

# “Neural basis of emotions and social interaction”

## Research topic

We will study human behaviours emerging from the interplay of cognitive and emotional systems. Our research concerns the role of emotions in social and individual decision making.

The research outlined below examines the deficits of individuals with prefrontal cortex damage (pfc) through the lens of behavioural decision theory.

People who have suffered damage to the pre-frontal cortex behave differently than do people whose brains are intact. Nevertheless, precisely how their behaviour departs from normal is curious. While individuals with ventromedial pre-frontal cortex damage appear capable of generating plans and options, identifying their features or consequences, and attributing judgments of merit or value to those consequences -- they seem to make consistently poor choices. From a decision theoretic perspective this is paradoxical since the ability to generate options and to identify and assign values to their consequences are widely seen as keys to good, not bad, decision-making. Indeed, models of rational decision-making are predicated on peoples' ability to do such tasks.

People with lesions in the prefrontal cortex are impaired in many aspects of social and individual decision-making. The consequences of their behaviour are often disadvantageous and socially inappropriate. Examples are the tendency to lose their jobs, the inability to maintain stable personal relationships, and the repeated engagement in disastrous financial investments. The major anomaly consists in the fact that their behaviour is not due to lack of knowledge or limited intelligence (Saver and Damasio, 1991). They are, indeed, able to represent and judge correctly abstract social and individual contexts, while failing in analogous real-life situations. This is the description given by Eslinger and Damasio (1985), of the patient called EVR who at age 35 had a bilateral surgical excision of ventromedial frontal cortices in the course of treatment of a meningioma: *“EVR's social conduct was profoundly altered following his operation. In succeeding years his personal relationships deteriorated and his financial decisions led to bankruptcy. He has not been able to hold steady employment, and now lives in a sheltered environment, unable to support himself or his family.*

*On matters of minimal consequence, for example when deciding on clothes to wear or restaurants in which to dine, EVR may become consumed by extended deliberations, unable to reach a reasonably prompt and efficient decision. In contrast to this pattern of inappropriate social decision-making in real-life/real-time settings, however, during clinical interviews EVR appears able to proffer sensible social insights, make subtle distinctions between ambiguous concepts, and comments and discernment upon daily events.”*

The deficits exhibited by individuals with pfc are usually described as tendencies to be “impulsive”, “emotionally flat,” and “socially incompetent.” In this research, choice experiments and games drawn from the behavioural decision theory literature are presented to brain-damaged and normal subjects. These experiments are intended to clarify the role of the pre-frontal cortex in individual decision-making and social interaction, and to gain insight into the process whereby decisions are made by normal, as well as brain damaged, individuals.

## Project objectives

Results from the experiments will enable us to gauge in a direct way whether there are differences in the way losses and gains and immediate versus deferred outcomes are treated by pfc individuals as compared to normals, as well as what, precisely, is meant by “emotional flatness.”

The second battery of experiments examines the general claim that pfc individuals are socially incompetent (e.g., prone to behaving in socially inappropriate manners and unable to sustain relationships). One explanation for why pfc individuals behave in socially inappropriate ways is that, as a consequence of their impairment, they don't know / learn / apply social norms. As second possibility, more in line with the hypothesis that individuals with pfc don't sufficiently account for losses or deferred outcomes, is that while the norms of behaviour are understood, the deleterious

consequences of violating those norms simply don't matter (i.e., they aren't marked). To investigate these possibilities, we draw again on social and behavioural research -- in this case from work in experimental game theory.

### **Scientific originality and Innovation**

The objective of this research is to apply robust methods and findings from behavioural decision theory to bear on questions associated with brain function and, specifically, to better understand the role of the prefrontal cortex in decision making and judgment. However, additional benefits are anticipated. To the extent that many of the methods described below are used in the conduct of formal decision analysis, therapies or procedures for helping pfd individuals lead more normal or successful lives may be suggested by this research. In addition, these methods or others drawn from the decision making literature may provide additional means for diagnosis / measurement of the extent of impairment due to pre-frontal lobe damage. Finally, on a more philosophical note, the results of this research in conjunction with other work in this area, may give us greater insight into precisely what it means to be rational - a question which, as the divergence between the prescriptions of game theory and actual play in games suggests, is far from clear.

### **Research method**

#### **Part 1: The role of emotions in individual decision making: Somatic markers revisited**

People who have suffered damage to the ventromedial pre-frontal cortex behave differently than do people whose brains are intact. Nevertheless, precisely how their behaviour departs from normal is curious. While individuals with ventromedial pre-frontal cortex damage appear capable of generating plans and options, identifying their features or consequences, and attributing judgments of merit or value to those consequences -- they seem to make consistently poor choices. From a decision theoretic perspective this is paradoxical since the ability to generate options and to identify and assign values to their consequences are widely seen as keys to good, not bad, decision-making. The neurologist Antonio Damasio and colleagues explain this evidence in the context of what he refers to as the *somatic marker hypothesis*. Crudely put, this hypothesis argues that, for the purposes of deciding between options, those options (or their individual consequences) are unconsciously assigned markers denoting their goodness or badness. These markers provide the basis for choosing some options and eliminating others from further consideration. Damasio proposes that damage to the pre-frontal cortex interferes with this marking process, resulting in insensitivity to some of the outcomes associated with given courses of action. The experiments conducted by this group (Bechara et al, 1997) study the nature of the pre-frontal cortex in decision making. For instance, in one such experiment, called gambling task, normal and pfd individuals were asked to choose a card from any of four decks, then another card, and so forth. Each draw results in a gain but also, occasionally, a penalty. Decks A and B offer \$100 gains each time while C and D offer \$50 gains. However, periodic penalties associated with draws from A and B result in these decks having a negative expected payoff whereas small penalties for C and D result in these decks having positive expected payoffs. Subjects are asked to draw from the decks so as to maximise their profits. What Damasio and his colleagues found was that while normals develop skin conductivity responses (SCR) to the high penalty decks and end up playing the more favourable ones, pfd individuals don't develop SCR and continue to play the high risk - high penalty decks. Damasio explains this result in the context of what he refers to as the somatic marker hypothesis. Crudely put, this hypothesis argues that, for the purposes of deciding between options, those options (or their individual consequences) are unconsciously assigned markers denoting their goodness or badness. These markers provide the basis for choosing some options and eliminating others from further consideration. Damasio (1994) proposes that damage to the pre-frontal cortex interferes with this marking process, resulting in insensitivity to some of the outcomes associated with given courses of action.

However, what is not clear from these experiments is which markers are not being assigned, is it ones associated with losses or ones associated with future consequences?, or whether the problem is not insensitivity to losses or future outcomes but rather hypersensitivity to gains or current outcomes. To large extent these difficulties arise because the experiments conducted, though exceedingly clever, confound the concepts of preference for outcomes (i.e., gains and losses) and attitude toward risk.

A recent experiment by Tomb et al. (2002) shows opposite results to Damasio's, in a variation of the gambling task. In Tomb et al the good decks (with expected gains) have the larger rewards and the larger punishments along the sequence of cards (fixed every 10 cards), whereas in the original version of the gambling task the bad decks (with expected losses) had the larger punishments and the larger rewards, with losses exceeding gains. Tomb et al. report similar results of Damasio for the behaviour of normal subjects, i.e. they have chosen more often the good decks; and opposite results for the SCR, i.e. higher anticipatory SCR when the good decks were chosen. This finding confirms that the original gambling task is confounded. But still, the new version of the gambling task introduced by Tomb et al. does not disentangle between hypersensitivity to rewards or punishment and risk attitude in terms of variability of possible outcomes.

In a first experiment, we propose another variation of the gambling task in which we have a combined situation between good and bad decks in terms of risk, i.e., we have high variability (higher rewards and higher losses) in one of the two good decks and in one of the bad decks. In this way we can disentangle the effect of preference for outcomes and risk. Indeed, if the anticipatory SCR would be higher for the two high-risk gambles, then we will be able to conclude that the marker is attached to the variance between possible punishment and reward. Whereas, if the anticipatory SCR is still correlated with the type of decks (good or bad), then we could conclude that the marker is attached to the levels of the outcomes. In this case we would proceed with another variation of the gambling task in which we have the same variability between the good (with higher rewards) and the bad (higher punishments) decks. In this way we would be able to find if there is hypersensitivity to higher gains or to higher losses.

This experiment has the purpose of benchmarking the behaviour of individuals with pfc patients relative to normal subjects and also for corroborating, refining, or refuting the somatic marker hypothesis.

Results from the experiments described to this point will enable us to gauge in a direct way whether there are differences in the way losses and gains and immediate versus deferred outcomes are treated by pfc individuals as compared to normals along lines suggested in the somatic marker hypothesis, as well as what, precisely, is meant by "emotional flatness."

### **The role of emotion in social interaction: "Reciprocity-based emotions"**

The second part of the project examines the general claim that pfc individuals are socially incompetent (e.g., prone to behaving in socially inappropriate manners and unable to sustain relationships.) One explanation for why pfc individuals behave in socially inappropriate ways is that, as a consequence of their impairment, they don't know/learn/apply social norms. As second possibility, more in line with the hypothesis that individuals with pfc don't sufficiently account for losses or deferred outcomes, is that while the norms of behaviour are understood, the deleterious consequences of violating those norms simply don't matter (i.e., they aren't marked). To investigate these possibilities, we draw again on social and behavioural research -- in this case from work in experimental game theory.

We intend to study pfc and normal subjects' behaviour in games with underlying reciprocity responses. We use the explication of reciprocity given by Fehr and Gächter: "Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently much more nasty and even brutal" (cf. Fehr and Gächter, 2000 a, pp159.) In this definition there are two implicit concepts: positive reciprocity and negative reciprocity. People respond to friendly or hostile actions disregarding material incentives. We assume that emotions matter. The specific emotions that are related with negative and positive reciprocity are anger and elation. Anger (elation) is related with counterfactual thinking (see Kahneman and Miller, 1986) that arises when individuals compare the obtained outcome with better (worse) outcomes that might have been realised (see Zeelenberg et al., 1999). We use the terms upward or downward counterfactual when

the obtained outcome is compared with better outcomes or worse outcomes, respectively. Anger and elation result from upward and downward counterfactual, respectively.

In the case of reciprocal interaction the obtained outcome depends on the decision of two or more agents. We refer to anger (or elation) if the individual will feel just partially responsible of her final outcome. Indeed, she will attribute the responsibility of her outcome to another individual or group of individuals. The counterfactual reasoning allows the individual to focus on the alternative outcome that she would have obtained if the other individual (or individuals) had behaved differently. This process is based on the attribution of intentionality to the other individual's behaviour (see Rabin, 1993).

In our experiment subjects will participate to two types of extensive-form games, called punishment game and trust game. These two types of games are introduced in order to study negative and positive reciprocity. In the punishment game the first player can end the game and equally split an amount of money (30 experimental dollars), or give to the other player the opportunity to choose and possibly obtain a greater payoff for herself (\$25) and a smaller payoff for the second player (\$5), if the second player does not punish her. Punishment occurs if the second player ends the game with zero payoffs for both players.

The subgame perfect Nash equilibrium solution is always to play opportunistically for the first player and always not to punish for second player. Different behaviours indicate deviation from the standard game theoretical solution.

The second type of game is the trust game. In the trust game, if the first player does not trust the second player to reciprocate, she can end the game with a small payoff for both (\$10 for the first and \$10 for the second player, respectively), or she can move giving to the second player the opportunity to choose between reciprocating and defecting. If the second player reciprocates, she and the first player get more (\$15 for player 1 and \$25 for player 2) than the first possible outcome; whereas the game ends with her maximum payoff (\$40) and with a zero payoff for the first player if she defects. In this game if the first player gives to the second player the opportunity to choose, means that she trusts the second player to reciprocate. The subgame perfect Nash equilibrium solution is always "not-trust" for the first player and always "defect" for the second player.

These two types of games are introduced in order to study negative and positive reciprocity. The experiment is computerized. Both players (player 1 and player 2) are seated in front of a computer. They know (through the instructions) that they are interacting with another person, i.e. their counterpart. Subjects play a (random) sequence of trust and punishment games with different payoffs structure, i.e. the payoffs change at each round as a result of a linear transformation of the payoffs structure of the two basic games. In this way we maintain invariant the "nature" of the two games and we reduce the effect of habituation due to repeating the same payoff structure. Subjects' role (as player 1 or player 2) is random assigned at each round. They are informed, previous to play, about their role.

Both players know all the possible outcomes of the game, and they are conscious about the consequences of their choice and the possible choices of the other player. Additionally to the data concerning subjects' choices, we record the galvanic responses (SCR) of both players in each phase of the experiment.

People respond to friendly or hostile actions disregarding material incentives. In this context, we assume that emotions matter. The specific emotions that are related with negative and positive reciprocity are anger and elation. Anger (elation) is related with counterfactual thinking (see Lewis, 1973; Roese and Olson, 1995; Kahneman and Miller, 1986) that arises when individuals compare the obtained outcome with better (worse) outcomes that might have been realized (see Niedenthal et al., 1994; Zeelenberg et al., 1999). We use the terms upward or downward counterfactual when the obtained outcome is compared with better outcomes or worse outcomes, respectively. Anger and elation result from upward and downward counterfactual, respectively.

We assume that the experience of anger and elation affects individual choice. We name these emotions as "reciprocity-based" emotions. In particular, experienced reciprocity-based anger increases the willingness to punish (even if this is costly), and experienced elation increases the willingness to trust and reciprocate.

Results from normal subjects show the presence of trust (and trustworthiness) and punishment behaviour (Coricelli, McCabe, and Smith, 2000; Fehr and Gächter, 2002; Bowles and Gintis, 2001). This divergence between the play of normal subjects and that predicted by game theory is attributed

to the acceptance of social norms (e.g. fairness) and their enforcement driven by emotional response to others' behaviour.

In light of speculations that pfc individuals are tempted by the possibility of immediate gains, we might anticipate that their responses in these games will be, somewhat paradoxically, more in line with the prescriptions of "rationality" embodied in game theory. This prediction is based on the assumption that they will be impaired in expressing "reciprocity-based emotions". Whether they appreciate strategic features of social interactions and/or recognize appropriate norms of behaviour may also be indicated by their responses.

### *Bibliography*

- Bechara, A., Damasio, H., Tranel, D., Damasio, A. (1997). "Deciding Advantageously Before Knowing the Advantageous Strategy." *Science*, 275, 1293-1294.
- Coricelli, G., McCabe, K.A., and Smith, V.L. (2000). Theory-of-mind mechanism in personal exchange. In Hatano, et al. (Eds.), *Affective Mind*. Elsevier Science Publisher, 250-259.
- Eslinger and Damasio (1995). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology*, 35, 1731-41.
- Fehr, E., and Gächter, S. (2000 a). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14, 159-181.
- Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature*. 415, 137-140.
- Frijda, N.H., Kuipers, P., and ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*. 57, 212-228.
- Kahneman, D., and Miller, D. (1986). Norm Theory: Comparing reality to its alternatives. *Psychological Review*. 93, 136-153.
- Loomes, G., and Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *Review of Economic Studies*, 53, 271-282.
- Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83, 1281-1302.
- Saver J.L., and Damasio, A.R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, 29, 1241-49.
- Selten, R. (1975). "Re-examination of the perfectness concept for equilibrium points in extensive games." *International Journal of Game Theory*, 4: 25-55.
- Tomb, I., Hauser, M., Deldin, P., Caramazza, A. (2002). "Do somatic markers mediate decisions on the gambling task?". *Nature Neuroscience*, vol. 5, November, 1103-4.
- Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*. 12, 93-106.