by Cognitive Neuroscience. In this chapter, basic concepts, such as reward and value, are introduced. These concepts are at the basis of the actual definition of decision making process in neuroscience. In chapter II, different sorts of decisions are described, such as economic decisions, social decisions and moral decisions. Moreover, a definition of emotions is provided and how these are interconnected with the process of decision making. In chapter III a model is proposed that is able to describe optimal economic decision making. These chapters introduce important concepts and definitions that are used in the experimental part.

In conclusion, in my research work I investigated the role that emotion plays in the process of decision making, by using relatively complex decisions such as those regarding

moral, social and economic behaviour. Results of the experiments offer the possibility to better understand the interaction between cognition and emotions in decision making. However, our current knowledge of neural mechanisms underlying social decision making is still limited. Further research is necessary to better understand this interesting and complex human behaviour.

Chapter I

STARTING POINTS FOR DECISION MAKING

1.1 Decision Making

Almost everything that a human being does involves decisions. Therefore, to theorise about decisions is almost the same as to theorise about human activity. Modern theories about human activity have developed since the middle of the 20th century through contributions from several academic disciplines. Nowadays, a specific new academic subject called Decision Theory (DT) exists that studies decision processing, but the subject of DT still remains a not very unified one. There are many different ways to theorise about decisions, and therefore also many different research traditions. Economists, statisticians, psychologists, political and social scientists and philosopher are all contributing to a better understanding of decision making, and recently neuroscientists have been added to this long list; with the specific purpose of understanding the brain mechanisms underlying decisions.

Decision theorists place significant effort in defining how we should decide in order to be rational, especially when there is uncertainty and lack of information. If a general wants to win a war, the decision theorist tries to tell him how to achieve this goal in the most rational way. In this sense, a decision is a goal-directed behaviour in the presence of options (Hansson, 1994). Alternatives are typically courses of action that are open to the decision-maker at the time of the decision. The set of alternatives can be more or less well-defined. In some decision problems, it is *open* in the sense that new alternatives can be invented or discovered by the decision-maker.

The question of whether the general should try to win the war at all is not regarded as a decision-theoretical issue. The focus of decision theory is on the norms for a rational

decision to achieve whatever goals. The motivations/reasons to achieve the goal do not interest DT. DT approaches all decisions as rational processes as constituted by two essential dimensions: *choice*, the evaluation of options and selection of actions; and *judgement*, information processing and probability estimation. To portray the matter fully accurately, it must be noted that DT encompasses a large number of models describing decision making process but the majority contain these two components (Hansson, 1994).

Economists added to the DT approach a general and unified goal: maximisation of utility (Sanfey, 2006). Maximisation of utility means always selecting the options with the highest economic benefits and the lowest costs. This approach dominated the research in decision making for the entire XXth century, banning decision making processing in unrealistic rational/mathematical dimensions and forgetting the role of emotions, contexts, moral motivations, values etc. The lack of realism evident in rational choice theory was first demonstrated by Kahnema & Tversky and their colleagues' work (1979, 2000), when they showed that choices under experimental conditions differ strongly from what rational choice theory suggests: maximisation of utility (further details in paragraph 2.1).

However the partial approach of economics and DT, gave to the actual research in neuroscience two important staring points: i) the existence of a finite number of processes regarding decision making and ii) a unified goal - motivating system, applicable to all kinds of decision (Sanfey et al 2006). Introducing these assumptions into a psychological approach, we can define "deciding" as goal-directed behaviour in the presence of options, with the goal of achieving the highest benefits/rewards and lowest costs.

Before explaining further the concept of reward, I will make a short digression on the definition of decision making. In the philosophy of mind, the standard conception of decision making equates *deciding* and *forming an intention before an action* (Davidson, 1980). This intention can be equivalent to, inferred from or accompanied by desires and beliefs. Even if this definition seems more exhaustive, words like desire, beliefs, intention seem too far away from a possible and easy neuronal translation. On the contrary the abstract and simplified formalised definition from DT and economy seems more suitable for neuroscientists studying decision making. Neuroscience explains decision-making as the product of brain processes involved in the representation, anticipation, evaluation, and selection of choice opportunities. It breaks down the whole process of a decision into mechanistic components.

The general goal for one's choice is the achievement of the highest benefits/rewards and lowest costs. Achieving the highest reward and avoiding costs seems to motivate our choice. The concept of motivations has undergone several changes (for a review see Berridge, 2004 and Anselme, 2010). Until now reward has been perceived as a motivational phenomena of its own, involving its own active brain mechanisms (Berrige 2003, 2004). Reward means a very basic bio-psychological phenomenon, and most research on reward has involved animals, and used basic rewards such as food and water. However this concept of reward is used also for more complex and abstract rewards such as: social, economic and moral. The unified incentive motivational model is proposed as a complex neuronal system for appetitive ''desire", which mediates a coherent organismic urge to explore the environment and seek resources in response to bodily needs and external incentives (Ikemoto & Panksepp, 1999; Panksepp 2005). It has been developed into a dopamine-centered ''wanting'' or ''incentive salience'' model by Berridge and Robinson (2003). Specific brain activations have suggested the existence of two separate components of reward: the Liking and Wanting dimensions (Berridge, 2003). The Liking dimension of reward concerns essentially hedonic-impact which evokes subjective feelings of pleasure and contributes to positive emotions. The Wanting dimension of reward is a motivational dimension, which concerns the effort that a subject wants to invest in order to obtain a certain reward. Together with Liking and Wanting dimensions, reward is also considered a positive reinforcer because it increases the frequency and intensity of behaviour that leads to the acquisition of goal objects (Montague & King-Casas, 2007), as described in classical and instrumental conditioning procedures.

Investigating decision making processing inevitably requires one to understand how reward perception could drive our choice and how reward shapes our criteria to perceive and select the best option. Neuroscientists, whether employing electrophysiological studies in animals to fMRI studies in humans, are actively investigating the brain's circuits of rewards. The first step in this research has been to clarify the differences between 'reward' and 'value'. While reward refers to the immediate advantage accrued from the outcome of a decision/action (e.g., food, sex, or water), the value of a choice represents an anticipated estimate about how much reward (or punishment) will result from a decision, both now and into the future (Rangel et al, 2008). Thus, value incorporates both immediate and long-term rewards expected from the decision. Reward is more like immediate feedback, whereas value is more like a judgment about what to expect. Furthermore, the value of an

option is a sort of abstract measure, a kind of ‘‘common currency’’ for all choices. Using this abstract “representation of value”, humans can in fact compare apples to oranges when they buy fruit. This ‘‘common currency” of the judgment of value is particularly important because it supports one fundamental assumption that is made both in economy and in psychology: the existence of unique reasoning systems applicable to a wide range of problems.

1.2 Reward: perception and detection

Although there are no specialized peripheral receptors for rewards, neurons in several brain structures seem to be particularly sensitive to reward. Prominent examples are the dopamine neurons in the pars compacta of substantia nigra and in the ventral tegmental area. In various behavioural situations, including classical and instrumental conditioning in monkey, most dopamine neurons show short, phasic activation in a rather homogeneous fashion after the presentation of liquid and solid rewards, and visual or auditory stimuli that predict reward (Shultz, 1986; Schultz & Romo, 1990). These phasic neural responses are common (70–80% of neurons) in medial tegmental regions that project to the nucleus accumbens and frontal cortex, but are also found in intermediate and lateral sectors that project to the caudate and putamen (Ljungberg, 1992). These same dopamine neurons are also activated by novel or intense stimuli that have attentional and rewarding properties (Ljungberg, 1992; Horvitz, 2000).

Neurons that respond to the delivery of rewards are also found in brain structures other than the dopamine system described above. These include the striatum (caudate nucleus, putamen, ventral striatum including the nucleus accumbens) (Shidara et al., 1998;), subthalamic nucleus, pars reticulata of the substantia nigra (Schultz, 1986), dorsolateral and orbital prefrontal cortex, anterior cingulate cortex, amygdala, and lateral hypothalamus (Schultz, 2000). Some reward detecting neurons can determine the magnitude of rewards (amygdala) or distinguish between rewards and punishers, orbitofrontal cortex (Schultz, 2000).

In animal studies, food is generally used as a positive reinforcer. Caudal regions of the orbitofrontal cortex (OFC) codes physical attributes of rewarding stimuli, in particular the taste and smell of food (Rolls, 1999, 2000). However, it has been demonstrated that OFC responses to taste and smell depend on the reward value of the stimulus. Specifically

OFC neurons fire more strongly to tastes and smells in hungry animals than in animals that are satiated (Critchley & Rolls, 1996). These data suggest that the orbitofrontal neurons responding to taste and smell do not simply code sensory properties, but code also the current incentive value of stimuli. The existence of unique brain circuits able to attribute an abstract value to different kinds of options and able to compare them with each other brings neuroscience near to one of the fundamental assumptions evident in the classic economic model: the existence of a unique reasoning system with a consistent and stable set of preferences.

1.3 From reward to value

A well-learned reward-predicting stimulus evokes a state of expectation in the subject. The neuronal correlate of this expectation of reward may be the sustained neuronal activity that follows the presentation of a reward predicting stimulus and persists for several seconds until the reward is delivered. This activity seems to reflect access to neuronal representations of reward that were established through previous experience (Schultz, 2000). Mechanisms regarding the expectation of reward are studied by classic conditioning experiments, in which a cue stimulus is followed by a reward or a punishment. Studies involving animals have shownthat the activation of areas that receive dopaminergic projection (striatum and OFC) is modulated not only by reward but also by the expectation of reward. Reward-expectation neurons are found in monkey and rat striatum orbitofrontal cortex, and amygdala (Shultz, 2000). These neurons discriminate between trials with reward and trials without and they are active just before the reward (Hollerman, 1998).

Another important aspect is evident in that this expectation signal changes systematically with experience. Neurons in the striatum and OFC initially show reward-expectation activity during all trials with novel stimuli. With experience, this activity is progressively restricted to rewarded rather than unrewarded trials (Tremblay & Schultz, 2000 a,b). Another important aspect of these reward-expectation neurons is sensitive to the relative value context. In an elegant experiment, Tremblay and Schultz (2000) demonstrated that neuronal firing to the same stimuli depended on the relative value context in which they were experienced. Animals were taught to associate three visual cues with raisins, apple, and cereal. When the ‘‘apple’’ cue was presented in conjunction with the ‘‘raisin’’ cue, it was relatively less preferred by the monkey; when paired with the ‘‘cereal’’ cue, it was relatively more preferred. Orbitofrontal firing to the apple cue

depended on the context, being greater in the apple–cereal pairing (where it was associated with the preferred food), than in the apple–raisin pairing (where it was associated with the less preferred food). These findings, suggest that OFC neurons code the relative value of reward, rather than simply reward itself. This can be interpreted as a role for the OFC in representing motivational or incentive values of rewarding stimuli (Baxter et al., 2000; Schoenbaum et al., 1998; Watanabe, 1996).

One set of studies has documented responses in OFC related to values of different rewards. Padoa-Schioppa and Assad (2006) recorded from area 13 of the OFC while monkeys chose between pairs of juices. The amount of each type of juice offered to the animals varied from trial to trial, and the types of juices offered changed across sessions. Based on each monkey's actual choices, they calculated a subjective value for each juice reward, based on type and quantity of juice, which could explain these choices as resulting from a common value scale. They then searched for neurons that showed evidence of this hypothesized common scale for subjective value. They found three dominant patterns of responding, which accounted for 80% of the neuronal responses in OFC. First and most importantly they identified *offer value neurons*, cells with firing rates that were linearly correlated with the subjective value of one of the offered rewards, as computed from behaviour. Second, they observed *chosen value neurons*, which tracked the subjective value of the chosen reward in a single common currency that was independent of type of juice. Finally, they observed *taste neurons*, which showed a categorical response when a particular juice was chosen.

Another set of studies has documented similar responses for subjective values of choice options in the striatum and in putamen. Lau and Glimcher (2008) recorded from the caudate nucleus while monkeys dynamically adjusted the proportion of their responses to each target to match the relative magnitudes of the rewards earned for looking at those targets. They found three kinds of task-related responses that were closely related to the orbitofrontal signals of Padoa Schioppa and Assad (2006, 2008): action value neurons, which tracked the value of one of the actions, independent of whether it was chosen; chosen value neurons, which tracked the value of a chosen action; and choice neurons, which produced a categorical response when a particular action was taken. Action value responses occurred primarily early in the trial, at the time of the monkey's choice, while chosen value responses occurred later in the trial, near the time of reward receipt. Samejima and colleagues (2005) recorded from putamen while monkeys performed a

manual choice task, turning a lever leftward or rightward to obtain rewards. Across different blocks, the probability that each turn would be rewarded with a large (as opposed to a small) magnitude of juice was changed. Recording from the putamen, they found that one-third of all modulated neurons tracked action value. Thus, the responses in the caudate and putamen in these two studies mirror those found in orbitofrontal cortex, except that neural responses were anchored to the actions produced by the animals rather than to a more abstract goods-based framework as observed in orbitofrontal cortex.

1.4 Learning value

Solid evidence now indicates that dopaminergic neurons in the midbrain encode a teaching signal that can be used to learn the subjective value of actions (Montague & King-Casas, 2007). How does this subjective value signal arise? One of the most critical sources of value information is undoubtedly past experience. We learn from the past experience to attribute a value and then we use this subjective value during the decision making process. Following reinforcement learning theories, subjective values are learned through iterative updating based on experience. The theories rest on the idea that each time a subject experiences the outcome of her choice, an updated value estimate is calculated from the old value estimate. The difference between the experienced outcome of an action and the outcome that was forecast is revealed as a signal of an error: reward prediction error. Pioneering studies of Schultz and colleagues (1997) provided the initial evidence that dopaminergic neurons encode a reward prediction error signal. They demonstrated that, during conditioning tasks, dopaminergic neurons (1) responded to the receipt of unexpected rewards, (2) responded to the first reliable predictor of reward after conditioning, (3) did not respond to the receipt of fully predicted rewards, and (4) showed a decrease in firing when a predicted reward was omitted. This error signal is scaled by a learning rate, which determines the weight given to recent versus remote experience. In simple terms, if a reward occurs unpredictably after a given action then the prediction error is positive, and learning about the consequences of the action that produced the reward occurs. However, once the consequences of that action have been learned (so that the reward that follows subsequent repetition of the action is now predicted), the prediction error falls to zero and no new information about the consequences of the action is learned. By contrast, if the expected reward is not received after repeating a learned action then the prediction error falls to a negative value and the behaviour is extinguished.

To summarise, the starting point of our decision begins from a basic function such as: i) perceive and detect a signal of reward, ii) create a subjective value, iii) detect contextual signal anticipating reward and iv) update the expectation of reward with the actual reward signal.

Subsequent paragraphs will show how reward's signal and subjective value is not only related to basic needs like, food, water and sex, but is also related to more social concepts at the root of our life such as money, equality, fairness, responsibility, etc.

CHAPTER II

DIFFERENT DECISIONS

In chapter I concepts such as value and reward were introduced that are at the basis of all kinds of decisions. This chapter introduces concepts related to specific kinds of decisions: economic decisions under uncertainty, and social and moral decisions. In all these types of decisions, emotions seem to play a significant role. Therefore, in the last part of the chapter the concept of emotion and its components is introduced.

2.1 Economic decision under uncertainty

Decision making is the process of selecting an action from a set of available options. From this general definition, a *rational* decision according to the classic economic model means selecting the action with the highest reward and lowest cost (e.g., highest utility). However, it is not always possible to easily discriminate which option has the highest reward, especially in uncertain environment where it is not possible foresee future events and consequences of one's choice. Most of our decisions in everyday life involve risk and are supported only by probabilistic knowledge. As reported in the first chapter, economy and psychology are trying to understand what guides our decisions by starting from two basic assumptions: (i) the existence of a general and unique reasoning system (ii) a consistent and stable set of preferences. On the base of this finite consistent stable set of preferences, most of the effort of economists has been to describe the mathematical function of our choice heuristic. In this prospective, a fundamental preference of our choice is the selection of an action with the highest expected utility (EU). Bernoulli's

Expected Utility (EU) theory assumes that people assign subjective values to consequences and weigh them according to their probabilities (Bernoulli, 1738). This means that subjects in uncertain contexts do not choose the option with the highest absolute value, but they choose the option with the more likely value (von Neuman & Morgestern 1944). Subjects attribute a value (x) to each of the options. This value is associated with the probability (p) with which it can be obtained (Bernoulli, 1738). Following this definition, the EU is the product between the option value and the probability to get this value:

$$EU = U(x)*p.$$

Even if this formalization summarizes well the process of attributing value and how this is strongly dependent on probability or more generally on learning, the EU function does not fully describe real choice, and several sets of behavioural data contradict the EU theory. Well known examples in this direction are the Framing Effect (Tversky & Kahneman (1981) and the Allais's paradox (1953). The framing effect[1] shows how the format rather than the content of two identical options can alter people's decisions. Specifically, individuals have a tendency to show inconsistent choices, depending on whether the question is framed to concentrate on losses or gains (Plous, 1993). The Allais paradox[2] shows that the significant majority of real decision makers order uncertain prospects in a way that is inconsistent with the postulate that choices are independent of irrelevant alternatives.

---

[1] **Framing effect.** Tversky and Kahneman (1981) asked to choose between a certain (i.e. sure) or a probabilistic (i.e. risky) option to save lives (positive frame) or minimize deaths (negative frame).

Imagine that the United States is preparing for an outbreak of an unusual Asian disease that is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Scientific estimates of the consequences of the programs are as follows. ***Positive frame:*** *If Program A is adopted, exactly 200 people will be saved. If Program B is adopted, there is a 1 in 3 probability that all 600 people will be saved and a 2 in 3 probability that no people will be saved.* ***Negative frame****: If Program C is adopted, exactly 400 people will die. If Program D is adopted, there is a 1 in 3 probability that nobody will die and a 2 in 3 probability that all 600 will die.*

Most people chose options A and D, despite the fact that in terms of consequences, these choices are contradictory (A is equivalent to C, as B is to D). People appear to exhibit a general tendency to be risk seeking when confronted with negatively framed problems and risk averse when presented with positively framed problems.

[2]**Allais paradox.** *Consider this choice among two lotteries: A)Lottery L1 promises a sure win of €100; B) lottery L2 is a 10% chance to win 500€, 89% to win 100€ and 1% to win 0€. Second choice is between: C) Lottery L3 is a 11% chance to win 100€ and 89% chance to win 0€ and D) Lottery L4 is a 10% chance to win 500€ and 90% chance to win 0€.*
Following the EU theory subjects should prefer A to B and C to D. Mathematically in the first choice U(100)*1 > U(500)*0.10+ U(100)*0.89+ U(0)*0.01 whereas in the second choice U(100)*0.11+U(0)*0.89> U(500)*10+U(0)*0.90. Behavioral data showed that most subject prefer A to B but D to C, contradicting the assumption of EU.

Other theories tried to develop and improve the EU theory, taking in consideration the real behaviour of subjects (e.g. prospect theory; Kahneman & Tvresky, 1979), but only recently has the new cooperation between different disciplines (psychology, economy, sociology) and more attention on behavioural and neuronal data, allowed a better understanding of decision making processes. From this prospective, new elements such as emotion and social values are showing their role in explaining our choice.

A nice example of economic decision, quite far away from a classical definition of rational decision, is the 'endowment effect' concerning the fact that people place a higher value on objects they own than objects that they do not. Lerner and colleagues (2004) investigated the impact of specific emotions on the 'endowment effect'. Participants before the real task watched a film clip with a neutral, disgust or sadness content. Half the participants were endowed with an object and then given the opportunity to sell it back at a range of prices (sell condition); the other half were shown, but not given, the object and then asked whether they would prefer to receive the object or to receive various cash amounts (choice condition). Results showed that the sadness induction reduces selling prices but increases choice prices, and disgust reduces both selling and choice prices. Interestingly, these data are in accord with the experience of depression of compulsive shoppers. Probably in these subjects the choice price is the normal or greater motivating purchase, whereas the selling price is greatly underestimated in inducing depressive state.

In expected utility theory, a decision maker selects an action by maximizing the sum of utilities for various outcomes weighted by their probabilities. Unfortunately, this theory was too simplistic to account for the pattern of choice behaviour that people displayed in reality. Furthermore, economists and psychologists have long recognized that a variety of contextual factors influence the utility of a physical stimulus (Kahneman et al, 1979), and that emotions play an important role in decision making (Bechara, 2005).

2.2 Social Decision

The nature of the decision making process changes when individuals begin to interact in a social setting, making the outcome of a decision also dependent on the decisions of others. Social decision making is one of the most complex animal behaviors.

It often requires subjects to recognize the intentions of other subjects correctly, and to adjust behavioural strategies rapidly. In addition, humans can cooperate or compete with one another, and various contextual factors influence the extent to which humans are willing to sacrifice their personal gains to increase or decrease the well-being of others. The basis of social decision making can be investigated quantitatively by applying Game Theory. Game theory aims to help us understand situations in which decision-makers interact. Game Theory (von Neuman & Mongester, 1944) incorporates several models in explaining the choice of subjects interacting with other decision makers in a specific situation called the 'game'. A game is a specific context of interaction with precise roles, rules, and fixed time of actions between players. There are two central assumptions in (classical) rational game theory: (i) players are self-interested and (ii) reach an "equilibrium" in which everyone is choosing (or planning) strategies that yield the best outcome, anticipating that others are doing the same (Nash equilibrium, 1950). The first assumption pertains to self-regarding actors that in a social situation want only to maximize their own payoff. A self-regarding actor thus cares about the choice and payoffs of other individuals only insofar as these influence his own payoff. One reason for the prevalence of the self-interest hypothesis in economics is that it has served the profession quite well because self-interest is without doubt one important motivational force, and some people indeed display very self-interested behaviours. In some domains, such as competitive experimental markets, models based on the self-interest hypothesis even make very accurate quantitative predictions (Smith, 1982). However, in strategic interactions, where individuals' actions typically have a direct impact on other individuals' payoffs, the self-interest hypothesis often fails to predict correctly players' behavior (Fehr and Gächter, 2002; Camerer & Fehr, 2006). The experimentally observed failures of the self-interest model gave rise to the development of alternative models where the social component of interactions is also taken in consideration. A clear example in this sense is the game called Ultimatum Game (Güth et al., 1982).

The ultimatum game is a social situation between two anonymous individuals: one player, the proposer, makes a take-it-or-leave-it offer, dividing some amount of money, say €10, between herself and another person. If the second person, the responder, accepts the division, then both people earn the specified amounts. If however the responder rejects it, they both get nothing. Following the rational interpretation, the responder interested only in material gains (i.e. money) should accept any positive offer (1€ is better than nothing),

and the proposer, anticipating this, will offer the smallest offer that is possible. Nevertheless, evidence (see review Camerer, 2003, Fehr and Gächter, 2000; Camerer & Fehr, 2006): suggests that not only approximately 50% of the offers that fall below 30% of the initial endowment are rejected by responder but also the proposer usually proposes a fair offer around 50% of the initial amount. This evidence is generally interpreted by saying that people are not driven exclusively by self-regarding preferences based on material gains, but also by other-regarding preferences guided by fairness and equality motives (Fehr & Gatcher, 2002; Bowles, 2006). A fair decision is the decision, from the proposer's point of view, to offer around 50% of one's own initial amount. Equal division of the initial amount between proposer and responder is what the responder wants (equity). If the responder receives an unfair offer (for example £1), then he can reject the proposed allocation (Güth et al, 1982). If the responder rejects it, both players receive nothing. Rejections are usually taken as evidence of willingness to punish (Rabin, 1993) those who have behaved unfairly (Fehr & Schmidt, 1999). The amount a responder loses by rejecting a proposed allocation serves as a measure of the strength of these motives (Fehr & Schmidt, 1999). Economists have largely used the concept of fairness, considering fairness as a fundamental normative category of social decision without, however, considering the psychological nature of this category.

A possible interpretation in psychological terms is that in cases of low offers (say 1€ or 2€ out of the 10€ available), the responder faces a conflict between accepting the money, due to its expected reward value, and rejecting it because of both anger and indignation due to the perceived unfairness of the allocation (Elster, 1998). Proposers must possess a reasonably accurate prior model of what the responder is likely to reject, and the responders must be willing to reinforce such models by enforcing substantial rejection levels, thus ensuring fair (higher) offers from proposers.

Another example of social behaviour is investigated by the prisoner's dilemma. The Prisoner's dilemma (PD) game is used to study " social dilemmas " that arise when the welfare of a group conflicts with the narrow self-interest of each individual group member. For example, in a typical two-player PD, each player can choose either to " cooperate " or " defect. " Payoffs are symmetric, and chosen so that the sum of the payoffs is greatest when both choose "cooperate " and least when both players choose " defect. " However, each player earns the most if he chooses to "defect " when the other cooperates. Thus, the

perfect Nash Equilibrium of this environment is for both players to defect. In this game, two players at the same time can each choose between cooperation and defection. If this game is played only once and the players care only about their own payoffs, both players should defect, which corresponds to the Nash equilibrium for this game. In reality and in laboratory experiments (see, for example, Davis and Holt, 1993; Ledyard, 1995), both of these assumptions are frequently violated. Humans often cooperate in prisoner's dilemma games, even when the game is one shot. The key finding was that, in aggregate, cooperation occurs about half of the time in PD games (McCabe, 2001). Therefore, for humans, decision making in social contexts may not be entirely driven by self-interest, but at least partially by preferences regarding the well-being of other individuals or society in general (Fehr & Fischbacher, 2003). Interestingly, cooperation and altruistic behaviors abound in human societies but seem to occur also in non-human primates (Hauser et al., 2003), suggesting a basic mechanism supporting social behaviour. They proposed an example of altruistic food giving among unrelated individuals in cotton-top tamarins (Saguinus oedipus). This study showed altruistic food giving among genetically unrelated individuals that is extremely rare in nature. Authors provided evidence that tamarins are more likely to give food to a conspecific who unilaterally gives food back, as opposed to a conspecific who unilaterally refrains from giving food; suggesting that food giving may be based on others' contingent behaviour. Moreover authors showed that tamarins distinguish between altruistic food giving as opposed to by-products of selfishness or simple reinforcement contingencies; individuals are more likely to give food to an individual who has altruistically given food in the past than to individuals who give food as a result of selfish attempts to procure their own food. These results suggest that human cooperation may have evolved from psychologically ancient and primitive mechanisms, present in closely and related animals (see Millinski 1987; Hauser 2000).

Evidence for altruistic social preference in humans are also investigated in other experimental games, such as the Dictator Game (DG), and the Trust Game (TG) (Fehr & Camerer, 2007). In the DG, a dictator receives a fixed amount of money and donates a part of it to the recipient. This ends the game, so there is no opportunity for the recipient to react. Any amount of donation reduces the payoff to the dictator, so the amount provides a measure of altruism. During dictator games, people tend to donate on average about 25% of their money (Camerer, 2003). It could be that not only inequity aversion drives the dictator game, but also altruism, which is an important aspect of our social life.

Anonymous donation to charitable organizations is an outstanding example of this unique aspect of human altruism, which relies on our ability to directly link motivational significance to abstract moral beliefs and societal causes.

In this above paragraph, trust is not described because is the topic of two experimental studies presented in the experimental part where trust relation and cooperation are covered fully.

2.3 Moral Decision

The last kind of decision introduced in this chapter regards decision with moral meaning. Recent interest in social cognitive neuroscience has led to a growing body of research aimed at elucidating the neural and cognitive mechanisms that underlie human moral behavior (Moll et al., 2005; Beer & Ochsner, 2006; Lieberman, 2007). Moral behavior refers to what individuals should do based on principles and judgments (i.e. moral values) shared with other members of their social environment (Ciaramelli et al., 2007). The basic assumption underlying a moral behavior is that individuals choose and control their own actions. Further, they are aware of what they are doing as they do it, and can normally predict the consequences of their actions, at least the proximate consequences. Therefore, their conscious knowledge of what they are doing should allow them to choose between right and wrong actions. Right and wrong pertain to the sets of customs and values that are embraced by a cultural group to guide social conduct (Moll et al., 2005). Moral decisions first involve a concept of agency that in moral terms becomes: responsibility. In most of legal tradition, indeed, criminal conviction depends on both a harmful consequence and the intent to arm, where the intention to assume responsibility is a key point.

In a person able to control his/her behaviour and aware of what he/she is doing, what does support his/her moral behaviour? For the next 15 years during the "cognitive revolution" (1960-1980s), the Kolberg's (1976) approach dominated moral psychology. In his point of view (close to Kant's position), moral decision was focused only on conscious verbal reasoning. In the same time a long tradition from Hume (1739, 1984) to more recent "affective revolution" of 1980s (Haidt 2001, 2007) supported the idea that an inextricable relation of actions as the object of moral sentiments exist. Recent investigation in

neuroscience attempted to solve this dichotomy by showing how our brain is able to integrate both cognitive and emotional drives in moral decision.

When we are agents of social actions conforming to our values, we may feel pride, whereas when another person is perceived as the agent, we may feel gratitude. On the negative side, when we act counter to our values, we may feel guilt, and when another person acts in the same way towards us, we feel indignation or anger instead (Moll et al., 2007). Furthermore, the anticipation of these moral sentiments in particular social situations often guides our behavior (Tangney et al., 2007). Areas of the brain supporting moral behaviour are several: moral phenomena could emerge from the integration of contextual social knowledge, represented as event knowledge in the prefrontal cortex (PFC); from social semantic knowledge, stored in the anterior and posterior temporal cortex; and from motivational and basic emotional states, which depend on cortical-limbic circuits. In particular, two opposite components could be at the basis of our moral behaviour: a cold and cognitive component that is mostly cortical and related to control and functions, and a hot and emotive component that is subcortical, related to emotive and reward function (Greene et al., 2001; Haidt, 2007; Moll et al. 2007).

A classic method to investigate moral behaviour is to request subjects to solve complex moral situations that incorporate dilemmas. Famous examples of such situations are the *footbridge dilemma* and the *trolley dilemma* (Thomson, 1986). The *footbridge dilemma* is defined as a personal dilemma because it implies i) *a* clear notion of agency, and an action that ii) causes serious bodily harm to iii) a victim vividly represented as an individual. The *trolley dilemma* is defined as an impersonal dilemma because it fails to meet one of these criteria. In both dilemmas there is a trolley which is about to run over and kill five people. In the trolley dilemma one can save them by hitting a switch that will divert the track onto another track where there is only a person. In the footbridge dilemma one can save the five people by pushing a large man off the bridge above the trolley path, killing him but stopping the trolley. A cold economic solution suggests a sacrifice of one life to save 5 in both dilemmas, but most people approve of sacrificing the life of one person only in the trolley dilemma but not in the footbridge dilemma (Greene 2001, 2004). This apparent contradiction was illuminated recently by neuroimaging and neuropsychological study (Greene et al. 2001, 2004; Ciaramelli et al., 2007; Koenigs et al., 2007). The action in the footbridge dilemma could elicit a stronger negative emotional response involving emotive areas such as the medial prefrontal cortex, posterior cingulate

area and anterior insula. This emotive activation forces subjects to avoid sacrificing the large man. In contrast the trolley dilemma recruits more areas associated with problem solving and deliberative reasoning including dorsolateral prefrontal cortex (dlPFC) and inferior parietal lobe. This cognitive, rather than the emotional approach, induces subjects to take the more rational or utilitarian position of shifting the trolley direction in order to save 5 people but killing one. Neuropsychological data confirm this dual process theory (emotion vs cognition). Patients with lesion of the vmPFC usually present high levels of aggressiveness, lack of concern for social and moral rules and irresponsibility (Eslinger & Damasio, 1985; Stuss et al., 1992; Damasio, 1994; Blair & Cipolotti, 2000). These patients, facing the same moral task concerning person and impersonal dilemma, show a greater number of utilitarian choices in personal dilemma than for healthy controls (Ciaramelli et al., 2007; Koenigs et al., 2007). The data shows how lesion of vmPFC could reduce the influence of emotion on moral reasoning but leave intact the cognitive/rational solution. However none of the existing studies has systematically measured subjects-emotional responses emerging during (and, presumably, having an impact on) evaluation of moral dilemmas.

It is clear that no single moral circuit exists in the brain, but our moral behaviour is dependent on a combination of several circuits with different functions. For example, orbitofrontal areas appear to be primarily involved in on-line representation of reward and punishment value (Bechara, 1995). The contribution of this area is not necessarily moral but the regulator function in which affective information guides approach and avoidance behaviour in both social and non social context. Greene tested this dual process theory while controlling for the effect of cognitive load on moral judgment Greene (2008). In contrast, Valdesolo (2006) tested the influence of the induced mood on moral judgment. In Greene's study only personal moral dilemmas were used and the reaction time (RT) for utilitarian and non-utilitarian responses were compared. Data showed that cognitive load increases RT only for utilitarian judgment but not for non-utilitarian judgments. This data confirm that utilitarian solution in personal dilemma engages a cognitive control function.

2.4 Emotion-based decision

2.4.1 Components of emotion

All the kinds of decisions described above suggest that emotions have some role in driving our choices. In the following paragraph the concept of emotion and its components

are introduced.

Although the term *emotion* is commonly used to capture all affective experience, in neuroscience the term emotion is proposed to reflect the *discrete response* to an external or internal event that entails a range of synchronized features, including subjective experience, expression, bodily response, and action tendencies (Phelps, 2009). This definition reduces the emotion at discrete responses offering a real pragmatic point of view in which emotion can be measured and manipulated.

One of the features of emotion is subjective experience. One of the primary sources of confusion in emotion research is the relation between emotion and feeling (Damasio, 1999). The subjective experience of emotion, called *feeling*, is just one of the features that affective scientists consider a component of emotion. However most emotion researchers today acknowledge that there are several components of emotion that do not necessarily depend on feeling (LeDoux, 1996). For research in non-human animals this distinction is critical, since subjective experience is not accessible in other species, but other types of emotional responses, such as physiological changes, are easily assessed and have characteristic patterns across species.

One of these physiological changes is *expression*; it refers here to motor responses in the face, voice or body that convey the emotion to others in a social environment. The primary function of expression is the communication of emotion. The expression of emotion has most often been studied in the characteristic motor response of the face when a person is experiencing an emotion.

Another physiological change regards the *bodily responses*. One of the unique characteristics of emotion is the patterned behavioural, hormonal, and autonomic response that follows the perception of an emotion eliciting event. In contrast to emotional expressions, such as face expression in which a primary function might be the communication of emotion, bodily reactions are thought to be adaptive in preparing the organism to respond (Sinker et al, 2009; De Gelder, 2009). The characteristic patterns of expression and bodily response of emotion provide a powerful means to assess emotional reactions using psychophysiological techniques that are non-intrusive and do not depend on subjective experience or verbal report.

In contrast to bodily responses, in which any behavioural motor actions that occur may be best characterized as automatic, reflexive reactions, emotion also elicits a tendency towards action that does not have a predictable motor pattern and is expressed as instrumental responses. These action tendencies motivate the organism towards a particular class of actions. For example, to move away from and avoid a stimulus that predicts potential punishment or a tendency to approach a stimulus that is rewarding is an instrumental response to an emotion-eliciting event. This component clearly suggests a more complex role for emotion in action selection than a simple reaction.

An additional important component of emotion is the evaluation and appraisal of an event. The primary function of emotion is to highlight the significance or importance of events so that these events receive priority in further processing. The *evaluation* of the relevance or significance of event can occur rapidly, without conscious awareness or cognitive interpretation (Zajonc, 1984; LeDoux, 1996). More often than not, we are aware of the emotional significance of an event. This awareness and the cognitive interpretation of the meaning of the event can initiate and alter an emotional response (Lazarus, 1984).

2.4.2 Details on autonomic response component of emotion

Autonomic activity corresponds to the physiological expressions, mostly electrical and hormonal, under the control of the autonomic nervous system (ANS) and represents the neural activity related to the brain and body regulation. Historically, the use of the term “autonomic” implied that the given part of the nervous system was functionally independent of any voluntary nervous or cognitive control. ANS controls visceral targets such as cardiovascular tissues (heart, blood vessels), smooth muscles (most visceral organs), glands (endocrine and exocrine) and sensory systems (eyes, skin); having specific roles in physiological and behavioural adaptation. Autonomic control mainly regulates the internal environment in order to maintain the body homeostasis. However, the role of autonomic activity cannot be reduced to maintain brain or body immediate homeostasis; it provides support to complex behaviours, such as emotional reactions. In that sense, autonomic activity is reactive to a stimulus but also contains an anticipatory dimension. Peripheral electrical measures that have been used successfully as indicators of mind expressions are: pupillometry, electrodermal activity, cardiovascular resposes (heart rate variability, blood pressure, peripheral blood flow). Pupillary responses have been correlated with emotional processing, cognitive load or degrees of alertness, and task-evoked changes in pupil size can be observed within the first several hundred milliseconds

after the stimulus presentation (Beatty, 1986). Electrodermal and cardiovascular responses are acknowledged to index respectively the activation level and the valence of emotional stimuli (Bradley and Lang, 2000; Solbakk et al., 2005). Various aspects of cardiovascular functioning (heart rate variability, blood pressure, peripheral blood flow) can also be correlated with emotions (Lang et al., 1993), passive and active attention (Öhman et al., 2000), or motor processes (Sequeira and Ba- M'Hamed, 1999). For instance, heart rate acceleration often accompanies perception of unpleasant compared with pleasant pictures (Solbakk et al., 2005) and heart rate deceleration can be observed during detection tasks (Lacey and Lacey, 1978).

A particularly important autonomic activity in studies of emotion is the skin conductance, defined by Sequeira and colleagues (2009) as *'a window on the arousal dimension of emotion'*. Skin conductance variations depend on the quantity of sweat secreted by eccrine sweat glands. Such secretion is under the control of sympathetic innervation which transmits influences from the central nervous system to the eccrine glands. Sweating variations are sensitive markers of events having a particular signification for individuals, usually related to emotional, novelty or attentional fields (Sequeira et al, 2009). Skin conductance is a good indicator of reticular activation and therefore seems to reflect the energetic dimension of behaviour and particularly of emotion. Indeed, the amplitude of electrodermal responses increased linearly as ratings of arousal increased, regardless of emotional valence (Bradley and Lang, 2000). This effect is observed when emotional pictures (Winton et al., 1984) or emotional words (Manning et al, 1974) are used.

2.4.3 Details on endocrine response component of emotion

Life history construction is mediated through a suite of neuro-endocrine mechanisms that regulate resource partitioning among competing acute and life course goals and demands. Stress responses illustrate this point. Perceived threat or challenge triggers dual signals from brain to body transmitted via endocrine and neural pathways (Boyce and Ellis, 2005). One route involves the hypothalamo-pituitary-adrenocortical (HPA) pathway that triggers cortisol release, while the other involves autonomic nervous system activation that triggers both ‘‘fight or flight’’ organ responses and direct sympathetic neural release of catecholamines in the adrenal medulla (SAM). Each component of the suite of coordinated central and peripheral responses has deep evolutionary roots grounded in common survival demands (Porges, 1995) that shift

priorities to immediate survival (accelerated heart rate, attentional focusing, elevated glucose), and away from deferrable activity (immunity and repair, digestion, growth, and reproduction). Fear-related behavioral responses are orchestrated by diverse neurochemical signals; one set of these signals involves the regulation of corticotrophin-releasing hormone (CRH) by glucocorticoids in various regions of the brain. Several clinical observations have linked alteration of CRH and cortisol levels to anxious and/or fearful depression (e.g. Gold et al 2002; Nemeroff et al 2004). Other chemical signals in the brain, such as 5-hydroxytryptamine (5-HT) and noradrenaline, are important in fear and its pathologies (Gold et al 2002; Nemeroff et al 2004).

Positive social interactions and emotions are associated with a unified pattern of physiological events. Suckling or breastfeeding, aspects of maternal behaviour, represent examples of positive social interactionwhich have been explored in depth from a physiological and neuroendocrine point of view. Lactating rats have a decreased sympathetic nervous tone, manifest a lowered blood pressure when compared with non-lactating rats. They also have an enhanced vagal nerve tone, resulting, for instance, in an increased release of insulin and of other gastrointestinal hormones. Together, these results document the fact that a shift in autonomic nervous tone—from sympathetic to parasympathetic, vagal nerve dominance has occurred. Furthermore, lactating rats are less responsive to certain stressful stimuli than are non-lactating animals (Uvna¨s-Moberg, 1996; Uvna¨s-Moberg and Eriksson, 1996). A key hormone and neurotransmittor in breastfeeding is the Oxytocin. Oxytocin has emerged as a core component of the mechanisms mediating the health benefits and anti-stress effects of positive social interactions. It has been shown that oxytocin treatments increase social contact in several animal species (Carter et al., 1995; Witt et al., 1992). In addition, in monogamous voles, oxytocin is essential for selective social behaviours and the formation of pair bond (Williams et al., 1994).

2.4.4 Basic emotions

In his seminal work *The Expression of Emotion in Man and Animal*, Charles Darwin (1872/2002) proposed that there are a limited number of basic, universal human emotions. More recently, Paul Ekman and his colleagues studied the facial expression of emotion and suggested that there are six basic emotional expressions: happy, sad, fear, anger, disgust, and surprise (Ekman & Friesen, 1971). Each of these expressions is

characterized by a unique subset of facial muscle movements. The ability to convey these emotional expressions appears to be innate. Studies examining the vocal expression of the emotion also provide some evidence for basic emotions (Johnstone & Scherer, 2000). Despite years of significant effort, there is relatively little evidence to suggest that these basic emotions are reflected in corresponding, unique patterns of autonomic responding (Cacioppo *et al*, 2000). In research on emotion, these basic facial expressions have proved useful in both assessing emotion perception and evoking corresponding emotional responses in others. However, it is important to acknowledge that these six basic emotions do not capture the full range of human emotional experience. There are several more complex emotions, such as guilt and love, which are less clearly linked to specific facial or vocal displays. Social or moral emotions, such as pride, guilt, shame, or embarrassment, differ from the basic emotions in their external triggers (Haidt, 2003), and both the perception and expression of social emotions differ culturally between individualistic and collectivistic nations (Eid & Diener, 2001). Emotions form the omnipresent background for behaviour and attitudes, and they serve a critical role in social interaction (Forgas, 2003). Damasio (1999) proposes also another kind of emotion called: a class of background emotions such as well-being or malaise, calm or tension, fatigue or energy, anticipation or dread. In the background emotions, the inducer is normally internal and the focus of response is mainly the "internal milieu" of the body.

2.4.5 Nature of emotions

In the late eighteenth century, William James and Carl Lange suggested with the now famous James-Lange theory that changes in bodily responses are a necessary condition for emotional experience to arise (James, 1894). They argued that emotions could not be experienced in the absence of these bodily feelings. This theory still remains a milestone in the theory of emotion. From this theory, Damasio (1999) developed his own theory defining emotions as patterns of chemical and neural responses, the function of which is to assist the organism in maintaining life by prompting adaptive behaviours. Emotions are due to the activation of a set of brain structures, most of which also monitor and regulate bodily states around optimal physiological values, in processes known as homeostasis or homeodynamics. Emotions are biologically determined, stereotypical, and automatic, although it is acknowledged that both culture and individual development may influence the set of inducers and can inhibit or modify overt expressions.

This approach is actually the most used in neuroscience, and especially in the study of decision making. However the approach proposed by Davidson and colleagues (1990) based on the distinction between approach/withdrawal is also relevant. This approach classifies different emotions according to motivation. One of the primary functions of emotion is to motivate action, and different emotional states lead to different goals for action. Some emotional states, such as happiness, surprise, and anger, are referred to as approach emotions – that is, they evoke a motive or goal to approach a situation. Other emotional states, such as sadness, disgust, or fear, are withdrawal emotions, in that there is a natural tendency is to withdraw from situations linked to these emotions. Interestingly this approach does not consider emotion as a simple reaction, but considers also that emotions are able to motivate our choice.

2.5. Emotion and decision making

Although emotions have been considered as important variables affecting decisions, the role of emotion in decision making has rarely been coupled with the detailed investigation of the range of components, factors, and measures that have characterized the psychological study of emotion and affect. There are surprisingly few studies on social and economic decision making that have explicitly manipulated or measured emotion or affect variables. In many studies and theories, especially in neuroeconomics, emotion is inferred, but not directly altered or assessed. The aim of the experimental part of this research work is to increase the data available about the role of emotions in decision making, proposing several tasks where I attempted to manipulate and measure emotions in order to understand how this can affect the social and moral decision making process.

The distinction between emotion and cognition has been prominent since early philosophical writings, and this simple dichotomy continues to influence folk psychological theories and scientific thought, including economic and moral research on decision making (Damasio, 1994). In neuroscience, this dual systems approach has gained prominence in studies attempting to characterize the impact of emotion on decision making (e.g., Cohen, 2005; Shiv, 2005; Damasio, 1994; Greene, 2007). This is also the prospective assumed in this research work. This point of view has suggested the existence of two different systems for the decision making process: a deliberative system and an affective system. The deliberative one is a complex, reflective and slow system

(Lowenstein, 2004) corresponding roughly to expected utility theory or more in general to the rational paradigm. The affective system cares primarily of short-term outcomes; it is simple, reactive and fast (Lowenstein, 2004; Greene, 2007).

Even if the emotion-cognition dichotomy theory seems the more common and better supported theory in current research on decision making, the relation between emotion and cognition in cognitive neuroscience cannot be considered in relation to the impact of emotion or affect on cognition without further specifying and assessing the specific emotion or affective process engaged (Phelps, 2006). Specifically, it seems necessary to further investigate the relationship between the concepts of value, emotion and choice in order to better understand how emotion and cognition interact in decision-making.

2.5.1 Emotions and decision two systems in the brain.

Following the dichotomy theory, emotive process are localized in limbic and paralimibic structure such as striatum and amygdala and in cortical structure such insula and OFC. The activation of these areas are linked to stimuli with specific value (reward or punishment), and give fast and automatic responses to the stimuli (such as threat). The cognitive and deliberative processes take place in the cortical area; specifically in the dorsolateral prefrontal cortex (DLPFC) and the parietal cortex. These areas are involved in several different functions such as working memory, abstract reasoning and problem solving (Miller & Cohen, 2001). The interaction between two systems gives the possibility of a flexible and rapid solution to complex problems and consents goal directed behaviour, with the capacity to solve conflict between different stimuli (McClure et al, 2007). There is an influential series of neuropsychological studies by Damasio, Bechara, and colleagues (Bechara et al., 1997; Damasio, 1994; Bechara, 2005) investigating the interaction between these two systems, and, in particular, how the intuitive/emotive process can support and help the rational one in a probabilistic context. In the Iowa Gambling Task (IGT), subjects are presented with a choice of four decks of cards, face down. On each trial, they are required to choose a card from one of the four decks. When they turn the card over, either a financial reward or a penalty is revealed. At the outset of the experiment, subjects are told nothing about the contingencies pertaining to the individual decks. In fact, two of the four decks are ‘‘high risk’’ while the other two are ‘‘low risk.’’ The high risk decks offer a prospect of immediate large rewards but carry a cost of even larger long term penalties.

Over the duration of the task, choosing predominantly from these two decks results in subjects losing money. The low risk decks offer smaller immediate rewards but even smaller long term penalties. Over the duration of the task, choosing from these decks results in subjects steadily accumulating money. This task was designed to model real life decision making situations where subjects must weigh the potential benefits and possible risks associated with choosing particular courses of action. When healthy subjects perform this task, they gradually learn the contingencies over the first 20, 30 trials and then choose the low risk decks on most subsequent trials. During the experiment, skin conductance responses (SCRs) were monitored, both before and after decisions. Healthy participants showed an increase of SCR, occurring in the 5-s window before selecting a card, when they pondered risky decisions, and began to prefer the good decks before having adequate conscious knowledge of the situation. The SCRs could be considered as a physiological index of emotional arousal. Such arousal is related to the sympathetic division of the autonomic nervous system (Boucsein, 1992), and is widely used as a sensitive and objective measure of emotional processing and attention (see Dawson, Soulières, Gernsbacher, & Mottron, 2007 for a recent review). The vmPFC seems to be the critical area implicated in the generation and feedback representation of bodily states of arousal (Damasio, 1994; Tranel & Damasio, 1994; Bechara et al., 1996; Bechara et al., 1999; see also Nagai et al., 2004). Patients with ventral frontal lesions, involving part of the OFC, show pronounced deficits on this task (Damasio, 1994; Bechara, 2005). Instead of tending to choose the low risk decks after learning the contingencies, patients continue to opt for the high risk decks on the majority of the trials. Interestingly these patients failed to generate anticipatory SCRs observed in healthy subjects. The contrast between defective emotion on the one hand and preserved intellect on the other hand in vmPFC patients suggested to Bechara and colleagues that, somehow, disturbed emotional signalling could explain the decision deficits in this neurological population.

This study is particularly interesting because the task provided the first laboratory diagnostic procedure for patients with ventromedial prefrontal damage – a rather useful advance, given that these patients generally passed all other neuropsychologic tests and only exhibited their defects in everyday life outside the laboratory. The task was also instrumental in showing a persuasive correlation between indices of emotional change (skin conductance responses) and the advantageous or disadvantageous playing of the card game (Bechara et al., 1997).

CHAPTER III

BRAIN CIRCUITS OF DECISIONS

In the first paragraph of this chapter, I summarized current knowledge and propose a model that is able to describe optimal economic decision making. The model proposed follows mainly the components suggested in the Back-pocket Model proposed by Glimcher (2009). The Two-Stage model, described in paragraph 2.1, suggests that medial frontal cortex has a relevant role at the stage of attribution of value to the options available in order to choose the most valuable among them. Paragraphs, 3.2 and 3.3 expose in detail the role of this area in decision making, and in particular the possible involvement of emotions in decision making through the involvement of medial prefrontal and orbitofrontal cortex

Although neuroscientists and economists are trying to describe a common and unique model for decisions, actual knowledge is far away from allowing an exhaustive description of this model. However, recent studies (see paragraph below) have begun to suggest that some specific components of decision making could be distinguished both neurochemically and anatomically.

3.1 A common model for economic decision making: Two-Stage model

As introduced in chapter I, the basic mechanism for producing decisions involves two stages: evaluation and choice. The first of these stages is concerned with the evaluation of all goods and actions; the second is concerned with choosing amongst the goods or actions presented in a given choice set. At a very basic level, one can think of the evaluation mechanism as being associated with learning and representing the values of objects and actions. Comparisons between different kinds of options (either goods or actions) rely on this abstract measure of subjective value, a kind of ‘‘common currency’’ for choice. There is now growing evidence (Kable &, Glimcher, 2009; Padoa-Schioppa and Assad, 2006, Lau and Glimcher, 2008) that subjective value representations do in fact play a role at the neural algorithmic level, and that these representations are encoded primarily in medial prefrontal cortex and striatum. Medial prefrontal cortex (specifically the ventomedial aspect) and striatum encode the subjective value of different goods or actions during decision making in a way that could guide choice. But how do these subjective value signals arise? A critical source of value information is past experience. Dopaminergic neurons in the midbrain encode a teaching signal that can be used to learn the subjective value of actions (Niv and Montague, 2009). Indeed, these kinds of signals can be shown to be sufficient for learning the values of different actions from experience. Since these same dopaminergic neurons project primarily to prefrontal and striatal regions (Haber, 2003), it seems likely that these neurons play a critical role in subjective value learning. Subjective values are learned through iterative updating based on experience. The theories rest on the idea that each time a subject experiences the outcome of her choice, an updated value estimate is calculated from the old value estimate and a reward prediction error, the difference between the experienced outcome of an action and the outcome that was forecast. This reward prediction error is scaled by a learning rate, which determines the weight given to recent versus remote experience (Schultz et al, 1997).

Learning and encoding subjective value in a common currency is not sufficient for decision making; one action still needs to be chosen from among the set of alternatives and passed to the motor system for implementation. What is the process by which a highly valued option in a choice set is selected and implemented? Unlike evaluation, which has been extensively studied in both humans and other animals, choice has been the subject of

study principally in awake-behaving monkeys in neuroscience. This may reflect the fact that the temporal dynamics of choice make it difficult to study with fMRI. The knowledge for these last components of the decision process are limited and strongly dependent on the experimental tasks, as for the example the model of decision making based on the saccadic-control system in monkeys (Andersen and Buneo, 2002; Glimcher, 2003). Findings (review, Kable et al., 2009) suggest that the areas for choosing, based on values, are the Lateral Prefontal Cortex and Parietal Cortex. These areas could be responsible for the selection and implementation of choices from among any set of available options.

At a theoretical level, the process of choice must involve a mechanism for comparing two or more options and identifying the most valuable of those options. This is also true for the action involving the choice; moreover the system that performed such a comparison must be able to represent the values of each option before a choice is made. Basso and Wurtz, (1998) established that activity at the two candidate movement sites, during the period before the burst, was graded. If the probability that a movement (a saccade) would yield a reward was increased, firing rates associated with that saccade increased; and if the probability that a saccade would yield a reward was decreased, then the firing rate was decreased. Platt and Glimcher (1999) found that firing rates in area LIP before the collicular burst occurred were a nearly linear function of both magnitude and probability of reward. Other studies (Dorris and Glimcher, 2004; Janssen and Shadlen, 2005; Kim et al., 2008) showed that various manipulations that increase (or decrease) the subjective value of a given saccade also increase (or decrease) the firing rate of neurons within the frontal-parietal maps associated with that saccade. The fronto-parietal map encodes the subjective value of a particular saccade relative to the values of all other saccades under consideration. This suggests that whereas the orbitofrontal and striatal neurons appear to encode absolute (and hence transitive) subjective values, parietal neurons, presumably using a normalization mechanism, rescale the absolute values so as to maximize the differences between the available options before choice is attempted.

There are obviously many open questions about the details of this mechanism, as well as many vigorous debates that go beyond the general outline presented. With regard to evaluation, some important open questions, especially for this research work, concern how the function of medial prefrontal cortex and striatum might differ and the role of medial

prefrontal specifically the orbito and ventromedial one, in peculiar human decision such as social and moral decisions.

3.2 Neuronal basis of social decision making

The basic building blocks of decision making (see the two stage model above) that underlie the process of learning and evaluation also play important roles for decision making in social contexts. However, interactions among multiple decision makers in a social group display some new features. In social context is still true that the 'rational' view suggests that human seek to maximize their self-interest according to the information available, but at the same time social interactions open the possibility of competition and cooperation. Humans and animals indeed act not only to maximize their own self-interest, but sometimes also to increase or decrease the well-beings of others around them. These unique aspects of social decision making are reflected in the activity of brain areas involved in learning and evaluation. As anticipated in chapter II, a good starting point for studies of social decision making is game theory (von Neumann and Morgenstern, 1944).

Socially interactive decision making tends to be dynamic and the process of discovering an optimal strategy can be further complicated by the fact that decision makers often act according to their other-regarding preferences. Nevertheless, the basic neural processes involved in outcome evaluation and reinforcement learning might be generally applicable, regardless of whether the outcome of choice is determined socially or not. One of the areas that plays a key role in socially interactive decision making is the striatum. During decision making without any social interactions, activity in the striatum is influenced by both real and fictive reward prediction errors (O'Doherty, 2003; Lohrenz, 2007). Reward prediction errors during social decision making also lead to activity changes in the striatum. For example, during the prisoner's dilemma game, cooperation results in a positive BOLD response in the ventral striatum, when this was reciprocated by the partner, but produces a negative BOLD response in the same areas when the cooperation was not reciprocated (Rilling et al, 2002; Rilling et al 2007). In addition, the caudate nucleus of who receive money from an investor (the trustee) in repeated trust

game[3] displays activity correlated with the reputation of the investor (King-Casas et al, 2005).

Other data related to the brain activation during social decisions come from a study conducted during the ultimatum game (see description of the game in chapter II). A functional magnetic resonance imaging study (Sanfey et al., 2003) examined unfair behavior in the ultimatum game (UG) and found that anterior insula exhibits greater activation for unfairness offers. The activation of this area predicted the player's decision to either accept or reject the offer, with rejections associated with significantly higher activation than acceptances. The presence of anterior insula activations in rejection of unfair offers is particularly interesting because this brain region is also responsive to physically painful (Derbyshire et al., 1997) and disgusting stimuli (Calder et al., 2001). An unfair offer generates the same activation as physical pain related to the emotion of disgust; physical disgust plays a similar role in the unfair offer to moral disgust. Despite the differences between contexts, brain activation and emotional reaction are essentially the same. Anterior insula and associated emotion-processing areas may play a role in marking a social interaction as aversive signal discouraging social availability. Separate measures of emotional arousal provide support for this hypothesis. In a study measuring skin-conductance responses as an autonomic index of affective state (van 't Wout, 2006), the author found higher skin conductance activity for unfair offers than for fair ones, and, in a like finding with insular activation, the SCR discriminates between acceptances and rejections of these offers. The influence of emotions on social decisions is not only an ongoing process but can have also an initial motivational function, preceding the actual decision. Moretti & di Pellegrino (in press) primed subjects by pictures with disgusting or neutral contents before several shots of ultimatum game. Subjects primed with disgusting pictures rejected unfair offers more frequently than subjects primed with neutral stimuli. These data support the hypothesis of the involvement of emotions in social decision making.

[3] The game starts when both players receive an initial endowment for example: €9. The first mover (investor) has to decide how much she wants transfer to the trustee. Any amount invested by the investor is tripled by the experimenter and sent to the trustee. When the trustee receives the amount he decides whether he wants return any to the investor. The amount of money that the investor decides to send to the trustee is a measure of trust, whereas the amount that the trustee sends back to the investor is a measure of reciprocity. A complete description of trust game is provided in Study I.

In a study reported by de-Quervain and colleagues (2004) the investor has the possibility to punish a trustee who betrays the trust received. Authors observed that subjects punish the betrayer even if this implies an economic cost to themselves (de-Quervain et al, 2004). Such punishment may have some hedonic value for the investors, since activity in the caudate nucleus of the investor was correlated with the magnitude of punishment and increased only when this punishment was effective. Because the caudate nucleus has been associated with experience of gain and pleasure, activation of this nucleus before punishing selfish others reveals the motivation (hedonic) role of emotion in the decision to punish someone.

These findings suggest the possibility that the striatal response to the reward received by others might change depending on whether a particular social interaction is perceived as competition or cooperation. Indeed, during a board game in which subjects were required to interact with each other competitively or cooperatively, a number of brain areas were activated differentially depending on the nature of interaction (Decety et al, 2004). For example, compared to competition, cooperation resulted in stronger activation in the anterior frontal cortex and medial orbitofrontal cortex. However, whether and how these cortical areas influence the striatal activity related to social preference is currently not known.

The last aspect for a description of social decision making mechanism regards the fact that social decisions requires a theory of mind, namely, the ability to predict the actions of other players based on their knowledge and intentions (Gallagher, 2003). Many neuroimaging studies on experimental games have found that social interactions with human players produce stronger activations in several brain areas, often in the anterior paracingulate cortex, compared to similar interactions with computer players (McCabe, et al 2001; Rilling et al 2004). There is some data in this sense related to the decision making process using game theory tasks. However using the trust game, a recent study has identified a unique role for the cingulate cortex in representing the information about the agent responsible for a particular outcome (Tomlin et al, 2006).

Our current knowledge of neural mechanisms for social decision making is still limited. Social decisions, in comparison to basic decisions such as which food to eat, imply additional features: the ability to predict the actions and emotional reaction of others, the

basic and complex emotions related to the social interaction, and moral and social rules intrinsic to social relations. All these aspects complicate the not fully completed basic model of decision making.

In the experimental part of this work, I will try to better understand the role of emotions in social decision making. With regard to the evaluation stage, it is still not completely clear how emotions could be enrolled in the evaluation process and, more in general, in the attribution of value during a social relation. The orbitofrontal cortex (OFC) seems be a key area in this sense. The orbitofrontal cortex (OFC) seems to have an important role in the generation of subjective value especially in contexts that concern abstract goods such as the social and moral meaning of human relations. Before introducing a summary of studies regarding the activation of OFC in social decision making, the next paragraph will explain the anatomical correlate of OFC.

3.3 "The mysterious orbitofrontal cortex "

The orbitofrontal cortex (OFC), located above the orbits of the eyes, is a part of the prefrontal cortex, often is defined topographically as the cortex on ventral surface of the frontal lobe. The OFC has more clear anatomical definition in primate involving area 14 medially, area 13/25 caudally, areas 11 and 12 around the inferior convexity and the ventral part of area 10, toward the frontal pole (see figure 3.1 lower part). However these areas do not have a perfect homologue in human brain and for this reason there is still a deep debate between authors on the anatomical definition of OFC. We report here the anatomical localization proposed by Price (2007) (see figure 3.1 upper part), including the lateral surface area 47/12 (incorporating the human equivalent of the primate area 13), the most caudal region of the medial OFC, area 25, extending to area 10 toward the frontal pole. Area 11 extends both medially and laterally on the ventral surface.

The OFC is densely interconnected with many other brain regions, suggesting it may subserve multiple functional roles. The orbital network is special in that: it receives input from the cortical areas associated with most of the sensory systems, including olfaction, taste/visceral afferents, somatic sensation, and vision (Price, 2007). In addition to the sensory inputs, the orbital network has specific connections with the thalamus and the striatum, and, furthermore, is connected with a number of limbic structures, some of which are assigned a key role in emotional processes: including the amygdala, hippocampus, entorhinal cortex, and parahippo-campal gyrus (Price, 2007).

3.4 Putative functions of orbitofrontal cortex

The OFC has multiple functional roles, and its activation was observed in numerous and completely different tasks. The OFC seems critically involved in reward processing, however its role appears to be a complex one, mediating the interaction between reinforce value, predictability and behavioral choice. First OFC responses have been associated with a variety of rewarding stimuli in functional Magnetic Resonance Imaging (fMRI) studies. Pleasant tastes (Berns et al., 2001; De Araujo et al., 2003; O'Doherty et al., 2000; Small et al., 2001) and smells (Gottfried et al., 2003; Rolls et al., 2003) associated with food reward have been shown to elicit responses in the OFC. Sexual stimuli (in the form of erotic film clips) also activate the OFC (Arnow et al., 2002), as do drug stimuli in drug abusers (London et al., 2000). These OFC responses to primary reinforces confirm findings from experimental animals that the OFC interacts with the brain reward circuitry. The capacity to detect a reward by OFC is not only for concrete reward (such as: food, water, sex, drug) but also for more abstract reward such as money and social value. O'Doherty and colleagues (2001) showed that medial OFC response in an fMRI study was correlated with the amount of abstract (''play'') money won on a probabilistic decision making task, while lateral OFC response correlated with the amount of money lost.

Neuroimaging data confirm the findings of animal studies (see chapter I) that the incentive values of reinforcers are coded in the OFC. Regions of the human OFC respond differentially to varying reward value, and showed to be sensitive to reward anticipation. In another fMRI study O'Doherty and colleagues (2002) presented subjects with visual cues that they had learned to associate with pleasant, unpleasant, or neutral tastes. Cue-related OFC response was enhanced to the cue predicting a pleasant taste reward. Kirsch et al. (2003) demonstrated a similar effect for monetary reinforcers. Rather more complex tasks have also been used to demonstrate anticipatory OFC responses to reward. Breiter et al. (2001) developed a complex task using ''spinners'' of fortune depicting various reward probabilities and values. The OFC was one of a number of regions that responded to both the expectation and the experience of rewards. Other imaging studies have further suggested that anticipatory responses of OFC are dependent on the degree of uncertainty; one observes an increasing anticipatory activation in posterior–lateral OFC response with increasing uncertainty (Critchley et al., 2001). The OFC is a key brain area to translate knowledge of reinforcement contingencies into appropriate behavioral choices. Patients

with OFC damage, even if they can understand the contingencies of the task, fail to translate this knowledge into advantageous decision making (see gambling task in chapter II). This suggests that OFC is involved in specific decision making aspects of reinforcement processing tasks. Disinherited and socially inappropriate behaviors can often be sequelae of OFC damage in humans. The impulsivity and disinhibition that characterize real-life behavior of patients have also been observed in experimental cognitive paradigms (Berlin et al. 2004). Like animals with OFC lesions, patients with damage to this region display perseverative impairments on a reversal learning task (Rolls et al., 1994). Such perseverative response was highly correlated with scores on a questionnaire concerning disinhibited behaviors in everyday life. The authors argued that a difficulty in modifying responses in the face of negative information might underpin the behavioral problems of these patients. It has subsequently been demonstrated (Hornak et al., 2004) that the perseverative impairments seen in OFC patients are not due to a simple failure to inhibit motor responses, but a failure to reverse associations between stimuli and reinforcers. Thus, these authors propose that the inhibitory control problem observed in these patients is specific to reinforced contexts.

Behavioral selection and decision making not only involve the ability to relate different courses of action to potential reinforcing outcomes, but also the ability to change one's course of action as motivational contingencies change. This is particularly true in complex contexts that change in relation to abstract social and moral values, often in opposition to more salient and immediate material values. An abstract value could drive our decision and behaviour, but also more simple social stimuli such as emotional expression of face can induce a sudden change of decision or behavior. Children are particularly accurate in understanding from their parents' facial expression if what they are doing is appropriate or not. In everyday life, indeed, negative facial expressions provide important reinforcing cues signalling the potential need for a change in behavior. Patients with OFC lesions show impairments in processing negative emotional expressions (face and voice) (Hornak et al., 1996). By contrast, such patients possess a normal ability to identify positive expressions (e.g. happiness). Processing negative expressions can be a socially and biologically important component of the inhibitory control system. Impairments in processing expressions in OFC patients could reflect dysfunction of a system responsible for changing behavior in response to these socially/emotional salient cues. These findings suggest that inhibitory deficits observed in patients with OFC damage

may in fact reflect a specific form of reinforcement-processing deficit, namely, impairment in inhibiting previously appropriate behavior in the light of changing motivational contingencies. Rolls (2004) argues that the OFC plays ‘‘a special role’’ in stimulus-reinforcer learning, based on its ability to perform rapid reversals of associations. Gorno and Tempini (2001) reported enhanced OFC response to expressions of happiness relative to disgust as in the case laughing and smiling expressions (Iwase et al., 2002). Emotive expressions both positive and negative seem important reinforcing cues signalling the more appropriate behavior in a social context. In this sense OFC is involved in choosing responses and making decisions based on motivationally salient information. To do this effectively, the OFC must code the current incentive value of external reinforcing cues and be able to respond rapidly to changes in the environment. OFC must be able to process uncertainty anticipating the expected outcome, as well as be able to change response quickly in the face of unexpected negative outcome. Emotional processes seem to reliably engage a set of structures including reward-processing mechanisms and areas of the midbrain and cortex to which they project, such as vmPFC, orbitofrontal cortex, and anterior cingulate cortex, as well as other areas such as the amygdala and insula (Dalgleish, 2004). Neuroscientific studies offer the potential to examine the causal relationship between an emotional reaction and a subsequent social decision, as well as to investigate whether areas specialized for the processing of basic emotions may be co-opted for more complex affective reactions.

Further evidence is necessary to understand how basic affective processes can be involved in more complex emotional social behavior. This is particularly true for positive social-emotional dispositions such as trust, altruism, cooperation and moral behavior. To my knowledge, no studies exist that have tried to manipulate emotions during positive social decisions, such as whether to trust someone or for moral decisions. The following studies are attempts to provide new data in relation to these matters.

STUDY I

PREFRONTAL DAMAGE REDUCES BETRAYAL AVERSION IN ECONOMIC EXCHANGES

1.1 Defining trust and reciprocity

Trust is ubiquitous in society and an essential ingredient of human exchange (Arrow, 1974); it lubricates social and economic transactions, and has been long recognized as a critical antecedent of cooperative behaviour (Ostrom & Walker, 2003). Trust can be defined as one's willingness to place resources at the disposal of another party in situations in which there is uncertainty regarding the other party's motive, intentions and actions (Mayer et al., 1995; Rousseau et al., 1998). An action that is trusting of another is one that creates the possibility of mutual benefit, if the other person is cooperative. Yet trusting behaviours also imply the risk of injury or loss to oneself if the other person defects. Overriding aversion to such risks is required for trust to emerge (Kosfeld et al., 2005). In other words, trust is the willingness to accept vulnerability based upon positive expectations about another's behavior. In general, trust can exist between individuals, groups, and institutions, and can represent either a global belief in humanity or a situation-specific and/or trustee-specific attitude (Butler, 1991). In this study the dyadic-level interpersonal trust is investigated.

From these definitions it is clear how trust has a double nature of being both desirable and risky (Roderick, 2001). On the one hand, trust is desirable because it creates a cooperative atmosphere that is opportune for the human being (Guth, 1982), avoiding the generation of negative emotional states such as fear and insecurity. Moreover, trust is

fundamental in building common and generally stable (but not certain) rules of relationship between people. This shared tenet reduces the cognitive and emotional effort connected with an ambiguous and unsecure environment. On the other hand, trust is risky because we cannot be sure that our trust will be repaid suitably: to trust someone always involves taking the risk of being betrayed. This risk is due to a lack of full knowledge of others, as to their motives, their intentions, and their responses to endogenous as well as exogenous changes (Gambetta, 1988). Although theoretical work has identified a number of factors at the base of trust and what influences trust (Mayer et al., 1995), fundamental questions still remain about how trust actually operates.

Another concept related to trust regards the normal reaction that trust induces. In real life, in response to a friendly action, people usually are favourably disposed to repay back with a similarfriendly action, which is often called reciprocity. To trust someone can be considered a friendly action. The tendency to repay the favour of trust is a behavioural propensity to cooperate conditionally with other group members' (Fehr & Fischbache, 2003). Reciprocity behaviour is based on two important motivational drives: "reciprocal fairness" (Rabin, 1993; Falk & Fischbacher, 2006) and "inequity aversion" (Fehr & Schmidt, 1999). A reciprocally fair subject is motivated by the desire to respond to kind acts with kindness and to hostile acts with hostility. An inequity-adverse subject is motivated by the desire to avoid inequity and to implement equitable outcomes (Fehr & Schmidt, 2000).

1.2 Rational and emotional processes in Trust

A commonly held view suggests that trust is a result of rational calculation and higher cognitive processes (Coleman, 1990); however, in some other accounts trust is held to be founded on social-emotional processes (Hardin, 2002). Consistent with this latter account, behavioural studies suggest that incidental emotions significantly influence trust (Dunn & Schweitzer, 2005). Moreover, several neuroimaging studies have shown that tasks that require social evaluation (Winston et al., 2002; Somerville et al., 2006), or cooperation with another individual (McCabe et al., 2001; Gallagher et al., 2002; Rilling et al., 2002, 2004; Tomlin et al., 2006) activate brain regions known to process social emotions, including the anterior cingulate cortex and adjacent medial frontal cortex. Importantly, when subjects interact with partners they know to be just computers, these activations are not seen, suggesting that they reflect the interpersonal nature of the task

(McCabe et al., 2001; Rilling et al., 2004; Tomlin et al., 2006; van den Bos et al., 2007). Moreover, a recent study on trust offered the chance of gaining a deeper understanding of the neural mechanisms underlying cooperative behaviour using a specific neuropetide, Oxytocin (OT) plays a central role in the ability to form social attachment and affiliation. OT has been shown to increase the ability to infer the mental states of other (Domes et al., 2007), and has a specific effect on trust decisions, increasing the willingness to take social risks. Interesting, OT does not increase the tendency to take risks in general, but rather increases only the tendency to taking social risks (Kosfeld et al., 2005). This finding suggests a possible role of OT in reducing social fears associated with betrayal (Kosfeld et al., 2005). In another recent study (Baumgarten et al, 2008), it has been shown that OT reduces subjects' behavioural responses to breaches of trust. This result could suggest that OT changes the equilibrium between risky and desirable aspects of trust in favour of cooperation.

Only recently, a study by Krajbich and colleagues (2009) explored the behaviour of patients with lesions in the vmPFC in order to understand the role of this area in economic exchanges with other individuals. They required 6 vmPFC patients to take part in three games: the Ultimatum game, the trust game, both as investor and trustee, and the Dictator game. All three games were performed in one session, and by phone with the constant presence of the experimenter. During the tasks, the experimenter spoke with the 'fictitious' other person on the phone and described the subject's decisions as well as asked for the other player's decision (except for the dictator game). The aim of the work was not investigate the trust behaviour directly, but describe from patients' choice a formal model of how, and to what degree, vmPFC lesions affect an individual's social decision-making in general. In that study vmPFC patients, as compared to healthy subjects and non-frontal control patients, showed a significantly lower level of trustworthiness (e.g. reciprocity), whereas trust behaviour (money invested) remained unaltered.

There are three critical aspects regarding this first evidence about the involvement of vmpFC in decisions to trust a stranger. First the setting of the experiment was not realistic; the experimenter mediated all interactions. The experimenter's presence could have increased the social desirability bias of participants, especially in controls groups, forcing them to reply in a manner that would be viewed favourably by the experimenter. On this hypothesis, the difference between vmPFC and controls could be related to the

different sensitivity to the social desirability bias and not to trust behaviour per se. The second aspect regards the presence of feedback; even if the trust choice was one shot, feedback information has a clear impact on the next choice (Berg, 1995). In a task with feedback it becomes difficult to understand if the choice is a reliable measure of trust or is the reaction to the previous feedback. The third point regards the lack of a control condition in trust measurement; ultimatum and dictator game provide different measurements of social behaviour and cannot be easily compared to the trust game. To understand if there is any difference in trust behaviour between subjects it is important to compare the participant's choice in the trust game with choices made in a similar economic situation but without a social counterpart.

In conclusion, even if the Krajbich and colleagues (2009) study is the first to measure the trust behaviour in vmPFC lesioned subjects, their findings did not investigate the trust decision adequately, and their method has some important limitations.

1.3 Prefrontal damage reduces betrayal aversion in economic exchanges

In this study, we examined whether emotions, specifically social emotions subserved by the ventromedial prefrontal cortex (vmPFC), affect people's willingness to trust others. Several evidences suggest this possibility. First, the vmPFC is densely interconnected with basolateral amygdala, ventral striatum, and subcortical structures that control autonomic and visceral responses (Carmichael & Price, 1995; Haber et al., 2006), and is therefore ideally located for generating emotional responses, and guiding social interactions. Second, neuroimaging studies in humans have implicated the vmPFC in guiding behavioural choice under uncertainty (De Martino et al., 2006; Hsu et al., 2005), and have argued that this region is critical for balancing potential gains against losses to ensure optimal decision-making in social context (De Quervain et al., 2004). Finally, damage to the vmPFC in humans can be associated with strikingly poor judgement and decision-making (Eslinger & Damasio,1985; Bechara et al., 1994, 1997; Koenigs et al., 2007), due to markedly reduced (Koenigs et al., 2007; Ciaramelli et al., 2007, Krajbich et al., 2009;) or poorly regulated (Koenigs & Tranel, 2007) social emotions.

To address whether the vmPFC plays a necessary role in the decision to trust a stranger, a sample of patients with adult-onset vmPFC lesions, as well as healthy control

subjects (HC) and patients with lesions outside the frontal lobe (non-FC patients), played the role of investor in a one-round trust game (Berg et al., 1995). This game involves real monetary exchanges between two anonymous individuals, the investor and the trustee, who receive a sum of money from the experimenter (see figure 1.1 for a graphical example of trust game). The investor can keep all the money or decide to invest some amount, which is tripled by the experimenter and sent to the trustee. Next, the trustee decides how much of the tripled amount to return. Money sent by the investor is used to measure her trust, while money returned by the trustee is used to measure her reciprocity. The investor thus faces a motivational conflict between the prospect of increasing her payoffs, which motivates her to invest and cooperate, and the perceived probability of loss and betrayal, which drives her toward distrust.

Clearly, the decision to trust entails a risk (Rousseau et al., 1998). Uncertainty regarding whether the trustee intends to, and will, honour the investor's trust is the source of risk. This raises an important concern over whether a person's attitude toward risk in general influences trust (Eckel & Wilson, 2004; Karlan, 2005; Schecter, 2007). To control for between-group differences in risk attitudes, we therefore also implemented a risk game offering the same options and payoffs as the trust game, but in which a random device (e.g., a computer, see figure 1.2), not a human partner, determined the investor's risk. The risk game constitutes a critical control condition because recent behavioural (Bohnet & Zeckhauser, 2004; Hong & Bohnet, 2007; Bohnet et al., 2008; Houser et al ., 2009) and neurobiological (Kosfeld et al., 2005; Baumgartner et al., 2008) evidence strongly indicates that the decision to trust is not only determined by risk aversion (i.e., the negative emotion associated with the possibility of losing objects or money) but also by betrayal aversion, that is, the fear to be betrayed by another in social exchange. These observations are consistent with theoretical models (Rabin, 1993; Fehr & Schmidt, 1999; Charness & Rabin, 2002; Falk & Fischbacher, 2006) positing that, in addition to material outcome, people also value other agents' intentions. Betrayal aversion plays no role in the risk game, since random devices are incapable of intentionality or awareness, and they cannot really betray our trust. Therefore, the contrast between the trust game the and risk game is ideal to assess whether vmPFC damage specifically affects trusting behaviour in social exchanges (rather than risk-taking behavior in general), because – except for the type of opponent partner (human vs. computerized partner) – everything else remains constant across these two games.

Based on previous findings showing that activity in the vmPFC may be critical for making prediction and anticipating the unpleasant state of loss in decision-making (Damasio, 1994; Bechara et al.,1997), when the implications of another individual' intentions must be taken into account before choosing (Rudebeck et al., 2008; Behrens et al., 2009), we hypothesized that investors in the vmPFC-lesioned group would show higher money transfers than those in the control groups, particularly in the trust game in which both social and non-social risks operate to inhibit trust. In other words, if negative emotion responses mediated by the vmPFC represent the proximate mechanisms behind betrayal aversion, then damage to this neural structure ought to result in diminished levels of betrayal aversion and, therefore, lead to more trusting behaviour.

Several researchers (Andreoni & Miller, 2002; Cox, 2004) have argued that measures of trust taken from the trust game do not discriminate between actions motivated by trust and actions motivated by altruism or generosity. To address this question, we measured the amount of money participants returned when they played the role of trustee in a separate session. If lesion to the vmPFC increases generosity rather than trusting behaviour, then one might hypothesize that a player will send more as investor and return more as trustee, thus appearing both more trusting and trustworthy.

Finally, we included a measure of the investor's subjective expectation about the trustee's back transfer at different investment levels. This was in order to control whether vmPFC patients trust more because they are more optimistic about the trustee's trustworthiness (e.g., they show higher expected back transfers).

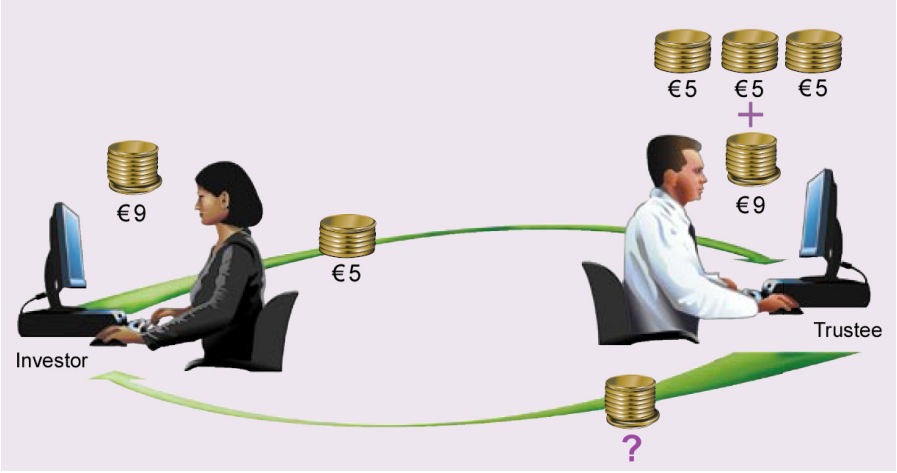


Figure 1.1. Graphical example of the trust game. The interaction is one shot and completely anonymous. The game starts when both player sreceive an initial endowment that in the picture is represented by €9. The first mover, the investor has to decide how much she wants transfer to the trustee. Any amount invested by the investor is tripled by the experimenter and sent to the trustee. In the example reported the investor sends €5, and the trustee receives €15. At this point the trustee decides whether he wants return to the investor. The amount of money that the investor decides to send to the trustee is a measure of trust, whereas the amount that the trustee sends back to the investor is a measure of reciprocity.

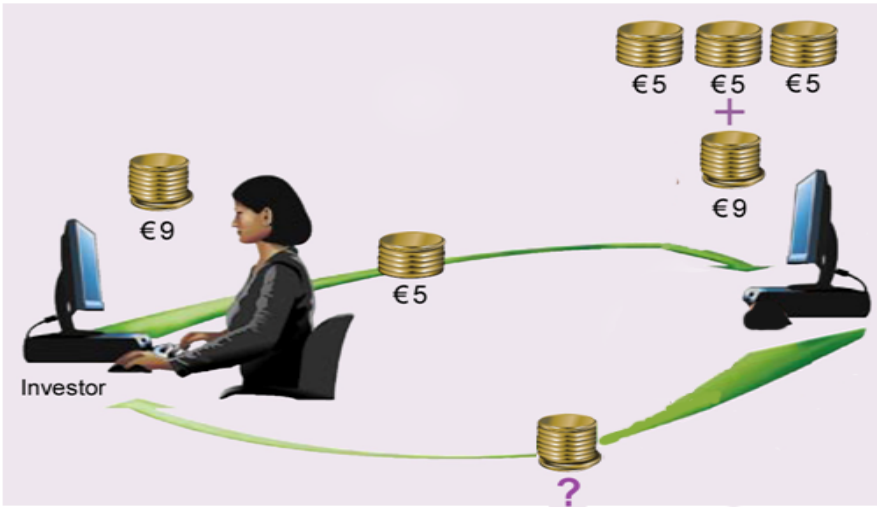


Figure 1.2. Graphical example of the risk game. In the risk game, the investor plays with a non-human counterpart (computer). The rules and the amounts of money are the same as those presented in the trust game. The amount that the investor bets in this game is a measure of risk propensity.

METHOD

Participants

Three groups of subjects participated in the study: (a) a group of patients with focal lesions involving the vmPFC (the vmPFC group, n = 10), (b) a control group of patients with damage sparing the frontal cortex (the non-FC group, n = 10), and (c) a control group of healthy subjects (the HC group, n = 10), who were matched on age, education and sex with the vmPFC group. Brain-damaged patients were recruited from the Centre for Studies and Research in Cognitive Neuroscience in Cesena and from Azienda Ospedaliera "Spedali Civili" in Brescia. They were selected on the basis of the location of their lesion evident on CT or MRI scans.

Table 1.1 shows demographic and clinical data, as well as the Mini-Mental Status Examination score (MMSE, Folstein, et al., 1983) There were no significant differences between vmPFC patients and comparison groups with regard to age, education, and clinical variables (p >.05 in all cases). In the vmPFC group, lesions principally involved

the vmPFC, which is defined as the medial one-third of the orbital surface and the ventral one-third of the medial surface of the frontal lobe, following the boundaries laid out by Stuss and Levine (2002). Lesion aetiology was haemorrhage due to ruptured aneurysm of the anterior communicating artery in 9 out of 10 vmPFC patients, and to traumatic brain injury in 1. The vmPFC damage was bilateral (although often asymmetrically so) in 6 cases, right unilateral in 2 cases, and left unilateral in 2 cases. All vmPFC patients presented with clinical evidence of a decline in social interpersonal conduct, impaired decision-making and emotional functioning, but had generally intact intellectual abilities (see table 1.2).

The non-FC patients were selected on the basis of having damage that did not involve the mesial orbital/ventromedial prefrontal cortex and frontal pole, and also spared the amygdala in both hemispheres. In this group, lesions were unilateral in 9 patients (in the left hemisphere in 5 cases, and in the right hemisphere in 4 cases) and bilateral in 1 patient, and were caused by ischemic or hemorrhagic stroke in 9 cases, and by traumatic brain injury in 1 cases. In the non-FC group, lesion sites included the lateral aspect of the temporal lobe in 6 patients, the lateral occipital area in 2 patients, and the occipito-parietal junction in the remaining 2 patients (see table 1.3 for more details).

All subject groups were administered a short neuropsychological battery including tests with potential sensitivity to frontal damage, as well as intelligence and memory tests (results are provided in table 1.2). The groups differed significantly only in their performance on the Stroop task, with vmPFC subjects making more errors than both non-FC patients and healthy controls (Mann–Whitney U-test, $p < .05$). Patients were not receiving psychoactive drugs at the time of testing, and had no other diagnosis likely to affect cognition or interfere with participation in the study (e.g., significant psychiatric disease, alcohol misuse, history of cerebrovascular disease or focal neurological examination). Neuropsychological and experimental studies were all conducted in the chronic phase of recovery, more than a year post-onset. All lesions were acquired in adulthood. Patients gave informed consent to participate in the study according to the Declaration of Helsinki (International Committee of Medical Journal Editors, 1991) and the Ethical Committee of the Department of Psychology, University of Bologna.

Normal participants were healthy volunteers who were not taking psychoactive medication, and were free of current or past psychiatric or neurological illness as determined by history. Normal controls scored at least 28 out of 30 on the MMSE.

Table 1.1 Summary data for participants [mean (standard deviation)]

Group	Sex (M/F)	Age at test (year)	Education (year)	Time since lesion (year)	Lesion volume (cc)
vmPFC (n=10)	7/3	57.8 (6.6)	10.4 (4.5)	4.6 (2.8)	32.6 (19)
non-FC (n=10)	7/3	54 (13.4)	10.3 (3.9)	3.8 (3.5)	26.5 (11.4)
HC (n=10)	7/3	57.3 (7.3)	9.5 (4.2)	-	-

Table 1.2 Results of selected neuropsychological tests [mean (standard deviation)]

Group	SRM	Digit Span Forward	Phonemic Fluency	Semantic Fluency	Corsi	Stroop Task Errors°	Rotter	PNR	BIS
vmPFC	35.5 (13)	5 (0.8)	20.2 (9.3)	36.6 (14)	3.7 (0.2)	6.5 (7.3)	2.2 (0.5)	2.9 (1.5)	1.9 (0.2)
non-FC	30.6 (4.8)	4 (0.9)	28.2 (10)	42.8 (15)	4.2 (0.7)	1 (1.4)	1.9 (0.4)	2.9 (1.1)	1.5 (0.4)
HC	32.2 (3.4)	5.7 (1)	29.2 (9.2)	49.5 (18)	4.8 (0.7)	0 (0)	1.9 (0.4)	2.9 (1.2)	1.9 (0.3)

SRM = Standard Raven Matrices (scores in percentile values), WMS = Wechsler Memory Scale, Rotter = Rotter Interpersonal Trust Scale, Rotter 1967), PNR = Personal norm of Reciprocity PNR, Perugini 2003, BIS = Barratt Impulsivity Scale, BIS-11,Barratt 1996).

Lesion analysis

Lesion analysis was based on the most recent clinical computerized tomography (CT) or magnetic resonance imaging (MRI). The location and extent of each lesion were mapped by using MRIcro software (Rorden & Brett, 2000). The lesions were manually drawn by a neurologist with experience in image analysis onto standard brain template from the Montreal Neurological Institute (MNI), which is based on T1-weighted MRI scans, normalized to Talairach space. This scan is distributed with SPM99 and has become a popular template for normalization in functional brain imaging. For superimposing of the individual brain lesions, the same MRIcro software was used. Figure 1.3 shows the extent and overlap of the brain lesions in the brain-damaged patients. Brodmann's areas (BA) affected in vmPFC group were areas 10, 11, 12, 32 (subgenual portion), and 24, with region of maximal overlap occurring in BA 10 and 11.

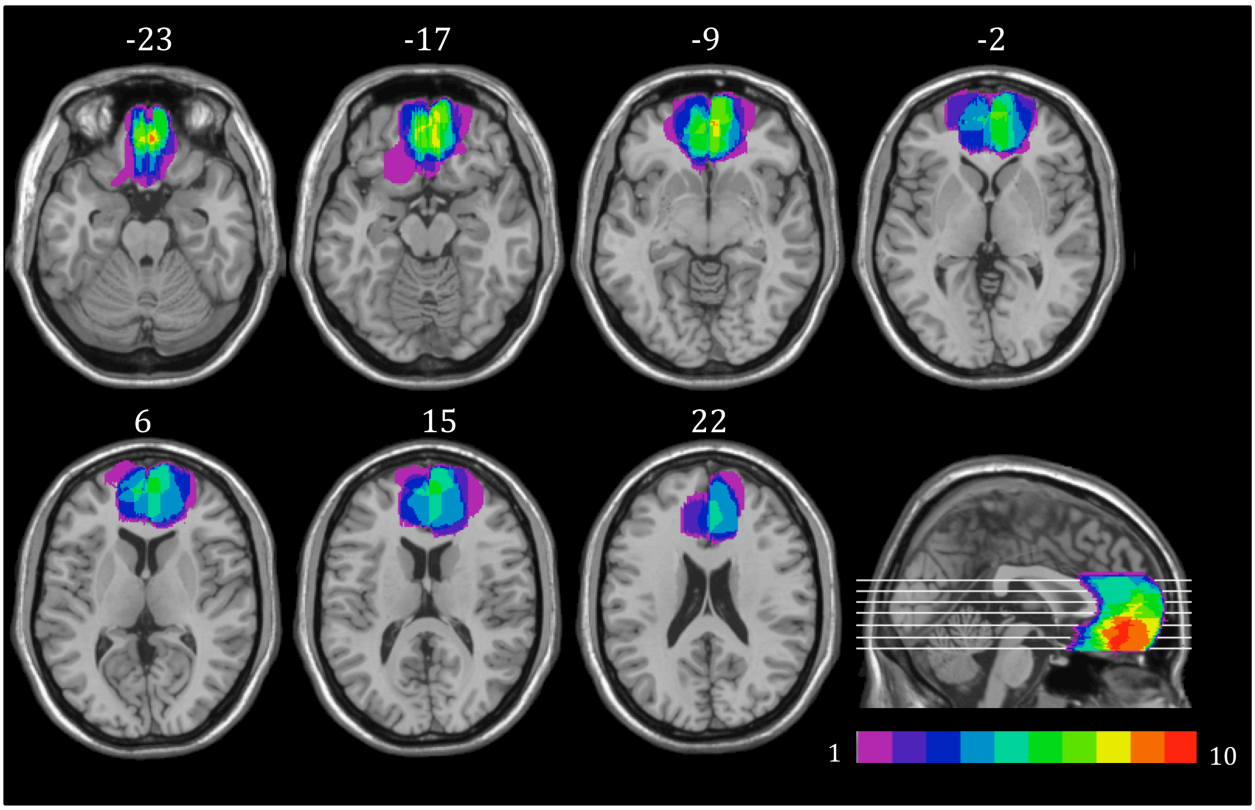


Figure 1.3 Location and overlap of brain lesions. The panel shows the lesions of the 10 patients with vmPFC damage projected on the same 7 axial slices and on the mesial view of the standard Montreal Neurological Institute brain.

		Trust	Crawford's modified t-test*	Risk	Crawford's modified t-test*	Belief	Crawford's modified t-test*	Trustee	Crawford's modified t-test*	Gender	Etiology	Site of lesion: Broadman Area
1	vmPFC	0.78	*t = 3.08, p < .01*	0.81	t = 2.49, p < .05	1.19	*t = -0.92, p = .36*	4.04	*t = -0.65, p = .51*	M	ACoA	10, 11 (left)
2	vmPFC	0.71	*t = 2.54, p < .01*	0.51	t = -0.12, p = .90	1.24	*t = -0.51, p = .61*	7.04	*t = 1.33, p = .19*	M	ACoA	10, 11, 24, 12, 32 (B>right)
3	vmPFC	0.76	*t = 2.99, p < .01*	0.62	t = 0.85, p = .40	1.19	*t = -0.92, p = .36*	3.06	*t = -1.18, p = .24*	M	ACoA	10, 11, 12, 25, 32 BA (B>right)
4	vmPFC	0.49	*t = 0.62, p = .53*	0.63	t = 0.96, p = .34	1.29	*t = -0.10, p = .92*	3.04	*t = -1.32, p = .20*	M	ACoA	10,11, 24, 32,47, (right);
5	vmPFC	0.68	*t = 2.30, p < .05*	0.67	t = 1,32 p = .20	1.49	*t = -1.54, p = .14*	2.04	*t = -1.98, p = .06*	M	ACoA	10, 11, 32 (B)
6	vmPFC	0.72	*t = 2.65, p < .01*	0.72	t = 1,86 p = .07	1.14	*t = -1.33, p = .19*	4.08	*t = -0.39, p = .69*	M	ACoA	10, 11, 12, 32, 47 (left)
7	vmPFC	0.74	*t = 2.84, p < .01*	0.62	t = 0.96, p = .34	1.30	*t = -0.01, p = .98*	4.04	*t = -0.65, p = .51*	F	ACoA	11, 12, 24, 32, 46 (B)
8	vmPFC	0.44	*t = -0.18, p = .85*	0.40	t = -1.01, p = .32	1.27	*t = -0.26, p = .79*	2.06	*t = -1.85, p = .07*	F	ACoA	10, 11 (B)
9	vmPFC	0.62	*t = 1.74, p = .09*	0.63	t = 1,05, p = .30	1.26	*t = -0.34, p = .73*	3.08	*t = -1.05, p = .30*	F	ACoA	10,11 (B)
10	vmPFC	0.55	*t = 1.82, p = .27*	0.60	t = 0.79, p = .44	1.23	*t = -0.59, p = .56*	4.04	*t = -0.65, p = .51*	M	TBI	10, 11, 12, 24, 32 (B)
	mean	0.65		0.62		1.26		4.12				
	sd	0.12		0.11		0.10		1.40				
	Controls (n 20)									14 M		
	mean	0.43		0,5		1.31		5.39				
	sd	0.11		0.11		0.12		1.47				

Table 1.3. The t testa are calculated with the last upgraded version of the program Singlims.exe (Crawford & Garthwaite, 2007)
U = unilateral; B = bilateral, TBI: Traumatic brain injury

		Gender	Site of lesion*	Type of lesion**
1	non-FC	F	T,IC,ic (lef)	H
2	non-FC	F	OP (righ)	I
3	non-FC	M	TP,IC (lef)	H
4	non-FC	M	T, IC, bg (right)	H
5	non-FC	M	T, IC (left)	H
6	non-FC	F	T (left)	H
7	non-FC	M	TP (left)	I
8	non-FC	M	O (bilateral)	I
9	non-FC	M	O (right)	Traumatic brain injury
10	non-FC	M	OP (righ)	I

Table 1.4. Lesion's details in non-FC group.
* bg = basal ganglia; ic = internal capsule; P = parietal lobe; T = temporal lobe; O = occipital lobe; IC = insula cortex.
** I = ischaemic; H = haemorrhagic.
TBI = Traumatic brain injury

Experimental design and procedures

Every participant in the experiment played the role of investor in two treatment conditions: a trust game and a risk game. In the trust game, the subject played a standard trust game and she knew her counterpart was human; we call this the human interaction treatment. In the risk game, the subject knew her counterpart was a computer making random decisions; we call this the computer interaction treatment. Trust and risk games were played in separate sessions with an interval of at least 1 week between them. Half of

the participants in each group played the trust game in the first session, and half the risk game in the first session.

In the separate third session, all participants played also in the trustee's role in a trust game, whereas in a fourth session participants completed the questionnaire and received the feedback of previous sessions. Our main interest was the comparison between trust and risk behaviour, and for this reason we balanced the order of the two games. On the contrary, the trustee role has a different meaning: it does not measure trust or risk, but reciprocity, the willingness to sacrifice one's own economic gain to repay another's friendly (e.g., trustful) action. Moreover, when subjects play in trustee role it is easy for them to calculate the income for the game, and thus it is not possible to separate the genuine level of reciprocity from a reaction due to an investor's action. For these reasons the trustee role is always undertaken after the risk and trust game. The lack of feedback after each interaction in trust and risk game has the purpose of avoiding any influence from feedback in order to measure a general level of trust and risk. Finally, all interactions were one shot in order to avoid reputation and order effects.

All experiments took place in a quiet room in which an opaque, removable partition wall was used to create two separate settings. On either side of the wall, we placed a desk with a computer. Participants sat at one desk in front of the computer, while at the other desk sat either an actor who played in the role of the trustee (trust game), or no one (risk game). As a result, playing partners could be separated visually, thereby providing between-subject anonymity, without separating them audibly, thus lending our set-up credibility. Before each session, instructions about the nature and rules of the game were presented on the computer, and the experimenter verbalised them to ensure that participants understood them. In the instructions, it was emphasized that participants in the trust game would play the game anonymously and only once with each opponent player, and that they would receive the money earned in the game. Differently, in the risk game it was emphasized that participants would play with a computer counterpart. After reading the instructions, subjects were required to complete a quiz that required them to state the amount of money that each player would receive under various hypothetical circumstances. The game started once the subject successfully finished the quiz.

Subjects in the role of the investor received no feedback about their partner's decision between the different interactions. At the end of each session, the experimenter

put the cash payoff earned by subject during the game into an opaque envelope that was sealed and signed by the participant. Earnings envelopes were kept by the experimenter between games. Subjects did not receive feedback about the outcome of any game until the end of the experiment in order to avoid income effects and the possibility that current decisions were influenced by an opponent's previous decisions. All games were paid out at the end.

Human interaction treatment.

Participants acted as the investor in a series of 9 rounds of a trust game against 9 different anonymous human partners via a computer interface. At the beginning of each round, the actor that played the role of the trustee entered the room and sat at her position. When both investor and trustee were ready, the interaction started. Each round was presented as text through a series of five screens (see Figure 1.4 for a schematic illustration of a typical round). A 6-s initial screen depicted a silhouette of a human figure and indicated the endowment (E) available for both players in the current round. There were three equiprobable initial E, €6, €9 and €12, presented in random order during the game. The second screen posed the question "How many Euros between 0 and E do you transfer to Participant B?" and remained visible until a response was given. Participants were given the opportunity to send any integer amount from zero to their entire endowment available, and were instructed to indicate their decision by pressing the numeric keys of the computer keyboard. Following the response, a screen indicating the investor's transfer and the amount received by the trustee (three times the amount invested) was presented for 4 s. Then, a variable 5- to 15-s waiting screen informed the subject that the trustee (Participant B) was deciding how much of the tripled amount to send back. Subjects were informed that Participant B could choose the amount from any integer between zero and the tripled amount they had transferred to her/him. Finally, a screen signalled the end of the round. The trustee went out of the room and after a short break was replaced by another actor to begin the next round. When the trustee was out of the room, the investor was asked about her expectation in relation to the trustee's back transfer.

Computer interaction treatment.

Participants were instructed that they would play 9 rounds of a risk game in which a random mechanism determined the outcome of the game. In the risk game, everything was

identical to the trust game, except that subjects played against a computerized partner. A silhouette of a computer was displayed in the initial screen to indicate the computer interaction. Participants were informed that, in each round, the computer would randomly choose the amount to transfer back from any number between zero and the tripled amount they have transferred to it (figure 1.6).

In a separate session, participants played 5 rounds of a trust game in the role of trustee against 5 different anonymous investors via a computer interface. The experimental setup was as before, except that participants were assigned the role of trustee (Participant B), and an endowment of €9 was available for both players in every round.

Each new round began with a 6-s initial screen that depicted a silhouette of a human figure and indicated that €9 was available for both players in the current round (see figure 1.5 for a schematic illustration of a typical round). Then, a variable 5- to 15-s waiting screen informed that the investor (Participant A) was deciding how much between €0 and €9 to transfer to the trustee (Participant B). Next, a screen indicating the investor's transfer and the amount received by the trustee was presented for 4-s. The investor's transfers, X, were predetermined and presented randomly, and included one transfer of each €0, €3, €5, €7 and €9, so that the trustee received €0, €9, €15, €21 and €27, respectively. Then, the question "How many Euros between 0 and 3X do you transfer back to Participant A?" appeared on the screen, and remained visible until a response was given. Participants were given the opportunity to send back any integer amount from zero to the tripled amount received, and were instructed to indicate their decision by pressing the numeric keys of the computer keyboard. Following the response, a screen signalled the end of the round. The trustee went out of the room and after a short break was replaced by another actor to begin the next round.

Note that participants in all groups faced exactly the same set of investors' transfers. Thus, behavioural differences across these three groups cannot be attributed to differences in the distribution of investors' transfers.

Questionnaires.

Approximately two weeks after the experiment, participants also completed three self-report questionnaires that assessed selected personality traits. The Barratt

Impulsiveness Scale (BIS-11) is a 30-item, self-report measure of impatience and trait impulsiveness (Patton et al., 1995). The Personal Norm of Reciprocity (PNR) scale is 27 items questionnaire measuring three dimensions (9 items each) of reciprocity (i.e., the propensity to reward those who have behaved nicely and punish those who behaved badly): positive reciprocity, negative reciprocity, and beliefs in reciprocity (Perugini et al., 2003). Finally, the Interpersonal Trust Scale (Rotter, 1967) includes 25 component questions requiring subjects to express their trust expectations across a variety of social situations and with diverse social agents.

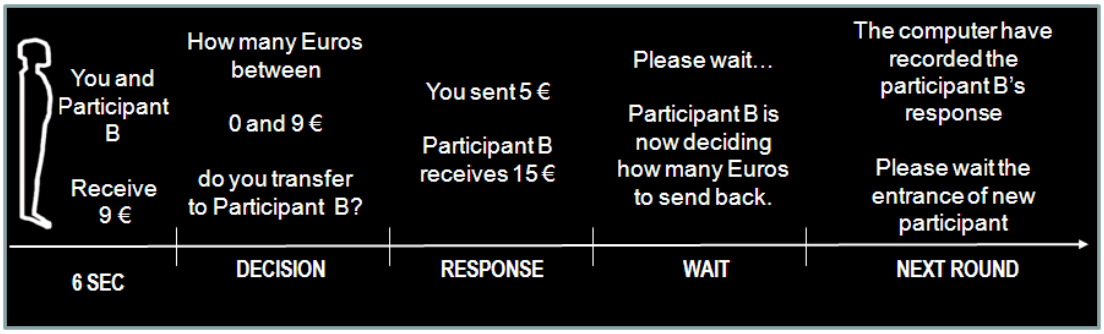


Figure 1.4 Schematic diagram of a single round of trust game where subjects played the investor's role. In the example presented the subjects in investor's role sends to the trustee €5. The original screens were in Italian.

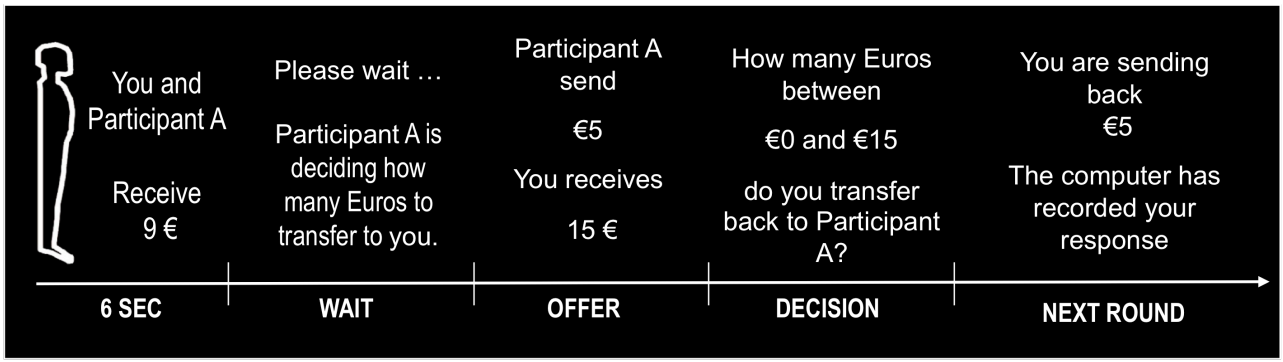


Figure 1.5 Schematic diagram of a single round of trust game where subjects played the trustee role. In the example presented the participant in trustee's role sends back to the investor €5. The original screens were in Italian.

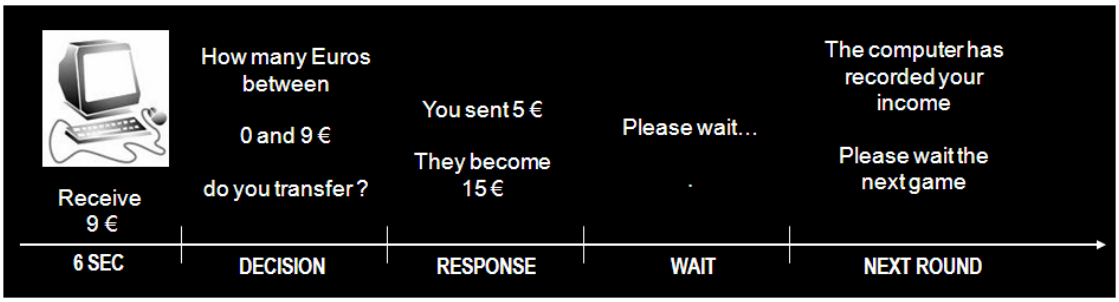


Figure 1.6 Schematic diagram of a single round of risk game where subjects played in the role of investor with a non-human counterpart (computer). In the example presented the subject in the investor's role bets €5. The original screens were in Italian.

Design Section

In the first analysis, a mixed design 4x3x2 ANOVA on all transfers amount was performed. Transfers amounts in this case concern the average transfer of each subject in each interactions/games: trust game, risk game, trustee role and expectation. In the ANOVA, the treatment (trust, risk, reciprocity and subjective expectation) is the within-subjects factors and groups (vmPFC, non-FC, HC) and gender (male female) are between-subjects factor. This analysis has as a main aim to observe if there is any gender, group effect and other preliminary indications.

Subsequent analysis concerned a more specific comparison between transfer amounts in trust and risk game. The transfer amounts are measured in three different groups (vmPFC, non-FC, HC) in two different conditions (treatment: human and computer) with three different kind of starting endowment available (€6, €9, €12). A mixed design 3x2x3 ANOVA on transfer amounts was performed with groups (vmPFC, non-FC, HC) as between-subjects factor, treatment (human and computer) and endowment (€6, €9, €12) as within-subjects factors. When necessary, pairwise comparisons were conducted using the Fisher LSD test, which is considered the most powerful technique for post hoc tests involving three groups (Cardinal & Aitken, 2006). On the same data the non parametric analysis is performed.

Thirdly, analysis was conducted that concerns the subjective expectation: the participant' responses to the experimenter's question about her expectation regarding the trustee's back transfer in the trust game. Subjective expectation was measured in three groups (vmPFC, non-FC, HC) with three different kinds of starting endowment available (€6, €9, €12) and was calculated on expected back transfers divided by the amount sent (a value > 1 indicates expected gain, whereas a value < 1 indicates expected loss from the exchange). A mixed design 3x3 ANOVA on subjective expectation is performed with group (vmPFC, non-FC, HC) as between-subjects factor and endowment (€6, €9, €12) as within-subjects factor

Fourthly, analysis involved participants' responses in the trustee role. This variable was measured in three different groups (vmPFC, non-FC, HC) in five different trials. A one-way ANOVA with group as between-subjects factor was performed.

Finally, analysis was conducted regarding the personality questionnaires; a nonparametric test (Kruskal-Wallis test) was used to compare the three groups (vmPFC, non-FC, HC). PNR questionnaire is divided in three subscales each subscale is analyzed by a Kruskal-Wallis test.

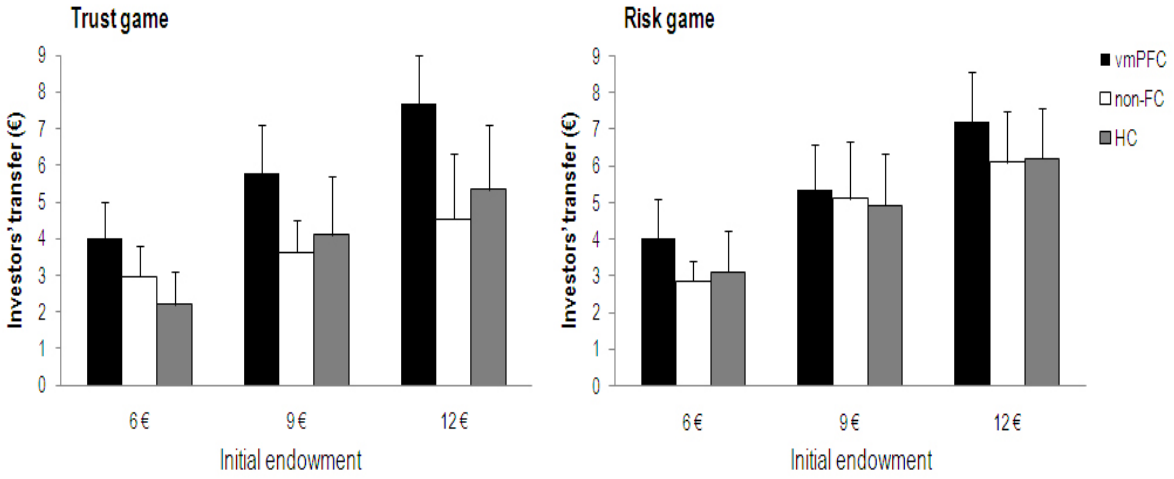


Figure 1.7 Investors' average transfer as a function of initial endowment, separately for the trust and for the risk games. Error bars report standard deviation.

RESULTS

A mixed design ANOVA is performed with group (vmPFC, non-FC, and HC) and Gender (Male, Female) as a between-subjects factors and kind of decisions as a within-subjects factor (investors' transfer with human, investor's transfer with machine, trustee transfer, subjective expectations). The ANOVA shows no significant main group factor [F (2, 24) = 1.70, p = .20]. Moreover there is no significant gender main factor [F(1, 24) = 1.15 p = .37]. On the contrary there is a significant main effect of Treatment, [F (3, 72) = 175, $p < .001$], and, interestingly, a significant Treatment by Group interaction,[F (3, 72) = 2.96, $p = .01$]. No other interactions are significant. The lack of a significant main group factor but a significant interaction between treatment and group strongly suggests that

there is not a generalized and constant difference in decisions between groups but a specific differenced related to particular decisions. The aim of the analysis that follows is to understand the difference and similarities between groups in the different roles and decisions.

The second analysis compared the level of trust and level of risk between groups. Figure 1.7 illustrates investors' average transfer as a function of initial endowment, separately for the trust and risk game. We performed a mixed design ANOVA on transfer amounts with Group (vmPFC, non-FC, and HC) as a between-subjects factor, and Treatment (human, and computer) and Endowment (€6, €9, and €12) as within-subjects factors[4]. The ANOVA showed a significant main effect of Group, [F (2, 27) = 9.41, p < .001], revealing that investors in the vmPFC group had overall significantly higher transfer levels (€5.7 out of a mean endowment of €9) than had investors in the HC (€4.3) and non-FC group (€4.2; both *ps* < .001 pairwise comparisons Fisher LSD test). There was also a significant main effect of Treatment, [F(1, 27) = 7.69, p < .01], indicating slightly higher transfers in the computer (€5) than in the human (€4.5) interaction, and a significant main effect of Endowment, [F(2, 54) = 108.07, p < .001], demonstrating that investors' transfer was modulated by initial endowment available.

More critically, analysis showed a significant Treatment by Group interaction, [F (2, 27) = 4.04, p = .02], indicating that the between-group differences in amount sent depended on the human *vs.* computer interaction. Pairwise comparisons showed that when participants played against a human partner, the average transfer was significantly higher in the vmPFC group (€5.8) than in both non-FC (€3.7) and HC group (€3.9; both *ps* < .05) [5], while transfers of the control groups did not differ (endowment €6: p= .10, endowment €9: p = .47; endowment €12 p = .35). When participants played against a computerized partner, only in trials with the smallest endowment there was a significant difference between vmPFC and other groups (non-FC p < .01, HC p = .03), whereas with greater

[4]. The same ANOVA with the gender as a between subject factor showed no significant effect of gender factor (p = .12), and no other and significant interaction of gender factor with other factors: group*gender (p = .78); treatments*gender (p = .48); endowments*gender (p = .72); group*gender*treatments (p = .88); group*gender*endowments (p = .73); gender* endowments*treatments (p = .67); group*gender* endowments*treatments (p =.65).

[5]. vmPFC vs non-FC: Trust Game, endowment €6: p= .02, endowment €9: p = .001; endowment €12 p = .005. VmPFC vs HC: Trust Game; endowment €6: p= .001, endowment €9: p = .01; endowment €12 p = .009.

amount (€9, €12) there was no significant difference (endowment €9, non-FC p = .68; HC p = .50; endowment €12 non-FC p = .20; HC p = .25, no difference between non-FC and HC all: endowment €6, p = .59; endowment €9 p = .79; endowment €12 p = .90.

In order to better understand the difference between the risk and trust games the average transfer was calculated from all initial endowments, then the same ANOVA described before with treatment (risk and trust game) as within-subjects factor and group (vmPFC, non-FC, and HC) as between-subjects factors was applied. Results about group and treatment factors were the same as shown in the previous ANOVA. The post hoc analysis (Fisher LSD test) revealed that there was no difference between risk and trust treatment for the vmPFC group (p = .48), whereas there was a significant difference between treatments both for nonFC (p = .02) and HC (p = .009) subjects.

A similar pattern of results was found when the data were analyzed using nonparametric methods. The Kruskal-Wallis test showed a significant difference amongst the 3 groups in the trust game (H = 12.8, df = 2, p < .002), but no difference in the risk game (H = 4.78, df = 2, p = .09). These data confirmed the non significant difference between controls and vmPFC in risk propensity, which was also confirmed by the Crawford's modified t test (table 1.3) analysis. Out of 10 subjects in each group, 8 vmPFC patients showed mean transfer levels higher than 50% of initial endowment in the trust game, whereas only 3 non-FC patients and 4 healthy controls displayed such transfers in the trust game. Conversely, in the risk game, 9 vmPFC patients, 7 non-FC patients and 7 healthy controls displayed mean transfers higher than 50% of initial amount.

Thus, results suggest that vmPFC damage leads to a substantial increase in transfer levels in the trust experiment but not in the risk experiment. Remarkably, following vmPFC damage, investors' transfers were not modulated at all by the type of opponent player present in the environment (€5.82 and €5.53, for the trust and risk game, respectively, p = .48). In sharp contrast**,** both control participants were more reluctant to invest in the trust game (€3.71 and €3.88, for non-FC and HC group, respectively), in which interpersonal interactions determines the risk, than in the risk game (€4.69 and €4.74; p = .04, and p = .01, for non-FC and HC group, respectively), in which a non-social, random mechanism constitutes the risk. This latter result is highly consistent with previous literature in healthy subjects (see Bohnet et al., 2008; de Quervain et al., 2004; Aimone and

Houser, 2008, Houser et al., 2009) suggesting that the prospect of betrayal plays a role in trusting decisions well beyond aversion towards monetary loss.

Next, we performed an analysis to explore whether vmPFC patients differed from control groups in their subjective expectations about trustee back transfers in the trust game (figure 1.8). An ANOVA, with Group (vmPFC, non-FC, and HC) as a between-subjects factor, and Endowment (€6, €9, and €12) as a within-subjects factor is performed. Results revealed a significant main effect of Endowment, $F(2, 54) = 28.26$, $p < .001$. More importantly, however, there was no main effect of Group ($F < 1$), nor any interaction between Group and Endowment ($F < 1$), revealing that the three groups of participants believed to obtain on average the same return for their money transferred as investor[6]. Thus, results suggest that damage to the vmPFC increases trusting behavior but does not significantly alter subjects' beliefs about others' trustworthiness.

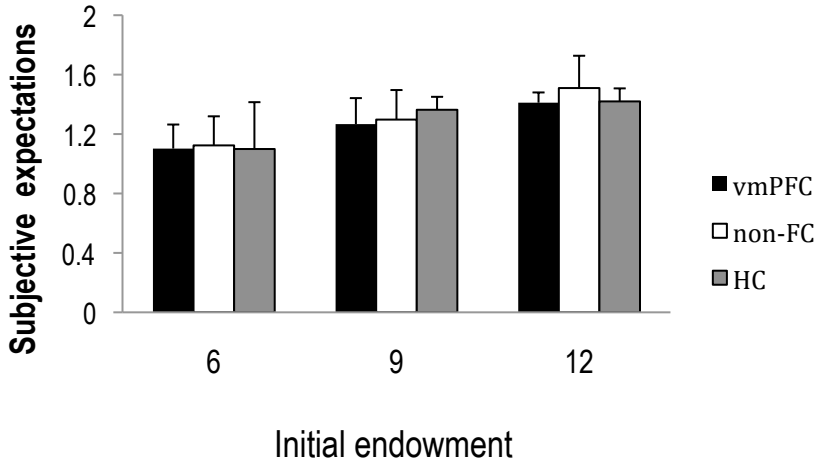


Figure 1.8 Subjective expectation value was conducted on expected back transfers divided by the amount sent (a value > 1 indicates expected gain, whereas a value < 1 indicates expected loss from the exchange). Error bars report standard deviation.

We next tested whether trustees' repayments to their investor in the trust game differed across the three groups of participants (Figure 1.9). A one-way ANOVA on trustees' average back transfers showed a marginally significant effect of Group, *[*$F(2, 27) = 3.06$, $p = .06$]. Pairwise comparisons revealed that vmPFC trustees made significantly lower back transfers than HC trustees (mean back transfer: €4.12 and €5.72, for the vmPFC and HC group, respectively, $p = .02$). The non-FC group (mean back transfer:

[6]The same ANOVA with the gender as a between subject factor showed no significant effect of gender factor (p = .21) and no significant interaction between gender, group and endowment: gender*endowment (p =.24); gender*group (p = .08); gender*group*endowment (p =.21).

€5.06) was not significantly different from the vmPFC (p = .16) or HC (p = .32) groups, possibly due to higher variance in performance observed in this group. Thus, results indicate that individuals with vmPFC damage do not show more trustworthy or altruistic behavior than control groups.

We further tested whether trustees' repayments to their investor in the trust game differed across the gender. The one-way ANOVA on trustees' average back transfers showed no general effect of Gender, *[F* (5, 24) =1.55, p = .21]. Investigating the gender effect on trustees' repayments for each groups (one –way ANOVA), no significant effect was found: vmPFC, p = .98; nonFC, p = .77; HC, p = .74.

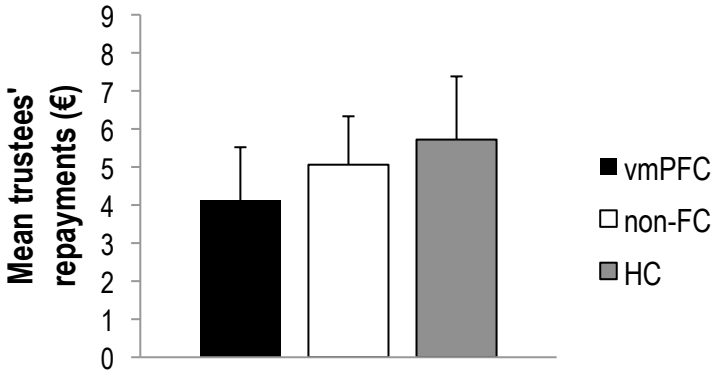


Figure 1.9 Mean trustees' repayment to their investor in the trust game. Error bars report standard deviation.

In the next analysis we propose a comparison of a single-case's score to scores obtained in a control sample[7], as suggested using the classical methods of Crawford and Garthwaite (2007). Crawford's modified t test (table 1.3) analysis showed that 6/10 vmPFC patients choose significantly different amounts from controls in the trust game, whereas only 1/10 patients has significantly different behavior in the risk game. There was no difference in back transfers between patients and controls when they acted as trustees, or when they were questioned about their beliefs about likely back transfers.

All vmPFC patients had lesions involving BA 10 and 11, and the lesions were quite homogeneous; thus preventing the possibility of relating performance to specific area

[7] The control sample in this case regards both HC and nonFC patients.

lesioned in the vmPFC patients. Trying to link performance with a specific lesion was only possible in relation to the side of the vmPFC damage. Subjects with a left-sided lesion (only 2 patients) showed significant impairment both in the trust and in the risk games, whereas subjects with bilateral or right-sided lesion showed impairment only in the trust game. Left-sided damage is associated with more impulsive behavior (Goyer et al 1994), and thus could cause a general impairment in decision-making; however this was not supported by the self report questionnaire regarding impulsivity of these two patients.

Personality questionnaires

Table 1.2 shows self-report measures of impulsivity, trust, and reciprocity for all three groups of subjects. There were no statistical differences across the three experimental groups on either Barratt Impulsiveness Scale (BIS-11) scores (Kruskal-Wallis test, $H = .83$, $df = 2$, $p = .65$)[8], or Interpersonal Trust Scale scores (Kruskal-Wallis test, $H = 4.07$, $df = 2$, $p = .09$)[9]. Likewise, we found no significant differences amongst the three groups in either positive reciprocity scores (Kruskal-Wallis test, $H = 2.09$, $df = 2$, $p = .35$), negative reciprocity scores (Kruskal-Wallis test, $H = .83$, $df = 2$, $p = .65$) or beliefs of reciprocity scores (Kruskal-Wallis test, $H = .75$, $df = 2$, $p = .69$) for the Personal Norm of Reciprocity (PNR) scale[10].

DISCUSSION

The present study was designed to understand whether emotions generated by the vmPFC play a role in the decision to trust a stranger. Like moral choice, trust involves a decision that incorporates a dilemma: if the investor trusts and her partner reciprocates, the investor can increases her payoffs. However, she is also subject to the risk that the trustee will abuse this trust. In the latter case, the investor is worse off than if she had not trusted at all and, furthermore, the trustee has an unfair payoff advantage relative to the investor. Substantial evidence exists to show that humans are averse to such risks (Bohnet & Zeckhauser, 2004: Holt & Laury, 2002; Fehr & Schmidt, 1999). Investors have to

[8] For each groups no statistical differences across gender: vmPFC p=.38; nonFC p = .82; HC p = .28.

[9] For each groups no statistical differences across gender: vmPFC p = .40; nonFC p = .83; HC p = .73.

[10] In each groups no statistical differences across gender in all three subscale (all p >.12).

overcome their negative emotion or aversion against these risks in order to trust, which allowed us to examine whether vmPFC, a brain region necessary for the normal generation of emotions and, in particular, social emotions, may play a critical role in trusting behaviour in humans. Previous findings have shown that activity in the vmPFC may be critical for making predictions and for anticipating negative emotions and the unpleasant state of loss in decision-making (Damasio, 1994; Bechara et al., 1997), especially in social contexts, when the implications of another individual' intentions must be taken into account before choosing (Rudebeck et al., 2008; Behrens et al., 2009; Moretti et al., 2009). In the present study, we found that investors in the vmPFC group showed higher money transfers to their partners than those in both control groups, thereby suggesting that damage to this brain area increases investors' trusting behaviour considerably.

First of all, we need to exclude other factors that could generate the pattern of choice shown by vmPFC patients in the trust game. Our findings revealed a significant endowment factor, that is a systematic modulation of investment by initial endowment. All subjects, either vmPFC and controls subjects, invested more when they were initially endowed with more money. Like control subjects, vmpFC patients were able to adapt their investment to the initial endowment available, showing normal cognitive ability to change their investment in relation to money available. Therefore, it appears that for vmPFC patients trust decisions cannot be accounted for in term of perseveration, since perseveration is the tendency to continue or repeat a previously rewarded act or activity even when no longer appropriate (Wallis, 2007). The fact that patients flexibly adapt (e.g., change) their investment and repayment in relation to the money they have been offered at the beginning of each game works against the hypothesis that they are simply repeating the same investment or repayment in every transaction. Furthermore, the absence of feedback (e.g., gain or loss) at the end of each trial, also helps to avoid vmPFC patients perseverating in a previously rewarded action.

Furthermore, data indicates a more complex pattern of decisions than that predicted by a simple environmental dependency syndrome in vmPFC patients. This hypothesis would suggest that vmPFC patients rely on environmental cues to make their choice, whatever the 'social' circumstances of their decision making context. That is, the more money vmPFC patients are given (e.g., environmental cues) the more they send to their anonymous (computer or human) partner. Firstly, I found that vmPFC sent more money

than controls when they acted as investor (increased trust or risk), but sent less or similar amounts of money than control groups when they played as trustee. This clearly indicates that they are not only relying on the money they are given to make their choice (investment or repayment), but are also sensitive to the specific context and role they play in the game (investor vs. trustee) to make their decision. These findings are in line with recent studies involving the Ultimatum game (Koenigs and Tranel, 2007; Moretti et al, 2009) showing that vmPFC patients' rejection of unfair offers, although higher than control participants, was modulated by the amount of money received from another during the game. Secondly, the difference between vmPFC patients and controls subjects in the risk game was significant only when subjects were playing with the lowest sum (6€). More specifically, vmPFC patients invested more than control subjects when they had a lower amount available, but not when they had a greater amount available (9€ and 12€).

Another possibility is that vmPFC patients suffer from a general emotional disinhibition. If emotional disinhibition of vmPFC could be a plausible explanation for the greater level of trust and risk with low endowment observed in vmPFC, the same explanation does not fit with the data regarding the risk choice with medium (9€) and high endowment (12€). In case of a pure emotional disinhibition with greater endowment the level of risk in vmPFC should be even greater and not at normal level.

Apparently, the normal level of risk showed by vmPFC patients in the risk game is in contradiction with their tendency to have a greater level of risk than normal found by Bechara and colleagues (1994) using the Iowa Gambling task. However, Fellows and Farah (2005) in their "shuffled" version of the gambling task[11], which used reward contingencies that did not initially bias the subject toward any of the decks, found that patients with vmPFC damage were not impaired on this version of the task. The results suggested that the deficits on the gambling task might have arisen from the problems that these patients have in reversing stimulus-reward associations. vmPFC more than risky behaviour showed inability in modifying their behaviour in response to negative feedback (perseveration). However in our tasks there was no negative or positive feedback on which it would be possible to balance the goodness of strategy of choice; on the contrary the task measures a pure internal and general disposition to take risks in an ambiguous situation.

[11] See description of Iowa Gambling task (Bechara et al. 1994) in chapter II paragraph 2.5.

The lack of feedback and the measurement of the initial and not biased disposition to risk showed that these patients are not different from controls in risk game.

What mechanisms might be involved in generating the effect of vmPFC damage on trusting behaviour? One possibility could be that vmPFC damage causes a general increase in altruism and prosocial inclinations. On this account, vmPFC damage should affect not only the prosocial behaviour of the investors but also that of the trustees. However, the data concerning the trustees' repayments to their investors in the trust game failed to show in the vmPFC group more trustworthy or altruistic behavior than control groups. On the contrary, data showed reduced generosity in the trustees' repayment in the vmPFC than in the control groups, thereby indicating that effect of vmPFC damage on trust is not caused by increased generosity or inclination to behave prosocially. There are two opposite patterns of behaviour for vmPFC subjects. In the trust game they showed a greater inclination to behave prosocially, whereas when taking on the trustee role they behave antisocially. In the case of the trust game, the main component is the social risk, whereas in the case of the trustee the motivational component is that moral rules 'not betray the trust received'. Their 'antisocial behavior' is consistent with a recent neuropsychological study (Krajbich et al. 2009) demonstrating that vmPFC damage significantly reduces trustworthiness, possibly due to impaired sense of guilt, a sociomoral emotion that plays a critical role also in moral decisions.

Another possible mechanism behind the effect of vmPFC on trust is that damage to this region alters patients' subjective expectations about others' trustworthiness or positive reciprocity. In other words, lesions to the vmPFC may render patients more optimistic about the probability of a good return from the investment. However, results showed these expectations do not differ significantly between vmPFC and control groups, therefore ruling out the possibility that vmPFC patients show more trusting behaviours because of unusual beliefs about the other players. Furthermore, also self-report measures of trust (Interpersonal Trust Scale, Rotter, 1967) and reciprocity (Personal Norm of Reciprocity (PNR), Perugini et al., 2003) indicate that vmPFC patients and control groups hold similar beliefs about others' trustworthiness and reciprocity. That is, when vmPFC subjects are involved in abstract questions concerning their level of trust or reciprocity they are able to answer not differently from controls. This finding is perfectly consistent with results from several other studies (Koenigs et al., 2007; Moretti et al., 2009; Krajbich et al., 2009)

showing that an explicit knowledge of social rules, as well as expectations and beliefs, remains intact and normally accessible following vmPFC damage. Despite this retained knowledge, however, vmPFC patients fail in valuing social information in social interaction and decision-making (Damasio, 1994).

A critical finding of this study emerges when comparing mean investors' transfer in the trust and risk games across the three groups of participants. We found, that following vmPFC damage, patients showed higher and similar investments in both games. That is, wmPFC patients did not distinguish between interactions with an intentional agent and those with a computer program that randomly generated outcomes. In striking contrast, control participants were less likely to invest when they believed that they were interacting with people than a computer opponent (Bohnet & Zeckhauser, 2004; Houser et al., 2009), revealing that normal economic decisions are driven by factors beyond mere probability, and that "people care not only about the payoff outcome but also about how the outcome came to be" (Bohnet & Zeckhauser, 2004). Accordingly, trust decisions, relative to risk decisions, entail additional costs, costs shown to be above and beyond mere monetary losses, which diverse authors (Bohnet & Zeckhauser, 2004; Bohnet et al., 2008; Houser et al., 2009; Fehr, 2009) have explained as due to betrayal aversion, namely, the fear to be exploited by others in social interactions. Here, we suggest that, after vmPFC damage, people lack such exploitation aversion, due to impaired social emotions, which makes them more willing to take risks arising from interpersonal exchanges. Concerns about 'others' do not matter for vmPFC patients, so that they perceive the decision of whether or not to trust basically as a risky choice and decide based on their expectations of trustworthiness and their propensity to risk. That is, it does not matter whether the risk is constituted through the uncertain behavior by the trustee, or through a random mechanism. In this sense, vmPFC patients behave more "rationally" than control participants in our trust games: they only care about their own payoffs and are hardly betrayal averse, as predicted by the standard economic model. The present study showed how subjects with vmPFC lesion may take extreme positions in social relation (prosocial or antisocial) because they are unable to consider all possible implications and effects of their choice both from an economic and social/emotional point of view. Specifically, they seem unable to consider the emotions involved in social situations and use them for shaping their decision.

The greater level of trust in vmPFC patients could be related to their incapacity to consider negative anticipatory emotional responses related to trusting behaviour, specifically they could fail to anticipate in their decision process the value of negative emotional responses associated with the risk of betrayal. Obviously, vmPFC patients' neglect of potential betrayal and increased willingness to take social risk may invite exploitation and attract selfish actors, which may explain, in part, why their social and financial investment are bound to fail.

Recently, Jenkins and colleagues (2007), by using a repetition suppression paradigm, found that vmPFC fail to discriminate between self-referential thought and mentalizing about a similar other, suggesting that thinking about the mind of another person may rely importantly on reference to one's own mental characteristics. Our data showed that in the case of patients with vmPFC damage there is a complete dissociation between their beliefs about possible trustee repayment and their real repayment when they play in the role of trustee. They do not attribute to the other players their own strategy, suggesting a complete distinction between themselves and the anonymous player. Probably this non consideration of the mental and emotional processes of others reduces the factors that normally are taken into account when we make decisions.

In conclusion this data has shown that vmPFC, and in parallel emotions generated by this brain region, could have a critical role in trusting decisions and, in general, is essential for the normal evaluation of social stimuli during an economic exchange with another person. These findings are highly compatible with current theories maintaining that vmPFC is a critical neural substrate for forecasting the (positive and negative) emotional consequences of available options in order to guide future behaviuor, both in personal and societal decision-making (Bechara & Damasio, 2005). Finally, the reported findings provide evidence for theoretical approaches to social cognition and decision-making that emphasize the pivotal role of medial prefrontal cortex in the integration of multiple signals to generate adaptive behaviour (Montague & Berns, 2002).

This study provides an extensive investigation about trust behaviour in patients with vmPFC damage. Findings showed how these patients are vulnerable because they have distorted level of trust which induces them to expose their self in dangerous social and economic relations. On the other hand the data showed a lack of moral obligation in

case of social relations, which suggests the view that they can be potentially dangerous for others.

STUDY II

AFFECTIVE MODULATION OF SOCIAL DECISION-MAKING

The aim study II is to investigate whether emotional stimuli, both consciously and unconsciously perceived, can influence a fundamental social disposition such as trust. Before presenting the experimental study, a brief session describes how psychology has traditionally investigated visual awareness. The first tree paragraphs (2.1, 2.2, and 2.3) introduce important terms and methodological aspects that are used in the experimental study investigating visual awareness. In paragraph 2.4 some examples about the influence of emotions on choice behaviour are introduced; subsequently (paragraph 2.5) a study investigating how incidental emotions can affect the decision of trust is presented.

2.1 Visual Awareness

In everyday life, we believe that we are fully aware of what we see; however this is not completely true. Some recent studies show that visual information can be processed correctly and almost fully without any accompanying conscious experience. One example of this unawareness is the phenomenon called blindsight (de Gelder, 1999). This phenomenon becomes evident in neurological patients with circumscribed brain damage to the primary visual cortex, which renders them blind in the associated (contralesional) part of the visual field. Nevertheless, if a stimulus is rapidly moved, for example, across the blind field, some of these hemianopic patients are able to 'guess' the direction of the stimulus considerably better than chance. Other data on healthy subjects support the idea that visual information can correctly processed without any accompanying conscious experience. Several studies have demonstrated that we are aware of far less of the visual world than we realize (e.g. change blindness, Rensink et al., 1997; inattentional blindness, Mack & Rock, 1998). However, this information processing without awareness helps us in

programming actions such as reaching and grasping before we become aware of the stimuli that are eliciting these actions (Castiello et al., 1991). Not only, our actions can be influenced by the meanings of words that we are not aware of having seen (Marcel, 1983).

2.2 Automatic processing of emotional stimuli

Unconscious routes to action seem to apply particularly when responses need to be made with great rapidity. All of us have had the experience of reacting to something that scares us by jumping or running away, and become aware of this reaction only when we are far away from danger. These automatic reactions elicited by strong emotional stimuli offer a significant survival advantage because they permit immediate reactions without losing time in engaging costly and slow access to conscious experience.

Several studies support the notion that emotional processing can be largely automatic and take place both irrespective of the focus of attention and independent of visual awareness (see for a review, Pessoa, 2005). The emotional processing may be overriding (Globisch, 1999), it could interfere with the ongoing processing of other information (Hartikainen, 2004). Moreover, unconscious emotional stimuli can elicit basic affective reactions with both behavioral and physiological consequences without reaching the level of consciousness. Specific examples regarding the unconscious processing of emotional stimuli come from neuropsychological literature. Patient GY has right hemianopia caused by left occipital lobe damage. This patient is able to discriminate between emotional facial expressions presented in his blind hemifield (de Gelder, 1999) showing what is called: affective blindsight. This famous single case has been recently confirmed by a study conducted by Bertini and colleagues (2010) that observed affective blindsight for fearful faces in patients with hemianopia (oral communication at the Society of Italian Neuropsychology, 2010).

2.3 Subliminal stimuli: methodological problems

There are several challenges associated with the measurement of awareness (see Merikle, 1992; Greenwald et al., 1996). These include the duration of presentation of stimuli in order to be perceived subliminally, and the criteria used to determine whether a participant is aware or unaware of the stimuli. Öhman (2002) suggested that the presentation of stimuli for less than 40 ms is the right time for subliminal perception. In a

recent work, Pessoa and colleagues (2005) showed that subjects differ widely in their sensitivity to fearful faces. In their study 36% of participants were able to detect a fearful face presented for 33 ms. Remarkably, some subjects could even detect fearful faces that were shown for 17 ms before masking. Some authors have suggested that these differences could be related to one's anxiety level (Etkin al., 2004), with more anxious subjects able to recognize fear even when presented for vey short time.

The second difficulty in measuring awareness regards the criteria used to determine whether a participant is aware or unaware of a stimulus. According to 'objective' criteria, unaware perception occurs when a subject's performance in a 'forced-choice' task is at chance, whereas in 'subjective' criteria, unaware perception occurs when subjects report that they are unable to perform the task better than chance (independent of their actual objective performance). An objective forced choice task remains the "gold standard" for the definition of awareness in behavioral psychology, although not all studies employ such a method.

2.4 Example about the influence of emotions on choices.

Although emotions are a constant presence in our lives, relatively little is known about the role that emotions play in decision processes in social contexts. As introduced in chapter II, emotions are conceived as discrete responses to an external or internal event that entails a range of synchronized features, including subjective experience, expression, bodily response, and action tendencies (Phelps, 2009). Some recent research has shown that emotional stimuli can influence behavior such as: our propensity to consume goods (Winkielman et al., 2005), the speed of our gait (Bargh et al., 1996), one's level of patience (Bargh et al., 1996), the willingness to punish someone (Carven's et al., 1984) and cooperative behavior (Moretti & di Pellegrino, 2010). Recently, Winkielman and colleagues (2005) investigated how subliminal expressions (happy versus angry faces) could influence the pouring and consumption of a beverage, and how subliminal expression could modulate the monetary value of a beverage. They used a modified version of the subliminal affective priming paradigm in which subjects first took part in an apparently unrelated gender classification task and then they performed a beverage task and a rating of the monetary value of the beverage. Before the gender task, a questionnaire measured their actual level of thirst and their general feeling. During the gender task, subliminal emotional faces (happy, angry, neutral) were presented (one emotion for each

block). Results showed that thirsty participants poured 114% more of the beverage after happy face primes than after angry face primes. Moderately thirsty participants poured 32% more of the beverage after happy than angry face primes, whereas priming did not influence pouring of non-thirsty participants. These results suggest that subliminal facial expressions alter beverage consumption depending on the individual level of thirst. Moreover, authors measured a more abstract value by asking to subjects to evaluate the economic value of a beverage. Participants had to fill a scale ranging from 10 cents to 1 dollar (U.S.) indicating their willingness to pay for a hypothetical can of the beverage. Participants were willing to pay 37 cents after happy primes and only 19 cents after angry primes. Remarkably, despite these changes in behavior and judgment, participants reported no change in their subjective state (mood, arousal) after the task.

Is it possible that subliminal emotional stimuli could influence more complex behavior such as socio-economic exchanges with others? There is growing evidence of automaticity in social psychological phenomena (Moretti & di Pellegrino, 2010; Dunn & Schweitzer, 2005; Bargh et al., 1996), however it is widely assumed, especially in social psychology, that behavioral responses to social environments are mostly under conscious control (Bargh, 1989). Social responses might well be consciously chosen on the basis of automatic reaction and feelings, but the ultimate behavioral decisions themselves are believed to be made consciously. On this point of view, Devine (1989) proposed a two-stage model of prejudice in which the perceptual phase is automatic (i.e., activation of stereotypes by the target person's features), whereas the second phase of prejudiced behavior is a matter of conscious choice, driven by one's relevant and consciously accessible values. The two-stage model of prejudice of Devine (1989) does not seem too far away from more recent neuroscientific positions that postulate that “rational decisions” in social and economic domain result from balancing two opposite processes: an affective/intuitive process and a controlled/deliberative process (Greene, 2007). The affective/intuitive process consists of emotion-laden processes that automatically evaluate socially relevant stimuli along a right-wrong or like-dislike dimension (Haidt, 2003; Fehr & Camerer, 2007). The deliberative process consists of highly controlled processes that arrive at social judgement or decision through laborious steps of deductive reasoning and cost-benefit analysis and optimization. The areas of the brain associated with problem solving and deliberate reasoning include the dorsolateral prefrontal cortex (dlPFC) and the inferior parietal lobule, while areas that have been implicated in emotion/intuitive

processing and social cognition are the medial and ventromedial prefrontal cortex (vmPFC) and the posterior cingulate gyrus (Greene, 2007).

Some studies (Carven et al., 1984; Moretti & di Pellegrino. 2010; Dunn & Schweitzer, 2005; Bargh et al., 1996) have tried to investigate how incidental emotions affect social behavior. In the influential study by Carven et al. (1984), a concept of hostility/anger was primed subliminally in a group of participants, whereas a second group was exposed to neutral priming stimuli. Participants, in what they believed to be an unrelated second experiment, were instructed to give shocks to a "learner" participant (actually a confederate) whenever he or she gave an incorrect answer. Compared to participants who were exposed to neutral priming stimuli, those presented subliminally with hostility/anger-related primes gave longer shocks. On the same line, Bargh and colleagues (1996) observed that participants primed with rudeness-related stimuli in an ostensibly unrelated first experiment interrupted a conversation reliably faster and more frequently than did other participants exposed to a concept of politeness. More recently Moretti and di Pellegrino (2010) investigated how an emotion such as sadness and disgust could modulate the level of cooperation in the ultimatum game. Three different groups were initially exposed to three different kinds of social pictures (neutral, disgust, sadness). Subsequently, subjects in a second apparently nonrelated task played several rounds of the ultimatum game with anonymous counterparts. Results showed that subjects exposed to neutral and sad pictures showed the same level of acceptance of either, fair and unfair offers. Differently, subjects exposed to disgusting pictures were less willing to accept unfair offers than other subjects.

Previous literature (Winkielman et al., 2005; Bargh et al., 1996; Carven et al., 1984) suggests that subliminal emotional stimuli can modulate not only simple perceptual task but also more complex social behavior. Here, I report an investigation aimed at assessing whether emotional stimuli can modulate our disposition to trust someone. The next paragraph reports how I addressed this question.

2.5 Emotion and Trust

One fundamental component of our social behaviour is trust. Trust is an essential ingredient of human exchange (Arrow, 1974); it promotes social and economic

transactions, and has been long recognized as a critical antecedent of cooperative behaviour (Ostrom & Walker, 2002). An operational definition suggests that trust involves a voluntary transfer of resources (physical, financial, intellectual, or temporal) from the investor to the trustee with no real commitment from the trustee (Coleman, 1994).

The decision of trust is based on opposite expectations/anticipation regarding trustee decision. These opposite expectations have associated emotional reactions. Accordingly, the decision to trust a stranger concerns two different levels of consideration: one economic and the other social. With regard to the economic point of view, there is the possibility to have greater mutual benefit from cooperation (greater income in case of trust game, see study I), but at the same time there is a negative possibility regarding betrayal, that in case of trust games means losing money. On the social point of view, there is the chance to send an encouraging signal to start a comfortable and profitable relation and also to build a positive reputation. On the other hand, however, there is the possibility of being betrayed and thus humiliated by a selfish other. Even if there is new interest about trust behaviour, really little data exists on how emotions influence trust behavior. If the decision to trust is based on opposite expectations with opposite emotional reactions, it is possible to hypothesize that incidental emotions may unbalance this equilibrium of opposite emotional reactions present at the time of a trusting decision. In a recent study, Dunn & Schweitzer (2005), by using an apparently separated task, induced in participants specific emotional states (angry, sadness, happiness). Then, subjects had to fulfill a questionnaire regarding trust. They found that incidental emotions significantly influence subjective reports of trust in unrelated settings. Specifically participants in a happy emotional state showed a significantly greater number of trusting judgments than participants in other emotional states. A further interesting result reported by Dunn & Schweitzer (2005) concerns that emotions do not influence trust when individuals are aware of the source of these emotions.

The present study seeks to investigate if subliminal emotional stimuli could incidentally modulate a real, on-line trust decision. To my knowledge, the following study is the first attempt to investigate how incidental emotion can influence real, on-line decisions to trust a stranger. In this study, emotions are introduced into a real dyadic social interaction requiring trust: such as the case of the trust game. The use of the trust game paradigm is particularly relevant because subjects do not make theoretical or abstract

judgments about trust, but this game offers the possibility to measure trust by means of a more realistic interaction between two people. In the trust game, already introduced in study I, two players interact in an anonymous manner. At the beginning, each player is endowed with the same amount of money. The first mover, the investor, has to decide how much of her initial endowment she wants to transfer (e.g., invest) to the second player, the trustee. Any amount sent by the investor is tripled by the experimenter and given to the trustee; at this point the trustee has to decide which fraction of money received she wants to return to the investor. The amount of money that the investor decides to give to the trustee is a measure of trust, whereas the amount that the trustee sends back to the investor is a measure of reciprocity. Two different emotions were further introduced during the trust game: happiness and fear. These emotions were introduced by presenting during the game a human face with one of two emotional expressions. In a social context, indeed, the emotional expressions of faces are an important and immediate source of information that may promote or discourage certain behavior. The two emotions introduced in the game appear congruent with the two opposite expectations. A happy face could emphasize the positive expectations of a mutual cooperation and greater benefit (both social and economic) for investor and trustee. A face with a fearful expression, on the other hand, could emphasize the fear of being betrayed and losing one's own money. The emotional expression of others, such as for example disgust, seems less relevant to the emotional anticipation involved in the decision to trust a stranger. I prefer to investigate the effect of basic emotion on trust decision as opposed to non-basic emotions, because in my knowledge there is no data about the effect of basic emotions on trust decisions in real interactions. The effect of basic emotion thus could represent a basic and useful comparison in a further study involving more complex emotions. From the methodological point of view, I believe that introducing non-basic emotions (for example guilt or regret) inevitably requires changes in method and task.

We predicted that happy faces could evoke a greater level of trust, whereas fearful faces could reduce the level of trust. In order to understand if the level of trust is modulated not only by emotional content of the face but also by the level of awareness with which these emotional stimuli are perceived, half of the participants were exposed to subliminal emotional faces, and the other half were exposed to supraliminal (e.g., visible) emotional faces. In this case we predicted in accordance with Dunn & Schweitzer (2005)

that the effect of incidental emotions could be greater when subjects were not aware of emotional stimuli.

METHOD

Participants

One hundred and seven healthy subjects (35 males) took part in this study. Their ages ranged from 19 to 32 with a mean of 24.7 (*SD* = 4.09). Participants were students from the University of Bologna, all blind regarding the nature of the experiments. Informed consent was obtained from each subject prior to commencing the experiment. None of the participants reported neurological or psychiatric disorders. The experiment was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki, and was approved by the Ethical Committee of the Department of Psychology, University of Bologna.

Procedure

The experiment was run in individual sessions. On entering the laboratory, participants were greeted by an experimenter who dealt with informed consent and who collected some demographic information. Then, the experimenter informed them that they would participate in two separate studies: one concerning economic decisions with real monetary payoffs, and a second one concerning the perception and attention of briefly presented stimuli.

All participants were presented with the same sequence of tasks: i) three rounds of the trust game, in which they played the role of investor, ii) a discrimination task and iii) questionnaires and payment. During both the trust game and the discrimination task the face of the same man (Figure 2.1 from Ekman & Friesen, 1976) was shown with different emotional expression: neutral, fearful, happy. Before participants left the laboratory, they were closely questioned for suspicion. Stepwise debriefing revealed that no participant guessed the actual purpose of the study.

Trust Game

The trust game session took place in a quiet room in which an opaque, removable partition wall was used to create two separate settings. On either side of the wall, we placed a desk with a computer. Participants sat at one desk in front of the computer, while at the other sat an actor played the role of trustee. As a result, playing partners could be separated visually, thereby providing between-subject anonymity, without separating them audibly, thus lending our set-up credibility. The instructions about the nature and rules of the game were presented on the computer, and the experimenter verbalised them to ensure that participants understood them. In the instructions, it was emphasized that participants in the trust game would play the game anonymously and only once with each opponent player, and that they would receive the money earned in the game. Moreover instructions mentioned that the presentation of a face had the aim of reminding the participant that they were interact with another person. After reading the instructions, subjects were required to complete a quiz that required them to state the amount of money that each player would receive under various hypothetical circumstances. The game started once the subject successfully finished the quiz.

Participants acted as investor in a series of 3 rounds of a trust game against 3 different anonymous human partners via a computer interface. At the beginning of each round, the actor that played the role of the trustee entered the room and sat at her position. When both investor and trustee were ready, the interaction started. Each round was presented as text through a series of screens (see figure 2.1 for a schematic illustration of a typical round). A 6-s initial screen indicated the endowment (E) available for both players in the current round. All three games presented an endowment of 9€. The second screen presented a neutral face or an emotional face: fearful or happy. In the subliminal condition, the face (neutral, fearful or happy) was presented for 25 ms followed by a neutral face turned upside down for 1,000 ms. In the supraliminal condition, the first face (neutral, fearful or happy) was presented for 1,000 ms followed by a neutral face turned upside down and presented for 25 ms. The next screen posed the question “How many Euros between 0€ and 9€ do you transfer to the other participant?” and remained visible until a response was given. Participants were given the opportunity to send any integer amount from zero to their entire endowment available, and were instructed to indicate their decision by pressing the numeric keys of the computer keyboard. Following the response, a screen indicating the investor’s transfer and the amount received by the trustee (three times the amount invested) was presented for 4 s. Then, a variable 5- to 15-s waiting screen

informed that the trustee was deciding how much of the tripled amount to send back. Subjects were informed that the other participant could choose the amount from any integer between zero and the tripled amount they have transferred to her/him. Finally, a screen signalled the end of the round. The trustee went out of the room and after a short break was replaced by another actor to begin the next round. Subjects received no feedback between games; only at the end of the entire experimental session subjects could see a screen with overall money earned by the subject during all three interactions and receive the payment.

The expression of face (neutral, fearful, happy) presented was manipulated between groups. Subjects viewed the same emotional face for all three games immediately before deciding the amount of money to transfer to their opponent player. The level of awareness in perceiving the emotional faces was also manipulated between groups. For half of the subjects, the presentation of emotional faces was at subliminal level: an emotional face was presented for 25 ms and successively masked for 1000 ms with an inverted neutral face. For the other half of the participants, the presentation was supraliminal: an emotional face was presented for 1000 ms and successively masked by an inverted neutral face presented for 25 ms. Participants were randomly assigned to one of the following six conditions: subliminal-neutral, subliminal-fear, subliminal-happy, supraliminal-neutral, supraliminal-fear, supraliminal-happy (see table 2.1).

Discrimination task

Following the three rounds of the trust game, participants performed the discrimination task in order to check the level of accuracy in perceiving the emotional face subliminally presented (see figure 2.2 for a schematic presentation of the task). The instruction stressed that in each of 8 trials two faces were presented in rapid succession, and that subjects had to recognize the emotional expression of the first face presented in the sequence. Moreover, the instruction emphasized that the first face of each pair was presented for a short time, and therefore could be difficult to perceive. The task started with a screen presenting a fixation cross for a variable interval of time (75-1000 ms); the next screen presented an emotional face (either fearful or happy) for 25 ms and immediately masked with an inverted neutral face presented for 1000 ms. The next screen asked the subject to indicate if the first picture was a face expressing fear or happiness by pressing one of two keys. Half of the subjects were instructed to press the key ‘A’ when

they perceived a happy face, and the key 'F' when they perceived a fearful face. For the other half of participants, the instructions were the opposite. No feedback was provided during the task.

The discrimination task included 8 trials, 4 sequences of fear-neutral, and 4 sequences of happy-neutral, given in random order. In a pilot discrimination task (8 subjects) I required to subjects to discriminate emotional vs. non-emotional expression. However this task using the same face (emotional expression masked by neutral expression) was too easy for participants (89% of correct responses). In order to avoid easy recognition I proposed to subjects a more difficult judgment regarding discrimination between different emotions.

Questionnaires

At the end of the control tasks, three questionnaires on Trust (Rotter Interpersonal Trust Scale, Rotter, 1967), Reciprocity (Personal norm of Reciprocity PNR, Perugini, 2003), and Impulsiveness (Barratt Impulsivity Scale, BIS-11, Barratt, 1996), were administered. Overall feedback about subject's earnings was given at the end of the experiment.

To summarize, subjects were randomly assigned to one of 6 different conditions (2 perceptual conditions: subliminal-supraliminal; X 3 emotion conditions, neutral or fear or happy). Subjects took part first in 3 rounds as investor in the trust Game without receiving any feedback. Afterwards, subjects took part in a discrimination task and then they filled three questionnaires. At the end, participants could see the responses of the trustee and received their payment.

Design section

The variable measured is the level of trust. There are two factors: awareness (2 levels) and emotions (3 levels) see table 2.1. The design is a between groups comparison, see six different groups in table 2.1.

The analysis involves:

i) Trust level in first trials (one-way ANOVA analysis between the six groups)

ii) Mean of trust level across all three trials (one-way ANOVA analysis between the six groups)

iii) Level of accuracy in discrimination task (one-way ANOVA analysis between the six groups)

Economists (Berg et al., 1995) suggest that the first trial in the trust game is a more reliable measure of trust because there is no history and previous experience that can affect the decision to trust a stranger. Differently, subsequent decisions, even when there is no feedback, have some information, such as amount of money surely earned because not invested, that can affect in some way the second decision regarding whether or not to trust someone.

Groups	Awareness	Emotion
1	Subliminal	Neutral
2	Subliminal	Fear
3	Subliminal	Happy
4	Supraliminal	Neutral
5	Supraliminal	Fear
6	Supraliminal	Happy

Table 2.1. Groups

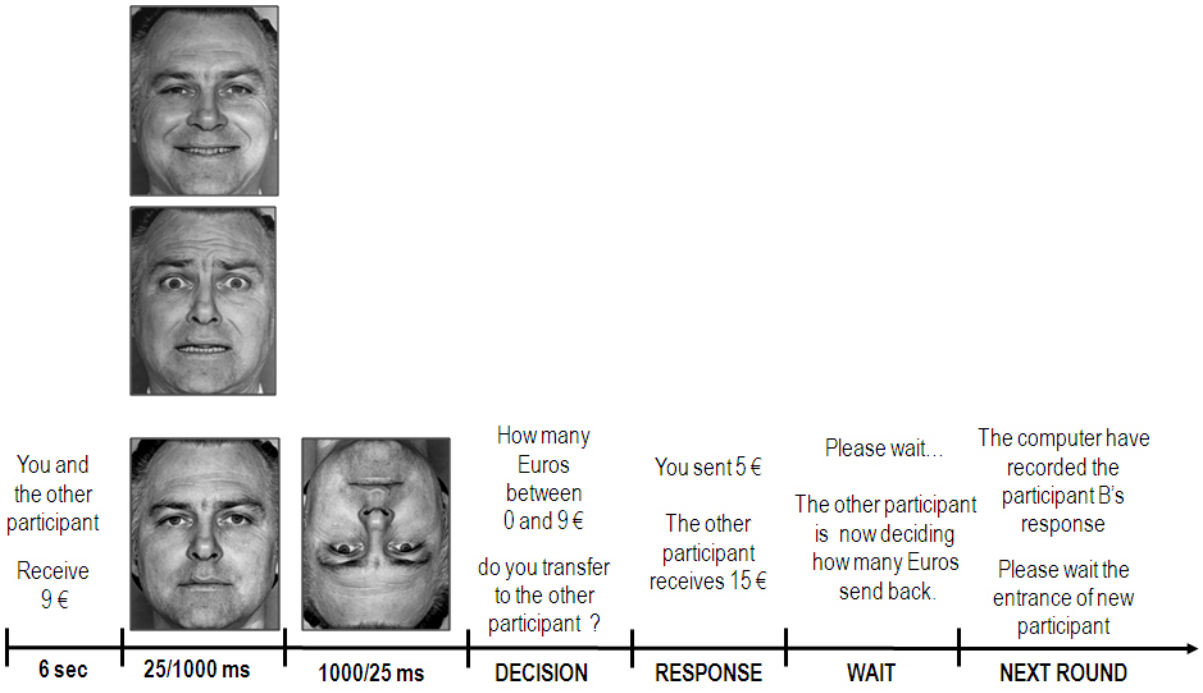


Figure 2.1. Example of 1 round of the trust game. In the subliminal condition, the first face (neutral, fearful, happy) was presented for 25 ms and followed by a reversed neutral face presented for 1000 ms. In the supraliminal condition, the first face was presented for 1000 ms and the following neutral face for 25 ms.

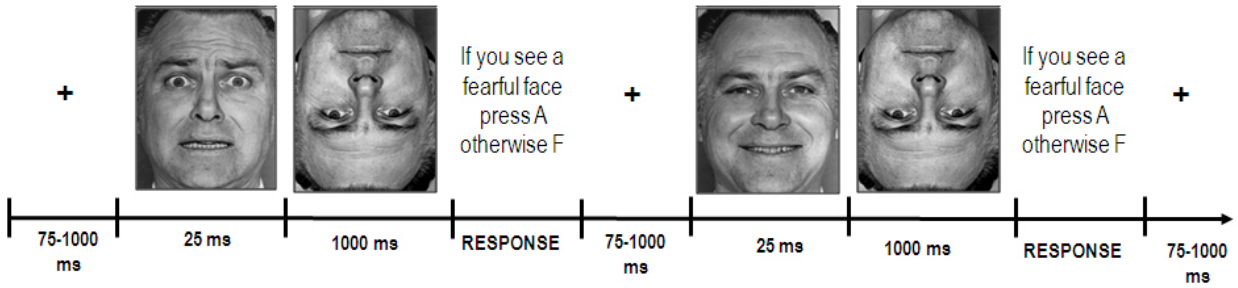


Figure 2.2. The figure presents two trials of the discrimination task, first sequence is fear-neutral, second sequence is happy-neutral.

RESULTS

Only subjects able to response correctly to less than 5 emotional expressions on 8 trials of the discrimination task were included in the study, except 4 subjects who recognised 5 expressions out of 8 but we could not exclude because this would have induced an unbalanced distribution of subjects between groups. As reported in table 2.2 below, 16 subjects recognized correctly more than 75% (≥ 6/8 correct responses) of the emotional expressions in the discrimination task. These subjects were excluded from the analysis because they were able to recognize the emotive expression. Table 2.2 reports the distribution of subjects excluded across conditions.

The remaining 91 subjects (32 males, mean age 24,0) recognized correctly 51% (SD: 16) of emotional expressions. This result represents the chance level showing that a subject failed to recognize at conscious level the emotional content of a face presented subliminally. One-way ANOVA analysis between six groups was performed on level of accuracy in recognizing the emotive expression. No difference in accuracy was found among all six groups: F (5, 85) = .85, p= .51.

Face	*Condition*	N. of subjects	Discrimination >5/8	Mean transfer in subjects with discrimination >5/8		Mean transfer in subjects with discrimination < 5/8	
Neutral	Subliminal	17	2	2,9		2,9*	
Fearful	Subliminal	19	4	3,5		3,4	
Happy	Subliminal	18	3	4,6		4,8	
Neutral	Supraliminal	16	0	3,8		3,7**	
Fearful	Supraliminal	19	4	3,7		3,7	
Happy	Supraliminal	18	3	3,8		3,8	
		107	**16**	**3,7**	**Mean**	**3,7**	**Mean**

Table 2.2. Table reports number of subjects able to recognize more than 5/8 of emotional expression in the discrimination task across the 6 different conditions of the experiment. The investors' average transfer is not different considering or excluding subjects with high rate of recognitions.
*1 subject recognized 5/8 expressions (mean of transfer in group without this subjects was 2,9)
** 3 subjects recognized 5/8 expressions (mean of transfer in group without these subjects was 3,8).
As reported in table the average of transfer considering subject recognizing 5/8 expressions does not differ from the average of transfer excluding subjects able to recognize less than 5/8 expression.

Figure 2.3 illustrates investors' average transfer in the first round of the trust game. In order to observe whether the content (neutral, fear, happy) of an emotional stimulus could affect trust behavior, one-way ANOVA analysis between six groups was performed on amounts invested in the first round of the trust game, separately in the subliminal and supraliminal conditions.

Investors in the subliminal-happy condition showed greater transfers (€4.6) than investors in subliminal-neutral (€3.2), and subliminal-fear conditions (€3.1). These differences between groups were statistically significant [F (2,42) = 3.75, p = .03]. T-test for independent samples revealed that investments in the subliminal-happy condition were significantly different from investments in the subliminal-fear [t(1, 28) = -2.32, p = .02], and the subliminal-neutral conditions [t(1, 28) = -2.21, p = .03], whereas there was no difference between subliminal-neutral and subliminal-fear conditions [t(1, 28) = -0.23, p = .81]. The analysis revealed that subliminal happy expressions could influence the amount of money that investors send to their trustees as opposed to fearful and neutral expressions.

In the supraliminal condition, the investors' average transfer in the first round was €3.6 when neutral and fear expressions were presented. When a supraliminal happy face was presented, the investors' average transfer was €3.8. There was no significant difference between groups in the supraliminal condition, F (2,43) = .37, p = .69.

The same pattern of results was observed when the mean transfer amount over the three rounds of the trust game was analyzed (see figure 2.4). In the subliminal condition, there was a significant difference between groups [F(2,42) = 4.30, p = .01]. Specifically the subliminal-happy condition was significantly different from both the subliminal-neutral [t(1, 28) = -2.50 p = .02] and the subliminal-fear [t(1, 28) = -2.43, p = .02], whereas there was no difference between these two latter conditions [t(1, 28) = .67, p = .50]. In the supraliminal condition, there was no difference between groups [F (2,43) = .00, p = .99].

To understand if the level of awareness in processing emotional expression can influence the transfer amount from investor to trustee, subliminal and supraliminal condition with the same emotion were compared. There was no difference between subliminal-fear (€3,2) and supraliminal-fear condition in the first round, [t(1, 28) = -.60, p = .59], or between subliminal-neutral and supraliminal-neutral [t(1, 28) = -.89, p = .37]. However, the subliminal-happy condition showed a greater transfer amount (€4.60) than supraliminal-happy (€3,13). This difference was statistically relevant: [t(1, 28) = 2.24, p = .03]. It must be noted, however, that when mean transfer amounts over the three rounds of the trust game were considered, the amount sent in the subliminal-happy condition was numerically greater (€4.7) than amount sent in the supraliminal-happy condition (€3.8), but the difference was not significant [t(1,28) =1.7, p = .10].

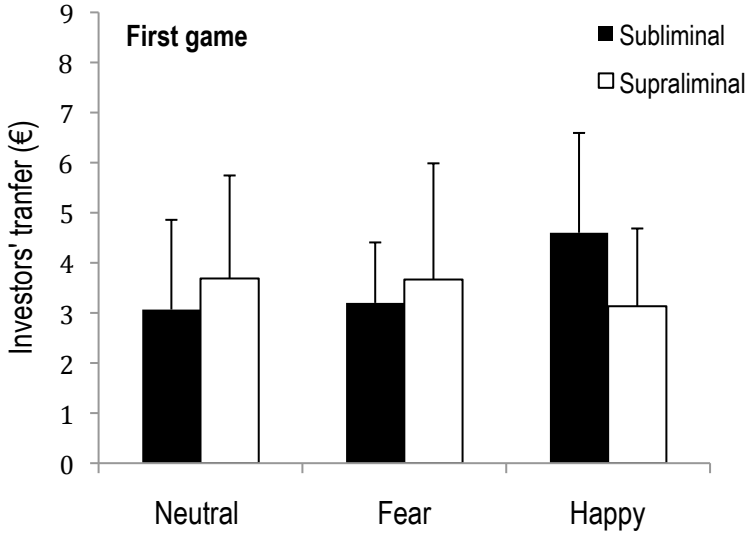


Figure 2.3. Investors' average transfer in first game, error bars refers to standard deviation

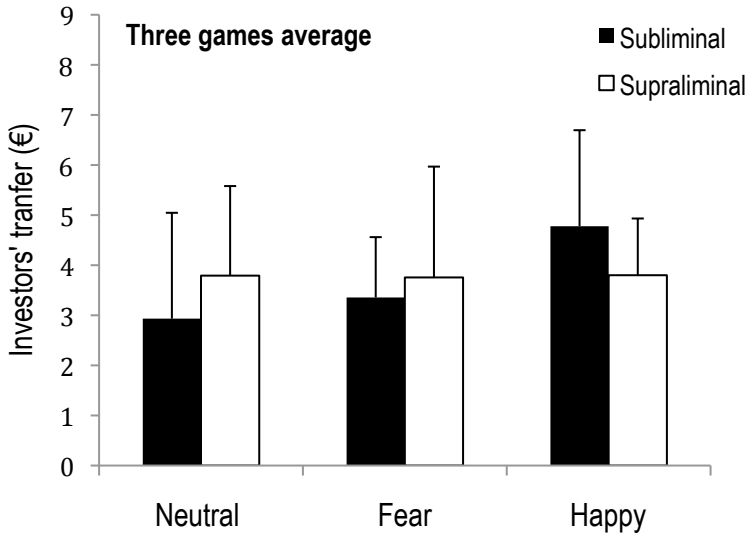


Figure 2.4. Investors' average transfer in all three games, error bars refers to standard deviation

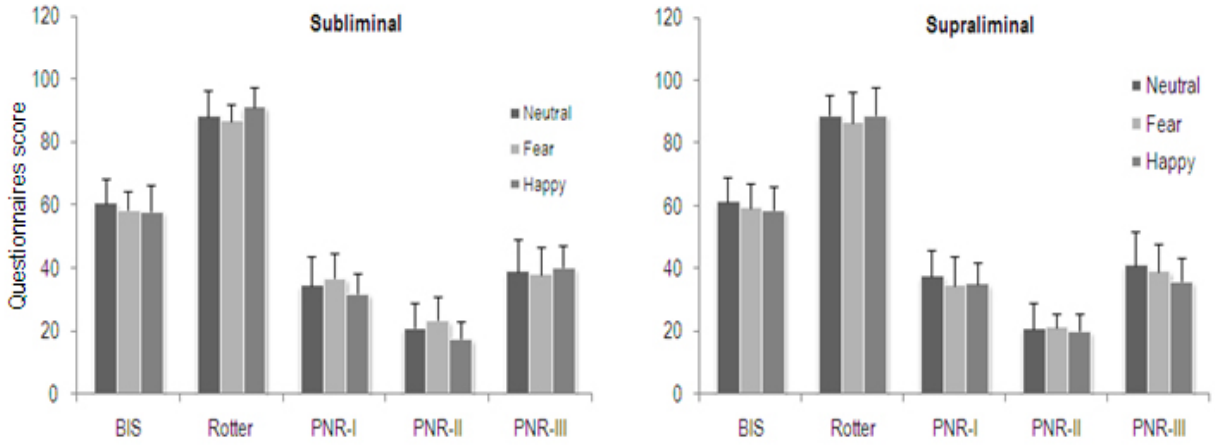


Figure 2.5. Self-report measures of impulsivity (Bis), trust (Rotter), and reciprocity (PNRI, II, III) between conditions.

Finally, results of the personality questionnaires (see figure 2.5) revealed no statistical differences across conditions for either propensity to reciprocate, trust or impulsivity, (all ps > .05), thus indicating that transfer amount differences across conditions cannot be attributed to pre-existing behavioral trait or social believes.

DISCUSSION ON METHOD

An important methodological limitation to this study concerns the use of one person's face only. There is good agreement among observers that some faces look more trustworthy than others (e.g. Winston et al. 2002; Todorov & Engell, 2008); to avoid any confounding effect regarding the trustworthiness of face, the same face was always presented for all participants.

Another limitation to this study is that we did not measure mood/emotion in our subjects after the tasks (e.g., emotion manipulation check). It is not possible to say whether some stimuli (e.g., fear) were ineffective on trusting behavior because they failed to affect participants' mood/emotion, or because they cannot penetrate trust decisions.

A further limitation concerns the between-subject design of the present study. In a previous pilot (30 participants), subjects were exposed to all three emotional expressions (neutral, fear, happy) in random order, thus creating a within-subjects design. However, average transfer amounts did not show any modulation by emotion cues. Only the first round of the game showed a significant difference between subjects. A possible interpretation of this null result is that the emotional activation is not an on-off process, but it shows carry-over effects from one trial to the next. Although a between groups design presents some limitations, it seems a better way to introduce specific incidental emotions into complex task concerning on-line social exchanges.

DISCUSSION

Trust requires one to balance economic and social motives, each with both positive and negative affective dimensions. The economic motive regards the possibility of gaining a higher payoff if the trustee shares, or to lose money if he does not share. The social motives concern the positive feeling associated with the possibility of a cooperative relationship, or the fear of being exploited by others in social context.

We have hypothesized that incidental emotions could modulate these motives that guide trust behavior. Moreover, the influence of incidental emotion could be greater when subjects are not aware of this emotional information. In accordance with Dunn &

Schweitzer (2005), our findings confirmed that incidental emotions can be effective in influencing trust. In particular, in our study, subliminal happy faces increased investors' transfer amounts relative to neutral faces, whereas fearful expressions failed to affect such amounts compared to neutral expressions. Studies investigating trust without any incidental emotional stimulation (Berg et al., 1994; McCabe et al., 2001) have shown that investors usually send to trustee approximately 30-40% of the money available. In the present study, participants exposed to fearful and neutral expression showed a level of trust similar to that reported in other studies without emotional stimulation. By contrast, subjects exposed to subliminal happy faces showed a greater level of trust.

A possible explanation of the null effect of fearful faces on trust behavior could concern a sort of 'floor effect'. In the trust game, risk aversion and betrayal aversion are powerful inhibitors of trust. As such, fearful faces cannot further reduce trust. Furthermore, extremely low level of trust in social exchanges may be not viewed as a cautious decision, but perceived as signs of hostility or punishment towards others. Therefore, it is possible that trust cannot be symmetrically influenced by positive and negative emotional stimuli, because in the anonymous one-shot interaction as those used here, trust is already very low.

The effect of happy face on trust level could occur because emotional primes temporarily changed the accessibility of knowledge relevant for interpreting the ambiguous situation (Higgins, 1996). In this sense the happy face could be used as a reinforcing stimulus supporting the positive components of trust. An alternative explanation could be related to general mood change due to presentation of emotional faces. This possibility, however, could not be tested because data concerning participants' mood before and after the task were not collected. However the measurement of mood before and after a task could be different not because participants saw emotional expressions, but because subjects were exposed to a social situation involving taking risks, such as that of sending money to a stranger for gain. Furthermore, the difference between priming of positive information and induction of positive is quite subtle.

A referee suggested a possible manipulation of facial appearance, or other characteristics of the 'trustee' face to increase the relevance of the mood manipulation. First we need to be sure that the findings are due to manipulation of accidental emotions,

and, second, manipulating the facial appearance of the trustee means changing the only information available about the trustee. In the present experiment, we controlled the role of facial appearance on trust by presenting always the same individual (a man), in order to measure level of trust independent from the facial appearance of the trustee. Manipulating the facial appearance of the trustee could help us to understand what is in the expression/face/parts of face that inspire more trust (see van 't Wout & Sanfey, 2008), but not whether emotions generated by the situation and/or incidental emotions are used in the decision to trust.

Another important result of this study concerns the fact that only unconsciously perceived happy faces were effective in modulating trust behavior. In accordance with Dunn & Schweitzer (2005), we found that subliminal happy faces increased trusting behaviour relative to subliminal neutral (and fearful) faces, but supraliminal happy faces failed to do so. It is possible to speculate that cognitive control mechanisms reduce the effects of supraliminal emotional stimulation, or that these control mechanisms select the relevant information and inhibit irrelevant information. This explanation is congruent with several fMRI studies (Hariri et al., 2003; Critchley, H. et al., 2000) that have shown that amygdala activation decreases when participants attend to faces in order to evaluate emotional features, relative to when participants make a non-emotional judgment of face gender, such that the emotion of the face is completely irrelevant to the subject's task. Interestingly, Hariri and colleagues (2003) showed that activation in the right prefrontal cortex (PFC) is negatively correlated with activity in the amygdala during conscious semantic processing of emotional stimuli. Hariri and colleagues (2003) suggested a subcortical, amygdala-based associative level of emotional processing that is subject to cognitive modulation by cortical networks involving the PFC. Thus, in case of supraliminal presentation of happy faces, cognitive control mechanisms operating on emotional processing can decrease the relevance of emotional information because this is not relevant for the task. By contrast, in case of subliminal happy faces, this information escapes control mechanisms because it is not consciously perceived, and therefore can reinforce those motives (e.g. pleasure of cooperation, expectation of a good outcome) that favour trusting behaviour.

Future studies could help us to understand if the influence of incidental emotions on trust is entirely dependent on the social aspect of the incidental stimulus (human face) or

can be dependent on more general meaning, for example positive or negative valence, irrespective of whether the incidental stimulus is social or not. Thus, might presenting a knife vs. a flower immediately before the decision of trust have the same effects as fearful and happy human expressions? Another interesting question regards how non-basic emotion can influence the decision to trust, and the difference between a basic and non-basic emotive effect on trust.

To conclude we examined how incidental emotion, both consciously and unconsciously perceived, could influence one's general willingness to rely on others in situations in which betrayal is possible, such as in situations involving trust. Our results indicate that, in one-shot anonymous interactions, positive emotions are more effective than negative emotions in modulating trust. These results provide important insight into the mechanics of trust, and identify incidental emotions as a robust and important determinant of trust. In many cases, emotions may play an important role in trust decisions precisely because people are unaware of the significant influence their emotional state has on their decisions.

STUDY III

MORAL JUDGMENT AFTER VENTROMEDIAL PREFRONTAL DAMAGE

3.1 Deliberative and intuitive process

For decades, moral psychology has been concerned with identifying a rational basis of human morality (Kohlberg, 1969; Piaget, 1965). A different and more recent approach, however, places strong emphasis on the causal power of affective and intuitive processes to drive our moral judgment and convictions (Haidt, 2001). Integrating these opposite views, recent work in psychology and neuroscience has suggested that moral judgments are mediated by two classes of computational processes (Greene et al., 2004; Greene, 2003; Greene & Haidt, 2002; Greene et al., 2001). One class, referred to as moral intuition, consists of emotion-laden processes that automatically evaluate socially relevant stimuli along a right–wrong or like–dislike dimension. A second class, moral reasoning, consists of controlled, deliberative processes that arrive at moral judgment or decision through laborious steps of deductive reasoning and cost–benefit analyses. For the most part, these processes work cooperatively to promote moral behavior. Certain ethical dilemmas, however, involve decisions in which the tension or conflict between intuitive and deliberative processes becomes apparent (Greene et al., 2001). One such dilemma is illustrated by the classic trolley problem (Thomson, 1986; Foot, 1978), in which two moral scenarios, impersonal versus personal, are contrasted. On the impersonal version (trolley dilemma), a bystander can use a switch to redirect a runaway trolley away from five victims and onto a single victim; on the personal version (footbridge dilemma), a bystander can push a single victim off of a bridge in front of a runaway trolley in order to stop its progress toward five victims. From a simple "economic" point of view, the two dilemmas are identical (i.e., killing one person to save five lives). Yet, numerous empirical studies

have demonstrated that a large majority of individuals consider it morally acceptable to sacrifice one person to save five in the impersonal dilemma, while they believe that it is wrong to push the large man to save the five victims (Ciaramelli et al., 2007; Koenigs et al., 2007; Mikhail, 2007; Cushman et al., 2006; Valdesolo & DeSteno, 2006; Greene et al., 2001; 2004; Petrinovich et al., 1993).

According to Greene et al. (2001), the reason for these seemingly contradictory responses lies in the stronger tendency of personal scenarios (i.e., the push case), compared to impersonal scenarios (i.e., the switch case), to engage emotional processes which would affect moral decisions. Supporting this proposal, neuroimaging has revealed that impersonal and personal moral dilemma yield dissociable patterns of neural activation (Greene et al., 2001). Specifically, impersonal moral scenarios characteristically yield greater activation in brain areas associated with problem solving and deliberate reasoning [including dorsolateral prefrontal cortex (dlPFC) and inferior parietal lobule], whereas personal moral scenarios yield greater activation in brain areas that have been implicated in emotion and social cognition (such as medial prefrontal cortex and posterior cingulate gyrus). On this view, the thought of pushing someone in front a trolley (i.e., a personal moral violation) elicits prepotent, seemingly negative, emotional responses that oppose or prohibit such repugnant act. In this case, making "more rational," "utilitarian" choices (i.e., deciding that is acceptable to make a harmful act in order to maximize overall utility) would require overriding a strongly aversive emotional response. Accordingly, in a later study, Greene et al. (2004) found that the (infrequent) selection of utilitarian responses in the context of personalmoral dilemmas elicits heightened activity in both "cognitive" (such as dlPFC) and emotional brain areas (including medial prefrontal cortex, posterior cingulate area, and anterior insula). Interestingly, utilitarian decisions were also associated with increased activity in anterior cingulate cortex, which is thought to reflect the conflict between competing processes (Botvinick et al., 2004), namely, cognitive processes favoring a utilitarian judgment and the emotional response to the prospect of doing harm to others.

3.2 vmPFC lesion and moral judgments

Perhaps the most direct evidence supporting a necessary role of emotion in shaping moral decisions has emerged from the neuropsychological investigation of individuals with selective deficits in affective processing. More specifically, patients with adult-onset lesions in ventromedial prefrontal cortex (vmPFC) develop a marked, albeit isolated,

impairment in social behavior that has been consistently attributed to a defective engagement of social emotions, such as guilt, embarrassment, and shame. Recent research has demonstrated that vmPFC patients respond normally to impersonal moral scenarios. However, they are more likely than control groups to endorse moral violations (i.e., inflicting serious harm to people) in personal moral scenarios (Ciaramelli et al., 2007), specifically, "high-conflict" personal scenarios, situations in which there are no clear social norms to decide whether a behavior is morally right or wrong (Koenigs et al., 2007; Hauser, 2006). One interpretation of this result is that vmPFC patients lack automatic affective responses, or aversion signals, impeding any personal moral violation. When affective reactions dissolve (due to brain damage), principled reasoning aimed at maximizing benefits and minimizing costs may prevail, thereby increasing the rate of "rationally appropriate" utilitarian choices (Greene, 2007; but see also Moll & de Oliveira-Souza, 2007 for a different view). Patient lesion studies, thus, strongly suggest that emotions, particularly those subserved by vmPFC, are integral constituents of our moral views. These conclusions, however, rest entirely on the assumption of general emotional blunting or flattened affect following vmPFC damage, a notion based on previous work concerning vmPFC and nonmoral (and nonsocial) decision-making (i.e., gambling task; Bechara & Damasio, 2005; Bechara et al., 1996; Damasio, 1994), or the evaluation of emotional responses to standardized social stimuli (see Koenigs et al., 2007). None of the existing studies has systematically measured subjects-emotional responses emerging during (and, presumably, having an impact on) evaluation of moral dilemmas. This is particularly relevant considering that in specific social circumstances, vmPFC patients have been found to exhibit increased, rather than reduced, emotional reactivity (Koenigs & Tranel, 2007; Barrash et., 2000; Grafman et al., 1996). Therefore, critical evidence linking vmPFC, emotion, and moral judgments is still lacking.

3.3 Searching psychophysiological evidence for moral judgment

The aim of the present study is the gathering of direct psychophysiological evidence, both in healthy and neurologically impaired individuals, that emotions are crucially involved in shaping moral judgment, by preventing personal moral violations. Toward this end, 8 patients with focal lesion involving the ventromedial sectors of prefrontal cortex (vmPFC patients), 7 control patients with lesions outside the frontal lobe (non-FC patients), and 18 healthy controls responded to personal as well as impersonal

moral dilemmas while skin conductance response (SCR) was recorded as a physiological index of affective state. The SCR is related to the sympathetic division of the autonomic nervous system (Boucsein, 1992), and is widely used as a sensitive and objective measure of emotional processing (Dawson et al., 2007; Naqvi & Bechara, 2006; Büchel et al., 1998). Moreover, among cortical regions, vmPFC is presumed to be critically implicated in the generation and feedback representation of bodily states of arousal (i.e., somatic markers), indexed by SCR, in the context of social, emotional, and motivational behavior (Nagai et al., 2004; Bechara et al., 1999; Bechara et al., 1996; Damasio, 1994; Tranel & Damasio, 1994). Consequently, SCR is a measure ideally suited to study the relationship among vmPFC, emotion, and moral decision-making. First, we expected to replicate previous evidence that, compared to normal controls, patients with vmPFC damage are more willing to judge moral violations as acceptable behaviors in personal moral dilemmas, whereas their performance in impersonal and nonmoral dilemmas is comparable to the controls. If emotional state activation mediated by vmPFC plays a critical and selective role in shaping personal moral judgments, then we should observe differences in SCRs between patients with vmPFC damage and comparison groups during contemplation of personal moral scenarios (such as the footbridge dilemma), but not during contemplation of impersonal moral scenarios (such as the trolley dilemma). An additional prediction, derived from the hypothesis that emotional reactions drive disapproval of harmful actions(even when aimed at promoting the greater good), was that skin conductance activity during contemplation of personal moral dilemmas would be negatively correlated with the tendency toward utilitarianism (i.e., percentage of utilitarian judgment made) in normal controls. In other words, we predicted that SCR would be higher in participants exhibiting fewer utilitarian choices than in those with a higher rate of utilitarian responses.

METHODS

Participants

Three groups of subjects participated in the study: (a) a group of patients with focal lesions involving vmPFC bilaterally (the vmPFC group, n = 8); (b) a control group of patients with damage sparing frontal cortex (the non- FC group, n = 7); and (c) a control group of healthy subjects (the HC group, n = 18), who were matched on age, education, and sex with the vmPFC group. Brain-damaged patients were recruited from the Centre for Studies and Researches in Cognitive Neuroscience in Cesena, and from the Azienda

Ospedaliera Spedali Civili in Brescia. They were selected on the basis of the location of their lesion evident on CT or MRI scans. Table 3.1 shows demographic and clinical data, as well as the Mini-Mental Status Examination score (MMSE; Folstein et al., 1983). There were no significant differences between vmPFC patients and comparison groups with regard to age, education, clinical and personality variables ($p > .05$ in all cases).

Eight vmPFC participants took part also in study I; specifically participants number: 2, 3, 4, 5, 7, 8, 9, 10 (see table 1.4 in study I for lesion site details). In the vmPFC group, lesions were caused by rupture and repair of anterior communicating artery (ACoA) aneurysm. Lesions involved vmPFC — defined as the medial one-third of the orbital surface and the ventral one-third of the medial surface of the frontal lobe, following the boundaries laid out by Stuss and Levine (2002) — and adjacent basal forebrain area.1. All vmPFC patients presented with clinical evidence of a decline in social interpersonal conduct, impaired decision-making, and emotional functioning, but had generally intact intellectual abilities (see Table 3.2). The non-FC patients (see Table 3.3) were selected on the basis of having damage that did not involve the frontal lobe, and also spared the amygdala and the insula in both hemispheres. In this group, lesions were unilateral in six patients (in the left hemisphere in 2 cases, and in the right hemisphere in 4 cases) and bilateral in one patient. Brain lesions were caused by arterial–venous malformation in one case, and by ischemic or hemorrhagic stroke in the remaining six cases. Lesion sites included the occipital lobe in two patients, the lateral occipito-temporal junction in three patients, and the lateral occipito-parietal junction in the remaining two patients. All subject groups were administered a short neuropsychological battery including tests with potential sensitivity to frontal damage, as well as intelligence and memory tests (results are provided in Table 3.2). The groups differed significantly only in their performance on the Stroop task, with vmPFC subjects making more errors than both non-FC patients and healthy controls (Mann–Whitney U test, $p < .05$). Patients were not receiving psychoactive drugs at the time of testing, and had no other diagnosis likely to affect cognition or interfere with participation in the study (e.g., significant psychiatric disease, alcohol misuse, history of cerebrovascular disease, focal neurological examination). Neuropsychological and experimental studies were all conducted in the chronic phase of recovery, more than a year post-onset. All lesions were acquired in adulthood. Patients gave informed consent to participate in the study according to the Declaration of Helsinki (International Committee of Medical Journal Editors, 1991) and the Ethical Committee of the Department of Psychology, University of Bologna. Normal participants were healthy volunteers who were

not taking psychoactive medication, and were free of current or past psychiatric or neurological illness as determined by history. Normal controls scored at least 28 out of 30 on the MMSE.

Lesion Analysis

Lesion analysis was based on the most recent clinical CT or MRI. The location and extent of each lesion were mapped by using MRIcro software (Rorden & Brett, 2000). The lesions were manually drawn by a neurologist with experience in image analysis onto standard brain template from the Montreal Neurological Institute, which is based on T1-weighted MRI scans, normalized to Talairach space. This scan is distributed with SPM99 and has become a popular template for normalization in functional brain imaging. For superimposing of the individual brain lesions, the same MRIcro software was used. figure 3.1 shows the extent and overlap of the brain lesions in the braindamaged patients. Brodmann's areas (BA) affected in vmPFC group were areas 10, 11, 32 (subgenual portion), and 24, with region of maximal overlap occurring in BA 10 and 11.

Table 3.1 Summary Data for Participants [Mean (Standard Deviation)]

Group	*Sex (M/F)*	*Age at Test (Year)*	*Education (Year)*	*Time since Lesion (Year)*	*Lesion Volume (cc)*	*MMSE*
vmPFC (n = 8)	7/1	53.1 (10.8)	13.3 (4.9)	5.1 (3.2)	35.3 (16.7)	27.1 (1.8)
non-FC (n = 7)	6/1	52.7 (16.6)	11.8 (4.5)	3.4 (2.6)	25.5 (10.4)	27.5 (1.3)
HC (n = 18)	16/2	53.5 (12.6)	13.5 (5.7)	–	–	28.7 (0.5)

MMSE = Mini-Mental State Examination.

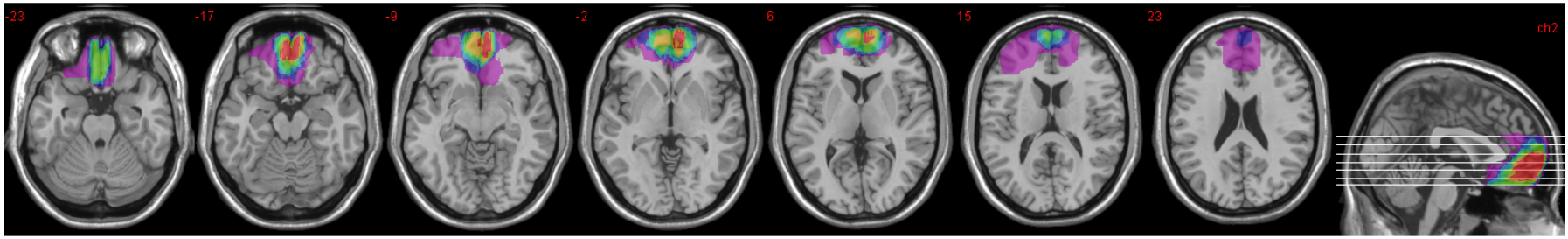
Table 3.2 Results of Selected Neuropsychological Tests [Mean (Standard Deviation)]

Group	*SRM*	*Digit Span Forward*	*Phonemic Fluency*	*Semantic Fluency*	*WMS*	*Stroop Task Errors*[a]
vmPFC	43.8 (4.7)	5 (0.8)	24.5 (7.8)	39.6 (7)	84.6 (5.4)	6 (3.4)[a]
non-FC	43.1 (4.4)	5.1 (0.9)	23.2 (6.3)	39.8 (2.8)	89.8 (4.1)	3.7 (0.8)
HC	46.3 (3.4)	5.1 (0.9)	26.9 (5)	43.1 (4.7)	96.9 (5.4)	1.9 (1.3)

SRM = Standard Raven Matrices (scores in percentile values); WMS = Wechsler Memory Scale.

[a]Values that differ significantly between groups.

vmPFC group



Non-Fc group

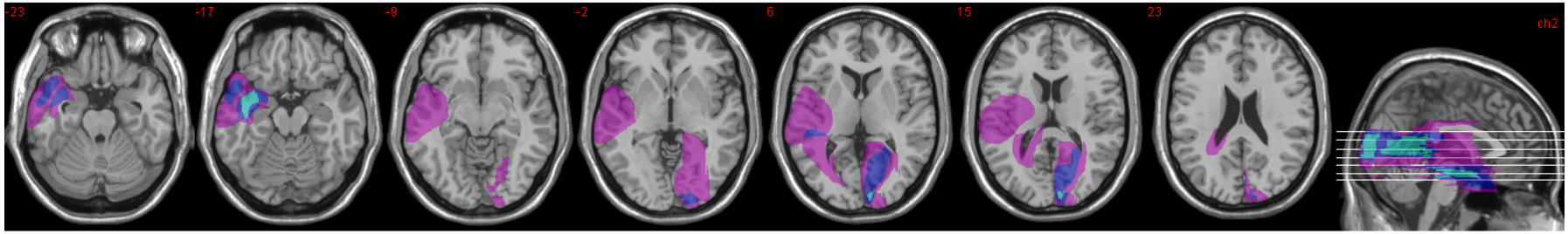


Figure 3.1 Location and overlap of brain lesions. The panels show the lesions of the eight patients with lesion at vmPFC and seven patients with lesions not involving frontal lobe. In the case of patients with vmPFC lesion, damage projected on the same seven axial slices and on the mesial view of the standard Montreal Neurological Institute brain. Maximal overlap occurs in the ventral and anterior portions of medial prefrontal cortex (Brodmann's areas 10, 11, and 32). In order to compare graphically the two groups the same axial slices used for vmPFC group are showed for Non-Frontal patients. The lateral view in Non-Fc reports only the overlapped damage at the same level than lesion for vmPFC patients.

		Gender	Site of lesion*	Type of lesion**
1	non-FC	M	OT	I
2	non-FC	M	OP	I
3	non-FC	M	O	H
4	non-FC	M	OT, bg, IC	AVM
5	non-FC	M	O	H
6	non-FC	F	OT, bg, IC	H
7	non-FC	M	OP	H

Table 3.3. Lesion's details in non-FC group.
* bg = basal ganglia; P = parietal lobe; T = temporal lobe; O = occipital lobe; IC = insula cortex.
** I = ischaemic; H = haemorrhagic, AVM = arterial-venous malformation

Materials

Stimuli in the present study were 15 personal moral dilemmas, 15 impersonal moral dilemmas, and 15 nonmoral dilemmas, randomly selected from a battery of 60 dilemmas developed by Greene et al. (2001), and used in previous study (Ciaramelli et al., 2007, see some examples of dilemma in appendix). Ten out of 15 personal moral scenarios were "high-conflict" dilemmas, whereas the remaining 5 were "low-conflict" dilemmas, as identified by Koenigs et al. (2007) on the basis of the reaction times and level of agreement among normal controls. Moral dilemmas are supposed to elicit moral emotions (i.e., emotions that respond to moral violations, or that motivate moral behavior, such as shame, guilt, pride, and compassion; Haidt, 2007; Tangney et al., 2007), whereas non-moral dilemmas are not (Greene et al., 2001). Typical examples of nonmoral dilemmas posed questions about whether to buy a new television or to have your old television repaired for the same price, or whether to travel by bus or train given certain time constraints.

Task Procedure

An IBM-compatible Pentium IV computer running E-Prime software (Psychology Software Tools, 2002, Pittsburgh, PA) controlled the presentation of dilemmas, timing operation, and behavioral data collection. Subjects sat in front of a computer screen (21-in. VGA monitor) in a quiet and dimly lit room. Each dilemma was presented as text through a series of two screens. The first screen described the scenario and was presented for 45 sec. The second screen posed a question about the appropriateness of an action one might performing that scenario, that is, the "dilemmatic question" (e.g., "Is it appropriate to save

the five persons by pushing the stranger to death?"). Participants indicated their judgments by pressing one of two different keys on the computer keyboard. There was no time limit. Participants were told to respond as soon as they had reached a decision. The intertrial interval, during which a blank allowing the psychophysiological response (see below) to return to baseline after each trial. For all dilemmas being tested, ("appropriate") affirmative responses implied the maximization of overall consequences (Greene, 2003), for instance, killing one instead of five persons (in a moral dilemma), or buying a new television instead of repairing the old one for the same price (in a nonmoral dilemma). However, only for moral dilemmas did "appropriate" responses result in moral violations. Note that "appropriate" and "inappropriate" is a value neutral description of what the participant said about the action in the dilemma and not an evaluation of the participant's decision. Both the number of "appropriate, affirmative responses and response times (RTs; i.e., the time from the onset of the dilemmatic question to the moment a response was given)" were collected. Dilemmas were presented in random order in a single session that lasted approximately 70 min.

Psychophysiological Data Acquisition and Reduction

We used the skin conductance activity as a dependent measure of emotional arousal and somatic state activation. For each participant, prewired Ag/AgCl electrodes (TSD203 Model; Biopac Systems, Goleta, CA), filled with isotonic hyposaturated conductant, were attached to the volar surface of the middle and index fingertip of the nondominant hand and held firmly in place with Velcro straps. Importantly, doing so left the dominant hand free for behavioral responses. The electrode pairs forming part of the input circuit were excited by a constant voltage of 0.5 V (Fowles et al., 1981; Lykken & Venables, 1971) and the current change representing conductance was recorded using a DC amplifier (Biopac GSR100) with a gain factor of 5 μS/V and low-pass filter set at 10 Hz. The analog signal was digitized using the MP-150 digital converter (Biopac Systems) at a rate of 200 Hz and fed into AcqKnowledge 3.9 recording software (Biopac Systems). As subjects performed the moral judgment task seated in front of the computer, SCR was collected continuously and stored for off-line analysis on a second PC. Each testing session began with a 10-min rest period during which the participants' SCR acclimated to the environment, and the experimenter ensured a correct attachment and conductance of the electrodes. Presentation of each dilemma was synchronized with the sampling computer to the nearest millisecond. Furthermore, each time the subject pressed a response key, this action coincided with a

mark on the SCR polygram. During acquisition of the psychophysiological data, the participants were asked to remain quiet and as still as possible to avoid confounding these measurements. After acquisition, skin conductance values were transformed to microsiemens values using the AcqKnowledge software. Also, this software provides an extensive array of measurements that can be applied to the collected data. Raw skin conductance data were low-pass filtered to remove high-frequency noise. The slow downward drift in baseline skin conductance level was removed using a moving difference function with a difference interval of 0.05 sec. Before the start of recording, we ensured that subjects were able to generate SCRs to external stimuli, such as loud sounds (i.e., hands clapping).

RESULTS

Behavioral Data

The proportion of affirmative responses (e.g., utilitarian choices in the context of personal moral dilemmas) for each type of dilemma and each participant group were computed (see figure 3.2). The data were subjected to a mixed-design ANOVA, with group (vmPFC, non-FC, HC) as a between-subject factor, and dilemma (personal, impersonal, nonmoral) as a within-subject factor. The ANOVA yielded a significant main effect of group [$F(2, 30) = 4.4$, $p < .05$], as well as of dilemma [$F(2, 30) = 6.3$, $p < .005$]. Critically, the two-way interaction between group and dilemma was significant [$F(4, 60) = 2.6$, $p = .05$]. Pairwise comparisons showed that both control groups gave fewer affirmative responses to personal (HC = 0.32, non- FC = 0.30) as compared to impersonal (HC = 0.51, non- FC=0.57) and nonmoral dilemmas (HC=0.52, non-FC= 0.51; all ps < .05). By contrast, vmPFC patients made a similar proportion of "appropriate," affirmative responses across all types of dilemma (0.59, 053, and 0.57, for personal, impersonal, and nommoral dilemma, respectively; all ps > .05). A more focused analysis on response patterns within the personal moral dilemmas revealed that vmPFC patients were more likely to endorse the "appropriate" (e.g., utilitarian) response than either comparison groups when high-conflict scenarios were presented (Kruskal–Wallis test, $H = 11.9$, $df= 2$, $p < .01$). In contrast, for low-conflict personal scenarios, the frequency of selecting the affirmative response was negligible and with no significant difference between vmPFC patients and control groups ($H = 1.3$, $df = 2$, $p = .5$). The RT data (figure 3.3) were also subjected to a mixed design ANOVA with group (vmPFC, non-FC, HC) as a between-subject factor, and dilemma (personal, impersonal, nonmoral) and response (affirmative,

negative) as within subject factors. As a violation of the ANOVA, assumption of sphericity was detected using the Mauchly sphericity test and the Greenhouse–Geisser correction for repeated measures was applied. The analysis revealed a significant main effect of group [$F(2, 30) = 5.4$, $p < .01$], as well as a significant two-way interaction between choice and dilemma [$F(2, 60) = 5.9$, $p < .01$]. Moreover, the ANOVA yielded a marginally significant three-way interaction [$F(4, 60) = 2.6$, $p = .07$]. Pairwise comparisons showed that healthy controls and nonfrontal, control patients took longer to make affirmative relative to negative responses in personal moral dilemmas (HC: 4996 vs. 3625 msec; non-FC: 8805 vs. 5709 msec; both ps<.01), but not in impersonalmoral dilemmas (HC: 3548 vs. 3654 msec; non-FC: 6352 vs. 6597; both ps > .5), and in nonmoral dilemma (HC: 3837 vs. 3759 msec; non- FC: 6140 vs. 6496 msec; both ps > .5). In stark contrast, vmPFC patients showed similar RTs for affirmative and negative responses in either personal (5315 vs. 6341msec), impersonal (6937 vs. 7365 msec), and nonmoral dilemmas (7725 vs. 7854 msec; all ps > .1).

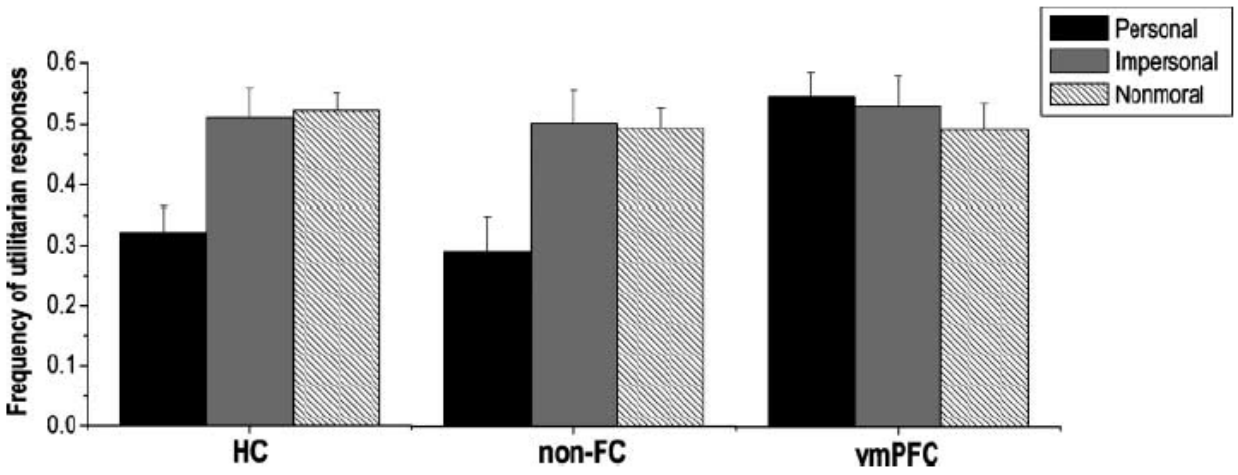


Figure 3.2 Proportion of affirmative responses to personal, impersonal, and nonmoral dilemmas in ventromedial prefrontal patients (vmPFC), nonfrontal patients (non-FC), and healthy controls (HC). Bars refer to 1 standard error of the mean.

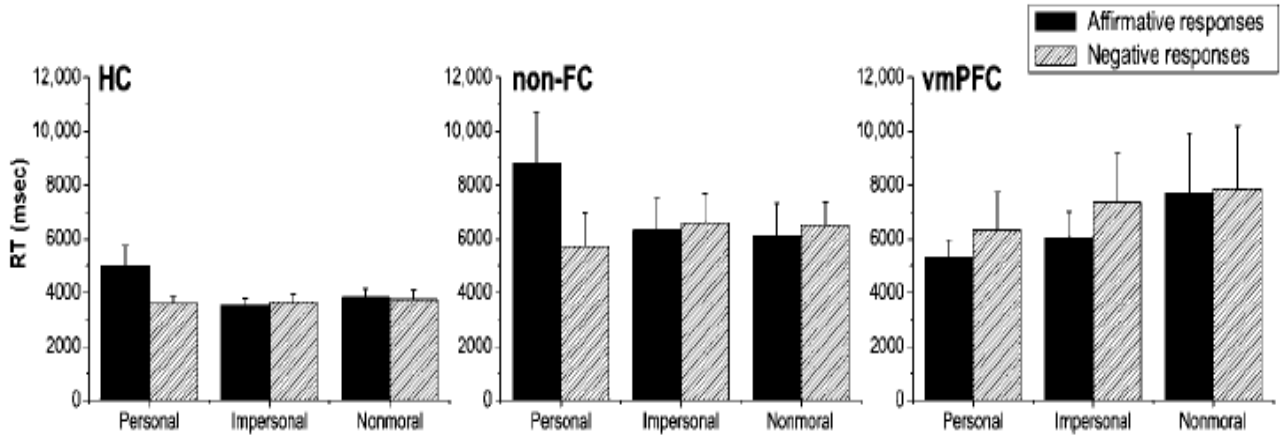


Figure 3.3 Mean response time for affirmative and negative responses to personal, impersonal, and nonmoral dilemmas in ventromedial prefrontal patients (vmPFC), nonfrontal patients (non-FC), and healthy controls (HC). Bars refer to 1 standard error of the mean.

Psychophysiological data

For analysis, each trial was divided off-line into four separate time periods: (a) baseline, the 15-sec time period immediately preceding each dilemma; (b) contemplation, the 45-sec time window during which participants viewed the dilemma; (c) decision, the time period comprised between the presentation of the dilemmatic question and the emission of a response; (d) post-response, the 5-sec time period following participants' response. To examine psychophysiological changes in more detail, the contemplation period was further divided into three consecutive epochs, lasting 15 sec each. SCRs were computed for each epoch of a trial as "area under the curve" (Naqvi & Bechara, 2006; Vianna & Tranel, 2006; Damasio et al., 2000). The "area under the curve" measurement is similar to the function of an "integral" except that, instead of using zero as a baseline for integration, a straight line is drawn between the endpoints of the selected area to function as the baseline. The area is expressed in terms of amplitude units (microsiemens, μS) per time interval (sec). All SCRs were square-root-transformed to attain statistical normality.

Baseline SCRs

Skin conductance levels during the baseline period were submitted to a mixed design ANOVA with group (vmPFC, non-FC, HC) as a between-subject factor, and dilemma (personal, impersonal, nonmoral) as a within-subject factor. Although baseline skin conductance level of vmPFC patients was somewhat lower than control groups, the analysis did not reveal a significant main effect of group, or a significant interaction between group and dilemma ($F < 1$ in both cases). Likewise, the main effect of dilemma was not significant ($F<1$).

Contemplation SCRs

Figure 3.4 shows mean SCRs elicited during each of three consecutive epochs of the contemplation period of personal, impersonal, and nonmoral dilemmas, separately for each participant group and type of response (affirmative vs. negative response). Psychophysiological responses were subjected to a mixed-design ANOVA with group (vmPFC, non-FC, HC) as a between-subject factor, and dilemma (personal, impersonal, nonmoral), epoch (I, II, III), and response (affirmative, negative) as within-subject factors. The analysis revealed a significant main effect of dilemma [$F(2, 60) = 5.3$, $p < .01$], indicating higher SCRs during contemplation of personal relative to impersonal and nonmoral scenarios, as well as a highly significant effect of response [$F(1, 30) = 22.4$, $p < .0001$], due to increased levels of skin conductance for affirmative versus negative responses. Also, there was a significant interaction between dilemma and response [$F(2, 60) = 9.9$, $p < .001$], and between group and response [$F(1, 30) = 5.7$, $p < .01$]. More important for the present purposes, however, the analysis showed a marginally significant three-way interaction between group, dilemma, and response [$F(4, 60) = 2.4$, $p = .058$], whereas the four-way interaction was not significant [$F(8, 120) = 0.9$, $p = .5$]. To uncover the source of the marginally significant three-way interaction, separate ANOVAs were conducted on contemplation SCRs (collapsing across epochs) for the different types of dilemma. For the personal dilemmas, both the main effect of response [$F(1, 30) = 19.4$, $p < .001$] and the two-way interaction between group and response [$F(2, 30) = 4.9$, $p < .01$] were significant. Pairwise comparisons using the Fisher LSD test, which is considered the most powerful technique for post hoc tests involving three groups (Cardinal & Aitken, 2006), revealed that both non-FC patients and healthy controls generated larger SCRs during contemplation of personal moral dilemmas that were associated with affirmative responses (e.g., utilitarian judgments) (all $ps < .01$); in contrast, vmPFC patients showed no differential skin conductance activity preceding affirmative and negative responses in personal moral dilemmas ($p = .91$). For both impersonal and nonmoral dilemmas, ANOVAs showed that the factor group did not result in a main effect; neither did it alter any of the interactions, suggesting that contemplation of impersonal and nonmoral scenarios resulted in similar skin conductance activity across all groups of participants. To ensure that our findings were not driven by group differences in tonic level of electrodermal activity, we repeated the main ANOVA with baseline skin conductance activity as a covariate. The previously (marginally) significant Group by Dilemma by

Response interaction remained significant [$F(4, 58) = 2.9$, $p = .026$], as did the Group by Response interaction [$F(2, 29) = 5.5$, $p < .01$]. The response deficit in the vmPFC patients is, therefore, not a function of lower baseline electrodermal activity.

Decision SCRs

Mean SCRs elicited during the 5-sec period following the dilemmatic question were subjected to a mixed-design ANOVA (figure 3.5), with group (vmPFC, non-FC, HC) as a between subject factor, and dilemma (personal, impersonal, nonmoral) and choice (utilitarian, nonutilitarian) as within-subject factors. Both the main factor of choice [$F(1, 30) = 19.4$, $p < .001$] and the interaction between choice and dilemma [$F(2, 60) = 4.5$, $p < .01$] were significant. In contrast, the three-way interaction was not significant [$F(2, 60) = 1.4$, $p=.2$]. Nevertheless, for completeness, we also conducted planned comparisons. Particularly, we found that normal controls and non-FC patients generated larger SCRs prior to utilitarian as compared to nonutilitarian judgments in personal moral dilemma ($p < .05$), whereas vmPFC patients showed similar skin conductance activity regardless of choice type ($p = .31$). Again, no group difference emerged when both impersonal and nonmoral dilemmas were considered. Finally, adding baseline skin conductance activity as a covariate in the ANOVA did not alter the pattern of results.

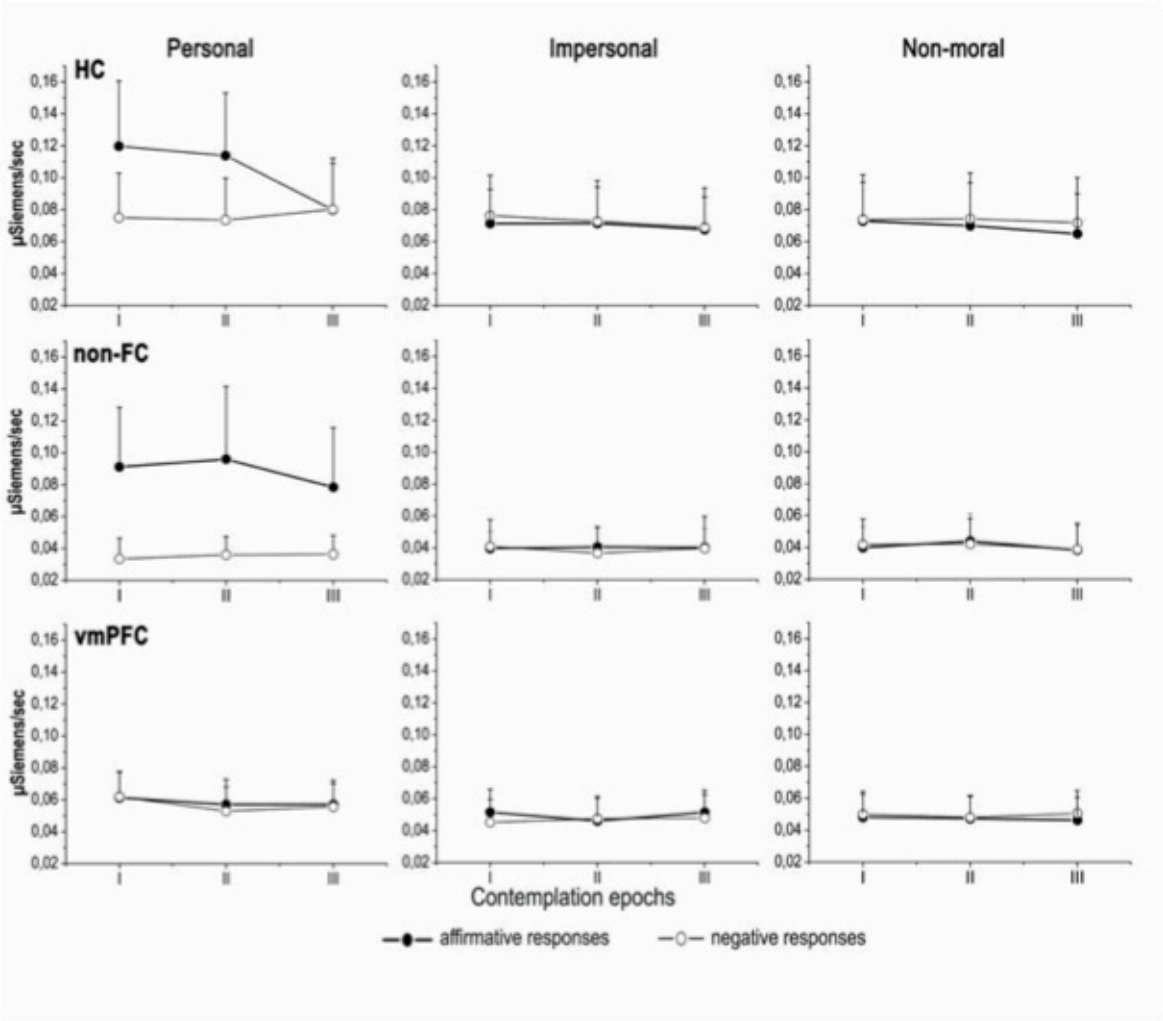


Figure 3.4 Mean SCRs elicited during each of three consecutive epochs of the contemplation period of personal, impersonal, and nonmoral dilemmas, separately for each participant group and type of response (affirmative vs. negative). SCR was measured as "area under the curve" in µS/sec. vmPFC = ventromedial prefrontal patients; non-FC = nonfrontal patients; HC = healthy controls. Bars refer to 1 standard error of the mean.

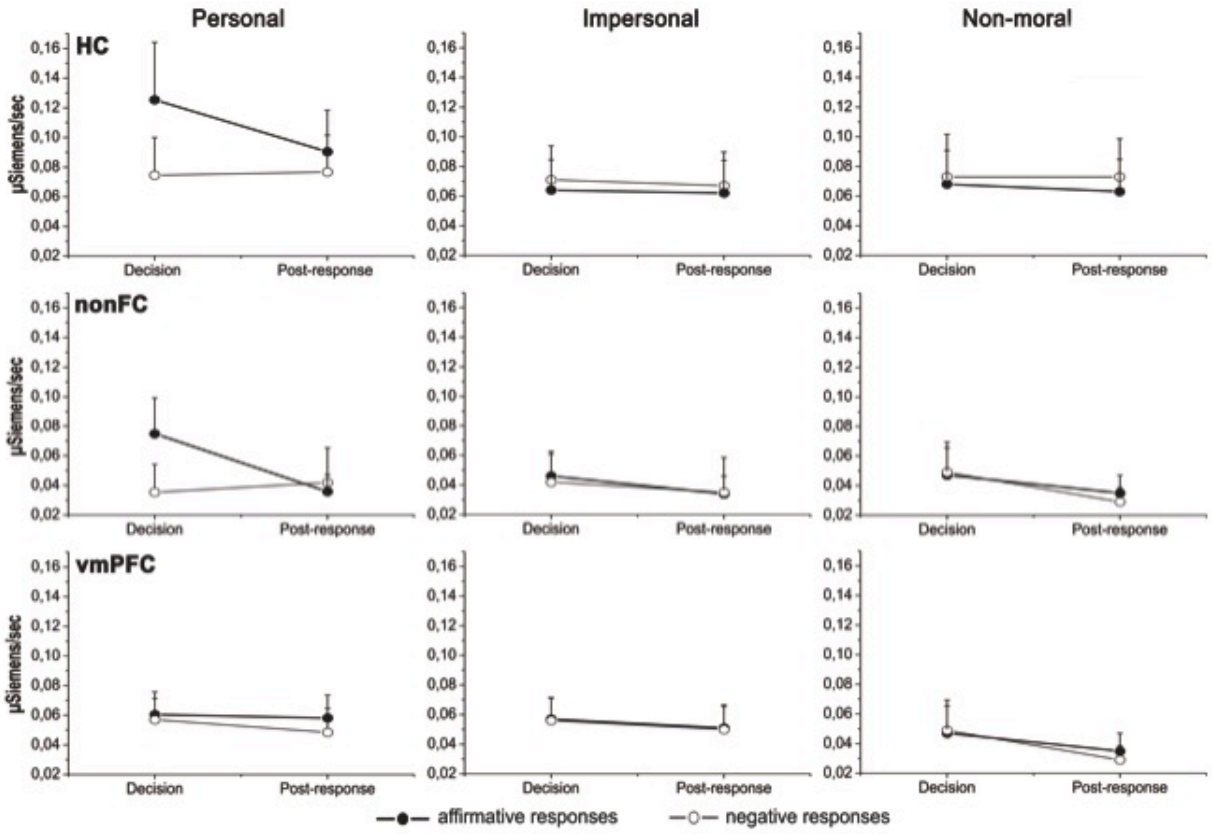


Figure 3.5 Mean SCRs elicited during Decision and Post-response period of personal, impersonal, and nonmoral dilemmas, separately for each participant group and type of response (affirmative vs. negative). SCR was measured as "area under the curve" in µS/sec. vmPFC = ventromedial prefrontal patients; non-FC = nonfrontal patients; HC = healthy controls. Bars refer to 1 standard error of the mean.

Post-response SCRs

Mean SCRs elicited during the 5-sec period following participants' response were subjected to a mixed-design ANOVA (figure 3.5), with group (vmPFC, non-FC, HC) as a between-subject factor, and dilemma (personal, impersonal, nonmoral) and choice (utilitarian, nonutilitarian) as within-subject factors. The main effect of choice was significant [$F(1, 30) = 7.1$, $p < .05$], indicating overall larger SCRs following utilitarian versus nonutilitarian choices. However, the factor group was not significant ($F < 1$), nor did it enter in any significant interactions (all $Fs < 1$). To sum up, the results from the ANOVAs revealed that, for healthy subjects and nonfrontal patients, SCRs were stronger during evaluation of personal moral dilemmas that subsequently attracted an affirmative (e.g., utilitarian) response than during evaluation of personal moral dilemmas that subsequently attracted a negative (e.g., nonutilitarian) response. These individuals, on average, selected nonutilitarian over utilitarian choices in personal moral dilemmas. In contrast, for vmPFC patients, who were more inclined toward utilitarian judgment, SCRs

did not change for personal moral dilemmas, subsequently attracting utilitarian versus nonutilitarian choices. This finding relates increases in SCR during the anticipation of a utilitarian choice (and therefore, a personal moral violation) with low tendency toward utilitarian judgment. One possibility is that anticipatory SCRs, by marking a particular option– outcome pair with a negative tag, bias individuals to avoid similar scenarios in the future (Bechara et al., 1996; Damasio et al., 1996). To investigate whether anticipatory skin conductance activity was predictive of the type of choice on personal moral dilemmas, a further analysis was performed. We computed an autonomic utilitarian index [(SCRs prior to utilitarian choices − SCRs prior to nonutilitarian choices)/ (SCRs prior to utilitarian choice+SCRs prior to nonutilitarian choices)] for each healthy control participant, and then entered into a regression analysis with the percent of utilitarian choices made by each subject in response to personal moral dilemmas. Results showed that the autonomic utilitarian index correlated negatively with the proportion of utilitarian judgments ($r = -.64$, $p < .005$) (figure 3.6). Indeed, the autonomic utilitarian index decreased linearly as the percent of utilitarian choices increased, indicating that low utilitarian participants exhibited higher skin conductance activity prior to utilitarian judgments of personal moral dilemmas, whereas high-utilitarian participants showed the opposite pattern. By contrast, this was not the case for impersonal moral dilemmas ($r = .10$, $p = .7$), thereby revealing that utilitarian judgments were not related to skin conductance activity for this type of moral dilemmas.

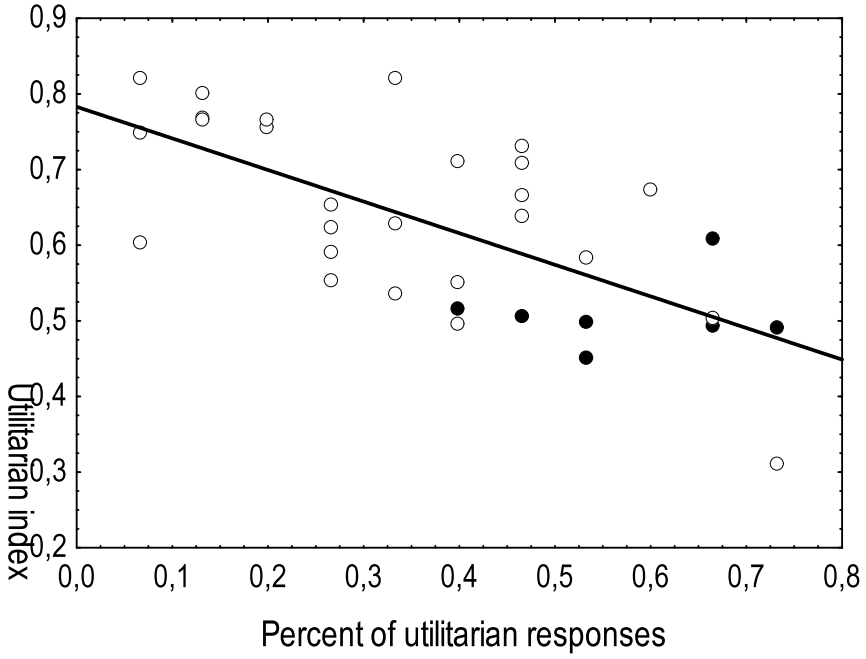


Figure 3.6 Graph reporting the regression analysis results. Filled circles indicate vmPFC patients.

DISCUSSION

Recent findings from human lesion (Ciaramelli et al., 2007; Koenigs et al., 2007) and brain imaging studies (Greene et al., 2001; 2004) converge to suggest that medial prefrontal cortex constitutes a critical neural underpinning of judgments about personal moral dilemmas, where one option involves directly inflicting serious harm to other persons. In particular, it has been found that vmPFC-lesioned patients, relative to healthy individuals and neurological patients with brain damage in other cerebral regions, are more likely to endorse personal moral violations in order to maximize good consequences (i.e., the utilitarian response). According to one account, this abnormally increased utilitarian pattern of moral judgment would result from impaired affective and intuitive processes, mediated by vmPFC, which normally oppose deviations from moral values and rules shared by a social group (Greene, 2007). Although these results strongly suggest a causally necessary role of emotions in morally relevant decision-making, a mechanistic account of how, and at which point, emotional states subserved by vmPFC influence moral judgment is still lacking (see Huebner et al., 2008 for a discussion). The present study was designed to examine the pattern of skin conductance changes, used as an autonomic index of individuals' affective responses, associated with personal versus impersonal moral judgments, both in vmPFC patients and control participants. This study would provide the first neurophysiological evidence of emotional activation before making a moral decision.

Eight vmPFC patients had lesion involving the rostral aspect of the anterior cingulated cortex (ACC). The ACC, specifically BA 32 and 24, is considered together with OFC the cortical components of the limbic system (David et al., 2005). ACC and medial OFC lesions lead to a reduction in emotional responsiveness and a reduction in the value that is attributed to social stimuli during decision making, respectively (Rushworth et al., 2007). In macaques, OFC lesions lead, most notably, to increased aggression and diminished fear (Izquierdo et al., 2005; Machado & Bachevalier, 2006). ACC activity has been prominent in neuroimaging experiments that have examined interactions between individuals (Rilling et al., 2002; Tomlin et al., 2006; Amodio & Frith., 2006). Social exchange experiments are necessarily complex and activity changes in the brain are widespread. Nevertheless, the ACC gyrus is critical for the normal valuation of social information. Usually, male

macaques value the opportunity to observe other macaques, and they are most interested in dominant males and in females (Deaner et al., 2005), but macaques that have lesions of the ACC gyrus do not exhibit the same social-valuation patterns (Rudebeck et al., 2006). Unlike with OFC lesions, changes in emotional responsiveness, such as increased aggression or diminished fear, and impairments in visual-discrimination learning are not seen consistently after ACC gyrus lesions.

Traditionally the basal ganglia have been associated with motor processes, although evidence for their role in parallel cognitive functions is mounting (for a review, see Middleton & Strick, 2000). Two non-FC patients have lesion at basal ganglia specifically on putamen, which appears to subserve cognitive functions limited to stimulus-response or habit, whereas caudate seems play a critical role in supporting the planning and execution of strategies and behavior required for achieving complex goals (Grahn et al., 2008). The main role of putamen seems related to the implementation of action (sensorimotor coordination), and, in the case of our patients, could be related to their specific deficit in movement coordination.

In complete agreement with previous data (Ciaramelli et al., 2007; Koenigs et al., 2007), our present findings reveal that patients with vmPFC damage made significantly more utilitarian choices in response to high-conflict personal moral scenarios, compared to patients with brain damage that spared vmPFC and to healthy controls. Moreover, patients with vmPFC lesions were also faster than control groups to approve personal moral violations. On the other hand, their behavior in low-conflict personal, impersonal, and non moral dilemmas was comparable to that of controls, both in terms of the quality of the choices they made and in the time they needed to make their decisions, further demonstrating the rather selective role played by vmPFC mediated emotions on personal moral judgments (Young & Koenigs, 2007; Hauser, 2006).

The psychophysiological data mirrored the behavioral results: whereas autonomic bodily signals during consideration of impersonal and non moral dilemmas did not differ across participant groups, skin conductance recordings during contemplation of personal moral scenarios differed considerably between patients with vmPFC damage and control groups. Both healthy subjects and brain-damaged control patients exhibited increased skin conductance activity several seconds before choosing the utilitarian option in personal moral dilemmas, for instance, deciding that it would be appropriate to kill one person in order to save others. In striking contrast, vmPFC patients did not generate SCRs in

anticipation of utilitarian choices in personal moral dilemmas. These findings indicate profound differences in the making of moral judgment between vmPFC patients and controls. In control groups, emotional/somatic signals were critically recruited during moral judgment, and characterized the anticipation of personal moral violations. In contrast, no apparent emotional/ somatic response accompanied personal moral violations in vmPFC patients. Importantly, somatic responses shaped personal moral judgment. A preliminary analysis showed a negative correlation between anticipatory skin conductance activity and frequency of utilitarian responses in normal controls, such that individuals with higher SCRs before utilitarian choices were more reluctant to judge moral infractions as acceptable behaviors than those with lower SCRs. One possibility, therefore, is that emotional responses mark utilitarian choices in personal moral dilemmas with a negative tag, discouraging the selection of those options in future decisions. Studies of patients with discrete brain lesions and, more recently, functional imaging techniques have strongly implicated vmPFC in both generation and feedback representation of states of bodily arousal, indexed by SCRs, which may influence cognition and bias motivational behavior (Nagai et al., 2004; Critchley et al., 2001; Damasio et al., 1990). In several studies, vmPFC patients often exhibit impaired autonomic arousal and subjective feeling in response to emotionally charged events (Roberts et al., 2004; Blair & Cipolotti, 2000; Tranel & Damasio, 1994; Damasio et al., 1990). Importantly, in a now seminal series of studies, Bechara et al. (1996, 1999), Bechara, Damasio, Tranel, and Damasio (1997) and Damasio et al. (1990) have shown that vmPFC lesioned patients perform poorly on a gambling task, and unlike normal controls, fail to show anticipatory SCRs immediately before selecting a high-risk option (i.e., one offering immediate gain but a high probability of long-term monetary loss). These findings have led to a proposal that central representations of bodily states of arousal guide social behavior and bias decision-making, formulated as the "somatic marker hypothesis" (Bechara & Damasio, 2005; Bechara et al., 1996; Damasio, 1990; 1994; 1996). According to this hypothesis, the SCR would operate as an alarm signal that, by marking a specific option–outcome combination with a negative tag, promotes the avoidance of similar options in the future. This interpretation of the SCR is also broadly consistent with the observation of anticipatory SCRs in aversive conditioning paradigms (Tabbert et al., 2005; Büchel et al., 1998), and with the proposal that the SCR might represent a "somatic marker of erring" (Hajcak et al., 2003; 2004). The lack of somatic marker in response to emotional events, as well as vmPFC patients' inability in modifying their behavior in response to negative feedback (Fellows & Farah, 2005) can

account for the present findings. It has been suggested that the somatic-emotional reaction before making a moral decision may constitute a warning signal alerting subjects they are making a potentially disadvantageous choice (Damasio, 1994). The vmPFC patients lack this negative somatic reaction before accepting moral violations in personal dilemma, whereas their somatic activation is completely normal for dilemmas not involving strong negative emotion such as those elicited by impersonal moral dilemma.

Our current finding of increased somatic arousal in control participants immediately before endorsing morally reprehensible actions (in the context of personal dilemmas) is highly consistent with the anticipatory SCR obtained with Bechara gambling task. In keeping with the somatic marker hypothesis, anticipatory somatic states of arousal, supported in part by circuits in vmPFC, may help forecast the negative emotional consequences (e.g., shame, guilt or remorse) of approving personal moral transgressions (e.g., utilitarian judgments), thereby motivating individuals to avoid actions that generate such negative somatic states in subsequent choices. Thus, the SCR signal could not only serve as an affective signal that alerts us to the moral relevance of a rule transgression (particularly if that transgression may cause serious harm to others), but also as a teaching signal aimed at decreasing the likelihood of morally impermissible behaviors. Accordingly, the absence of anticipatory SCRs in vmPFC patients may indicate that they fail to represent the affective expectations of highly aversive personal moral transgressions, thereby lacking a powerful biasing signal (e.g., a moral reinforcer) that is critical for driving changes of behavior and compliance with moral values (Tangney et al., 2007; Amodio & Frith, 2006; Frijda, 2005). This conclusion appears in accordance with current theories maintaining that vmPFC is a critical neural substrate for representing potential positive and negative action outcomes in order to promote approach/ avoidance learning and behavior flexibility (Murray et al., 2007; Montague et al., 2006; Oya et al., 2005). The interpretation that we offer is compatible with recent evidence from fMRI, showing that imagined socio-moral transgressions associated with sentiments of guilt elicited activation within medial sectors of prefrontal cortex (Zahn et al., 2009; Kédia et al., 2008). Moreover, data from economic games indicate that patients with vmPFC damage are abnormally insensitive to guilt in social and economic interactions (Krajbich et al., 2009).

One might even argue that whether moral judgment results impaired in patients with ventromedial prefrontal lesions would critically depend on the degree to which the

task taps emotional/self-focused processing. Accordingly, patients with ventromedial prefrontal lesions report reduced self-conscious emotions after engaging in socially inappropriate behaviors compared to patients with dorsolateral lesions (Eslinger & Damasio, 1985; Beer et al., 2006). This lack of self-conscious emotions could be the result of the lack of somatic activation, especially somatic negative activation. In the study by Beer and colleagues (2006) patients who had previously failed to feel that their behavior was socially inappropriate were able to recognize it as such on a later video recording. This finding suggests on one hand the preserved capacity to understand the social context and the reactions of others, but on the other it shows that vmPFC patients are not able to anticipate and feel the negative emotions implied in their own inappropriate behavior. The ventromedial prefrontal cortex would be crucial for self-focused, rather then externally-focused (or even knowledge-driven), social cognition mechanisms (Beer et al., 2006; see also Lieberman, 2006). The dissociation found in vmPFC patients between impaired personal moral judgment and preserved impersonal moral judgment, both at the behavioral and psychophysiological level, provides further support to this interpretation.

Social abilities, such as empathy, heavily rely on processing in medial prefrontal regions (e.g. Brothers & Ring, 1992; Eslinger, 1998), and may be impaired in patients with ventromedial prefrontal damage (Shamay-Tsoory et al., 2005), possibly resulting in reduced responsiveness and empathy to victims (see Blair & Cipolotti, 2000). From a neuroscientific perspective, however, it is important to demarcate empathy from cognitive perspective taking, on the basis of different neural networks for empathy and cognitive perspective taking outlined by de Vignemont & Singer (2005). We refer to cognitive perspective taking as the ability to understand intentions, desires, beliefs of another person, resulting from (cognitively) reasoning about the other's state. By contrast, we refer to empathy as an affective state, caused by sharing of the emotions or sensory states of another person (Hein & Singer, 2008). Clearly, the incapacity of vmPFC patients to anticipate and consider their own emotions or sensory states (especially the negative one) affects their empathy: how can they share their emotions with others when they seem unable to process their own emotions? However, it would be wrong to say that vmPFC patients are more inclined to judge personal moral violations as acceptable than normal controls, because of lack of empathy. As our findings suggest, vmPFC does not govern moral behavior or empathy *per se*, but its role is specifically related to the anticipation of (social) emotions in most of circumstance of life, especially in the social and moral

domains, in which the emotional value of event is critical to choose among alternative options (see Krajbich et al., 2009 for a similar conclusion).

Finally, our view is also in agreement with the finding that early damage to vmPFC can lead to severe deficits in moral sentiments, including guilt, remorse, and empathy, as well as profound impairments of moral reasoning (Anderson et al., 1999), thereby suggesting that emotional processing mediated by this area is developmentally necessary for the learning and acquisition of moral concepts.

A different, but not necessarily mutually exclusive, account of the present findings would instead invoke the concepts of attention regulation rather than emotion and affective valuation of consequences (Botvinick, 2007; Dawson et al., 2007). Indeed, SCR variability has been often used as an index of attention-related arousal (Boucsein, 1992). Notably, a recent study has shown elevation in skin conductance immediately before actions associated with a high demand of controlled cognitive processing (Botvinick & Rosen, 2009). On this view, the increase in arousal preceding the endorsement of personal moral violations could be related to the recruitment of cognitive control needed to solve the conflict between incompatible outcomes (e.g., utilitarian and nonutilitarian outcomes) in response to difficult (e.g., personal) moral dilemmas (Greene et al., 2004). One way of reconciling these two seemingly disparate accounts is to consider that the anticipatory SCRs obtained in our experiments may reflect both affective valuation of the degree of cognitive effort and conflict associated with utilitarian judgments, and the socially negative consequences of endorsing moral violations in personal moral dilemmas (Botvinick & Rosen, 2009; Botvinick, 2007). Such a conclusion would be consistent with model proposing that vmPFC represents the composite values of different predictions for subsequent decisions and judgments (Montague & Berns, 2002). To conclude, the present results suggest that emotion processing mediated by vmPFC plays a necessary role in guiding moral decisions about whether or not sacrificing an individual in order to save a greater number of persons (e.g., high-conflict personal moral dilemmas). In particular, we found that activation of somatic states (monitored through SCRs) prior of utilitarian moral judgments is impaired following vmPFC lesion. That is, contemplating morally impermissible actions was not emotionally taxing in patients with vmPFC damage. We argue that this deficit may prevent vmPFC patients to anticipate the negative emotional consequences of moral violations, and, as a consequence, to conform their behavior to moral norms and values shared by their social group.

The novelty of the present study concerns the pattern of autonomic activation observed in control subjects and the lack of this pattern in vmPFC patients before responding to personal moral dilemmas. Our findings provide critical evidence relating emotional activation, personal moral judgments and the vmPFC. These findings support a necessary role for emotion in the generation of those judgements.

STUDY IV

EXPERIENCE OF AGENCY AND SENSE OF RESPONSIBILITY

4.1 Voluntary action

There is a long tradition in psychology of studies about voluntary action and a more recent interest in moral decision, but these two themes are mostly independent. The aim of the present study is to show how these two apparently different topics are in reality rather overlapped, both from phenomenological point of view and in a neuroanatomical sense.

Most of the scientific studies about volition investigate only simple voluntary actions, such as pressing a button (Libet et al., 1983; Haggard et al., 1999; Brass et al., 2007; Sirigu et al., 2005; Moore et al., 2008; Moore et al., 2009). Although there is no moral implication in this kind of action, the first problem for neuroscientists is to understand the experience of causal relation between thought and action and the causal relation between action and effect.

One extreme and general solution to this problem proposes that there is no real causal relation between our thought, action and effect even though we perceive that there is. This solution defines free will as a delay experience that people have when they interpret their own thought as the cause of their action (Wegner 1999, 2002). In this sense free will is a simple *post hoc* illusion because the causal relation is a simple inference of constant conjunction (Hume, 1784) or something that the mind creates because we really do not know what is causing our actions. Several observations support this definition of illusion of conscious will, showing how conscious will is weak and malleable (Wegner

1999; 2004). An example of the manipulation of conscious will is hypnosis. The most profound effect of hypnosis is the feeling that your acts are happening to you, rather than that you are doing them (Lynn et al., 1990). Other examples come from neuropsychological disorder, for instance anarchic hand syndrome. People with this syndrome experience a conflict between their declared will and the action of one of their hands (Della Sala, 2005), claiming that the hand had a mind of its own and often did whatever 'pleased it' (Della Sala, 1994). Normal people often report losing consciousness of acting, particularly when the action is repetitive or well-learned (Scooler et al., 2002). The hypothesis that the experience of conscious will as mental causation are not genuine events but a *post hoc* reconstruction are not easily compatible with the social principle that all of us can control their action and so we are responsible of our behaviour.

Why do we have this constant, strong, illusion of conscious and free will? Why do we have this sensation of control of action, and agency? If there is no free will and no causal relation between our thoughts and our actions, how can we explain other experiences depending on the causal relation such as, for example, responsibility, guilt and blame?

Even if we are not aware of what really cause our behaviour, we perhaps need a concept like 'causal agency' and 'conscious will', because these are important way by which people understand human actions (Wegner et al., 2004). The illusion of conscious will could be a sort of strategic tool to give meaning to all unconscious causes that move the human actions. In this prospective conscious will is a sort of reaction to avoid the *horror vacui*, and it has many of the qualities of an emotion, one that reverberates through the mind and body to indicate when we sense having authored and intended an action. Conscious will could be a retrospective feeling of authorship telling us which events around us seem to be attributed to our authorship. This emotion allows us to develop a constant sense of who we are and are not. And, most important, this authorship emotion allows us to maintain the sense of responsibility for our actions that serve as a basis for morality.

This definition of volition as morally significant authorship suggests a new and useful perspective for neuroscientific investigation, and a possible dialog between illusory mental causation and causal responsibility. A more radical view (Dennett, 1992) suggests that conscious intention is not a bona fide mental state at all, but rather an inference that is

retrospectively inserted into the stream of consciousness as the hypothetical cause of the physical movement of our bodies.

Modern neuroscience rejects the traditional dualist view of volition as a casual chain from a conscious mind or 'soul' to the brain and body. Rather, volition involves brain networks making a series of complex, open decisions between alternative actions. Volition is thus defined as a set cognitive decisional processes, implemented by specific brain circuits, and giving rise to body movements. '*These processes jointly specify several kinds of information that determine our actions, so voluntary action is therefore a form of decision making*' (Haggard, 2008 pag 937). Brass and Haggard suggested a model with three main components of voluntary decision: *what, when* and *whether* (WWW) (Brass & Haggard 2008). Haggard (2008) distinguished two levels in decisions that concerned whether to act: an early, motivational decision whether to make any action at all, and a final predictive check before execution. Second – the 'what decision' specifies which goals or task to pursue and the selection of movements to achieve them. The 'when component' often depends on the combination of environmental circumstances and internal motivations.

On the basis of neurological analyses of patients with forebrain lesions, Antonio Damasio (2000, 1993) has advanced the 'somatic marker' hypothesis of consciousness. Humans are aware of their bodies, our 'selves', and this inner-directed attention forms the root of consciousness. Damasio argues that consciousness is based upon an awareness of the 'somatic' milieu, and that awareness of inner states evolved because this enables us to use somatic states (ie. emotions) to 'mark', and thereby 'evaluate', external perceptual information. He proposes that the subjective process of feeling emotions requires the participation of brain regions that are involved in the mapping and/or regulation of our continuously changing internal states. The feelings are grounded in the body itself, based on multi-tiered and evolutionarily developed neural mechanisms that control the body's state. These feelings help to guide behavioural decisions producing a 'perceptual landscape' that represents the emotional significance of a particular stimulus that is being experienced. These feelings distinguish between inner-world representations and outer-world representations, and allow the brain to build a meta-representational model of the relationship between outer and inner entities. So, the representational image of the body's state provides a neural basis for distinguishing self from non-self, and re-representations of this image enable the behavioural neural agent to project the effects of possible actions

onto the state of the body, as well as the resultant changes in such feeling states due to interactions with other (external) agents. A summary definition based on this description suggests that awareness of any object requires, first, a mental representation of oneself as a feeling (sentient) entity; second, a mental representation of that object; and third, a mental representation of the salient interrelationship between oneself and that object in the immediate moment (Craig, 2002).

4.2 Measuring volition

Voluntary action is fundamental to human existence. It involves voluntary control of bodily movement to achieve a desired goal. The main model describing voluntary action in cognitive psychology is the 'perception model'. In this model, first brain motor system produces a movement as a product of its different inputs and second the conscious experience of volition is informed of this movement, and it is perceived as being freely chosen (Hallett, 2007).

The 'perception model' considers the conscious will something that we perceive, but the perception of conscious will clearly differs from external senses like vision in important aspects. Traditional psychological studies generally deliver a known input or stimulus to a system and measure the system's reaction. This approach is clearly not suitable to measure the experience of intention or conscious will because the input, in the case of will, comes from the person themselves and not from any external stimulus. What distinguishes a simple reaction, e.g. a reflex, from a voluntary action is the level of dependency on external stimuli. The reflex is an immediate motor response, the form of which is determined by the form of stimulation. In contrast, the occurrence, the timing, and the form of voluntary action are not directly dependent on any stimulation. Voluntary actions involve the cerebral cortex, whereas some reflexes are purely spinal. Moreover volition matures late in individual development, whereas reflexes can be present at or before birth.

A useful strategy to detect the voluntary aspect of action is to give partial instruction to the subjects. Subjects can choose when they want to perform an action (Libet et al., 1983) whether or not perform an action (Brass et al., 2007) and which actions perform (Haggard et al., 1999). One problem concerning these studies on conscious will is that generally there is no reason or value to motivate the participants to choose one action over other. This problem is starting to be solved by introducing a reward component in relation to

selected action (Kuhn et al., 2008) but it is far away from being solved completely. In general the relation between decision making processing and voluntary action is mostly unexplored. Another empty field for the study of volition is the social action. Experiences like responsibility, guilt, are fundamental concepts that are necessary to understand and organize other people behaviour. However neuroscientists know little about the main characteristics of experience of voluntary and non voluntary social/moral actions.

4.3 The experience of voluntary action

Voluntary actions are characterized by two specific subjective components: the experience of intention and the experience of agency. The experience of intention is: 'planning to do or being about to do something'. '*Planning to do something'* is defined by Searle (1983) 'prior intention' and concerns translate desires, goals into behaviour (e.g. to telephone a friend this evening) whereas '*being about something'* is called 'intention in action', which would occur during the process of reaching for the phone (Haggard, 2005). Most of neuroscientific literature on voluntary actions, investigates conscious states associated with simple manual actions, corresponding to the Searl's intentions in action concept.

The normal experience of intentional action includes an implicit content that the action occurred because, and via the intention that the agent had to perform it. However in cases of anarchic hand syndrome and utilisation behavior (Boccardi et al., 2002) this experience of intention is absent or garbled. These patients, who typically show bilateral frontal mesial damage to the supplementary motor area (SMA), or unilateral SMA and callosal damage (Della Sala et al., 1991) magnetically respond to environmental objects without a specific intention to do so. For example, the mere presence of a pen on a table will lead them to pick it up and start writing, even if they have no particular intention to write. The patient does not deny authorship of the action, but they clearly have no conscious experience that their intentions are the source of the action: ''my hands move by themselves''. Another example of garbled experience of intention occurs in Tourette's syndrome (TS). TS is a neuropsychiatry disorder characterized by motor and vocal tics (Albin, 2006). TS patients often cannot say whether their tics are voluntary or involuntary (Hallett, 2007). TS patients report having a clear experience that tic is coming before the real occurrence of the tic. When they perceive this sensation they cannot suppress tic. However they can suppress the tics before the occurrence of this sensation but this requires strong afford and a successive sensation of tension. TS reveals an interesting possibility of

different levels of intention, since these patients seem to have normal intention for voluntary action but reduced sense of intention for their tics. This reduced experience of intention for tics could be related to the impossibility to control and suppress the action. There is no research about subjective experience of action, both voluntary and involuntary, in TS and in general little is know about the intensity of intention and its components.

A key experiment for subjective components of intention in voluntary action is the famous 'Libet experiment' (Libet et al., 1983). In this experiment participants watch a spot or clock hand rotating on a screen. At the time of their own choosing they spontaneously make a movement of the right hand. The clock stops after a random interval, and the participant reports the position of the clock hand at the moment when they first feel the urge to move their hand. At the same time electrodes placed on the scalp record the activity of prefrontal motor areas in preparing the movement. On average, participants reported the conscious intention to act (W: will judgment) around 200 ms before the onset of muscle activity. By contrast, preparatory brain activity could begin 1s or more before movement. Libet (1983) concluded "that cerebral initiation of a spontaneous, freely voluntary act can begin unconsciously, that is, before there is any (at least recallable) subjective awareness that a decision to act has already been initiated cerebrally". This suggests that the initiation of action involves an unconscious neural process, which eventually produces the conscious experience of action. Conscious intentions therefore occur as a result of brain activity, and do not cause brain activity. Haggard and Eimer (1999) found that the experience of conscious intention is tied to the specific body movement that is prepared, rather than to a general preparation to move. In this study they showed how judgments of conscious intention correlates with the onset of the lateralized readiness potential (later phase of preparation, in which brain activity contralateral to the selected hand exceeds ipsilateral activity).

The second subjective component of voluntary action is the experience of agency. Agency is the subsequent feeling that one's action has indeed caused a particular external event (Haggard, 2005) and necessarily involves the experience of external sensory consequences. Two levels of agency have been identified (Jannerod 2009; Synofzik 2008). One level is automatic/unconscious, where the agency judgments are outputs from internal prediction of motor system of what the consequences of an action are like to be (Frith et al., 2000). A second level is non automatic/conscious, where the agency seems to be a reconstructive process due to the constant conjunction of cause and effect (Jennerod 2009).

This distinction between automatic and non automatic agency resemble in part the useful distinction between judgment of agency (JoA) and the feeling of agency (FoA*)* by Synofzik (2008). JoA refers to explicit conceptual attribution of whether one did or did not make an action or cause an effect. FoA refers to the subjective experience of fluently controlling the action one is currently making, and is non-conceptual, unconscious and more related with the sensorimotor prediction (forward model Blakemore et al., 2002). Many studies have used explicit judgements of agency in cases where the facts of agency are uncertain (Wegner et al., 1999; Tsakiris et al., 2006). Several of these experimental studies have investigated agency by asking participants to judge whether they caused a particular sensory event (Farrer et al., 2003; Wegner et al., 2003). In these studies delay between action and effect reduce sense of agency. Moreover greater spatial distortion between the real action and the action's feedback reduce sense of agency. This data shows how agency is strongly dependent on the space-temporal correlation between action and its effects. However, these situations resemble judgements of agency rather than the feeling, or sense of agency (Synofzik, 2008).

Haggard and colleagues (2002) developed an interesting method to investigate one feature of sense of agency. When people make a voluntary action, which causes an external effect, such as a tone after a short delay, they perceive the action and the effect as closer together in time than would be expected from judgements of actions alone, or of the effects alone (figure 4.1). This perceptual attraction across time between actions and effects is not found if actions are replaced by involuntary movements (Haggard et al., 2002). For this reason this perceptual attraction is called 'intentional binding effect'. The intentional binding effect tell us that voluntary movement depends on a cascade of cognitive-motor process which links actions and effects across time, producing a temporal attraction between them (Haggard et al., 2002). This attraction seems an implicit and reproducible quantitative measure related to the feeling of agency.

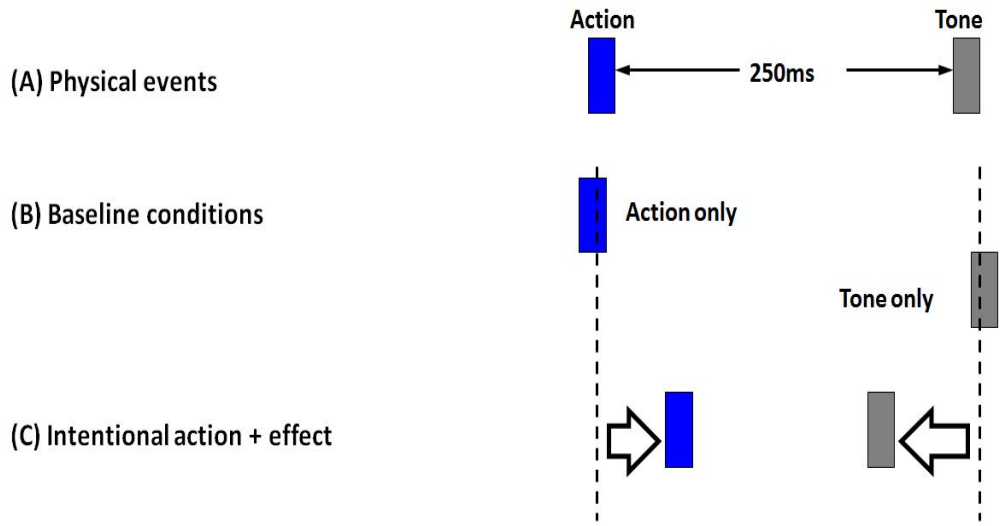


Figure 4.1. The intentional binding effect (adapted from Haggard *et al.* 2005). A) Participants' voluntary key-press actions are followed after 250 ms by an effect (a tone) B) Baseline estimates are obtained for actions occurring without a following tone, and tones occurring in the absence of actions. This controls for individual differences in the perception of these events, and provides a baseline against which to compare the time experience of the same events in an agency or passive context. C) In an agency context, intentional actions are perceived later and the effects are perceived earlier, than their respective baselines (hollow arrows).

Recent studies showed which factors can modulate the sense of agency. These studies are interesting because they show how the sense of agency depends on a causal model that subjects create inferring a statistical relation between action and effect. Moore and Haggard (2008), by varying the probability by which a simple manual action produced an auditory effect, they showed that both the actual and the predicted occurrence of the effect could influence agency. In the block where the predictability of the effect was low the binding effect occurred only on those trials where the auditory effect occurred. In contrast, when the predictability was high the temporal binding occurred even on trials where the action produced no effect. This data suggests that the binding of action towards their effects seems to have two components: a predictive component, which depends on the action reliably predicting the effect, and a postdictive or reconstructive component, whereby the occurrence of the effect triggers a revision of the temporal experience of the action (Moore et al., 2008). In the same way the contingency factor modulates the sense of agency. Contingency is an index of the casual relation between events, and predicts patterns of operant learning animals. Binding increases with the experienced strength of association (statistical contingency) between action and effect (Moore et al., 2009). In both Moore's studies is shown how events entirely independent of the motor system have a strong influence on the experience of action. They suggest that contingency learning makes

a contribution on causal knowledge and also to consciousness of agency. This data show how different information that clearly originates outside the sensorimotor system itself is nevertheless important for the sense of agency. This does not mean that the sensorimotor information is not necessary for the sense of agency but means that agency also depends on a complex and integrate pattern of information: sensorimotor, contextual, statistical and emotional.

Reconstruction implies that the effect of action is important in sense of agency, but no studies have systematically manipulated key features of responsibility, such as the severity of the effect, its moral status, and whether the action directly contributed to causing the effect. In the experimental part of this document, there is the first attempt to understand how social and moral meaning of action could influence the sense of agency. Additional future research could focus on emotive influence and cognitive load on experience of agency.

4.4 Brain circuits of volition

In a recent review, Haggard (2008) reports which cortical motor circuits may mainly contribute to voluntary actions. He shows two sub-circuits, differently both in an anatomical and a functional sense. The first sub-circuit regards basal ganglia and frontal cortex sending signals to pre-supplementary motor area (pre-SMA), which sends output to primary motor cortex (M1). M1 executes motor commands by transmitting them to the spinal cord and muscles. In this circuit pre-SMA seems to have a special role in the experience of conscious intention. Patients during neurosurgeons directly stimulate with low intensity current on pre-SMA have distinct conscious experience of an 'urge to move' (Fried et al., 1991). FMRI studies comparing freely chosen movement with non-freely movement show a strong activation in the anterior part of the pre-SMA (Passingham, 1987). Despite several lineax of evidences supporting the key role of pre-SMA for the internal generation of voluntary action, what is unclear is the independence of this activation from basal ganglia. Several neuropsychologcal and neuroanatomical data (Cunnington et al., 1996) show a subcortical loop starting from the basal ganglia integrating a wide range of cortical signals. Our understanding of cortico-basal ganglia networks is still primitive. In a study by Moore and colleagues (2009) the influence of Parkinson's disease and dopaminergic medication on the temporal experience of voluntary actions is investigated. The data show no difference between healthy volunteer participants

and PD patients *off* medication. This result suggests that the disease state itself is not associated with changes in sense of agency, at least in the earlier stages. However, a significant difference is founded between PD *on* medication and controls. Dopaminergic medication significantly strengthened the temporal binding between actions and effects, which is interpreted as a heightened sense of agency. Increased availability of dopamine strengthened the experience of association between actions and external events, enhancing the sense of agency. These data suggest that dopamine could have a key role on conscious experience of action.

Self control function has also a fundamental social function. From evolutionary point of view the self control mechanisms should have evolved especially for regulation of impulsive responses, as in the cases of delayed gratification, revenge, and what we call in general immoral or antisocial action.

4.5 *Volition in social context: feeling responsible*

All known human cultures have the concept that an individual is responsible for their actions. Responsibility takes both an individual dimension (we have to live with what we do), and a social dimension (society may praise or punish us for what we do). Responsibility for action in turn rests on a concept of voluntary actions: individuals choose and control their own actions.

There are two main approaches in cognitive psychology to investigate responsibility. One approach asks to subjects : 'are you responsible of this action?'. In this case participants perform voluntary action that causes a certain effect which is sometimes veridical and sometimes distorted (spatially, temporally: Farrer et al., 2003; Wegner et al., 2003). Sometimes another possible agent is present performing a similar movement at the same time (Tsakiris et al., 2005; Wegner et al., 2004). These studies generally find a bias to judge oneself to be the author of action. This data show how the experience of being responsible for an action depends on spatial-temporal correlations between one's actions and its effects. In these experiments the actions have trivial effects and there is no real motivation to act.

The second approach studying responsibility comes from the moral behavior literature and investigates the brain activation in moral action. These studies usually induce some sense of moral responsibility presenting script such as: "your mum called you and said she did not feel well. You ignored her, and the next day she died" (Moll et al., 2007) and compare

the brain activation in this kind of script with script without a first personal attribution or non moral meaning. Evidence from these kind of studies (Moll et al., 2002a; 2002b; 2005; 2007) suggests that these complex subjective experiences of moral agency arise from distributed activations in neocortical (anterior PFC) as well as phylogenetically older mesolimbic and orbitofrontal (OFC) regions (Moll et al., 2005). However these studies fail to capture a real sense of causal responsibility and rather seem localize a more general 'moral sensivity' area.

Imaging studies investigating patients with antisocial personality disorder and psychopaths have revealed reduction of grey matter in prefrontal cortex and abnormal brain activation in limbic regions, as well as in the prefrontal and temporal lobes (Muller, 2003). Interesting vmPFC lesions acquired at an early age led to impairments in moral reasoning and behavior, indicating that moral development can be arrested by early PFC damage.

4.6 Volition in social context: imputing responsibility

On the basis of recent finding and theory *dual process theory*, the word 'responsibility' is preferable to guilt because provide a good equilibrium between the emotive component and a cognitive components of moral action.

Recently, Buckholtz and colleagues (2009) used event related fMRI to study regions that were sensitive to information about criminal responsibility. They scanned participants while they determined the appropriate punishment for actions committed by a protagonist in a series of written scenarios. They presented three categories of scenario: Responsibility scenarios describing intentional crime such as rape and murder. Diminished-Responsibility included actions of comparable gravity to those described in the Responsibility set but also containing mitigating circumstances, and a No-Crime set which the protagonist was engaged in non-criminal actions. The data showed activation in the right dorsolateral prefrontal cortex (rDLPFC) that was significantly greater in the Responsibility as compared with the Diminished-Responsibility condition. Bilateral anterior intraparietal sulcus (aIPS) showed a pattern of responsibility-related activity that was similar to rDLPFC whereas the temporo-parietal junction (TPJ) showed the reverse pattern, with more activity in the Diminished-Responsibility as compared with the Responsibility condition. The author's speculate that the early rDLPFC deactivation may reflect a perspective-taking-based evaluation of the beliefs and intentions of the scenarios'

protagonist, which is followed by a robust rDLPFC activation as subjects go on to make a decision to punish based on assessed responsibility and blameworthiness.
Interesting, rDLPFC signal amplitude was not correlated with punishment ratings but was correlated instead with activation in the right amygdala, and other brain regions commonly associated with social and affective processing, including the posterior cingulate, temporal pole, dorsomedial and ventromedial prefrontal cortex, and inferior frontal gyrus.
These data seems suggest these areas can influence the decision regarding the amount of assigned punishment during legal decision-making showing the role of emotion in the attribution of legal responsibility. This data support the idea that the punishment is in first instance a deep emotive reaction and secondarily is used to prevent future harm to society (Withman, 2003). Thus, participants' decisions about punishment amount for each of the crimes depicted in the Responsibility scenarios were strongly correlated with the recommended prison sentences for those crimes, according to real sentencing guidelines. The higher activation of rDLPFC in the Responsibility compared to the Diminished-Responsability condition and during punished versus non-punished trials is therefore consistent with a role for rDLPFC in the suppression of emotional reactions (Hausehofter, 2009).

Experience of volition and moral behavior involve a common neural pattern. There is a motivation and basic emotional state which depends on cortical limbic state and a cognitive component that is mostly cortical and related with control and inhibitory function. On the basis of the limbic circuit conscious will, could have many of the qualities of an emotion. This emotion allows us to develop a constant and stable sense of who we are and could be perceived as a sort of positive emotive reward. Most important, this emotion of conscious will allows us to maintain a constant sense of responsibility for our actions that serve as a basis for morality.
Conscious will could be a sort of strategic tool to give meaning to all unconscious causes that move the human actions. In the same way morality is a tool that shapes social behaviour. Future research should investigate more systematically the interaction between emotional state and control function in experience of voluntary and moral action. Certainly culture and education represent a powerful learning signals for brain's emotive-cognitive and motor circuits.

4.7 Experience of agency and sense of responsibility

Folk psychology assumes that individuals choose and control their own actions. Further, they are aware of what they are doing as they do it, and can normally predict the consequences of their actions, or at least the proximate consequences. Therefore, their conscious knowledge of what they are doing should allow them to choose between right and wrong actions. This view of human action is pervasive in human life, since it forms the basis of the legal concept of a criminal action. For example, systems deriving from Roman Law require that a crime involve not only a physical act (*actus reus*), but also a corresponding intention (*mens rea*). Thus, a person may have diminished responsibility for their action if they did not intend or could not foresee the consequences of the action, or if they performed the action without intention under duress.

In normal circumstances, people readily associate actions with their outcomes. In particular, people have a distinctive 'sense of agency', or feeling of control, for events caused by their own actions, but not for other events (Synofzik et al., 2008). The sense of agency is clearly related to personal responsibility. However, the relations between choice, action and responsibility remain unclear, for three specific reasons. First, most studies of voluntary selection of action have studied actions devoid of any meaning or consequence, and often without any significant element of choice (Fleming et al., 2009). Second, quantifying the sense of agency is problematic. Most existing studies have relied on explicit judgements of authorship in cases where it is uncertain whose action caused a given effect (Wegner & Wheatley, 1999; Tsakiris et al., 2007). However, these situations resemble judgements of agency rather than the feeling, or sense of agency (Synofzik et al., 2008).

Here we have used the intentional binding effect (Haggard et al., 2002) as an implicit, quantitative measure related to the sense of agency. When people make a voluntary action, which causes an external effect, such as a tone, after a short delay, they perceive the action and the effect as closer together in time than would be expected from judgements of actions alone without tones, or of effects alone without actions. This temporal attraction across time between perceived actions and effects is not found for involuntary movements (Haggard et al., 2002). No previous studies have investigated whether this measure changes according to key features of responsibility, such as whether

the action directly contributed to causing the effect, the outcome value of action, or the moral significance of the effect.

We have therefore used the intentional binding effect to investigate the relation between experience of action, and responsibility. In particular, we set out to ask the following questions:

1. Does the experience of an action depend on the outcome value of action?
2. Are actions that have morally significant effects experienced differently from other actions?
3. Does the experience of an action vary according to whether the action directly causes the effect, or merely allows continuation of a chain of events which would anyway lead to the effect?

To investigate these questions, we have combined a low-level measure of action experience (intentional binding) with high-level action scripts previously used in studies of moral decision-making. These neuropsychological study on moral decision-making assume that the *experience* of action varies according to the anticipated moral-emotional consequences. Further, it is assumed that the experience of action and consequence forms a key part of homeostatic design (Damasio, 1994), influencing participants to choose actions in a way that avoids excessive feelings of responsibility for undesirable outcomes. However, very few studies have directly examined how moral-emotional consequences influence the basic phenomenology of action. To investigate this question, we used a low-level measure of the association between an action and a subsequent external effect, namely the temporal attraction, or "binding" between the perceived time of action and effect (Haggard et al., 2002). Such implicit measures have the advantage that they do not explicitly require any judgement about the significance or value of either action or effect.

We embedded the temporal binding task within several standard moral and economic dilemmas (Greene et al., 2001). We presented the dilemmas, initially as verbal scripts for familiarisation, and then represented them as visual schematics during time estimation trials. We required to subjects to choose between either acting to intervene or not in the circumstances of the dilemma, and to indicate their action choice by a left or right index finger movement. During re-presentation of the dilemmas, participants viewed a rotating clock hand, which they used to report either the time of their action, or the time

that they saw a visual representation of the consequence of their action (Haggard et al., 2002). We predicted that the experience of action would differ according to the moral content of the dilemma, the outcome value, and whether the action directly caused the effect, or merely allowed an existing causal chain to continue.

METHOD

Participants

Thirteen individuals (5 females, 8 male, aged 22-47 years: mean 30 years) took part in the study with ethical committee approval and on the basis of written informed consent. All participants had normal or corrected to normal vision. The data from 2 subjects were excluded because they failed to recognize the content of the schematic pictures during the experiment. A third subject had an unusually high variability in temporal judgements (standard deviation of judgement errors more than 2 times that of the mean of the other subjects) suggesting a particular difficulty in using the clock to report subjective experiences. The analyses presented here are therefore based on 10 subjects, (5 females, aged 23-47 years: mean 31 years).

Stimuli

Nine choice scenarios were presented as verbal scripts, and also as visual schematics. An example is the Trolley dilemma used by Greene and colleagues (2001). Each picture (97 x 69 mm) shows on the left, the content of the story, and on the right branching line representing choice of two different outcomes. Six scenarios involved morally significant choices, while three had purely economic significance (Moral – Non moral conditions). The overrepresentation of moral scenarios partly reflected our particular interest in this condition, and also aimed to correct for imbalances in subjects' choices (see below) that we expected on the basis of pilot data. Example pictures are given in figure 4.2, and the full scripts are given in supplementary material. Each choice scenario involved choice between two possible outcomes. While both outcomes were always negative, they differed in value (severe, moderate). For example, the trolley dilemma the severe outcome is the death of 5 workers whereas the moderate outcome is the death of only one. The pictorial representation of the two outcomes was balanced so that each appeared equally often as upper and lower branch of the dilemma on the screen.

Further, subjects could make their choice with two different levels of control over outcomes. On 90 trials, question marks beside the two dilemma branches indicated a blind decision: participants did not know which branch of the decision is preselected (figure 4.2). In these *unpredictable effects* trials, the participant chooses whether to STAY with or CHANGE the preselected outcome, without knowing in advance what the preselected outcome is. 250 ms after each keypress, a smaller picture (44 x 30 mm) showed participants a vignette of the effect that their CHANGE or STAY choice had on these trials. On six further trials (3 moral and 3 non-moral) an arrow appeared along one branch of the dilemma, to show which choice is preselected. In these *predictable* trials, the subject knows that pressing a 'STAY' key will cause the event shown by the arrow, while pressing CHANGE will cause the other outcome. The arrow always pointed at the most severe outcome (e.g. the track with 5 workers rather than 1 worker, see the first picture in figure 4.3). This small number of predictable trials had two control functions, first to check if subjects could understand the picture and choose the answer with less negative effect especially in non-Moral dilemma, and second to imply by context that the subject could really choose between outcomes in unpredictable trials.

Procedure

Participants were first familiarized with the choice scenarios by listening to the scripts for each scenario and watching the corresponding picture. During the experimental session only the pictures were presented. At the start of each trial a picture of the scenario and the possible outcomes appears (Figure 4.3). At the same time, a clock hand was seen superimposed on the picture, with a single hand rotating every 2560 ms. After 800 msec the picture disappeared but the clock continued to rotate. The participant then decided which of two keys to press: the 'STAY' key (f9 on the keyboard, pressed with the right hand for 6 participants, f4 pressed with the left hand for 4 participants), or the 'CHANGE' key (f4, pressed with the left hand for 6 participants, f4 with right hand for the other). Participants were told that they could press when they wished. Pressing caused a picture, showing the effect of the action, to appear after a delay of 250 ms, and remain visible for a random interval (1500-2500 ms). When a preselected action outcome was shown (predictable effects condition), pressing the STAY key allowed events to run their course towards this outcome, and the corresponding effect picture was shown after the keypress, while pressing the CHANGE key selected the alternative outcome. When no preselected outcome was shown (unpredictable effects condition), one of the two possible effect

pictures was shown at random, regardless of whether participants pressed the STAY or the CHANGE key (see fig 2). The incidence of moderate and severe outcomes was balanced between STAY and CHANGE responses.

The clock stopped a random interval (1500-2500 ms) after the effect picture. Participants then made one of 2 judgments in separate blocked conditions. In one block they judged the time of their keypress, in another the time of onset of the effect picture. There were 96 trials in each block, corresponding to 2 types of script (Moral, Non-moral), 2 levels of outcome values (Severe and Moderate). Six trials presented an arrow indicating the current setting of the dilemma (predictable trials),
while 90 trials presented question marks, indicating a blind decision. The six trials with arrow serve only as manipulation check, and timing data for these trials were not analysed.

At the beginning of each block there was a training session of 10 trials. Further, to ensure that participants considered the choice scenarios carefully, rather than just pressing keys at random, after occasional trials the experimenter asked the participant to report: the content of picture, the magnitude and presence of arrow or question mark indicating outcome predictability. Finally, participants also performed two baseline blocks. In one baseline block, they randomly selected between F4 and F9 keys, pressed the selected key at a time of their own choosing, and indicated the time of the keypress. No choice scenarios or effects of action were presented. In another baseline block, participants did not make any actions, but simply viewed the effect pictures occurring after a random delay, and reported their time of onset. There were 30 trials in each baseline block. The order of the two experimental and two baseline blocks was randomized. These blocks gave baseline estimates of the perceived time of actions and effects. Subtracting each subject's corresponding baseline from their judgments in the experimental condition gave an estimate of the binding between perceptions of actions and picture effects in experimental trials where the participant's action caused the subsequent display of the picture effect.

To summarise, we asked subjects to observe the picture and decide between two possible solutions of the dilemma by pressing the 'CHANGE' or 'STAY' key. The participant also had to report the position of the clock hand on each trial. The event that they judged varied between blocks. In one block the judged the time at which they pressed

the key. In another blocks they judged the time at which they saw the picture showing the effect of their choice.

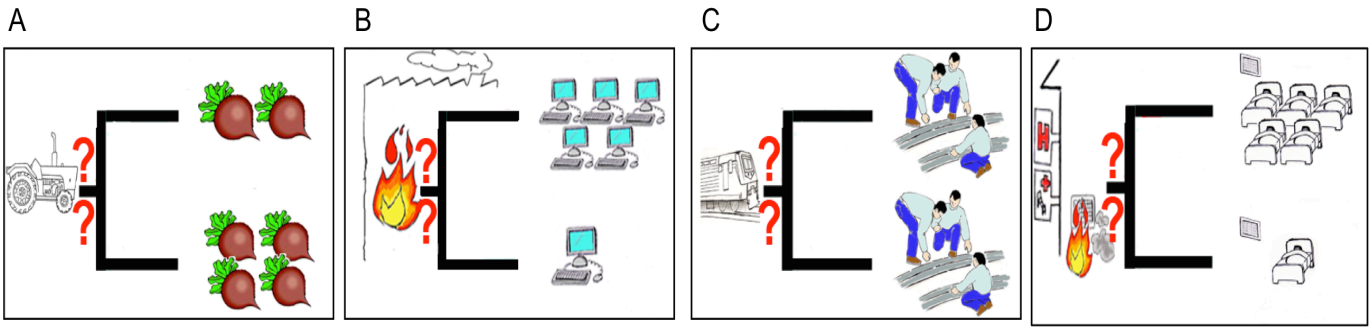


Figure 4.2 Examples of typical pictures shown in the experiment. Here are presented two non moral scripts , frame A e B and two moral scripts C, D. Letters indicate the corresponding scripts in supplementary materials. The pictures shown were presented briefly at the start of each trial, and then disappeared. Each picture shows a dilemma in which the event shown on the left will lead to either the upper or lower outcome shown on the right. The question marks indicate that the participant cannot know which outcome has been preselected. In control trials, a single arrow designated that the upper or lower outcome had been preselected. The participant pressed one of two keys to indicate whether to STAY with the current setting, or CHANGE to the other alternative. Nine illustrative scenarios are presented; 6 involve human life, while 3 involve material goods. The outcomes have two alternative values 1 or 5 units, and the participants make blind decisions regarding whether to STAY or CHANGE. A second picture corresponding to the outcome of the participant's action (shown as upper or lower branch of the right hand side of each picture) was shown 250 ms after the action. All picture stimuli were presented in colour.

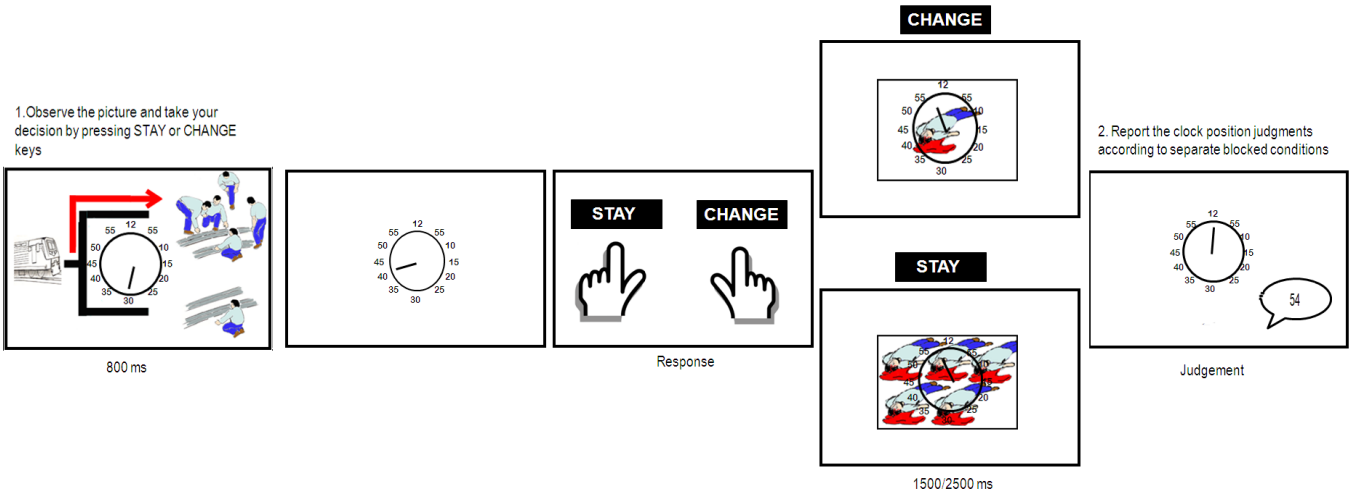


Figure 4.3 Schematic representation of a control trial in the experimental blocks. The arrow is pointing the worse effect (5 workers). If subjects press stay a picture with 5 dead worker appears, 250 ms later, otherwise if subjects press change a picture with 1 dead worker appear.
At the start of each trial a picture of the scenario and the possible outcomes appears. At the same time, a clock hand was seen superimposed on the picture, with a single hand rotating every 2560 ms. After 800msec the picture disappeared but the clock continued to rotate. Subjects had to decide which solution for the dilemma and in the same time pay attention to the clock hand position. In one block they have to report the clock hand position when they pressed the key, in another block they have to report where was the clock hand position when they picture with the effect appeared.

Design section

In the experiment, three variables are measured: choice between alternative action situations, and estimates of the time of actions and of their effects.There are two conditions, in which actions have predictable and unpredictable consequences respectively. However the predictable condition is only a check condition in order to control if subjects are able to recognize and respond properly at the task. The experimental design involves four within-subject factors: action chosen (CHANGE, STAY), choice context (moral, non-moral), outcome value (severe, moderate), and judgement type (action, effect).

For the analysis of choice, since the two responses are complementary alternatives, statistical analysis focused only on the number of CHANGE responses. CHANGE responses in predictable trials are analysed to check if subjects are able to recognize and respond properly at the task. 2x2 ANOVA was performed for CHANGE responses in unpredictable trials with judgement type (Action, Effect), choice context (moral and non-moral choices) as main factors. This analysis has the aim of observing if there is any difference in responses between blocks and choice contexts.

Time estimation was only analysed for unpredictable trials, in which participants made blind decisions. A preliminary analysis on baseline judgment time estimates is made in order to check if differences in visual salience of the effect pictures might have influenced judgements. 2x2 ANOVA was performed on baseline time estimates as a function of choice context (Moral, and Non Moral) and outcome values (severe, moderate).

The following analysis has the aim of exploring how outcome value and decision modulated the temporal experience of action and effect. Separate ANOVAs were performed for baseline-corrected time estimates, one for action judgments and one for effect judgements. 2x2x2 ANOVA was performed on baseline-corrected time estimates as a function of choice action (CHANGE, STAY), choice context (Moral, Non Moral) and outcome values (severe, moderate). The same analysis is performed on an overall binding measure, defined as action binding minus effect binding, to capture the perceived temporal association between action and effect.

RESULTS

Choice responses

Since the two responses are complementary alternatives, statistical analysis focused only on the number of CHANGE responses. In control trials, when an arrow showed participants the preselected outcome, CHANGE responses were overwhelmingly chosen because the arrow always indicated preselection of the most severe outcome. This effectively serves as a manipulation check: subjects clearly understood the dilemma pictures, and chose less negative outcomes when they could predict the outcome. Thus 'CHANGE' responses were made on 92% (SD across subjects 0.7%) of predictable trials.

When no preselected outcome was shown, participants effectively made blind decisions. In this case, 'CHANGE' responses were reduced, to 48% of trials in moral contexts, and 54% in non-moral contexts. 2x2 ANOVA was performed for 'CHANGE' responses with factors of judgement type (Action, Effect), choice context (moral and non-moral choices) as main factors. The ANOVA showed no significant effects or interaction (all $p>.1$). This suggests that there is no difference in number of 'CHANGE' between blocks and between choice contexts.

Time estimation results

First, in order to check if differences in visual salience of the effect pictures may have influenced judgements, the mean time estimates in baseline effect judgements for each participant were calculated as a function of choice context (Moral, and Non Moral) and outcome values (severe, moderate). The 2x2 ANOVA showed no significant results ($p >= .7$). Therefore, remaining analyses considered that any differences in time estimation for different classes of effect pictures reflected the pictures' meanings rather than their visual surface form only.

Time estimation was analysed only for unpredictable trials, in which participants made blind decisions. Baseline corrected time estimates were obtained for each participant's action judgments by subtracting average time estimates for actions in baseline conditions from time estimates in the experimental blocks. This subtraction gives a measure of binding for actions. By convention a positive value indicates a delay in action awareness of action, towards the subsequent effect. Similarly, time estimates of effect

onset were corrected by each subject's mean judgments in the baseline effect condition. A negative value indicates an anticipatory shift towards the preceding action. The baseline-corrected data are shown in Table 4.1. The table shows the standard intentional binding effect reported previously (Haggard, 2002; Clark, 2002; Kalogeras, 2002). That is, the perceived time of actions is shifted later than baseline values, towards the subsequent effects, while the perceived time of effects is, in general, shifted earlier than baseline, towards the preceding action. Mean time estimates in baseline action judgements were subtracted from estimates in experimental action judgements to measure binding for actions.

Further, the mean of each participant's estimates for all pictures in the baseline judgements was subtracted from estimates for each type of picture in the experimental conditions. This subtraction gives the shift in the perceived time of each picture due to the participant's action, and is used as a measure of binding for effects of action. Baseline-corrected time estimates for the unpredictable condition were analyzed as a function of outcome value (Severe, Moderate), choice context (Moral vs Non-moral), and action chosen (CHANGE vs STAY). Separate ANOVAs were performed for the action judgments and effect judgements. ANOVA of action judgments did not show any significant effects or interactions (all $p > .1$). In contrast, ANOVA of effect binding data showed significant main effects of context $F(1, 9) = 4.5$ $p = .05$. Specifically, moral contexts produced stronger effect binding than non-moral contexts $F(1, 9) = 12.6$: $p = .02$. There was also a main effect of outcome value $F(1, 9) = 8.04$ $p = .02$ (see Fig. 4). There was no significant main effect of action choice (CHANGE/STAY) $F(1,9) = .14$ $p = .7$, and no significant two-way interactions. The three-way interaction between chosen action, context and outcome value is close to significance $F(1, 9) = 4.2$ $p = .066$

Finally, we calculated an overall binding measure, defined as action binding minus effect binding, to capture the perceived temporal association between action and effect. The overall binding measure showed similar results to effect judgments (see figure 4.4). Specifically the 2x2x2 ANOVA of overall binding data showed significant main effects of moral context $F(1,9) = 7.8$ $p = .02$ and outcome values $F(1,9) = 7.07$ $p = .02$. There were no significant interactions between factors (all $p>.05$).

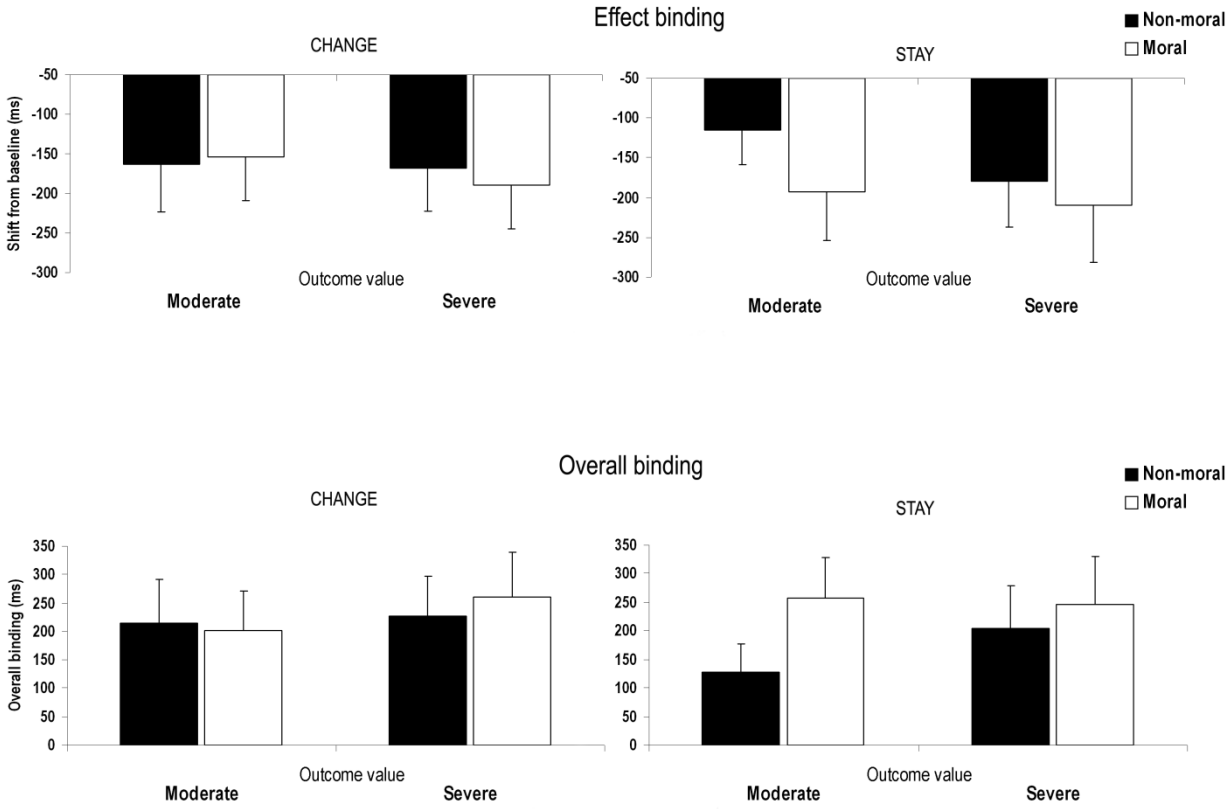


Figure 4.3 Effect binding; mean baseline-corrected effect estimates (ms). Overall binding, action binding was subtracted from effect binding. Bars show standard error across participants.

			Action binding ms (SE)	*Effect binding ms (SE)*	*Overall binding ms (SE)*
Moral	Moderate	CHANGE	-5 (28)	-154 (141)	202 (69)
Moral	Moderate	STAY	12 (24)	-192 (172)	257 (71)
Moral	Severe	CHANGE	17 (27)	-189 (169)	259 (80)
Moral	Severe	STAY	-16 (35)	-209 (172)	246 (83)
Non-Moral	Moderate	CHANGE	-2 (26)	-163 (121)	214 (77)
Non-Moral	Moderate	STAY	-41 (34)	-115 (144)	127 (50)
Non-Moral	Severe	CHANGE	6 (26)	-169 (152)	228 (69)
Non-Moral	Severe	STAY	-27 (29)	-179 (179)	205 (74)

Table 4.1 Mean binding effects (SE = standard error across participants)

DISCUSSION

To summarise, we found an expected effect of outcome values on action decisions. When participants chose whether or not to intervene to change a predicted outcome, they

consistently chose to produce the least negative outcome. This confirms that participants understood the choice scenarios and processed them in utilitarian way. Interestingly, they did so in both moral and non-moral dilemmas. However, when the prediction of outcome was not possible, the percentage of change and stay responses in moral and economic scenario was not statistically different.

Temporal estimates of action and effects showed the 'intentional binding' effect reported previously (Haggard, 2002; Clark, 2002, Kalogeras, 2002). Previous studies showed strong binding of auditory and somatic effects towards actions, with weaker binding of actions towards effects (Haggard, 2002; Clark, 2002; Kalogeras 2002; Haggard & Cole 2008; Tsakiris & Haggard, 2003). In our experiment, we confirmed the same pattern of intentional binding of action when effects of action are presented visually, and in the same location as the clock used for time estimation.

More important, our finding shows how the binding effect is modulated by factors relevant to responsibility, such us the moral context of action and the action outcomes. Our temporal measures of action awareness confirmed that a broad concept of responsibility pervades action awareness. Specifically, we found two results that are readily interpretable as enhancing the sense of responsibility. In each case, enhanced responsibility was associated with stronger binding, shifting the perceived time of effect towards the action. First, we found that moral contexts showed stronger effect binding than non-moral contexts. Second, we found stronger binding when the participant's action lead to the more negative of the two possible outcomes. Interestingly, the modulation of effect binding by outcome value was present even though participants made blind decisions, and thus could not actually control whether the outcome was severely negative or moderately negative. That is, subjects experienced a stronger temporal association between their actions and severely negative outcomes, despite knowing that they were not, in fact, responsible for these outcomes. Finally, these effects did not interact, suggesting that moral context and outcome value of action have additive effects on the experience of control.

Taken together these data suggest that action awareness is modulated by the *impact* of a participant's action in a situation. When the impact is increased, the link between action and effect is strengthened in subjective experience. This impact could depend both

on emotive activation due to the moral content of action, and also on outcome value of action.

An alternative explanation of our results could be related to differences in the visual salience of effect pictures used for different levels of the various experimental factors. However we can exclude this possibility because comparing the time estimation in baseline effect in moral and economic context there is no significant difference between them.
Our results can also be considered from the prospective of event predictability. Pariyadath and Eagleman (2007) recently demonstrated that duration judgments vary as a function of predictability, such that predictable events are judged to be shorter than unpredictable ones. This interpretation could be related to the subjects' attempt to create or infer a statistical relation between action and effect. In our case we did not formally debrief subjects to ask how they interpreted the relation between action and effect. They might, for example, have automatically associated the moral context of action with severe outcomes. This might in turn lead to stronger temporal association between action and effect observed, following Pariyadath and Eagleman's hypothesis. In this case the greater binding for moral contexts, and for severe outcome is not due to the nature of the context or outcomes per se, but instead to the subjective degree of predictability. However our stimuli were balanced to be equally repetitive and unpredictable.

A tight temporal association between action and effect seems to be a low-level phenomenal marker of the attribution of responsibility. Our results show that this impact has at least two separate dimensions: magnitude and moral/emotional significance. In contrast, whether the participant intervened (CHANGE choices) or merely let events run a preselected but unknown course (STAY choices), had no significant effect. The irrelevance of the decision factor for experience of action probably arises because STAY and CHANGE are effectively meaningless in 'blind' and unpredictable decisions. The two action choices in our experiment therefore had only the formal appearance of a 'what' decision (Brass & Haggard 2008), rather than the real content of 'what' decision, because the choice was not supported by any actual association with the action outcome.

Our findings showed that we experience the closest binding between action and effect when the situation has moral rather than simply material importance, and where the

consequences of intervention are potentially great. We suggest that these two factors could contribute to binding in different ways. The factor of moral significance of action presumably represents *contextual* modulations of action awareness. This factor should influence the experience of action and effect quite generally, from the very beginning of decision making process, right through to perceiving the effects of action. In contrast, the factor of outcome value is quite time-specific in our experiment. Participants did not know whether their action had produced a severe or moderate effect until the second picture was presented, 250 ms *after* action. Thus, the effect of outcomes' value on binding cannot be based on *prediction* or knowledge about the significance of action. Instead, it can only be retrospective. We speculate that the sense of responsibility involves two components: a general knowledge that one performs an action and is therefore responsible for all its effects, and a specific enhancement of this experience when the effects of action are particularly important. Interestingly, in our experiment, this enhanced, retrospective experience was entirely illusory, since the participants did not in fact have any influence over the outcome value of action outcomes in these unpredictable, 'blind decision' trials.

We needed to introduce the unpredictable trials in order to balance the number of responses between context and effect, but this imposes a great limitation. Stay/change with an unpredictable outcome is effectively like a guess, and the 50/50 responding suggests participants treated it as such, reducing the level of intention in the choice and in action. In a fully intentional choice condition, such as our predictable condition, participants generally chose the least severe outcome, making factorial analyses very unbalanced. Some method to control the level of intentionality while maintaining a balance of more and less severe outcomes, would have helped considerably with the interpretation of the findings.

Our results have interesting implications for both action awareness and for concepts of responsibility. Our results suggest that the temporal experience of linkage between voluntary actions and outcomes, or which forms part of the sense of agency, is not merely a 'cold' cognition based on learning associations through 'constant conjunctions of events' (Hume 1739/1888). Rather, subjects experience strong linkage of actions to their effects, when actions are morally and emotively important, and when actions are found to produce important outcomes. We used an implicit measure of sense of agency, namely the temporal structure of action-event relations, to reveal this enhancement.

One modern view treats moral principles and moral responsibility as an institutionalized social expression of basic emotions that are subsequently rationalized (Haidt, 2001), and identifies them with the orbitofrontal brain areas (Koenigs, 2007; Ciaramelli, 2007). However, the connection between moral judgements and action choices has rarely been considered. We show that the impacts of action decisions, including moral impacts, have a direct effect on the primary experience of action outcomes.

This opens the interesting possibility that a deficient sense of moral responsibility in individuals with developmental disorders or acquired brain damage may be caused by deficits in primary experiences of action-outcome linkage. We have demonstrated an enhanced connection between experience of actions and external events when actions are important, and when they have moral significance. We speculate that this 'impact enhancement' of the link between actions and outcomes could be produced by a specific cognitive module housed in a yet-unidentified brain area. This would make it logically possible that developmental pathology or acquired lesion could prevent some people from experiencing the moral impact of their own actions. This hypothetical patient would then raise an interesting and socially important neuroethical problem, since punishing them for 'immoral behaviour' would seem worryingly close to punishing them for their brain lesion. Currently, such individuals are often judged 'not guilty by reason of insanity', on the grounds that they do not understand the consequences of their own actions (Moran, 1981).

CONCLUSION

For long time, decision making has been considered as only influenced by cognitive process, a matter of estimating which of various alternatives actions would yield the most positive consequences for the agent. This explanation was particularly useful for the attempt of decision theorists to define how we should decide in order to be 'rational', where the word 'rational' often was used as a synonym of optimal decision. The rational decision was traditionally opposed to irrational decision, where irrational meant a decision depending not on rational/cognitive deliberation but on a passion beyond control that can override reason, deliberation and/or self interest. In most of recorded human intellectual history, as within literature and philosophical discussion, emotions have been viewed in largely negative terms, especially due to their unpredictable corrupting influence. Examples of this idea are still present in modern legal systems, in which "crimes of passion" are treated differently because the perpetrator is viewed as being "out of control". Two prejudices have driven previous research about decision making: i) decision making is only a cognitive rational process and ii) the influence of emotions is only negative and against rationality.

Only recently has there been a new appreciation of the positive functions served by emotions in governing decisions (Damasio 1994, Greene et al., 2001; Phelps, 2009; Loewestain & Lerner, 2003; Loewestain, 2004). Although emotion was considered an important variable in this new positive prospective, the role of emotion in decision making has rarely been coupled with the detailed investigation of the range of components, factors and measures that have characterized the psychological study of emotion and affect. Few studies of social and economic decision making have explicitly manipulated or measured

emotion or affect variables. In many studies and theories, especially in neuroeconomics, emotion is inferred, but not directly altered or assessed. The aim of the experimental part of this research work was to increase the availability of data about the role of emotions in decision making, proposing several tasks where it was attempted to manipulate and measure emotions in order to better understand the different ways in which emotions enter into decision making and understand which brain areas are mainly involved. In this last part of my thesis, using the new data provided in the experimental sections, I will try to discuss the possible role of emotions in the choice process.

Loewestein and Lerner (2003) suggested two different ways in which emotions could enter into the decision making process. The first influence consists of predictions about the emotional consequences of decision outcomes: expected or anticipated emotions. Expected emotions reintroduce the concept of expected utility model (see chapter I and II), assuming that people attempt to predict the emotional consequences associated with alternatives courses of action and then select actions that maximize positive emotions and minimize the negative emotions. The second kind of affective influence on decision making consists of emotions that are experienced at the time of decision making: immediate emotions. The model proposed regards decisions under uncertainty where subjects have to be able to predictions, the possible decision outcomes, the probability of each outcome and the emotional consequences of decision outcomes associated with them.

In experiment III we observed the role of expected emotions associated with non-ambiguous alternative solutions. We chose a 'non-uncertain' decision regarding moral content. In a non-ambiguous choice, subjects can foresee the certain effects of their decision. The moral context permits a modulation of affective response (personal and impersonal conditions). For example, in the footbridge dilemma, the personal dilemma with the highest affective activation, a subject can decide whether to push a single victim off of a bridge in front of a runaway trolley in order to stop its progress toward five victims or to leave the trolley killing five men. The expected consequences are clear: one dead man killed by the subject's direct action or five men killed by the trolley. The expected emotions can be: a sense of responsibility and guilt in the first case, and a general sense of impotence in the second case. The situations involving dilemmas that were proposed in the task induce a conflict between a rational solution (better one man than five men) and a moral solution (follow the commandment: not kill). In front of a not easily solvable

conflict, people could use their feeling associated with each possible solution to form their judgment. Emotions become thus a kind of information used to solve the conflict. Interestingly this implies that individuals care about emotional attributes of a choice alternative. Psycophysiological data presented in study III seems to confirm this interpretation. Both healthy subjects and brain-damaged control patients (non-FC) exhibited increased skin conductance activity several seconds before choosing the utilitarian option in personal moral dilemmas, for instance, deciding that it would be appropriate to kill one person in order to save others. In control groups, emotional/somatic signals were critically recruited during moral judgments, and characterized the anticipation of personal moral violations. Importantly, somatic responses seem to have shaped personal moral judgment. A preliminary analysis showed a negative correlation between anticipatory skin conductance activity and frequency of utilitarian responses in normal controls, such that individuals with higher SCRs before utilitarian choices were more reluctant to judge moral infractions as acceptable behaviours than those with lower SCRs. One possibility, therefore, is that emotional responses mark utilitarian choices in personal moral dilemmas with an emotive negative judgment, discouraging the selection of those options in future decisions. This negative emotive judgment in a personal moral dilemma could correspond to the expected emotions such as the sense of responsibility and guilt. In order to avoid this negative emotional experience, subjects prefer to take the non utilitarian solution. Another interpretation, applying a modified version of regret theories (Bell, 1982)[12], suggests that the intensity of an experienced emotional reaction could depend on the affective comparison between the two possible solutions. In our case the sense of impotence forecasted in a non utilitarian solution, compared to sense of guilt forecasted in a utilitarian solution, is less negative and less intense as showed by skin conductance activity. The selection of the less negative solution on the affective point of view suggests a modified version of utility function[13] involving mainly "affective forecasting" (Loewestein, 2004).

[12] The original regret theories are normally applied for decision under uncertainty. The theories assumed that the intensity of experienced regret depends on a simple comparison of the outcome one experiences against the outcome one would have experienced if one had made a different choice.

[13] We refer to 'utility function' and not to 'expected utility function' because the decisions required are not under uncertainty.

vmPFC patients did not generate SCRs in anticipation of utilitarian choices in personal moral dilemmas. These findings indicate profound differences in the making of moral judgments between vmPFC patients and controls. In contrast, no apparent emotional/somatic response accompanied personal moral violations in vmPFC patients. This suggests the incapacity of vmPFC patients to give different affective attributes to the utilitarian and non utilitarian solutions and thus the impossibility to follow a utility function based on 'affective forecasting'.

Not all situations involving choices have the entire set of necessary information for decision-making. On the contrary, most of our choices are taken under uncertainty. What is the role of emotion in decision under uncertainty and risk? We believe that in case of decision under uncertainty emotions could have a more intense influence on the decision making process. As reported in the case of non-uncertain decision (see above), there is an affective influence by the predictions about the emotional consequences of decision outcomes (expected emotions) that could guide the selection of option with the more positive forecasted emotion (affective utility function). The second influence regards the immediate emotions that are experienced at the time of decision making. Immediate emotions influence decision making via two routes, direct and indirect (Forgas, 1995; Loewestein & Lerner, 2003). Indirect effects are those that are mediated by changes in expected emotions or changes in the quality and/or quantity of information processing. Direct effects are those that are not mediated by changes in expected emotions or in cognitive processing (Forgas, 1995; Loewestein & Lerner, 2003). The immediate emotion appears to play a largely advisory role in relation to the first impact of a subject in facing the decision. According to the 'emotion as information' hypothesis (Damasio, 1994), in ambiguous situations people ask themselves "How do I feel about it? And then use their present feeling to form the judgment. Interestingly, the influence of immediate emotions depends on the level of knowledge and on the level of ambiguity of the situation. In unfamiliar contexts, immediate emotions influence our choice more than in familiar contexts (Srull, 1984).

I investigated choice in situations with very high levels of uncertainty. The trust game is indeed a risky situation in which is not possible to calculate or foresee the possible economic and social outcomes. The uncertainty of the situation, the absolute lack of information and the fact that you can lose money, could generally induce an immediate

emotion of insecurity (fear) and irritation. This immediate emotion could be the 'mood' where is grounded the process of affective utility function related to the expected emotions associated with the two possible outcomes (cooperation/increase income, betrayal/lose of money). In the study I, we found that investors in the vmPFC group showed higher money transfers to their partners than those in both control groups, thereby suggesting that damage to this brain area increases investors' 'trusting' behaviour considerably. This data suggest on one hand, that vmPFC subjects could have distorted immediate emotion, for example they could not feel the insecurity (fear) and irritation; and on the other hand they could be not able to consider in their forecasted emotion, especially the negative emotion related to the possibility of betrayal and loss of money.

Comparison between the same decisions in social (trust game) and non-social (risk game) contexts showed that vmPFC patients were unable to consider in their choice the specific affective information of social interaction. These findings are highly compatible with current theories maintaining that vmPFC is a critical neural substrate for forecasting the (negative) emotional consequences of available options in order to guide future behaviour, both in personal and societal decision-making (Bechara & Damasio, 2005). Emotions seem to reveal their fundamental guiding role especially in our social interactions.

Emotions exert not only a direct influence on behaviour but also an indirect influence via their impact on judgments of expected consequences and emotional reactions to them, as well as the quality and quantity of information processing. The indirect influence of emotion on decision making was investigated in study II, where emotional stimuli were introduced incidentally during a decision making process. In that study, the presentation of subliminal happy faces increased investors' transfer amounts relative to neutral faces, whereas fear expressions failed to affect such amounts compared to neutral expressions. Studies investigating trust without any incidental emotional stimulation (Berg et al., 1994; McCabe et al., 2001) have shown that investors usually send to the trustee approximately 30-40% of the money available. In study II, participants exposed to fearful and neutral expression showed a level of trust similar to that reported in other studies without emotional stimulation. By contrast, subjects exposed to subliminal happy faces showed a greater level of trust. Our findings confirmed that incidental emotions can be effective in influencing trust. More generally, the data can support the hypothesis that accidental emotions can influence people's judgments of the probability of positive and negative

outcomes (Loewestain & Lerner, 2003). A possible explanation of the null effect of fearful faces on trust behaviour could concern a sort of 'floor effect'. In the trust game, the immediate emotion of decision is insecurity (fear) and irritation. Moreover, there is negative expected emotion related to the possibility of being betrayed and losing money. As such, fearful faces cannot further reduce trust. Furthermore, extremely low levels of trust in social exchanges may be not viewed as pertaining to cautious decision, but perceived as signs of hostility or punishment towards others. Therefore, it is possible that trust cannot be symmetrically influenced by positive and negative emotional stimuli, because in the anonymous one-shot interaction as those used here, trust is already very low. The effect of a happy face on trust level could occur because the emotional primes temporarily change the accessibility of knowledge relevant for interpreting the ambiguous situation (Higgins, 1996). Specifically, incidental emotion can influence people's perception not only about the likelihood of different outcomes but also of how they will feel about those outcomes. Another important result of this study concerns the fact that only unconsciously perceived happy faces were effective in modulating trust behaviour. In accordance with Dunn & Schweitzer (2005), we found that subliminal happy faces increased trusting behaviour relative to subliminal neutral (and fearful) faces, but supraliminal happy faces failed to do so. It is possible to speculate that cognitive control mechanisms reduce the effects of supraliminal emotional stimulation, or that these control mechanisms select the relevant information and inhibit irrelevant information. This explanation is congruent with several fMRI studies (Hariri et al., 2003; Critchley, H. et al., 2000) that have shown that amygdala activation decreases when participants attend to faces in order to evaluate emotional features, relative to when participants make a non-emotional judgment of face gender, such that the emotion of the face is completely irrelevant to the subject's task.

In the previous three studies we presented how expected emotions, immediate emotions and incidental emotions can affect our decisions. However these studies do not consider the last component of a decision: the selection of an action representing our choice. The selection of action is not a merely instrumental realization of our choice, but is actually the real and probably the unique evidence of the choice and the possible linkage between decision and outcome. This voluntary action, more than a simple evaluation of option and the information processing, makes us responsible for a choice. The experience of responsibility/agency is the subsequent feeling that one's action has indeed caused a

particular external event (Haggard, 2005) and necessarily involves the experience of external sensory consequences. Choice and judgment do not incorporate the experience of external sensory consequences, so they do not include the sense of agency characterising the voluntary action. Study IV has showed how the experience of agency can be modulated by contextual factors such as the moral significance of action and by the magnitude of outcomes. This modulation resembles the distinction between immediate emotion due to the emotions that are experienced at the time of decision (context of choice) and expected emotions due to the consequences of decision outcomes. The awareness of action seems affected by emotions in the same way observed for decision making processing. Our results suggest an interesting alternative to the notion of awareness of action as 'merely cold cognition based on learning associations though constant conjunctions of event' (Hume, 1739/1888). This suggests that the emotional impact of action/decision, related to context and the effects, has a direct influence on the primary experience of awareness of action. Emotions as well as the predictability of the effect (Moore & Haggard, 2008) are factors able to strengthen the awareness of the last step of decision making process, the action.

Modern culture is based on the fracture between a scientific-rational knowledge and beliefs based on subjective affective experience. In the last ten years, the concept of emotion and the role of emotion on decision making has changed, introducing the possibility of a positive role for emotion in rational decisions. However the same revolution requires one to reshape our conception of rationality in order to overcome a dangerous dualistic definition of the human being (rational vs emotional). Actual definition of rationality suggests that rational means being able to follow the utility function in order to maximize self-interest and reward. This definition of rationality is partial and dangerous. If rationality is the best criteria that we possess to choose and act, then rationality has to be able to include all components of reality, especially the capacity of being stricken[14] and moved by what we really care about.

[14] In Latin, being striken is *affectus*.

APPENDIX

Study III

Examples of dilemmas

Non-Moral Dilemmas

1) *Train or Bus.* You need to travel from Bologna to Cesena in order to attend a meeting that starts at 2:00 PM. You can take either the train or the bus. The train will get you there just in time for your meeting no matter what. The bus is scheduled to arrive an hour before your meeting, but the bus is occasionally several hours late because of traffic. It would be nice to have an extra hour before the meeting, but you cannot afford to be late. Is it appropriate for you to take the train instead of the bus in order to ensure your not being late for your meeting?

2) *Generic Brand.* You have a headache. You go to the pharmacy with the intention of buying a particular name-brand headache medicine. When you get there you discover that the pharmacy is out of the brand you were looking for. The pharmacist, whom you've known for a long time and in whom you have a great deal of trust, tells you that he has in stock a generic product which is, in his words, "exactly the same" as the product you had originally intended to buy. Is it appropriate for you to purchase the generic brand instead of searching further for the name-brand product you were looking for?

Moral Impersonal Dilemmas

1) *Vaccine Policy.* You work for the Bureau of Health, a government agency. You are deciding whether or not your agency should encourage the use of a certain recently developed vaccine. The vast majority of people who take the vaccine develop an immunity to a certain deadly disease, but a very small number of people who take the vaccine will actually get the disease that the vaccine is designed to prevent. All the available evidence, which is very strong, suggests that the chances of getting the disease due to lack of vaccination are much higher than the chances of getting the disease by taking the vaccine. Is it appropriate for you to direct your agency to encourage the use of this vaccine in order to promote national health?

2) *Sculpture*. You are visiting the sculpture garden of a wealthy art collector. The garden overlooks a valley containing a set of train tracks. A railway workman is working on the tracks, and an empty runaway trolley is heading down the tracks toward the workman. The only way to save the workman's life is to push one of the art collector's prized sculptures down into the valley so that it will roll onto the tracks and block the trolley's passage. Doing this will destroy the sculpture. Is it appropriate for you to destroy the sculpture in order to save this workman's life?

Moral Impersonal Dilemmas

1) *Lifeboat.* You are on a cruise ship when there is a fire on board, and the ship has to be abandoned. The lifeboats are carrying many more people than they were designed to carry. The lifeboat you're in is sitting dangerously low in the water-a few inches lower and it will sink. The seas start to get rough, and the boat begins to fill with water. It seems to you that there is only one way to stop the boat from sinking, and that is to start throwing other passengers overboard, starting with old people who are too weak to resist. Is it appropriate for you to throw some of your fellow passengers overboard in order to save yourself and the remaining passengers?

2) *Submarine.* You are the captain of a military submarine travelling underneath a large iceberg. An onboard explosion has caused you to lose most of your oxygen supply and has injured one of your crew who is quickly losing blood. The injured crew member is going to die from his wounds no matter what happens. The remaining oxygen is not sufficient for the entire crew to make it to the surface. The only way to save the other crew members is to shoot dead the injured crew member so that there will be just enough oxygen for the rest of the crew to survive. Is it appropriate for you to kill the fatally injured crew member in order to save the lives of the remaining crew members?

STUDY IV

Scripts used for initial verbal presentation of action choices. Scripts were presented via audio recording only in familiarization session.

UNPREDICTABLE SCRIPTS

A) Non-moral script, unpredictable

Turnips. You are a farm worker monitoring an automatic turnip-harvesting machine. The machine is out of order and is approaching two diverging paths. On one path the machine will destroy 5 bushels of turnips. On the other path the machine will destroy up only 1 bushel of turnips. You cannot remember the original instruction for the machine, so you cannot foresee which path the machine will take. The only thing that you can do is to press a button to change the original direction. Do you want the machine to stay on its present course OR do you want to change the direction of the machine to the other path? Press the button to "change" or "stay".

B) Non-moral script, unpredictable

Factory.You are the late-night watchman in a factory. You see fumes and fires spreading into the factory. In one room of the factory there are 5 important electronic devices. In another room there is 1 important electronic device. You cannot foresee which room the fire will enter. The only thing that you can do is to hit a switch in front of you connected to a fire door which will cause the fire to change direction from one room to the other room. Do you want the fumes and fires to stay on its present course OR do you want to change the direction of the fumes and fires to the other room. Press the button to "change" or "stay".

C) Moral script, **unpredictable**

Trolley. You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks there is a group of 5 railway workmen. On the other track there is 1 workman. You do not know upon which path the trolley will go, but you can hit a switch on your dashboard and so change the original direction of the trolley. Do you want the trolley to stay on its present course OR do you want to change the original direction of the trolley? Press the button to "change" or "stay".

D) Moral script, **unpredictable**

Hospital. You are the late-night watchman in a hospital. Spilled chemicals are rising up through the hospital's ventilation system. In one room of the hospital there are 5 patients. In another room there is 1 patient. You cannot foresee which rooms the spilled chemicals will enter. The only thing that you can do is to hit a switch connected to a fire vent. This will cause the spilled chemicals to change direction and enter the other room. Do you want the fumes and fires to stay on their present course OR do you want to change the direction of the fumes and fires to the other room? Press the button to "change" or "stay".

E) Non moral script, unpredictable

TV Quiz. You are participating in a quiz on a TV show and you win 10 bars of gold. But then you make a wrong response and so lose part of your winnings. To define the amount of your loss the host starts another game. There are two keys, one for room A and one for room B. In one room there is a message that says 'LOSE 1 bar of gold', in the other room there is a message says 'LOSE 5 bars of gold. You have one key, but do not know which door this key will open. The host offers you the option to exchange your key for the other key. Do you want your original key or do you want to exchange it for the other key? Press the button to "change" or "stay".

F) Moral script, **unpredictable**

Commander. It is wartime. You and your 5 children are living in a territory that has been occupied by the enemy. You are taken to the headquarters with your family. The commander says you that there are two prison camps. In one camp you and 4 of your children will be killed and only the youngest will survive. In the second camp your family could survive but the youngest child will be killed. The commander tells you that if you pay him he can change the original destination but he will not tell you what the original destination is. Do you want to stay with the commander's original decision or do you want to change the original destination by bribing the commander? Press the button to "change" or "stay".

G) Moral script, unpredictable

Poison. A viral epidemic has spread in the hospital where you are working. You have developed two substances in your home laboratory. You know that one of them is a vaccine and the other is a deadly poison. Someone has removed the label and you don't know which test tube contains the vaccine. In one room there are five patients with severe symptoms. They will die soon if you do not inject them the vaccine. In another room there is only one patient with mild symptoms. You decide to inject the yellow substance to five patients in more severe condition. Do you want to stay in this decision or do you prefer change you previous idea

testing the yellow substance to patient with mild symptoms? Press the button to "change" or "stay".

H) Moral script, unpredictable

Modified trolley. You are on a footbridge over a runaway trolley quickly approaching a fork in the tracks. Next to you on this footbridge is a stranger who happens to be very large. There is a group of 5 railway workmen on the tracks. On the other track there is 1 workman. You do not know upon which path the trolley will go but you can push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The large man can die in order to stop the train. Do you want the trolley to stay on its present course OR do you want to change the situation pushing the stranger off the bridge in order to stop the train? Press the button to "change" the situation or "stay" with the current situation.

I) Moral script, unpredictable

Modified Commander. It is wartime. You and your 5 children are living in a territory that has been occupied by the enemy. You are taken to the headquarters with your family. The commander says you that there are two prison camps. In one camp you and 4 of your children will be killed and only the youngest will survive. In the second camp your family could survive but you have to kill the youngest by beheading. The commander tells you that your decision can influence his verdict and change the original destination but he will not tell you what the original destination is. Do you want to stay with the commander's original decision or do you want to stray to change the original destination killing the youngest child? Press the button to "change" or "stay".

PREDICTABLE SCRIPTS

1) Non-moral script, predictable

Turnips. You are a farm worker monitoring an automatic turnip-harvesting machine. The machine is out of order and is approaching two diverging paths. On one path the machine will destroy 5 bushels of turnips. On the other path the machine will destroy up only 1 bushel of turnips. If you do nothing your turnip-harvesting machine will go on the path with 5 bushels of turnips. If you want to change the direction of the machine you have to hit a switch on your remote control. Do you want the machine to stay on its present course and so mash up 5 bushels of turnips OR do you want to change the direction of the machine to the other path and so lose only 1 bushel of turnips. Press the button to "change" or "stay".

2) Non-moral script, predictable

Factory. You are the late-night watchman in a factory. You see fumes and fires spreading into the factory. In one room of the factory there are 5 important electronic devices. In another room there is 1 important electronic device. If you do nothing the fumes and fires will rise up into the room containing the 5 electronic devices and will destroy them. The only way to avoid the destruction of these 5 electronic devices is to change the direction of the fumes and the fires. If you hit a switch connected to a fire door, the fumes and fires will change direction and enter the other room containing the single electronic device. If you do not hit the switch the 5 electronic devices will be destroyed. Do you want the fumes and fires to stay on its present course destroying 5 electric devices OR do you want to change the

direction of the fumes and fires to the other room where there is 1 device. Press the button to "change" or "stay".

3) Moral script, **predictable**

Trolley. You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. There is a group of 5 railway workmen on the tracks. On the other track there is 1 workman. If you do nothing the trolley will proceed on the track where the group of 5 workmen is, causing their deaths. The only way to avoid the deaths of these 5 workmen is to change the direction of the trolley. If you hit a switch on your dashboard that will cause the trolley to proceed to the other track, causing the death of 1 workman. Do you want the trolley to stay on its present course where there are 5 workmen OR do you want to change the direction of the trolley to the other track where there is 1 workman? Press the button to "change" or "stay".

4) Moral script, **predictable**

Hospital. You are the late-night watchman in a hospital. Spilled chemicals are rising up through the hospital's ventilation system. In one room of the hospital there are 5 patients. In another room there is 1 patient. If you do nothing the spilled chemicals will rise up into the room containing the 5 patients and cause their deaths. The only way to avoid the deaths of these patients is to change the direction of the spilled chemicals by hitting a switch connected to a fire vent. This will cause the spilled chemicals to change direction and enter the other room containing the 1 patient causing his death. Do you want the fumes and fires to stay on their present course where there are 5 patients OR do you want to change the direction of the fumes and fires to the other room where there is 1 patient? Press the button to "change" or "stay".

5) Non-moral script, predictable

TV quiz. You are participating in a quiz on a TV show and you win 10 bars of gold. But then you make a wrong response and so lose part of your winnings. To define the amount of your loss the host starts another game. There are two keys, one to room A and one to room B. You have the key to room B. In room A there is a message that says 'You LOSE 1 bar of gold', the other message in room B says 'You LOSE 5 bars of gold. The host offers you the key to room A. Do you want to stay with the original key to room B OR do you want to exchange it for the other key to room A? Press the button to "change" or "stay".

6) Moral script, **predictable**

Commander. It is wartime. You and your 5 children are living in territory that has been occupied by the enemy. You are taken to enemy headquarters with your family. The commander informs you that there are two prison camps, Camp A and Camp B. He has already decided that you and your family will go to the Camp A where you and 3 of your children will be killed and only the youngest will survive. However the commander also tells you that if you pay him he could change the original destination and move your family to Camp B where your youngest son will be killed but you and other 4 children will survive. Do you want to stay with the commander's original decision to send you to Camp A (4 die and 1 is saved) or do you want to change to Camp B (1 dies and 4 are saved) by bribing the commander? Press the button to "change" or "stay".

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature, 15, 669-72.

Albin, R.L., Mink, J.W. (2006). Recent advances in Tourette syndrome research. Trends in Neurosciences, 29, 175-82.

Allais, M. (1953). Le comportement de l'homme rationel devant le risque, critique des postulates et axiomes de l'ecole americaine. Econometrica, 21, 503-546.

Amodio, D. M., Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. Nature Reviews Neuroscience, 7, 268-277.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. Nature Neuroscience, 2, 1032-1037.

Andreoni J., Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. Econometrica, 70, 737–753.

Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm glow giving. Economic Journal, 100, 464-477.

Anselme, P. (2010). The uncertainty processing theory of motivation. Behavioural Brain Research. 208, 291–310.

Arnow, B. A., Desmond, J. E., Banner, L. L., Glover, G. H., Solomon, A., Polan, M. L., Lue, T. F., and Atlas, S. W. (2002). Brain activation and sexual arousal in healthy, heterosexual males. Brain, 125, 1014–1023.

Aron, A.R., Poldrack, R.A. (2005). The cognitive neuroscience of response inhibition: relevance for genetic research in attention-deficit/hyperactivity disorder. Biological Psychiatry, 1, 1285-92.

Arrow, K.J. (1974). The Limits of Organization. New York, NY, USA: Norton.

Axelrod, R. (1984). The Evolution of Cooperation. New York: Basic Books.

Bagozzi, R. P., Baumgartner, H., Pieters, R., & Zeelenberg, M. (2000). The role of emotions in goal-directed behavior. In S. Ratneshwar, D. G. Mick & C. Huffman (Eds.),

The why of consumption: Contemporary perspectives on consumer motives, goals, and desires (36–58). New York: Routledge.

Bargh, J.A., Chen, M., Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. Journal of Personality and Social Psychology, 71, 230-44.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In. Uleman, J.S., Bargh J.A. (Eds.), Unintended thought, 3-51. New York: Guilford Press.

Baron, J., Spranca, M. (1997). Protected values. Organizational Behavior and Human Decision Processes, 70, 1-16.

Barrash, J., Tranel, D., & Anderson, S. W. (2000). Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. Developmental Neuropsychogy, 18, 355-381.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron, 22, 639-50.

Baxter, M.G., Parker, A., Lindner, C.C., Izquierdo, A.D., Murray, E.A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. Journal of Neuroscience, 20, 4311-4319.

Beatty, J., 1986. The pupillary system. In: Coles, M.G.H., Donchin, E., Porges, S.W. (Eds.),
Psychophysiology: Systems, processes, and applications. Guilford Press, New York, 43–50.

Bechara, A. (2005a). Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive prospective. Nature Neuroscience, 8, 1458-1463

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

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

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

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

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

Beer., J.S., Ochsner, K.N. (2006). Social cognition: a multi level analysis. Brain Research, 24, 98-105

Beer, J.S., John, O.P., Scabini, D., Knight, R.T. (2006). Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. Journal of Cognitive Neuroscience, 18, 871–9.

Behrens, T.E, Hunt, L.T, Rushworth M.F. (2009). The computation of social behavior. Science, 29, 1160-4.

Bell, D. (1982). Regret in decision making under uncertainty. Operations Research, 30, 961-81.

Berg, J., Dickhaut, J., McCabe, K. (1995). Trust, Reciprocity, and Social History. Games and Economic Behavior, 10, 122-142.

Berlin, H.A., Rolls, E.T., Kischka, U. (2004). Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. Brain 127, 1108–1126.

Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk. (Transl. from Latin, Sommer, L. 1954). Econometrica, 22, 23–36.

Berns, G.S., McClure, S.M., Pagnoni, G., and Montague, P.R. (2001). Predictability modulates human brain response to reward. Journal of Neuroscience. 21, 82793–82798.

Berridge, K.C. (2004). Motivation concepts in behavioral neuroscience. Physiology & Behavior, 81, 179– 209.

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. Trends in Neurosciences, 9, 507–513.

Blair, R.J., Cipolotti L. (2000). Impaired social response reversal: a case of "acquired sociopathy". Brain, 123, 1122-41.

Blakemore, S.J., Wolpert, D.M., Frith, C.D. (2002). Abnormalities in the awareness of action. Trends in Cognitive Science, 6, 237-242.

Boccardi, E., Della Sala, S., Motto, C., Spinnler, H. (2002). Utilization behavior consequent to bilateral SMA softening. Cortex, 38, 289-308.

Bohnet, I., Zeckhauser, R. (2004). Trust, risk and betrayal. Journal of Economic Behavior & Organization, 55, 467-484.

Bohnet, I., Greig, F., Herrmann, B., Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. American Economic Review, 98, 294–310.

Boyce WT, Ellis BJ. 2005. Biological sensitivity to context. I. An evolutionary-developmental theory of the origins and functions of stress reactivity. Developmental Psychopathology 17:271–301.

Botvinick, M.M., Rosen, Z.B. (2008). Anticipation of cognitive demand during decision-making. Psychological Research, 73, 835-42.

Botvinick, M.M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. Cognitive, Affective & Behavioural Neuroscience, 7, 356-366.

Botvinick, M.M., Cohen, J.D., Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences, 8, 539-546.

Boucsein, W. (1992). Electrodermal activity. New York: Plenum Press.

Bowles, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. Science, 314, 1569-1572

Bradley, M.M., Lang, P.J., 2000. Measuring emotion: behavior, feeling and physiology. In: Lane, R.D., Nadel, L. (Eds.), Cognitive Neuroscience of Emotion. Oxford University Press, Oxford, 242–276.

Brass, M., Haggard, P. (2008). The what, when, whether model of intentional action. Neuroscientist, 14, 319-25.

Brass, M., Haggard, P. (2007). To do or not to do: the neural signature of self-control. The Journal of Neuroscience, 27, 9141-5.

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

Brothers, L., Ring, B. (1992). A neuroethological framework for the representation of minds. Journal of Cognitive Neuroscience, 4, 107–18.

Büchel, C., Morris, J., Dolan, R. J., Friston, K. J. (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron, 20, 947-957.

Buckholtz, J.W., Asplund, C.L., Dux, P.E., Zald, D.H., Gore, J.C., Jones, O.D., Marois, R.(2008). The neural correlates of third-party punishment. Neuron, 10, 738-40.

Buhner, M.J., Humpreys, G.R. (2009). Causal binding of actions to their effects. Psychological Science, 20, 1221-8.

Butler, J.K. (1991). Toward understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. Journal of Management, 17, 643-663.

Calder, A.J., Lawrence, A.D., Young, A.W. (2001). Neuropsychology of fear and loathing. Nature Reviews Neuroscience, 2, 352-363.

Camerer , C.F. and Fehr, E. (2006). When does " Economic Man " dominate social behavior? Science 311, 47-52 .

Camerer, C.F. (2003). Behavioral game theory. Princeton: Princeton University Press.

Camerer, C., Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. Econometrica, 56, 1-36.

Cardinal, R.N., Aitken, M.R. (2006). ANOVA for the behavioural sciences researcher. Lawrence Erlbaum Associates, New Jersey, USA.

Carmichael, S.T., Price, J.L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. The Journal of Comparative Neurology, 25, 615-641.

Carver, C.S., Ganellen, R.J., Framing, W.J., Chambers, W. (1983). Modeling: An analysis in terms of category accessibility. Journal of Experimental Social Psychology, 19, 403-421.

Cacioppo , J.T. , Berntson , G.G. , Larsen , J.T. et al . (2000). The psychophysiology of emotion . In: M. Lewis and J.M. Haviland-Jones (eds) , Handbook of Emotions , 2nd edn. New York, NY : Guilford Press , pp. 173 – 191

Carter, C. S., DeVries, A. C. and Getz, L. L. (1995) Physiological substrates of mammalian
monogamy: The prairie vole model. Neuroscience and biobehavioral reviews. 19, 303–314.

Castiello, U., Jeannerod, M. (1991). Measuring time to awareness. Neuroreport, 2, 797-800.

Ciaramelli, E., Muccioli, M., Ladavas, E., di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. Social, Cognitive, & Affective Neuroscience, 2, 84–92.

Coleman, J. (1990). Foundations of Social Choice Theory. Cambridge. MA: Harvard Univ. Press.

Cox, J.C. (2004). How to identify trust and reciprocity Games and Economic Behavior, 46, 260-281
Critchley, H.D., Mathias, C.J., Dolan, R.J. (2001). Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron, 29, 537–545.

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., Van Amelsvoort, T., Robertson, D., David, A., Murphy, D. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: a functional magnetic resonance imaging study. Human Brain Mapping, 9, 93–105.

Critchley, H.D., Rolls, E.T. (1996). Olfactory neuronal responses in the primate orbitofrontal cortex: analysis in an olfactory discrimination task. Journal of Neurophysiology, 75, 1659-1672.

Cunnington, R., Iansek, R., Bradshaw, J.L., Phillips, J.G. (1996). Movement-related potentials associated with movement preparation and motor imagery. Experimental Brain Research, 111, 29-36.

Cushman, F. (2008). Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. Cognition, 108, 353-80.

Cushman, F., Young, L., Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. Psychological Science, 1082-1089.

Gambetta, D. (1988). Trust: Making and breaking cooperative relations. Blackwell: New York
Dalgleish, T. (2004). The emotional brain. Nature Neuroscience, 5, 583-9.

Damasio, A. (2009). Neuroscience and the Emergence of Neuroeconomics. In Neuroconomics Academic Press.

Damasio, A.R., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L., Parvizi, J. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. Nature Neuroscience, 3, 1049-1056.

Damasio, A (1999). The Feeling of What Happens: Body and Emotion in the Making of Consciousness. Italian translation: Emozione e coscienza. Adelpfi editore.

Damasio, A.R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 351, 1413-1420.

Damasio, A.R. (1994). Descartes' error: Emotion, reason, and the human brain. New York: G.P. Putnam Sons

Damasio, A.R., Tranel. D., Damasio, H. (1990). Individuals with sociopathic behaviour caused by frontal damage fail to respond autonomically to social stimuli. Behavioural Brain Research, 41, 81-94.

Darwin, C. (1872/2002). The Expression of Emotion in Man and Animal , 3rd edn. New York, NY: Oxford University Press.

Davidson, R.J., Ekman, P., Saron, C. et al. (1990). Approach/ withdrawal and cerebral asymmetry: emotional expression and brain physiology. Journal Of Personality And Social Psychology. 38, 330 – 341.

Davidson, D. (1980) Essays on Actions and Events, Oxford University Press, (Oxford).

Davis, D. and Holt, C. (1993). Experimental Economics. Princeton, Princeton University Press.

David L. Clark, Nash N. Boutros, Mario F. Mendez (2005). The Brain and Behavior: An Introduction to Behavioral Neuroanatomy. Cambridge University Press

Dawson, M.E., Schell, A.M., & Filion, D.L. (2007). The electrodermal system. In Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.). Handbook of psychophysiology (pp. 159–181). Cambridge, UK: Cambridge University Press.

Deaner, R.O. et al. (2005) Monkeys pay per view: adaptive valuation of social images by rhesus macaques. Current Biology. 15, 543–548

De Araujo, I.E., Rolls, E.T., Kringelbach, M.L., McGlone, F., Phillips, N. (2003). Tasteolfactory convergence, and the representation of the pleasantness of flavor, in the human brain. European Journal of Neuroscience. 18, 72059–67208.

De Gelder, B., Vroomen, J., Pourtois, G., Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. Neuroreport, 10, 3759-3763.

De Gelder B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Philosophical Transactions of the Royal Society B: Biological Sciences. 364:3475-84.

De Martino, B., Kumaran, D., Seymour, B., Dolan, R.J. (2006). Frames, biases, and rational decision-making in the human brain. Science, 313, 684-687.

De Quervain, D.J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U.(2004). The neural basis of altruistic punishment. Science, 305, 1254-1258.

Della Sala, S., Marchetti, C., Spinnler, H. (1991). Right-sided anarchic (alien) hand: a longitudinal study. Neuropsychologia, 29, 1113-27.

Dennett, D.C. (1991). Consciousness Explained. Boston, Little, Brown and Company.

Derbyshire, S.W., Jones, A.K., Gyulai, F., Clark, S., Townsend, D., Firestone, L.L. (1997). Pain processing during three levels of noxious stimulation produces differential patterns of central activity. Pain, 73, 431–445

Desmurget, M., Reilly, K.T., Richard, N., Szathmari, A., Mottolese, C., Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. Science, 324, 811-3.

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 5-18.

Domes, G., Heinrichs, M., Michel, A., Berger, C., Herpertz, S.C. (2007) Oxytocin improves "mind-reading" in humans. Biological Psychiatry, 61, 731-3.

Dunn, J.R., Schweitzer, M.E.(2005). Feeling and believing: the influence of emotion on trust Journal of Personality and Social Psychology, 88, 736-48.

Ebert, J.P., Wegner, D.M. (2009) Time warp: Authorship shapes the perceived timing of actions and events. Consciousness and Cognition. In press.

Eckel, C., Wilson, R.K., (2004). Is trust a risky decision?. Journal of Economic Behavior & Organization, 55, 447-465.

Eid M, Diener E. (2001). Norms for experiencing emotions in different cultures: inter- and intranational differences. Journal of Personality and Social Psychology. 81, 869–885.

Elliott, R., Deakin. B. (2005) Role of the orbitofrontal cortex in reinforcement processing and inhibitory control: evidence from functional magnetic resonance imaging studies in healthy human subjects. International review of neurobiology, 65, 89-116.

Elliott, R., Newman, J.L., Longe, O.A., and Deakin, J.F.W. (2004). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: A parametric functional magnetic resonance imaging study. Journal of Neuroscience. 23, 1303–1307.

Elster, J. (1998). Emotions and economic theory. Journal of Economic Literature, 36, 47-74.

Ekman , P. and Friesen , W. (1971). Constants across cultures in the face and emotion . J. Pers. Social Psychol. 17 , 124 – 219 .

Engbert, K., Wohlschläger, A., Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. Cognition, 107, 693-704.

Ernst Fehr, E., Schmidt, K.M. (1999)European Economic Review, 44, 1057-1068.

Eslinger, P.J. (1998). Neurological and neuropsychological bases of empathy. European Journal of Neurology, 39, 193–9.

Eslinger, P.J., Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. Neurology, 35, 1731-1741.

Etkin, A., Klemenhagen, K.C., Dudman, J.T., Rogan, M.T., Hen, R., Kandel, E.R., Hirsch, J. (2004). Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. Neuron, 44, 1043-1055.

Falk, A., Fischbacher, U. (2006) A theory of reciprocity. Games and Economic Behavior, 54, 293-315

Farrer, C., Franck, N., Paillard, J., Jeannerod, M. (2003). The role of proprioception in action recognition. Consciousness and Cognition, 12, 609-19.

Fehr, E., Camerer, C.F. (2007). Social neuroeconomics: the neural circuitry of social preferences. Trends in Cognitive Science, 11, 419-427.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. Nature, 415, 137-140.

Fehr, E., Fischbacher, U. (2003). The nature of human altruism. Nature. 23, 785-91.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: the economics of reciprocity. Journal Economic Perspective, 14, 159-181.

Fehr, E., Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. The Quarterly Journal of Economics, 817-868.

Fellows LK, Farah MJ. 2003.Ventromedial frontal cortex mediates affective shifting in humans:
evidence from a reversal learning paradigm. Brain 126:1830–37

Fleming, S.M., Mars, R.B., Gladwin ,T.E., Haggard P. (2009). When the brain changes its mind: flexibility of action selection in instructed and free choices. Cerebral Cortex, 19, 2352-60.

Folstein, M.F., Robins, L.N., & Helzer, J.E. (1983). The mini-mental state examination. Archives of General Psychiatry, 40, 812.

Foot, P. (1978). Virtues and vices and other essays in moral philosophy. Oxford: Blackwell Publishers.

Forgas J.P. (2003). Affective influences on attitudes and judgments. In: Handbook of Affective Sciences, edited by Davidson RJ, Scherer KR, Goldsmith HH. New York: Oxford Univ. Press, 852–870.

Forgas, J.P. (1995) Mood and judgment: The affect infusion model (AIM). Psychological Bulletin. 117, 39-66.

Fowles, D.C., Christie, M.J., Edelberg, R., Grings, W.W., Lykken, D.T., & Venables, P.H. (1981). Committee report. Publication recommendations for electrodermal measurements. Psychophysiology, 18, 232-239 .

Fried, I., Katz, A., McCarthy, G., Sass, K.J., Williamson, P., Spencer, S.S., Spencer, D.D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. The Journal of Neuroscience, 11, 3656-66.

Frijda, N. H. (2006). The laws of emotion. Mawwah, NJ: Erlbaum.

Frijda, N. (2005). Emotion and experience. Cognition and Emotion, 19, 473-97.

Gallagher, H.L., Jack, A.I., Roepstroff, A., & Frith, C.D. (2002). Imaging the intentional stance in a competitive game. Neuroimage, 16, 814-821.

Charness, G., Rabin, M. (2002). Understanding Social Preferences With Simple Tests. The Quarterly Journal of Economics, MIT Press, 117, 817-869.

Gifford, A. (2002). Emotion and self-control. Journal of Economic Behavior & Organization, 49, 113–130.

Globisch, J., Hamm, A.O., Esteves, F., Ohman, A. (1999). Fear appears fast: temporal course of startle reflex potentiation in animal fearful subjects. Psychophysiology, 36, 66-75.

Gold, P.W. et al. (2002) New insights into the role of cortisol and the glucocorticoid receptor in severe depression. Biological Psychiatry, 52, 381–385.

Gorno Tempini, M.L., Pradelli, S., Serafini, M., Pagnoni, G., Baraldi, P., Porro, C., Nicoletti, R., Umita, C., Nichelli, P. (2001). Explicit and incidental facial expression processing: An fMRI study. Neuroimage, 14, 465–473

Gottfried, J.A., O'Doherty, J., Dolan, R.J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science, 22, 1104–1107.

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H.R., & Salazar, A.M. (1996). Frontal lobe injuries, violence, and aggression: a report of the Vietnam head injury study. Neurology, 46, 1231-1238.

Grahn, J.A., Parkinson, J.A., Owen, A.M.(2008). The cognitive functions of the caudate nucleus. Progress in Neurobiology, 86, 141-55.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. Cognition, 107, 1144-54.

Greene, J.D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. Trends in Cognitive Sciences, 11, 322-323.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. Neuron, 44, 389-400.

Greene, J. D. (2003). From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? Nature Reviews Neuroscience, 4, 846-849.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen. J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. Science, 293, 2105-2108.

Greene, J. D., Haidt, J. (2002). How (and where) does moral judgment work? Trends in Cognitive Sciences, 6, 517-523.

Greenwald, A.G., Draine, S.C., Abrams, R.L.(1996) Three cognitive markers of unconscious semantic activation. Science, 273, 1699-702.

Guth, W., Schmittberger, R., Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization, 3, 367–388

Haber, S.N., Kim, K.S., Mailly, P., Calzavara, R. (2006) Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. Journal of Neuroscience, 26, 8368-76.

Haidt J. The moral emotions. (2003) In: Handbook of Affective Sciences, edited by Davidson RJ, Scherer KR, Goldsmith HH. New York: Oxford University Press, 852–870.

Haggard, P., Clark, S., Kalogeras, J. (2002). Action, binding and awareness. In W. Prinz & B. Hommel (Eds.)

Haggard, P., Clark, S., Kalogeras, J. (2002). Volontary action and conscious awareness. Nature Neuroscience, 5, 382-385.

Haggard P, Cole J. (2008) Intention, attention and the temporal experience of action. Consciousness and Cognition, 16, 211-20.

Haggard, P. (2005) Conscious intention and motor cognition. Trends in Cognitive Sciences, 9, 290-5.

Haggard, P. (2008). Human volition: towards a neuroscience of will. Nature Review Neuroscience, 934-46.

Haggard, P. (2009). The sources of human volition. Science, 8, 731-3.

Haggard, P., Clark S. Kalogeras J. (2002). Voluntary action and conscious awareness. Nature Neuroscience, 5, 382–385.

Haidt, J. (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychological Review, 108, 814-34

Haidt, J. (2007). The new synthesis in moral psychology. Science, 316, 998-1002.

Hajcak, G., McDonald, N., Simons, R.F. (2003). To err is autonomic: error-related brain potentials, ANS activity, and post-error compensatory behaviour. Psychophysiology, 40, 895-903.

Hajcak, G., McDonald, N., Simons, R.F. (2004). Error-related psychophysiology and negative affect. Brain and Cognition, 56, 189-197.

Hallett, M. (2007). Volitional control of movement: the physiology of free will. Clinical Neurophysiology, 118, 1179-92.

Hanoch, Y. (2001). "Neither an angel nor an ant": Emotion as an aid to bounded rationality. Journal of Economic Psychology, 23, 1–25.

Hansson, O. (1994). The Structure of Values and Norms. Cambridge: Cambridge University Press

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

Hardin, R. (2002). Trust and Trustworthiness. The Russell Sage Foundation Series on Trust.

Hariri, A.R., Mattay, V.S., Tessitore, A., Fera, F., Weinberger, D.R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. Biological Psychiatry, 53, 494–501.

Hartikainen, K.M., Ogawa, K.H., Knight, R.T. (2000). Transient interference of right hemispheric function due to automatic emotional processing. Neuropsychologia, 38, 1576-1580.

Hauser, M.D. (2006). The liver and the moral organ. Social Cognitive and Affective Neuroscience, 1, 214-220.

Hauser, M.D., Chen, M.K., Chen, F., Chuang, E. (2003). Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. Proceedings. Biological Science, 22, 2363-70.

Hauser, M. D. 2000 Wild minds: what animals really think. New York: Henry Holt.

Haushofer, J., Fehr, E. (2008). You shouldn't have: your brain on others' crimes. Neuron. 10, 738-40.

Hitlin, S., and Piliavin, J.A. 2004. Current research, methods, and theory of values. Annual Review of Sociology (vol. 30).

Hollerman, J.R., Tremblay, L., Schultz, W. (1998). Influence of reward expectation on behavior-related neuronal activity in primate striatum. Journal of Neurophysiology, 80, 947–963
Holt, C.A., Laury, S.A. (2002). Risk Aversion and Incentive Effects The American Economic Review, 92, 1644-1655.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., and Polkey, C. E. (2004). Reward related reversal learning after surgical excisions in orbitofrontal or dorsolateral prefrontal cortex in humans. Journal of Cognitive Neuroscience, 16, 3463–3478.

Hornak, J., Rolls, E. T., and Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. Neuropsychologia 34, 247–261.

Horvitz, J.C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience, 96, 651-656

Houser, D., Schunk, D., Winter, J. (2009). Distinguishing Trust from Risk: An Anatomy of the Investment. Game Journal of Economic Behavior and Organization. In press.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C.F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. Science, 310, 1680-3.

Huebner, B., Dwyer, S., & Hauser, M. (2008). The role of emotion in moral psychology. Trends in Cognitive Sciences, 13, 1-6.

Hume, D. (1784). Enquiry concerning human understanding. P.F. Collier & Son.

Hume, D. (1739/1888). A treatise of human nature. London: Oxford University Press.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens DA in motivated behavior, a unifying interpretation with special reference to reward-seeking. Brain Research Reviews, 31, 6–41.

Izquierdo, A. Suda, R.K., Murray, E.A. (2005). Comparison of the effects of bilateral orbital prefrontal cortex lesions and amygdala lesions on emotional responses in rhesus monkeys. Journal of Neuroscience, 25, 8534–8542

Iwase, M., Ouchi, Y., Okada, H., Yokoyama, C., Nobezawa, S., Yoshikawa, E., Tsukada, H., Takeda, M., Yamashita, K., Takeda, M., Yamaguti, K., Kuratsune, H., Shimizu, A., and Watanabe, Y. (2002). Neural substrates of human facial expression of pleasant emotion induced by comic films: A PET Study. Neuroimage, 17, 758–768.

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: a reappraisal. Experimental Brain Research, 192, 527-32.

James , W. (1884). What is an emotion? Mind, 9, 188 – 205 .

Jenkins A.C., Macrae, C.N., Mitchell, J.P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. Proceeding the National Academy of Sciences (PNAS), 105, 4507-12.

Johnstone , T. and Scherer , K.R. (2000). Vocal communication of emotion . In: M. Lewis and J.M. Haviland–Jones (eds) , Handbook of Emotions , 2nd edn. New York, NY : Guilford Press, 220 – 235

Kahneman, D., Tversky, A. (1979). Prospect theory: an analysis of decision under risk. Econometrica, 4, 263– 291.

Karlan, D.S. (2005). "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." American Economic Review, 95, 1688–1699.

Kédia, G., Berthoz, S., Wessa, M., Hilton, D., & Martinot J. L. (2008). An agent harms a victim: a functional magnetic resonance imaging study on specific moral emotions. Journal of Cognitive Neuroscience, 10, 1788-1798.

Kessely H.K., Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion Journal of Economic Psychology, 28, 197-213

Ketelaar, T. (2004). Ancestral emotions, current decisions: Using evolutionary game theory to explore the role of emotions in decision-making. In C. Crawford & C. Salmon (Eds.), Darwinism, public policy and private decisions. Mahwah, N., J. Erlbaum, 145–168.

Kirsch, P., Schienle, A., Stark, R., Sammer, G., Blecker, C.,Walter, B., Ott, U., Burkart, J., and Vaitl, D. (2003). Anticipation of reward in a non aversive differential conditioning paradigm and the brain reward system: An event related fMRI study. Neuroimage, 20, 21086–21095.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. Nature, 19, 908-11.

Koenigs, M., Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. Journal of Neuroscience, 27, 951-956

Kohlberg, L. (1969). Stage and sequence: the cognitive developmental approach to socialization. In Goslin D.A. (Ed), Handbook of socialization theory and research, 347-480. Chicago: Rand McNally & Company.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive developmental approach. Moral Development and Behavior: Theory, Research and Social Issues. Holt, NY: Rinehart and Winston.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E. (2005). Oxytocin increases trust in humans. Nature. 2, 673-6.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N.L., Camerer, C.F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. Journal of Neuroscience, 18, 2188-92.

Kreps, D.M. (1990). A Course in Microeconomic Theory. Princeton University Press.

Kühn, S., Haggard, P., Brass, M. (2008). Intentional inhibition: How the "veto-area" exerts control. Human Brain Mapping, 30, 2834-43

Lacey, B.C., Lacey, J.I., 1978. Two way communication between the heart and the brain. American Psychologist, 33, 99–113.

Lagnado, D.A., Channon, S. (2008). Judgments of cause and blame: the effects of intentionality and foreseeability. Cognition, 108, 754-70.

Lang, P.J., Greenwald, M.K., Hamm, A.O., 1993. Looking at pictures: affective, facial., visceral, and behavioural reactions. Psychophysiology, 30, 261–273.

Lau, B., Glimcher, P.W. (2008). Value representations in the primate striatum during matching behavior. Neuron, 8, 451-63.

Lazarus , R.S. (1984). On the primacy of cognition. American Psychologist. 39 , 124-129.

LeDoux , J.E. (1996). The Emotional Brain . New York, NY : Simon and Schuster .

Ledyard , J. (1995). Public goods: a survey of experimental research . In: J. Kagel and A. Roth (eds), H andbook of Experimental Economics . Princeton, NJ : Princeton University Press, 111-194.

Lerner, J., Small, D., Loewenstein, G. (2004). Heart Strings and Purse Strings. Carryover Effects of Emotions on Economic Decisions. Psychological Science, 15, 337-341.

Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K.(1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. Brain, 106, 623-42.

Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. Annual Review Psychology, 58,259-89.

Ljunberg, T. Apicella, P. Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. Journal of Neurophysiology, 67(1), 145-163.

Loewenstein, G., O'Donoghue, T. (2004). Animal spirits: Affective and deliberative processes in economic behavior. Working paper

Loewenstein, G., Lerner, J., S. (2003). The role of affect in decision making. In: Handbook of Affective Sciences, edited by Davidson RJ, Scherer KR, Goldsmith HH. New York: Oxford University Press, 619-637.

London, E. D., Ernst, M., Grant, S., Bonson, K., and Weinstein, A. (2000). Orbitofrontal cortex and human drug abuse: Functional imaging. Cerebral Cortex, 10, 3334–3342.

Lykken, D. T., Venables, P. H. (1971). Direct measurement of skin conductance: a proposal for standardization. Psychophysiology, 8, 656-672.

Mack, A., Rock, I. (1998). Inattentional blindness. Cambridge, MA: MIT Press.

McCabe, K., Houser, D., Ryan, L. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. Proceeds Natural Academy of Science, 11832 – 11835 .

Machado, C.J. and Bachevalier, J. (2006) The impact of selective amygdala, orbital frontal cortex, or hippocampal formation lesions on established social relationships in rhesus monkeys (Macaca mulatta). Behavioral Neuroscience, 120, 761–786

Manning, S.K., Melchiori, M.P.,1974.Words that upset urban college students: measured with GSRs and rating scales. Journal of Social Psychology, 94, 305–306.

Marcel, J. A. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. Cognitive Psychology, 15, 197–237.

Mayer, R.C., Davis, J.H., Schoorman, F.D. (1995) An integrative model of organizational trust. Academy of Management Review, 20, 709-734

McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T. (2001) A functional imaging study of cooperation in two-person reciprocal exchange. Proceedings of National Academy of Science USA, 98, 11832-5

McClure, S.M., Botvinick, M.M., Yeung, N., Greene, J.D., Cohen, J.D. (2007). Conflict monitoring in cognition-emotion competition. In Gross, J.J. (Ed.), Handbook of Emotion Regulation, 204-228. New York: Guilford

Mellers, B. A. (2000). Choice and the relative pleasure of consequences. Psychological Bulletin, 126, 910–924.

Merikle, P.M. (1992). Perception without awareness. Critical issues. American Psychology, 47, 792-5.

Middleton, F.A., Strick, P.L. (2000). Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. Brain Cognition, 42, 183-200.

Mikhail, J. (2007). Universal Moral Grammar: Theory, Evidence, and the Future. Trends in Cognitive Sciences, 11, 143-152.

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

Milinski, M., Semmann, D. & Krambeck, H.-J. 2002 Reputation helps solve the 'tragedy of the commons'. Nature 415, 424–426.

Moll, J., de Oliveira-Souza, R., Bramati, I.E., Grafman, J. (2002a). Functional networks in emotional moral and nonmoral social judgments. Neuroimage, 16, 696-703.

Moll, J., de Oliveira-Souza, R., Eslinger, P.J., Bramati, I.E, Mourao-Miranda, J., Andreiuolo, P.A., Pessoa, L. (2002b). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. Journal of Neuroscience, 22, 2730-2736.

Moll, J., de Oliveira-Souza, R., Garrido, G.J, Bramati, I.E, Caparelli-Daquer, E.M, Paiva, M.L, Zahn, R., Grafman, J. (2007). The self as a moral agent: linking the neural bases of social agency and moral sensitivity. Social Neuroscience, 2, 336-52.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., Grafman, J. (2005). The neural basis of human moral cognition. Nature Reviews Neuroscience, 6, 799-809.

Moll, J., de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. Trends in Cognitive Sciences, 11, 319-321.

Montague, P.R. King-Casas, B. (2007). Efficient statistics, common currencies and the problem of reward-harvesting. Trends in Cognitive Science, 11, 514-9.

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

Montague, P.R., King-Casas, B., Cohen, J.D. (2006). Imaging valuation models in human choice. Annual Review of Neuroscience, 29, 417-448.

Moore, J., Haggard, P. (2008). Awareness of action: Inference and prediction. Consciousness and cognition. 17, 136-44.

Moore, J.W., Lagnado, D., Deal, D.C., Haggard, P. (2009). Feelings of control: contingency determines experience of action. Cognition. 110, 279-83.

Moran, R. (1981). Knowing Right from Wrong: the insanity defense of Daniel McNaughtan. The Free Press.

Moretti, L., di Pellegrino G. (2010). Disgust Selectively Modulates Reciprocal Fairness in Economic Interactions, Emotion, in press.

Moretti, L., Dragone, D., di Pellegrino, G. (2009). Reward and social valuation deficits following ventromedial prefrontal damage. Journal of Cognitive Neuroscience, 21, 128-40.

Müller, J.L., Sommer, M., Wagner, V., Lange, K., Taschler, H., Röder, C.H., Schuierer, G., Klein, H.E., Hajak, G. (2003). Abnormalities in emotion processing within cortical and subcortical regions in criminal psychopaths: evidence from a functional magnetic resonance imaging study using pictures with emotional content. Biological Psychiatry, 54, 152-62.

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

Nagai, Y. Critchley, H.D., Featherstone, E., Trimble, M. R., Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: A physiological account of a "default mode" of brain function. Neuroimage, 22, 243–251.

Naqvi, N., Bechara, A. (2006). Psychophysiological approaches to the study of decision-making. In C. Senior, T. Russel, & M. Gazzaniga (Eds.), Methods in mind: the study of human cognition (pp. 103-122). Cambridge, MA, MIT Press.

Nash, J.F. (1950). Equilibrium Points in N-Person Games. Proceedings of the National Academy of Science USA, 36, 48-49.

Nemeroff, C.B. and Owens, M.J. (2004) Pharmacologic differences among the SSRIs: focus onmonoamine transporters and the HPA axis. CNS Spectr, 9, 23–31

Nowak, D.A., Fink, G.R. (2009). Psychogenic movement disorders: Aetiology, phenomenology, neuroanatomical correlates and therapeutic approaches. Neuroimage, 47, 1015-25.

Ochsner, K.N., Gross, J.J. (2005). The cognitive control of emotion. Trends in Cognitive Sciences, 9, 242-249.

O'Doherty, J. P., Deichmann, R., Critchley, H. D., Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. Neuron 33, 815–826.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science, 304, 452-454.

O'Doherty, J., Rolls, E. T., Francis, S., Bowtell, R., McGlone, F., Kobal, G., Renner, B., and Ahne, G. (2000). Sensory specific satiety related olfactory activation of the human orbitofrontal cortex. Neuroreport 11, 893–897.

Ohman, A. (2002). Automaticity and the amygdala: nonconscious responses to emotional faces. Current Direction in Psychological Scienc, 11:62-66.

Öhman, A., Hamm, A., Hugdahl, K., 2000. Cognition and the autonomic nervous system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.), Handbook of Psychophysiology. Cambridge University Press, New York, 533–575.

Ostrom, E., Walker, J. (2003). Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research. Volume VI in the Russell Sage Foundation Series on Trust, Russell Sage Foundation.

Oya, H., Adolphs, R., Kawasaki, H., Bechara, A., Damasio, A., & Howard, M. A. 3rd. (2005). Electrophysiological correlates of reward prediction error recorded in the human prefrontal cortex. Proceedings of the National Academy of Sciences of the USA, 102, 8351-8356.

Padoa-Schioppa, C., Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature, 11, 223-226

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

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans Consciousness and Cognition, 14, 30-80.

Porges SW. 1995. Orienting in a defensive world: mammalian modifications of our evolutionary heritage. A polyvagal theory. Psychophysiology, 32, 301–318.

Passingham, R.E. (1987). Two cortical systems for directing movement. Ciba Found Symp, 132, 151-64.

Patton, J.H., Stanford, M.S., Barratt, E.S. (1995). Factor structure of the Barratt impulsiveness scale, Journal of Clinical Psychology, 51, 768–774.

Perugini, M., Gallucci, M., Presaghi, F., Ercolani, A.P. (2003). The personal norm of reciprocity. European Journal of Personality, 17, 251-283.

Pessoa, L. (2008). On the relationship between emotion and cognition. Nature Neuroscience, 9, 148-58.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? Current Opinion in Neurobiology, 15, 88-96.

Petrinovich, L., O'Neill, P., Jorgensen, M. (1993). An empirical study of moral intuitions : toward and evolutionary ethics. Journal of Personality and Social Psychology, 64, 467-478.

Phelps, E. (2009). The Study of Emotion in Neuroeconomics. In Neuroconomics. Decision Making and the Brain. Academic Press, 233-250.

Phillips, M.L., Williams, L.M., Heining, M., Herba, C.M., Russell, T., Andrew, C., Bullmore, E.T., Brammer, M.J., Williams, S.C., Morgan, M. (2004). Differential neural responses to overt and covert presentations of facial expressions of fear and disgust. Neuroimage, 21, 1484-1496.

Piaget, J. (1965/1932). The moral judgment of the child. New York: Free Press.

Plous, S. (1993). The psychology of judgment and decision making. New York: McGraw-Hill.

Price, J.L. (2007). Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Annals of the New York Academy Science, 1121:54-71

Rangel A, Camerer C, Montague PR. (2008). A framework for studying the neurobiology of value-based decision making, Nature Neuroscience, 9, 545-56.

Rensink, R.A, et al. (1997). To see or not to see: the need for attention to perceive changes in scenes, Psychological Science, 8, 368–373.

Rabin, M., (1993). Incorporating fairness into game theory and economics. American Economic Review. 83, 1281–1302.

Rapoport, A. Chammah. A.M. (1965). Prisoner's Dilemma. Ann Arbor, University of Michigan Press

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

Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D. (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. Neuroreport, 15, 2539-43.

Roberts, N. A., Beer, J. S., Werner, K. H., Scabini, D., Levens, S. M., Knight, R. T. (2004). The impact of orbital prefrontal cortex damage on emotional activation to unanticipated and anticipated acoustic startle stimuli. Cognitive, Affective & Behavioural Neuroscience, 4, 307-316

Roderick, M.K. (2001). Trust Rules for Trust Dilemmas: How Decision Makers Think and Act in the Shadow of Doubt Lecture notes in computer science, 2001 – Springer

Rohan MJ. 2000. A rose by any name? The values construct. Personality Social Psychology Review, 4:255–77

Rolls, E. T., Hornak, J., Wade, D., and McGrath, J. (1994). Emotion related learning in patients with social and emotional changes associated with frontal lobe damage. Journal of Neurology and Neurosurgery and Psychiatry, 57, 1518–1524.

Rolls, E. T., Kringelback, M. L., and de Araujo, I. E. T. (2003). Different representations of pleasant and unpleasant odors in the human brain. European Journal of Neuroscience, 18, 695–703.

Rolls, E.T. (2000). The Brain and Emotion. Oxford, UK: Oxford Univ. Press

Rolls, E.T. Critchley, H.D. Browning, A.S. Hernadi, I. Lenard, L. (1999). Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex. Journal of Neurosciences, 19, 1532-1540.

Rorden, C., Brett, M. (2000). Stereotaxic display of brain lesions. Behavioral Neurology, 12, 191–200.

Rotter, J. (1967). A new scale for measurement of interpersonal trust. Journal of Personality, 35, 651-665.

Rousseau, D., Sitkin, M., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23, 393-404.

Rudebeck, P.H., Bannerman, D.M., Rushworth, M.F. (2008). The contribution of distinct subregions of the ventromedial frontal cortex to emotion, social behavior, and decision making. Cognitive Affective & Behavioral Neuroscience, 8, 485-97

Rudebeck, P.H. et al. (2006) A role for the macaque anterior cingulated gyrus in social valuation. Science 313, 1310–1312

Rushworth, M.F., Behrens, T.E., Rudebeck, P.H., Walton, M.E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends in Cognitive Science 11, 168-76.

Samejima, K., Ueda, Y., Doya, K., Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science, 25, 1337-40.

Sanfey, A.G., Loewenstein, G., McClure, S.M., Cohen, J.D. (2006). Neuroeconomics: cross-currents in research on decision-making. Trends in Cognitive Science 10, 108-16.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. Science, 13, 1755-8.

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay Journal of Economic Behavior & Organization 62, 272-292.

Schoenbaum, G., Chiba, A.A., Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nature Neuroscience. 1, 155–159.

Schooler, J.W. (2002). Re-representing consciousness: dissociations between experience and meta-consciousness. Trends in Cognitive Sciences, 6, 339-344.

Schultz, W. (2000). Multiple reward signals in the brain. Nature reviews Neuroscience, 1, 199-207.

Schultz, W., Dickinson, A. (2000). Neuronal coding of prediction errors. Annual Review Neuroscience, 23, 473–500 .

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

Schultz, W., Apicella, P., Scarnati, E. & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. Journal of Neuroscience, 12, 4595–4610.

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

Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioural trigger stimuli in the monkey. Journal of Neurophysiology, 56, 1439–1462 .

Schwartz, S. H. and W. Bilsky (1987). 'Toward a Universal Psychological Structure of Human Values'. Journal of Personality and Social Psychology, 53: 550-562.

Searle, J.R (1983). Intentionality, CUP.

Sequeira, H., Ba-M'Hamed, S., 1999. Pyramidal control of heart rate and arterial pressure in cats. Archives Italiennes de Biologie, 137, 1–16.

Sequeira H, Hot P, Silvert L, Delplanque S.(2009) Electrical autonomic correlates of emotion. International Journal of Psychophysiology, 71, 50-6.

Shamay-Tsoory, S.G., Tomer, R., Berger, B.D., Goldsher, D., Aharon- Peretz, J. (2005). Impaired 'affective theory of mind' is associated with right ventromedial prefrontal damage. Cognitive and Behavioral Neurology, 18, 55–67.

Shidara, M., Aigner, T.G., Richmond, B.J. (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. Journal of Neuroscience, 18, 2613-25.

Simon, H. A. (1955). A behavioral model of rational choice. Quarterly Journal of Economics, 69, 99–118.

Simon, H. A. (1956). Rational choice, and the structure of the environment. Psychological Review, 63, 129–138.

Sinke, C. B., Sorger, B., Goebel, R., de Gelder, B. (2009). Tease or threat? Judging social interactions from bodily expressions. Neuroimage, 15, 1717-27.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., Haggard, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. Nature Neuroscience, 7, 80-4.

Small, D.M., Zatorre, R.J., Daghar, A., Evans, A.C., Gotman, M. (2001). Changes in brain activity related to eating chocolate: From pleasure to aversion. Brain 124, 1720–1733.

Smith , V.L. (1982). Microeconomic Systems as an experimental science . American Economic Review. 72 , 923-955 .

Solbakk, A.K., Reinyang, I., Nielsen, C.S., 2005. ERP indices of resource allocation difficulties in mild head injury. ournal of Clinical and Experimental Neuropsychology, 22, 743–760.

Srull, T.K (1984). The effects of subjective affective states on memory and judgment. Advances in consumer research. 11, 530-533.

Stuss, D.T., Levine, B. (2002). Adult clinical neuropsychology: lessons from studies of the frontal lobes. Annual Review of Psychology, 53, 401-433.

Stuss, D., Gow, C.A., Hetherington, C.R. (1992). "No longer Cage": frontal lobe dysfunction and emotional changes. Journal of Consulting and Clinical Psychology, 60, 349-59.

Synofzik, M., Vosgerau, G., Newen, A. (2008). I move, therefore I am: a new theoretical framework to investigate agency and ownership. Consciousness and Cognition, 17, 411-24.

Tabbert, K., Stark, R., Kirsch, P., Vaitl, D. (2005). Hemodynamic responses of the amygdala, the orbitofrontal cortex and the visual cortex during a fear conditioning paradigm. International Journal of Psychophysiology, 57, 15-23.

Tangney, J. P., Stuewig, J., Mashek, & D. J. (2007). Moral emotions and moral behavior. Annual Review of Psychology, 58, 345-372.

Taylor, S.F., Phan, K.L., Decker, L.R., Liberzon, I. (2003). Subjective rating of emotionally salient stimuli modulates neural activity. Neuroimage 18, 650–659.

Thomson, J.J. (1986). Rights, restitution and risk:essays in moral theory. Cambridge. MA, Harvard Universitu Press.

Thorpe, S.J., Rolls, E.T., Maddison, S. (1983). The orbitofrontal cortex: Neuronal activity in the behaving monkey. Experimental Brain Research, 49, 93–115.

Tomlin, D., Kayali, M.A., King-Casas, B., Anen, C., Camerer, C.F, Quartz, S.R., Montague, P.R.. (2006) Agent-specific responses in the cingulate cortex during economic exchanges. Science, 19, 1047-50.

Tranel, D., Damasio, H. (1994). Neuroanatomical correlates of electrodermal skin conductance responses. Psychophysiology, 31, 427–438.

Tremblay, L., Schultz, W. (2000a). Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. Journal of Neurophysiology, 83, 1877–1885.

Tremblay, L., Schultz, W. (2000b). Reward-related neuronal activity during go–no go task performance in primate orbitofrontal cortex. Journal of Neurophysiology, 83, 1864–1876.

Tremblay, L., Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. Nature 398, 704–708.

Tsakiris, M., Prabhu, G., Haggard, P. (2006). Having a body versus moving your body: How agency structures body-ownership. Consciousness and Cognition, 15, 423-32.

Tsakiris, M., Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. Experimental Brain Research, 149, 439-46.

Tsakiris, M., Hesse, MD., Boy, C., Haggard, P., Fink, GR. (2007). Neural signatures of body ownership: a sensory network for bodily self-consciousness. Cerebral Cortex,17, 2235-44.

Tversky, A., Kahneman, D. (1981). The framing of decisions and psychology of choice. Science, 211, 453-458.

Uvna¨s-Moberg, K. (1996) Neuroendocrinology of the mother-child interaction. Trends in Endocrinology and Metabolism 7, 126–131.

Uvna¨s-Moberg, K. and Eriksson, M. (1996) Breastfeeding: physiological, endocrine and behavioural adaptations caused by oxytocin and local neurogenic activity in the nipple and the mammary gland. Acta Paediatrica 85, 525–530.

Valdesolo, P., DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. Psychological Science, 17, 476-7.

van den Bos, W., McClure, S.M., Harris, L.T., Fiske, S.T., Cohen, J.D. (2007). Dissociating affective evaluation and social cognitive processes in the ventral medial prefrontal cortex. Cognitive Affective Behavioral Neuroscience, 7, 337-46.

van't Wout, M., Kahn, R.S., Sanfey, A.G., Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. Experimental Brain Research, 169, 564-568.

Vianna, E.P., & Tranel, D. (2006). Gastric myoelectrical activity as an index of emotional arousal. International Journal of Psychophysiology, 61, 70-76.

von Neumann, J., Morgenstern, O. (1944). Theory of Games and Economic Behavior. Princeton, NJ:Princeton Univ. Press

Watanabe, M. (2008). Reward expectancy in primate prefrontal neurons. Nature 382, 629–632 (1996).

Weekes, J.R., Lynn, S.J. (1990). Hypnosis, suggestion type, and subjective experience--the order-effects hypothesis revisited: a brief communication. The International Journal of Clinical and Experimental Hypnosis, 38, 95-100.

Wegner, D.M. (2004). Précis of the illusion of conscious will. The Behavioral and Brain Sciences, 27, 649-59.

Wegner, D.M. (2002). The illusion of conscious will. Cambridge, MA: MIT Press.

Wegner, D.M., Wheatley, T. (1999). Apparent mental causation. Sources of the experience of will. The American Psychologist, 54, 480-92.

Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. Journal of Neuroscience, 18, 411-418.

Williams, J. R., Insel, T. R., Harbaugh, C. R. and Carter, C. S. (1994) Oxytocin centrally administered facilitates formation of a partner preference in female prairie voles (Microtus ochrogaster). Journal of Neuroendocrinology 6, 247–250.

Winkielman,P., Berridge, K.C., Wilbarger, J.L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value, Personality and Social Psychology Bulletin, 31, 121–135.

Winton, W.M., Putnam, L.E., Krauss, R.M., 1984. Facial and autonomic manifestations of the dimensional structure of emotion. Journal of Experimental Social Psychology, 20, 195–216.

Winston, J.S., Strange, B.A., O'Doherty, J., Dolan, R.J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nature Neuroscience, 5, 277-83.

Withman, J.Q. (2003). Bufalo Criminal Law review 7, 85-107.

Witt, D. M., Winslow, J. T. and Insel, T. (1992) Enhanced social interaction in rats following chronic, centrally infused oxytocin. Pharmacology Biochemistry and Behavior, 43, 855–861.

Young, L., Koenigs, M. (2007). Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. British Medical Bulletin, 84, 69-79.

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E.D., Grafman, J. (2009) The neural basis of human social values: evidence from functional MRI. Cerebral Cortex. 19, 276-83.

Zajonc, R.B. (1984). On the primacy of affect. American Psychologist. 39, 117 – 123.