Encoding of Financial Signals in the Human Brain

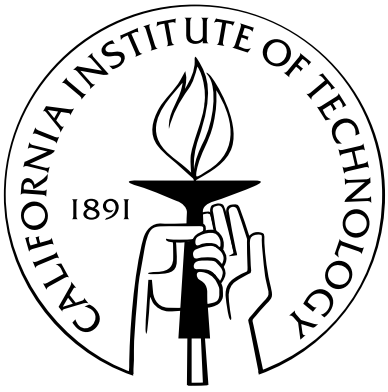
Thesis by

Antoine Jean Bruguier

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2007

(Defended October 19, 2007)

Acknowledgements

First, I would like to thank the various sources of funding that made all this work possible. During my first year as a master's student, I received funding as a teaching assistant directly form the California Institute of Technology (`http://www.caltech.edu`). As a graduate research assistant, I received funding from the Gordon and Betty Moore Foundation (`http://www.moore.org`) and the David and Lucile Packard Foundation (`http://www.packard.org`). The main funding came from a grant from the Social and Economic Sciences program (SES-0527491) of the National Science Foundation (`http://www.nsf.gov`).

I would also like to thank my two research advisors, Peter Bossaerts and Steven Quartz, for giving me a chance to work with them and prove myself. Hiring someone involves some risk and I hope that the rewards were large enough. It is extremely rare to be given such an opportunity and I consider myself very lucky to have studied at Caltech. In addition, I learned a lot both from Peter and Steve.

I would also thank the people who collaborated with me by sharing their fMRI data (chapters 2 and 3) and market data (chapter 4). Without their help, the results we present would have taken much more time to obtain and would have cost much more. I would also like to thank my labmates Cédric Anen, Ulrik Beierholm, Sayuri Desai, and Kerstin Preuschoff for their help.

Even before I started, I received help and support from Linda Dosza at Caltech and Diann Gruber and Jean de Largentaye who taught me English at ESIEE (`http://www.esiee.fr`). I would also like to thank James Lindauer and Tricia Brownstein for their advice, guidance, and support when I thought about quitting. Finally, I would like to acknowledge Steven Youra's help with the redaction of various manuscripts; all remaining typos, grammar mistakes, and obscure sentences are my fault.

The present document is available at: `http://www.bruguier.com/thesis.html` and `http://etd.caltech.edu`.

Abstract

Neuroeconomists investigate how the human brain analyzes and makes decisions about financial situations. They use functional magnetic resonance imaging (fMRI) of subjects who participate in economic games. Here we present three such experiments.

In the first experiment, we investigate how the brain recombines expected reward (ER) and risk. Recent fMRI results show that the brain decomposes a gamble in terms of these two metrics. However, economic theory predicts that the brain must recombine them in order to obtain an effective evaluation of the gamble. It was not clear what biological mechanism directs such recombination. Here we show that the brain uses the correlation of noise to recombine signals. We implement a new technique based on canonical correlation analysis and we show that ER is added to risk to form a metric that activates the medial prefrontal cortex.

In the second experiment, we investigate how the brain encodes two gambles instead of one. The brain is likely to encode the utility of each gamble in a common area but in separate groups of neurons. However, it is unknown how the brain indexes the gambles. Indeed, which group of neuron encodes which gamble can be decided in many ways. We hypothesized that the brain would use either the physical position of the gambles or an idiosyncratic parameter, such as ER or risk. Here we introduce a new analysis technique based on Hotelling T-squared statistics and we show that the brain uses risk as an index.

In the third experiment, we investigate a much more complex situation: a stock market. Contrary to what standard finance theory predicts, we hypothesize that the brain does not use mathematical models but instead heuristically uses a social cognition approach. Specifically, we posit that humans understand stock markets by using Theory of Mind (ToM), the ability to attribute to others mental states different from one's own. Here we show that humans engage brain structures related to ToM (paracingulate cortex, anterior cingulate cortex, insula, and amygdala). Subsequent behavioral tests show that ToM, rather than mathematical, abilities are better predictors of success in forecasting stock markets.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

We are interested in *neuroeconomics*. Neuroeconomists study how the human brain analyzes and reacts to economic and financial situations. Their aim is both to improve the understanding of human behavior and to understand how the human brain analyzes economic situations. A prime data acquisition method is functional magnetic resonance imaging (fMRI). In the following sections, we briefly describe how fMRI allows us to investigate the human brain and we introduce the basic economic concepts that we will use later.

After this brief introduction, three questions will emerge. The economic signals of *expected reward* and *risk* are encoded separately in the brain but economic theory predicts that they must be recombined. In chapter 2, we will show how the brain performs such recombination. The *utility* signals of two gambles are encoded in the same area of the brain but we do not know how the brain classifies these signals. In chapter 3, we will show that it uses risk. Finally, we do not know how information is transmitted across traders in a stock market. In chapter 4, we will show that successful traders use *Theory of Mind*.

1.2 A primer on fMRI

1.2.1 Data acquisition

fMRI is a data acquisition technique that allows researchers to look at the brain in action. We will not describe in detail the physics and biology behind fMRI and we refer readers to Jezzard et al. [2001] and Huettel et al. [2004] for a more complete introduction. It is an extension of clinical MRI (Gillespie and Jackson [2000]) in the sense that it does not

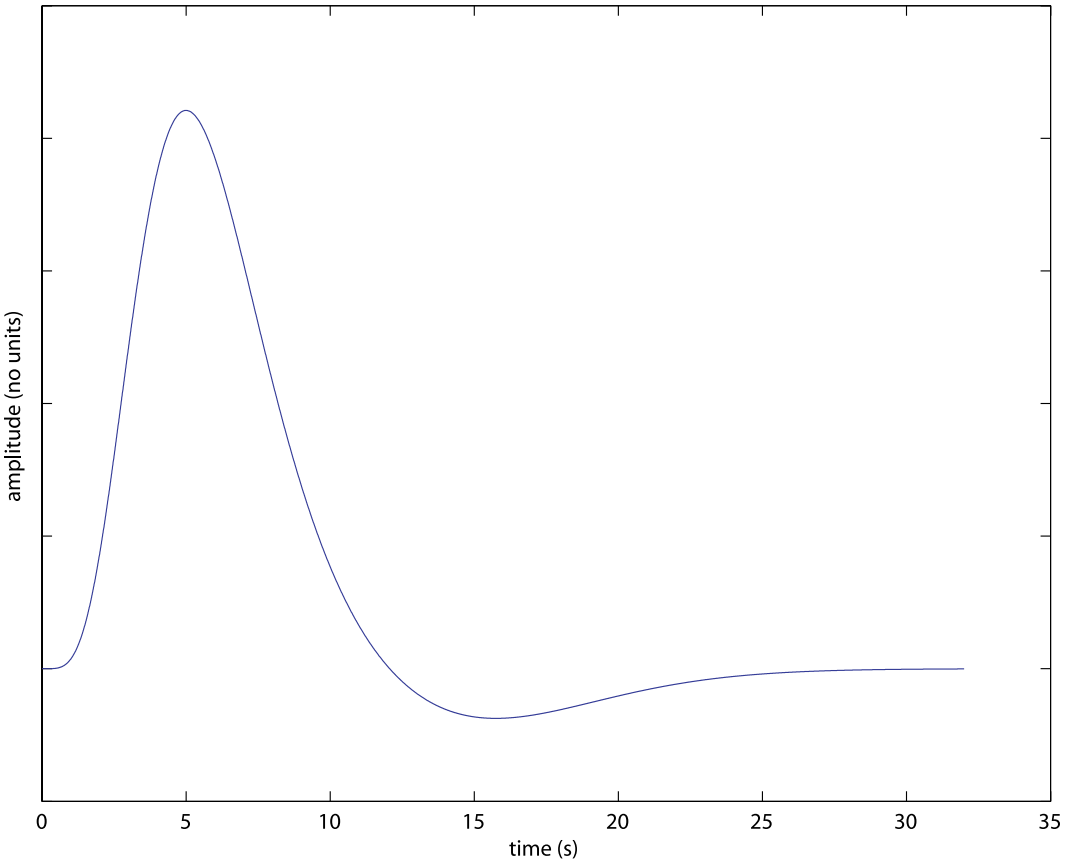


Figure 1.1: Canonical HRF. The HRF characterizes the dynamics of the blood flow in the brain (see also section 1.2). When an area is activated, there is a delay in the fMRI signal and an HRF can characterize this delay. It is the impulse response of a linear filter that models the delay.

deal with static 3D images but with a series of 3D images indexed by time. Both MRI and fMRI are safe, non-invasive procedures (Shellock [2001a]) provided that no magnetic metal is present in the subjects/patients' body (Shellock [2001b, 2005]).

MRI is based on a quantum physics principle. Particles have a magnetic spin that can be excited by a magnetic field. When magnetic particles are relaxed, they emit an electromagnetic radiation that can be recorded. Atoms of a given type emit a specific frequency and in the case of fMRI, we tuned our recording to track oxygen atoms. If an area of the brain is more active, it will consume more oxygen. Higher concentrations of oxygen (specifically the difference in magnetic susceptibility between oxyhemoglobin and deoxyhemoglobin) will result in more power recorded by an MRI scanner. The effect is referred to as the *BOLD* effect (Blood-Oxygen-Level Dependent). The change in oxygen level is not instantaneous

and the delayed response to a stimuli is represented by an Hemodynamic Response Function (*HRF*, see figure 1.1). In effect, observing the brain with fMRI is similar to observing a microprocessor with an infrared camera; we do not directly record the activity, just the evidence that energy was used.

An fMRI data set is a four-dimensional array. Three dimensions index space; we record amplitude of the fMRI signal in cubes called *voxels* (short for "volume element"). Typically each cube is 3 by 3 by 3 mm wide and contains hundreds of thousands of neurons. The fourth dimension is time; we record a new amplitude at regular intervals, called *TR* (time of relaxation). Typically, we acquire new data every 2 s; in one TR, a single neuron may fire hundreds of times.

Investigation with fMRI has several advantages. First, it is non-invasive and we can use this technique on healthy human subjects. Second, we can record the activity of the entire brain at the same time. However, it has poor spatial and time resolutions.

What neuronal process is the BOLD signal related to? A flurry of biological mechanisms occurs in the brain and it is not clear yet which ones are responsible for changes in neurons' oxygen consumption. Several hypotheses have been advanced. Some experimental results suggest that fMRI reflects synaptic activity (Logothetis et al. [2001], Niessing et al. [2005]). In this view, fMRI would reflect the input signal of neurons. Other results suggest that fMRI reflects the firing activity (action potentials) of neurons (Mukamel et al. [2005]). In this view, fMRI would reflect the output signals of neurons. All these studies correlated the recorded fMRI signals with other signals recorded with more invasive techniques. The difficulty in deciding between input and output comes from the fact that a single fMRI voxel contains many more neurons than we can record with any invasive technique and thus we cannot fairly decide whether fMRI reflects input or output. Additionally, the time resolution of fMRI is limited, making these comparisons more difficult. The distinction between input and output has not hampered research however, as a voxel reflects more the activity of a population of neurons than the activity of individual neurons. Moreover, fMRI may very well reflect a mix of both input and output. This distinction is important however when we link neuronal activity with fMRI data (as we will do in chapter 2). Fortunately, in the context of the experiments we will present, inputs and outputs of the neuronal area that we observe are the same and the distinction is irrelevant.

Finally, we need to localize structures in the brain. We use a popular coordinate system

established by Talairach and Tournoux [1988]. We normalize every subjects' brain so that coordinates are common across subjects. We also refer to some areas as defined by the cytoarchitecture (Brodmann [1905], Gazzaniga et al. [2002]).

1.2.2 Data analysis with general linear models

First, each of the analyses start with the preprocessing of the data, which we will not describe here (see Penny et al. [2003]). Then, fMRI data is traditionally analyzed with general linear models (GLMs). A GLM is a linear regression onto which we attach statistics (chapter 11 of Penny et al. [2003]).

The first step of the analysis is to create predictors. Predictors are time-indexed vectors that reflect an *a priori* hypothesis on the way the brain responds to a set of stimuli. For example, let us imagine an experiment where subjects see alternatively pictures of houses and pictures of animals, with blank screens in between. The experiment lasts for 1000 s and we have a TR=2 s; we therefore acquire 500 scans. We will create two predictors (one for the houses, one for the animals) that are vectors containing 500 elements. We fill the predictor corresponding to the houses (respectively animals) with zeros for every scans except for the scans when the subjects saw houses (respectively animals). Finally, to take into account the effect of the hemodynamic, we linearly convolve each predictor with an HRF (non-linear models can also be used, see Buxton and Wong [1998]).

Two main types of predictors can be used to analyze fMRI data. Block predictors are modulated by zeros and ones. When using them, we need to contrast two predictors corresponding to two carefully controlled conditions. In the example above, we would contrast the predictors for the animals and for the houses. Parametric predictors are modulated by a real variable and usually have zero mean. We do not always have to use contrasts with them.

Contrasts are useful to control for unrelated activations. In the example above, if we only looked at the predictor for the houses, we would also observe an area that just responds to light (known as V1), even if this area does not specifically process houses. By contrasting the two predictors, we avoid this confusion.

The second step of the analysis is to compute a linear regression. We create a *design matrix* that contains the predictors we created. In the example above, we would have a 500 by 2 matrix. However, we also add other predictors that reflect aspects of the experiments

that we are not interested in. In the example above, a key press for the subjects' response will activate the motor area in the brain. Thus, we need to create a predictor that models the key presses. We will refer to predictors that we investigate as *predictors of interest* and to predictors that we include in the design matrix but do not inspect as *extraneous predictors*.

The last step is to compute statistics. We can attach a p-value to predictors or contrasts of predictors (section A.1.2 or Weisberg [2005]). A p-value is the probability of a false positive. We are interested in voxels that have a low p-value, i.e., the probability that the activity of one voxel is due to noise instead of a predictor is low. The lower the p-value, the lower the probability that noise alone could have explained the fMRI data we observed, i.e., we need the predictor to explain the brain activity. We obtain a *p-map* that we threshold to discover interesting regions in the brain.

A problem arises from the fact that we inspect many voxels (*multiple comparison* problem). In science, a p-value below 0.05 is usually deemed sufficient to declare a result statistically significant. However, given the large number of voxels, having 5% of the voxels falsely significant is not acceptable. Two approaches are possible. One can lower the threshold to about 0.001 and only report clusters of more than 5 voxels. Another approach is to correct for multiple comparisons with algorithms such as False Discovery Rate (FDR, Genovese et al. [2002]).

Finally, serial time correlation appears in fMRI data sets. Because of the limited bandwith of the recording device, the noise on the time samples is correlated. If we do not take into account this correlation (in the case of fixed-effects analysis, Penny et al. [2003]), we will obtain p-values artifically low. To correct for this error, one can use auto-regressive moving-average (ARMA) models (Burock and Dale [2000]). With this procedure, we run two successive GLMs. The first GLM is computed as usual and is used to estimate the amount of correlation. We then remove the serial correlation from the data and run a second GLM that does not show any serial correlation.

1.3 A primer on finance and economics

In this section, we will provide a very basic description of the foundation of economic theory. Of course, this section will be very limited and will only give the necessary information for

the experiments to come[1].

Let us start with a simple example[2]. Would you rather receive \$50 for sure or receive \$100 with probability 0.5 (and nothing otherwise)? How about receiving \$30 for sure vs \$100 half of the time? Economists write these choices as (\$0,p=0;\$50,p=1) vs (\$0,p=0.5;\$100,p=0.5) for the first example, (\$0,p=0;\$30,p=1) vs (\$0,p=0.5;\$100,p=0.5) for the second example. Each possible choice is called a *gamble*.

Any economic and financial decision can be modeled as a combination of several gambles. For example, by combining basic gambles, we can, in theory, model the choice between investing retirement money in bonds that have a fixed return or in the stock market. Bonds have a certain return while stocks have random returns with various probabilities.

How do humans choose between two gambles? It was posited that humans make decisions according to four basic rules (Samuelson [1938]). These four *axioms of revealed preference* are:

- **completeness**: $p \preceq q$ or $p \succeq q$ for every gamble p and q. This axiom seems natural. One always prefers one gamble over another, unless the two gambles are equal. In other words, a human faced between a choice between two gambles always chooses one over the other; there is no indecision.

- **transitiveness**: $p \preceq q$ and $q \preceq r$ implies $p \preceq r$. This axiom seems also natural. If r is preferred over q and if q is preferred over p, then r is preferred over p.

- **Archimedean axiom**: This axiom is more technical. For every gamble p, q, and r such that $p \succ q \succ r$, then $\exists \alpha, \beta \in (0;1)$ such that $\alpha p + (1-\alpha) r \succ q$ and $q \succ \beta p + (1-\beta) r$. In effect, this axiom is the equivalent of the density of $\mathbb{R}$.

- **independence**: This axioms seems also natural. For any gamble p and q such that $p \succ q$, and for any gamble r and $\alpha \in [0;1]$ we have $\alpha p + (1-\alpha) r \succ \alpha q + (1-\alpha) r$. In effect, adding another gamble on both sides does not make any difference.

These four axioms seem natural but it has been shown that humans sometimes do not follow these rules. For now, however, we will ignore these problems and describe the concept of *utility*.

[1] A clear introduction can also be found at `http://cepa.newschool.edu/het/essays/uncert/vnmaxioms.htm`.

[2] We will use United States dollars for every experiment.

Given these four axioms, one can assign a unique (up to a linear transformation) real number to any gamble and we call this number *utility* (Neumann and Morgenstern [1953], Debreu [1964]). Mathematically, there exists a function that maps gambles onto $\mathbb{R}$. Humans are utility maximizers; when choosing between two gambles while following the axioms of preference, they act as if they computed the utility of these two gambles and chose the gamble with the highest utility. This utility function reflects a person's preferences and is therefore different from one person to another.

This utility function was also proven to be *affine*: given two gambles p and q and given $\alpha \in [0; 1]$, we have $U(\alpha p + (1-\alpha)q) = \alpha U(p) + (1-\alpha)U(q)$. This result is particularly useful. We can decompose a complex gamble consisting of several payoffs $\$x_i$ with probabilities p_i into a weighted sum: $\sum_i p_i U(x_i)$. We used a short-hand notation here: when a gamble pays an amount x with probability p=1, we just use the dollar amount as the argument of the function. In practice, actually, it is always the way the utility function is used. This view is also natural: we attach a certain utility to money.

What can we say about this utility function? First, it is increasing; more money is always more useful. Second, for many humans, this utility is mostly concave. Let us consider an example. A \$100 bill found on the street would be of higher marginal utility to a poor person than to a billionaire. Mathematically, the first derivative of the utility function, $U'(x)$, is decreasing, or in other words, the utility is a concave function, $U''(x) < 0$. This concavity is not a general rule and the utility can be concave or convex, depending on the person and the amount of money considered.

We will now introduce two central concepts: *expected reward* (ER) and *risk*. Given the affine property of the utility function, the utility of a gamble r is the sum of the utilities of the various dollar amount at stake weighted by their respected probability; i.e., the utility of a gamble is the expected value of the outcomes' utilites:

$$U(r) = \sum_i p_i U(x_i). \tag{1.1}$$

In a continuous probability space:

$$U(r) = \int_{-\infty}^{\infty} f(x)U(x)dx. \tag{1.2}$$

Let us use a Taylor expansion of this expression around a fixed dollar amount x_0:

$$U(r) = \int_{-\infty}^{\infty} f(x_0 + \epsilon) U(x_0 + \epsilon) d\epsilon \tag{1.3}$$

$$U(r) \simeq \int_{-\infty}^{\infty} f(x_0 + \epsilon) \left[U(x_0) + \epsilon U'(x_0) + \frac{1}{2}\epsilon^2 U''(x_0) \right] d\epsilon \tag{1.4}$$

$$U(r) \simeq U(x_0) + U'(x_0) \int_{-\infty}^{\infty} f(x_0 + \epsilon)\epsilon d\epsilon + \frac{1}{2} U''(x_0) \int_{-\infty}^{\infty} f(x_0 + \epsilon)\epsilon^2 d\epsilon. \tag{1.5}$$

Let us inspect two terms of equation 1.5. The term $\int_{-\infty}^{\infty} f(x_0 + \epsilon)\epsilon d\epsilon$ is equivalent to the first moment (expected value) of a random variable, economists call this term *expected reward*. Note that the coefficient in front of the expected reward, $U'(x_0)$, is always positive as the utility is an increasing function. The term $\int_{-\infty}^{\infty} f(x_0 + \epsilon)\epsilon^2 d\epsilon$ is equivalent to the second moment (variance) of a random variable and economists call this term *risk*. Note that the coefficient in front of the risk, $\frac{1}{2}U''(x_0)$, is usually negative (concave utility).

This view of economic decisions in terms of expected reward and risk is very commonly used, even though this is an approximation[3]. Humans seek a high ER (making a lot of money) but do not like taking too many risks. For example, some investors sometimes shun stocks because even though they have a higher ER than bonds, they also have a higher risk. The utility function characterizes the tradeoff between ER and risk. In that view, the coefficient in front of the expected reward is positive while the one in front of the risk is negative: people are usually *risk averse*. However, in some situations, the utility may be convex (people are *risk seeking*).

For a more complex situation, finance predicts that humans have *rational expectations* (Lucas [1972], Muth [1961]). In this view, prices are fixed by a rational argument, such as an economic equilibrium (Samuelson [1983]). Deviations from perfect prediction are only due to random error. Additionaly, this random error is supposed to have zero mean, i.e., there is no systematic bias. In this view, humans do not make systematic cognitive mistakes.

The entire construction of this section was based on the four axioms of preference. If humans respect these axioms at all time, then a utility function can always be computed (and approximated by ER and risk). In a majority of the cases, humans do indeed follow the axioms, but axioms are violated during situations known as paradoxes. The most famous

[3]For example, the special report in the May 19th 2007 issue of *The Economist* on international banking is entitled: "Risk and Reward."

paradox was discovered by Allais [1953].

To correctly predict these paradoxes, Economists created more complex theories that fit observed behavior. Neuroeconomists use another approach. While still using the basic tools to measure gambles, they record both humans' decisions and how their brains react.

1.4 Questions and overview

The main goal of neuroeconomics is to understand how the brain evaluates and makes decisions about economic situations. The field is vast and we will restrict ourselves to three main questions:

First, recent work showed that the brain decomposes ER and risk in different areas (Preuschoff et al. [2006], Knutson et al. [2001], Huettel et al. [2005]). This is remarkable because this decomposition was used by economists prior to any fMRI study. However, efficient evaluation of a gamble requires that these two parameters be recombined. We will try to address the question of how the brain recombines ER and risk in chapter 2.

Second, a recent study showed that the neurons in monkeys' brains encode separately a utility index for liquid treats (Padoa-Schioppa and Assad [2006]). Do humans use similar techniques to index economic gambles? We will try to provide an answer in chapter 3.

Finally, we will investigate stock markets in chapter 4. We will study how information is dispersed inside a financial market.

Obviously, the field of neuroeconomics is much wider than these three simple studies, we just restricted ourselves to three studies of tasks of increasing complexity. There is much more to be asked between and beyond the three experiments we present. We nevertheless hope that the present work will be consistent and of some interest.

1.5 Tools used

For the data sets presented in chapters 2 and 4, we acquired data with the Siemens 3T Trio scanner (`http://www.medical.siemens.com`) located at the California Institute of Technology. For the data set presented in chapter 3, we acquired data with the General Electric 1.5T Excite HD scanner (`http://www.gehealthcare.com`) located at Columbia University.

The data analysis is usually done with integrated packages that allow researchers both to pre-process and to analyze their data. SPM (`http://www.fil.ion.ucl.ac.uk/spm/`) is one such package. It is free and runs on Matlab (`http://www.mathworks.com`). BrainVoyager (`http://www.brainvoyager.de`) is a commercial package that runs on Windows and UNIX systems (OSX and Linux). Finally, one can run FSL (`http://fsl.fmrib.ox.ac.uk/fsl/`). We also developed in-house analysis code that runs with Matlab and conversion techniques from one environment to another.

Specifically, we used for:

- Chapter 2: BrainVoyager for preprocessing, in-house Matlab code for CCA, BrainVoyager for subsequent GLM, and BrainVoyager for display.
- Chapter 3: FSL for preprocessing, in-house Matlab code for analysis, and BrainVoyager for display.
- Chapter 4: BrainVoyager for the entire analysis.

Chapter 2

How to Recombine Risk and Expected Reward

In this chapter[1], we investigate how expected reward (ER) and risk are recombined in the brain. Recent work has shown that the human brain encodes separately ER (Preuschoff et al. [2006], Knutson et al. [2001]) and risk (Preuschoff et al. [2006], Huettel et al. [2005]). Two questions arise. First, economic theory (section 1.3) suggests that ER and risk must be recombined into a single metric. What is this metric? Second, if the human brain performs such recombination, what is the biological mechanism at work?

2.1 Introduction

The present work is based on a hypothesis by Salinas and Sejnowski [2001]. In their review paper, they hypothesized that the correlation of neuronal activity plays an organizational role. By simultaneously recording the activity of several neurons, they tried to infer how several neurons work together and investigated how the activities of several neurons are correlated.

Neuronal activity is contaminated by noise. The activity of neurons is disrupted by random firings and it is not clear how the brain handles this noise. However, Salinas and Sejnowski [2001] observed that correlation in neuronal firing may help alleviate the effect of noise and Romo et al. [2003] verified the hypothesis with direct neuronal recording. Specifically, the hypothesis is that the brain uses the correlation of the neuronal activities to recombine signal in a way that minimizes noise. We will describe this hypothesis in more detail in section 2.3.

[1]This work was a collaboration with Peter Bossaerts, Kerstin Preuschoff, and Steve Quartz.

We cannot investigate correlation with GLMs (section 1.2.2). Instead, we use a method based on Canonical Correlation Analysis (CCA, Anderson [2003], Johnson and Wichern [2002], Hotelling [1936]). We apply CCA to a data set that was previously acquired.

How do we investigate this correlation with fMRI? At first, it was unclear whether the effect of the HRF (section 1.2) would prevent us from detecting correlation at the neuronal level. With simple mathematics, however, we prove that if the neuronal activity is directly related to fMRI data, then the correlations of neuronal activities are directly related to the correlations of fMRI data (appendix A.2).

2.2 Original experiment

We use the data provided by Preuschoff et al. [2006]. The experiment is a series of simple gambles (figure 2.1). In the first step, the subjects place a blind bet that has two options: "second card higher" or "second card lower." Then, subjects see the two cards, displayed sequentially and drawn uniformly from a deck of 10. The subjects earn \$1 if their bet corresponded to the outcome and lose \$1 otherwise. Finally, in order to make sure subjects pay attention to the game, we ask them to report the outcome of the gamble and we fine them 25¢ per mistake. Then we repeat the trial independently with a new deck.

This design has three main advantages. First, it involves no decision but still elicits a feeling of "randomness." Because we ask subjects to place a bet at the beginning of each gamble, there is a feeling of randomness. However, there is no strategy attached to this bet. Thus, the bet is blind. This is especially useful because the analysis of the brain signals will be easier

Second, we can observe both ER and risk. In between the display of the two cards, subjects experience a situation were both ER and risk vary as a function of their original bet and the value of the first card. ER varies linearly with the value of the first card (with positive or negative slope depending on the blind bet). Risk varies quadratically with the value of the first card.

Third, ER and risk are orthogonal. Because the former varies linearly with the value of the first card and the latter varies quadratically, the two signals are orthogonal. We will show in section 2.5.1 and appendix A.4 that we can use this orthogonality to our advantage.

Fourth, the areas of interest are already discovered. Preuschoff et al. [2006] identified

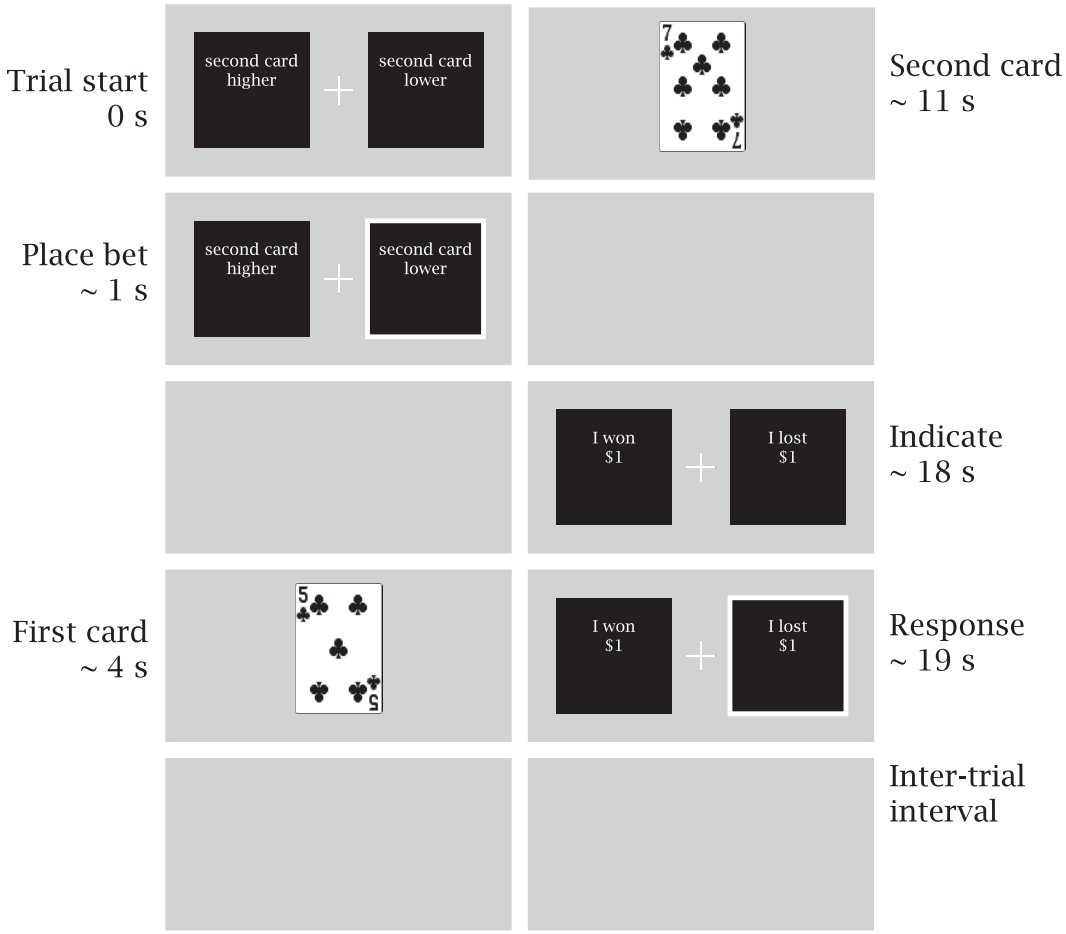


Figure 2.1: Stimulus set for the experiment. We used the data from Preuschoff et al. [2006], where the stimulus set was as follows: First, we asked subjects to place a blind $1 bet on one of two cards. We then displayed two cards, drawn randomly and without replacement from a deck of 10. We showed the cards sequentially so that subjects experienced various ERs and risks in the interval after the display of the first card but before the display of the second card. After we displayed the second card, the subjects knew whether they had earned or lost $1. In order to monitor attention, we asked subjects to report the outcome of the gamble and we fined them 25¢ for any mistake.

that the ventral striatum, putamen, and insula evaluated the gamble in terms of ER and risk (coordinates in table 2.1 on page 19). Thus, we can investigate how these three *upstream* areas export their signal to a *downstream* area.

We constructed the predictors in the same way as Preuschoff et al. [2006]. We created two boxcar functions with onset at the display of the first card and of duration 1 s. We modulated the first predictor with the ER and the second predictor with its risk.

2.3 Neurobiological foundations

Groups of neurons often display correlated activity (Averbeck et al. [2006], Salinas and Sejnowski [2001]). There are several types of correlation. First, correlation can be between the activities of two neighboring neurons or between two populations of neurons. Second, since the activity of neurons is the superimposition of a signal of interest and noise, both signal and noise can be correlated. Here we will only focus on the role of correlated noise between two populations.

In their review paper, Salinas and Sejnowski [2001] hypothesized that correlation serves a specific organizational purpose. Specifically, they hypothesized that neurons combine afferent signals in a way that minimizes the correlated noise. Romo et al. [2003] showed with neuronal recording that this hypothesis was plausible. We would like to find a method that we can apply to fMRI data. Let us mathematically model this hypothesis.

Let y_i be a time-indexed vector that represents the activity of an upstream population i. Salinas and Sejnowski [2001] proposed that it is the sum of a signal of interest, x_i, some correlated noise, $\widetilde{y_i}$, and some other noise, n_i (non-correlated neuronal activity and recording noise). Without loss of generality, we did not assign any weight, and mathematically:

$$y_i = x_i + \widetilde{y_i} + n_i. \tag{2.1}$$

The brain can take advantage of the correlation of the signals $\widetilde{y_i}$. For example, let us examine the case of two areas. If the signals $\widetilde{y_1}$ and $\widetilde{y_2}$ are positively correlated, then the brain should compute the difference of x_1 and x_2 to minimize the distortion due to noise. Generally, the brain should recombine several signals in a way that minimizes noise, i.e., it computes:

$$\min \left|\left| \sum_i a_i x_i - \sum_i b_i y_i \right|\right|^2 . \tag{2.2}$$

Unfortunately, the equation above does not properly take into account normalizations. Indeed, the error will be minimized by $a_i = 0$ and $b_i = 0$. Instead, we use an equivalent correlation form:

$$\max \left\{ \text{corr} \left(\sum_i a_i x_i ; \sum_i b_i y_i \right) \right\} \tag{2.3}$$

$$\max \left\{ \text{corr} \left(\sum_i a_i x_i ; \sum_i b_i \left(x_i + \widetilde{y_i} + n_i \right) \right) \right\} . \tag{2.4}$$

We need to specify the arguments of the max. Let us slightly modify equation 2.4 in two ways. The brain does not control a_i, it needs to adapt the correlation signal, $\widetilde{y_i}$, and the strength of the signals, b_i, in order to obtain the desired results., i.e., the brain computes:

$$\max_{b_i, \widetilde{y_i}} \left\{ \text{corr} \left(\sum_i a_i x_i ; \sum_i b_i \left(x_i + \widetilde{y_i} + n_i \right) \right) \right\} . \tag{2.5}$$

However, a simple inspection of equation 2.4 reveals that it is very close to an expression found in Canonical Correlation Analysis (CCA, appendix A.3, Anderson [2003], Hotelling [1936], Johnson and Wichern [2002]). We can modify equation 2.4 to match the expression from the CCA:

$$\max_{a_i, b_i} \left\{ \text{corr} \left(\sum_i a_i x_i ; \sum_i b_i \left(x_i + \widetilde{y_i} + n_i \right) \right) \right\} . \tag{2.6}$$

By solving equation 2.6 using CCA, we can reverse the brain's noise minimization modeled by equation 2.5.

The solution of 2.6 is called the *first row of the CCA*. However, a CCA computes other solutions (*rows*). These solutions have (generically) lower correlations (appendix A.3, Hotelling [1936], Anderson [2003], Johnson and Wichern [2002]). While these other solutions are not of interest *per se* for neuroscience, we will show that they are useful for the data analysis.

If there are n predictors and p regions, the CCA finds $\min(n, p)$ solutions reached with a_{ij} and b_{ij} ($j = 1..\min(n, p)$). For simplicity, we write the solutions of the first row as

$a_i = a_{i1}$ and $b_i = b_{i1}$.

There is one detail left to be addressed. The noise minimization described above is done at the neuronal level, but we do not have access to the neuronal activity, only to the fMRI data. However, with simple mathematics we show that if the fMRI time courses and the neuronal activities are directly related, then the correlations of neuronal activities and the correlations of the fMRI time courses are also directly related (appendix A.2).

2.4 Reverse engineering with CCA

Our method consists of three steps (figure 2.2). In the first step (black section), we compute *adjusted time courses*. Preuschoff et al. [2006] identified three upstream regions that encode ER and risk: insula, putamen, and ventral striatum (vst). We average the fMRI time courses of the 30 voxels closest to the center of these regions. However, the resulting time courses contain two types of signals. In addition to containing ER or risk, they contain other *extraneous* signals that we wish to ignore. Indeed, visual activation, motor activations, and win/lose signals contaminate the signals in the insula, putamen, and ventral striatum. To remove these extraneous signals, we build the predictors corresponding to these perturbations and compute linear regressions. The error of this linear regression (*residual*) is orthogonal to the extraneous signals but still contains the signals that we wish to investigate, namely ER and risk.

The second step is the heart of the method (blue section of figure 2.2). We use the adjusted time courses from the previous steps, y_i, and the predictors, x_i, as the input of a CCA. We reverse engineer the noise minimization method that the brain operates (see section 2.3) and obtain the weights a_i and b_i that minimize equation 2.6.

In the third step, we locate the downstream region (red section of figure 2.2). Since the second step computed the optimal weights, we can compute a new composite predictor, U_1, that optimally combines ER (x_1) and risk (x_2):

$$U_1 = a_1 x_1 + a_2 x_2. \tag{2.7}$$

We substitute the ER and risk with this new predictor and compute a GLM. If the new predictor significantly explains the activity in one area of the brain, then this area is the

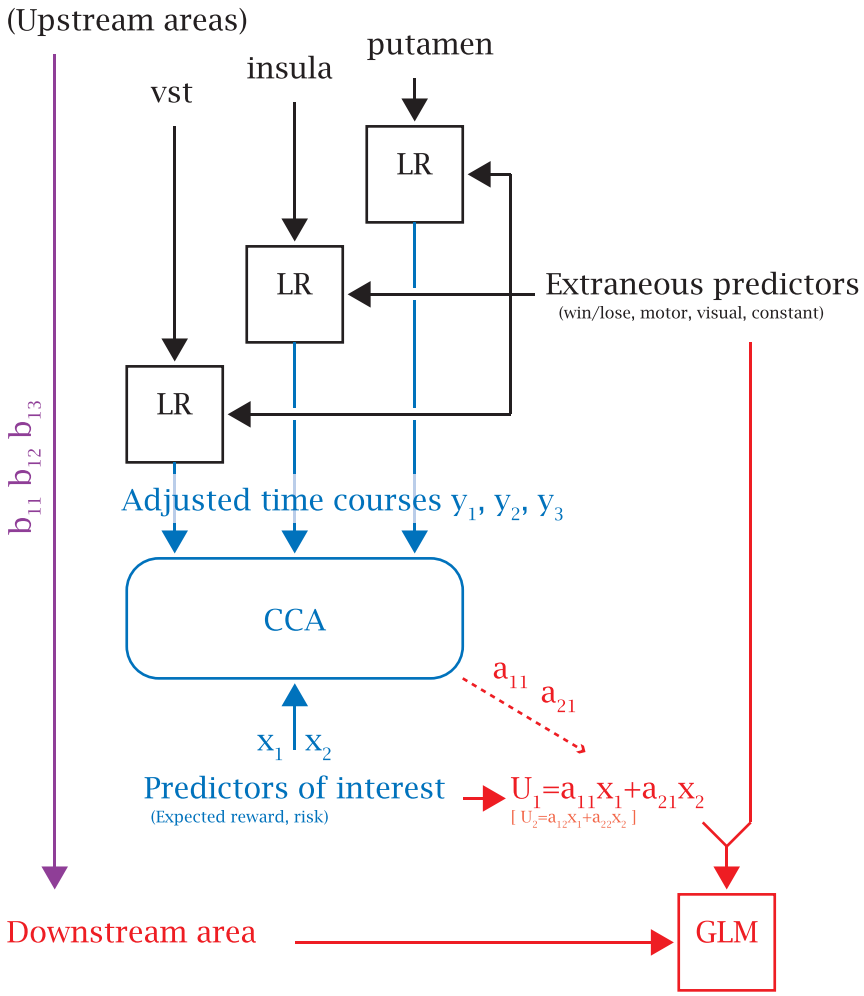


Figure 2.2: Overview of the CCA-based method. We perform the analysis in three steps: First (black section), we remove the influence of the extraneous predictor by computing a linear regression (LR). We create a design matrix with the extraneous predictors and we use the error (residuals) of the linear regression (labeled as *adjusted time courses*) as the input of the second step. In the second step (blue section), we use CCA to compute the weights on our predictors on interest. With these weights, we construct a new composite predictor. In the third step (red section), we compute a GLM and test if the composite predictor significantly explains the activity in the downstream region.

downstream region. Additionally, the third step provides a verification. If the results of the CCA step are conclusive but we do not discover any new area with the third step, it will cast a doubt on the validity of our approach. Indeed, it will mean that the correlations do exist but that the brain does not use them to combine signals in an optimal way.

One can wonder why we need a CCA at all. Indeed, the new composite predictor, U_1, is a linear combination of ER and risk (equation 2.7). A GLM is also a linear method that constructs a linear combination of ER and risk. Our approach has three essential advantages, however. First, it has a much higher statistical power. Indeed, to obtain the same statistical power, we would have to at least double the number of subjects in the study (appendix A.5.1). Second, if we had used a GLM approach with additional subjects, we could not have distinguished whether the downstream area encodes ER and risk in separate neurons located in a single area or whether the downstream region actually combines ER and risk in a single metric. Finally, with this approach, we confirm a second time the hypothesis of Salinas and Sejnowski [2001], this time with fMRI data instead of neuronal recording (Romo et al. [2003]). By observing the result from the CCA step, we discover how the recombination of several signals is done. Then, with the GLM step, we discover the area that performs the actual recombination.

In order to compare the GLM approach to the CCA one, we need to compute a second predictor, U_2. Indeed, we would like to have the same number of predictors in both cases to have a fair comparison with the original study (Preuschoff et al. [2006]). Several choices are possible, but we choose U_2 to be orthogonal to U_1 because this choice is optimal in terms of statistical power (appendix A.5.2); the second row of the CCA provides such an orthogonal predictor. It also provides a sanity check. If the downstream region truly encodes U_1, then the p-value attached to U_2 should be non significant.

Finally, we need to be able to perform statistical inference and hypothesis testing. Methods already exist to compute an overall p-value (Wilk's lambda, Anderson [2003], Johnson and Wichern [2002]). For the purpose of neuroscience though, we need a method that computes a p-value for each weight. Indeed, one cannot meaningfully talk of recombination of signals if one of the weights a_{ij} or b_{ij} is zero. We did not find any suitable method in the literature and instead devised a method that computes approximate p-values (appendix A.3).

Weight	p-value	Predictor / Region of Interest	Talairach		
			x	y	z
$a_{11} = 32$	$< 10^{-7}$	ER			
$a_{21} = 65$	$< 10^{-6}$	Risk			
$b_{11} = 0.24$	< 0.01	Putamen	-22	-8	8
$b_{21} = 0.45$	$< 10^{-7}$	Ventral Striatum	-12	5	-3
$b_{31} = -0.22$	< 0.0002	Insula	-31	21	9

Table 2.1: Result of the CCA computations. We concatenated the time courses and predictors for all the subjects and predictors before applying a CCA. We observe that ER and risk are added to form a new metric. Every coefficient is significant (tested with method described in appendix A.3).

2.5 Illustration

2.5.1 CCA results

We first ran a CCA on a subject-by-subject basis. We observed that out of the 19 subjects we used, 14 subjects had weights with identical signs (taken as positive) on the predictors for both ER and risk. This showed that the correlation of the noise did occur. Indeed, if there had been no correlation, we would have observed identical weights for roughly half the subjects and opposite signs for the rest (appendix A.4). Under the null hypothesis that no correlation exists, the probability that we observe a given number of subjects with identical signs followed a binomial distribution (with probability 0.5 for each case). The probability of observing 14 or more subjects with identical signs was $p = 0.032$ (binomial test).

Given the similarity of the subject-by-subject results, we ran an analysis with the signal from all the subjects concatenated. We also obtained identical weights (table 2.1). Indeed, we obtained a new composite predictor $U_1 = 32\,\mathrm{ER} + 65\,\mathrm{Risk}$. The three regions and the two predictors had significant p-values and the overall analysis had a p-value $p < 10^{-7}$ (Wilk's lambda).

In order to verify that these p-values were not due to serial correlation, we corrected for this effect with an ARMA method (Burock and Dale [2000]). Similarly, we used Singular Value Decomposition (SVD, Johnson and Wichern [2002]) to check that the correlations were not due to the recording system. With a linear regression, we removed the first eigenvector of an SVD from the time courses. In both cases, we observed similar results.

How can we interpret this new composite predictor? Since it is increasing in both ER

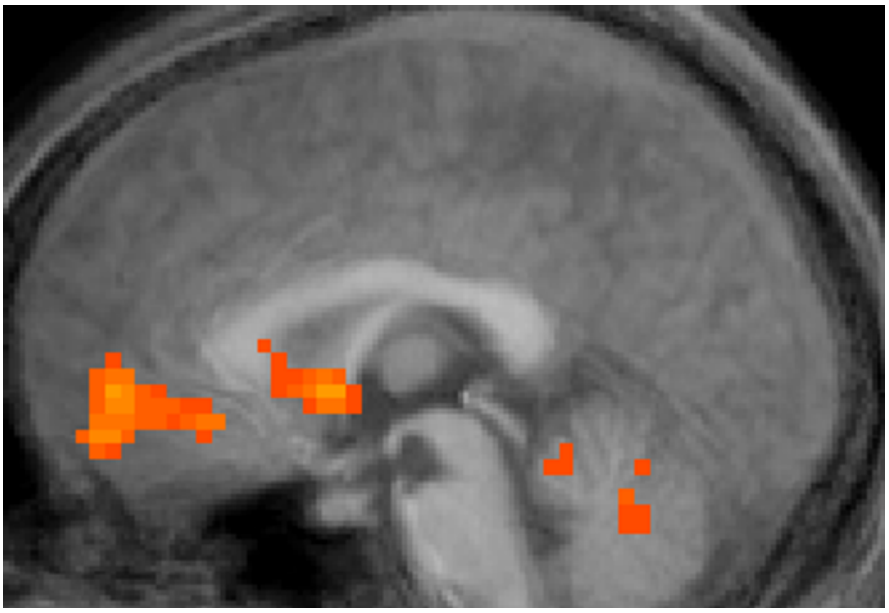


Figure 2.3: Activation to the new metric when we use a single set of weights. The area extends for 50 voxels around the center (0; 46; -2) in Talairach coordinates. We use the statistical threshold $q(FDR) < 0.025$.

and risk, it seems that it could represent a metric that reflects the impact of two opposing goals. The ER and risk are in opposition because humans seek high ER and low risk. In accordance with its role in evaluating gambles (Martino et al. [2006]), we hypothesize that the medial prefrontal cortex (mPFC) will be activated.

2.5.2 Subsequent GLM results

We ran two types of GLM analyses. In the first type, we used the single set of coefficients and ran a random-effect GLM for all the subjects. In the second type, we used a different set of coefficients for each subject and only used a subset of the subjects.

The analysis in the case of fixed weights was straightforward. We created a new composite predictor, U_1, from a single set of weights. We also added the second predictor, U_2, and computed a random-effect GLM. We observed that, as anticipated, the mPFC was activated by the new metric (figure 2.3). A large area extended around the Talairach coordinates (0; 46; -2). We also observed other areas. The putamen, insula, and ventral striatum appeared also activated. This was due to the fact that the new metric was not orthogonal to ER or risk. The mPFC was not reported active in Preuschoff et al. [2006]; our method had much

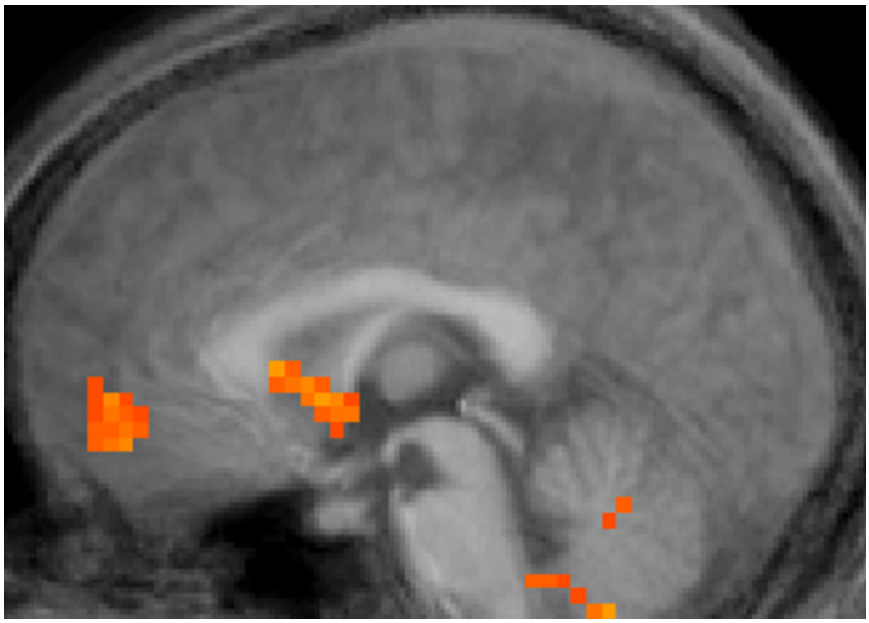


Figure 2.4: Activation to the new metric when we use a separate set of weights for each subject. We only used the subset of 13 subjects who had identical signs on both predictors with a Wilk's lambda $p < 0.10$. The area extends for 11 voxels around the center (1; 51; 13) in Talairach coordinates. We use the statistical threshold $p(uncorrected) < 0.001$.

greater statistical power (appendix A.5.1). Finally, as predicted, the predictor U_2 did not significantly explain the activity in the mPFC.

The analysis in the case of variable coefficients was more complex because two problems arose due to noise. First, how should we treat the subjects that displayed opposite signs? The noise may have inverted one coefficient, making estimates unreliable. Second, even if the two signs were identical, how could we have been sure that the estimates were precise?

We decided to exclude from the analysis the subjects that had opposite signs. To determine whether the estimates for the remaining subjects were of high enough quality, we used the p-value given by Wilk's lambda. If the p-value was above 0.10, we excluded the subject. We did not use this rather high statistical threshold to draw specific conclusions about the neuronal activity, but rather used this number as an estimate of the quality of the signal. Of the 19 original subjects, we kept only 13.

The results in the case of variable coefficients are very similar to the fixed coefficient case (figure 2.4). We observe a wide area in the mPFC in response to U_1 around the Talairach coordinates (1; 51; 13). The predictor U_2 did not significantly explain the activity in the

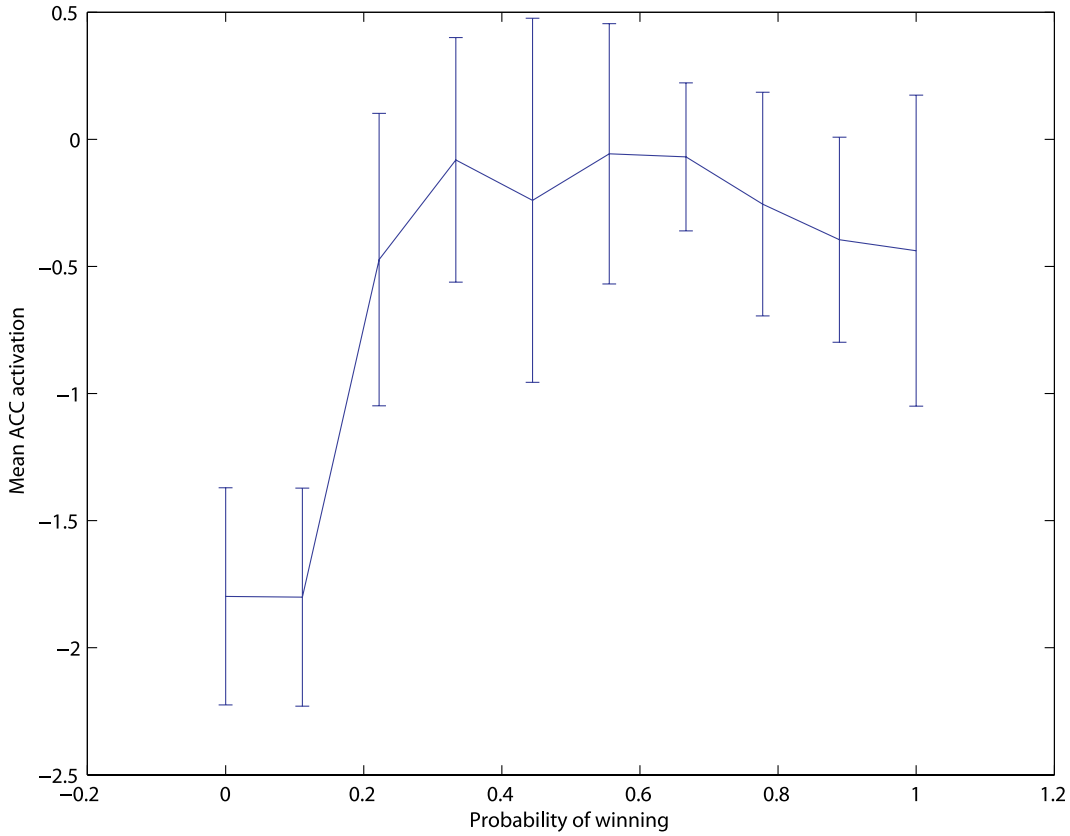


Figure 2.5: Region of interest analysis of the area found on figure 2.3. We average the signal of the region of interest and then we compute a GLM with a separate predictor for each probability of winning. The x axis represents the probability of winning and the y axis represents the coefficient attached to the predictor corresponding to this probability of winning. We also plotted ± 1 standard error. We confirm that the activation in the mPFC is positive in both ER and risk. It has the shape of an inverted U but is skewed by the ER (higher signal for higher probability of winning).

mPFC. We still saw the putamen, insula, and ventral striatum activated for the same reasons as before.

We then confirmed these results with a *region of interest analysis*. We averaged the fMRI time courses in the part of the mPFC that we just discovered, thus obtaining a single time course. We then ran a GLM that contained 10 predictors of interest. We created a separate predictor for each of the 10 probabilities of winning (corresponding to the 10 cards). Each predictor was a boxcar function with onset at the display of the first card and duration 1 s. We modulated the predictors by one when they corresponded to the current trial's probability of winning and by zero otherwise.

This region of interest analysis allowed us to check the previous results. It could be that

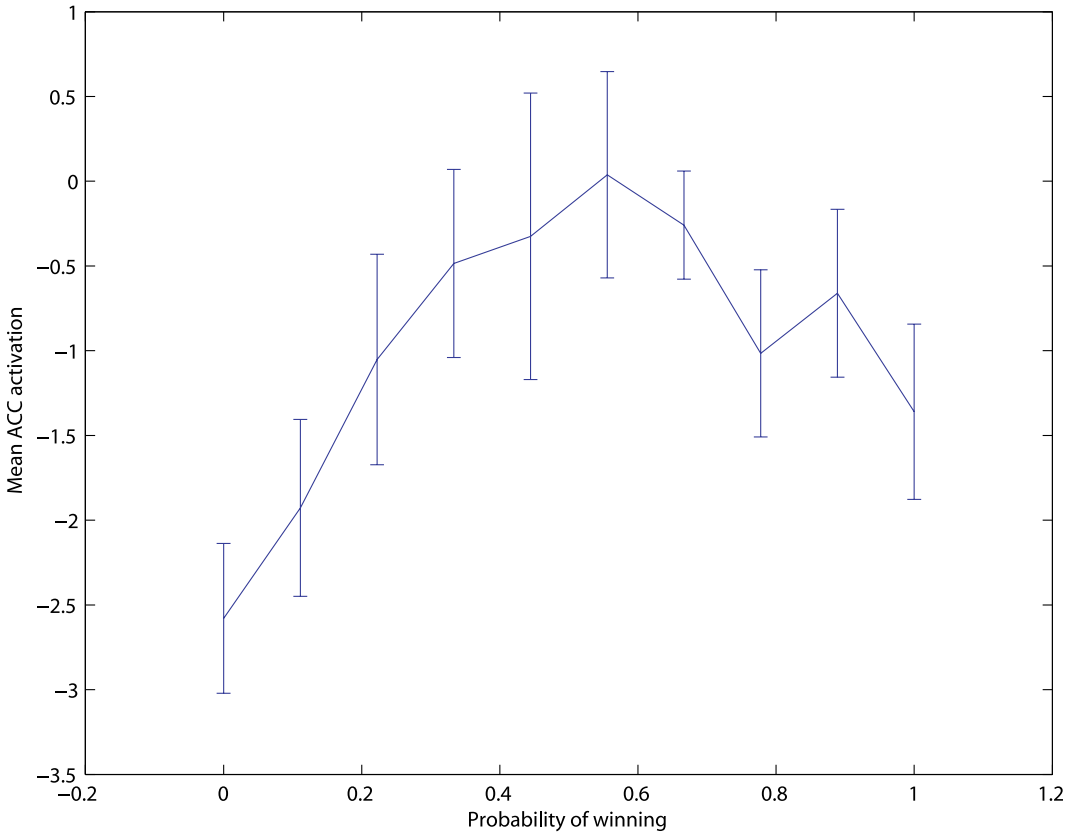


Figure 2.6: Region of interest analysis of the area found on figure 2.4. We average the signal of the region of interest and then we compute a GLM with a separate predictor for each probability of winning. The x axis represents the probability of winning and the y axis represents the coefficient attached to the predictor corresponding to this probability of winning. We also plotted ± 1 standard error. We confirm that the activation in the mPFC is positive in both ER and risk. It has the shape of an inverted U but is skewed by the ER (higher signal for higher probability of winning).

the mPFC responded in a way that was correlated with but different from the new metric that we created. In this case, we could still have observed a significant activation in the mPFC.

We predicted that we would observe a specific shape. Since the weight on risk is positive, we predicted that the overall shape would be an inverted U (quadratic function). However, this U would be skewed by the positive weight on the ER and we would observe a non-symmetrical response with a higher activity for higher probabilities of winning. The data confirmed our prediction, both in the case of fixed coefficients (figure 2.5) and variable coefficients (figure 2.6).

2.6 Discussion

2.6.1 Implications for neuroeconomics

It is surprising that risk is added to ER into a new metric. Indeed, one could have predicted that the correlation would have directed that risk is subtracted to ER to form a utility metric, as predicted by economic theory (section 1.3).

Several reasons can explain this result. First, computing a utility metric using only ER and risk is a poor choice because the result will only be an approximation. Second timing may explain why we did not obtain a utility signal. We created predictors as boxcar functions that lasted only 1 s from the display of the first card. It may be that the new metric we observed is an "advance" signal that tells the brain that more processing is needed. If we attempt to model other periods, we do not obtain any meaningful results that either confirm the use of the new metric or show that utility is used at a later stage. Third, it could be that new metric is indeed the only one used by the brain and future work could reveal why we observe paradoxes when using utility.

It is interesting that the addition of two signals into a new metric had previously been studied (Esteban and Ray [1999]). In this view, two competing groups lobby for a specific goal. The resulting metric is additive in the effort of both groups. In our experiment, both the regions that encode ER and the regions that encode risk "lobby" the mPFC and the resulting metric is additive in ER and risk. Our approach extends and clarifies the way the mPFC evaluates a gamble.

2.6.2 Implications for neuroscience

The use of the CCA was motivated by the hypothesis formulated by Salinas and Sejnowski [2001]. They posited that the correlations are used to direct an optimal integration of signals and we use CCA to investigate this correlation.

Several studies confirm this supposed role of the correlation. First, performance on a task is diminished when correlation is disrupted. Effective sensory perception in honeybees is impaired when the correlation is disrupted (Stopfer et al. [1997]). In rats, memory is also impaired (Robbe et al. [2006]). Indeed, if we disrupt correlation, we disrupt integration and the various neuronal processes that need it to combine signals. Second, in healthy humans, the perception of faces is concurrent with correlations that can be recorded with EEG (Rodriguez et al. [1999]). Finally, a lower attention level down-modulates correlation (Steinmetz et al. [2000]) and our hypothesis rightly predicts lower performance.

Our CCA-based method is not restricted to the particular experiment we presented. Indeed, we can use CCA with any experiment where the stimulus set elicits the activation of distinct areas to distinct predictors. The neuronal hypothesis is equally general. Even if this phenomenon has been observed in a specific experiment (Romo et al. [2003]), Salinas and Sejnowski [2001] do not prescribe any particular context in which the correlation operates. While it is possible that other neuronal mechanisms can concurrently direct integration, whenever the brain uses correlation, we will be able to use a CCA to discover how the integration is performed.

2.6.3 Advantages and limitations of our method

The method we propose is limited by the fact that we use fMRI data. First, we can investigate the role of correlation only between populations of neurons and not within a population. Indeed, the role of correlation within a population is limited to a group of about 100 neurons (Abbott and Dayan [1999], Schneidman et al. [2006], Shadlen and Newsome [1994]). Since fMRI does not have good enough a spatial resolution, we cannot investigate population correlation. Second, we used a linear model with an HRF function as a link between the neuronal and fMRI levels. While most studies use a standard HRF, researchers can use a balloon model if they want to use a non linear link (Buxton and Wong [1998]). Should the method we propose prove to be useful to others, future work could

adapt the balloon model to CCA.

While CCA has already been used in the context of fMRI under the name CVA (Friston et al. [1995]), our approach is fundamentally different. Instead of using a limited number of predefined time courses, Friston et al. [1995] use a data-reduction step to obtain a smaller data set. Then, they apply a canonical correlation. However, the results are difficult to link to the brain activity because of the data-reduction step. Additionally, they do not base their approach on biological observations. Instead, we base our approach on a specific biological hypothesis that neurons recombine signals in a way that minimizes noise (Salinas and Sejnowski [2001], Romo et al. [2003]).

Our approach can be seen as the combination of a GLM and a principal component analysis (PCA, Johnson and Wichern [2002]). A CCA is a generalization of a GLM. Instead of using a GLM to explain a single brain signal with several predictors, we use a CCA to explain several signals with several predictors. Indeed, if we had applied a CCA to a single brain signal, the results would have been the same as if we had used a GLM (appendix A.3). It is a generalization of PCA because it uses the same basic techniques but reintroduces predictors.

Our approach is more powerful than exploratory methods, such as independent component analysis (ICA, McKeown et al. [1998]). When using ICA, researchers do not create any *a priori* predictors. Instead, they let fMRI data describe the activity of the brain. This approach has two main drawbacks. First, it does not allow testing of falsifiable hypotheses as the brain activity can always be explained with *ad hoc* reasoning. Second, it may not always be possible to formulate a clear explanation of how the brain works. Precisely because this method forgoes predictors, its results are more sensitive to noise and harder to explain.

Our approach is complementary to other methods that investigate how several brain regions interact. With Psychophysiological Interactions (PPI, Friston et al. [1997]), researchers discover how areas work together. However, it is more difficult to investigate how one region integrates several signals. Researchers also investigate effective connectivity with structural equation models (SEM, Friston [2002]) and dynamic causal modeling (DCM, Goncalves and Hull [2003]). With DCM, they investigate at the neuronal level while with SEM they directly investigate fMRI data (Penny et al. [2004]).

These three approaches have three main drawbacks however. First, they suppose bi-

ological mechanisms that have not been observed independently with neuronal recording. Second, they are applied to predefined regions. Instead, our approach discovers a new downstream area. Third, they do not directly refer to predictors. While other methods offer more flexibility to explain fMRI data, the results are more difficult to relate to experimental conditions.

The method we propose has several advantages. First, it is based on a known biological mechanism instead of an off-the-shelf statistical method. Second, it creates a new predictor that is easy to interpret. Third, it allows researchers to test falsifiable hypotheses.

Chapter 3

How Utility is Indexed in the Cortex

In this chapter[1], we depart from chapter 2 in two ways. First, we investigate how two gambles are evaluated instead of one. Second, precisely because we have two gambles, we can investigate how subjects choose between them, instead of simply investigating how subjects evaluate a single gamble, as we did before. Furthermore, we introduce a new analysis technique; we investigate the overlapping activations of several types of neurons with a statistical test based on Hotelling's T^2 statistics.

3.1 Introduction

The present study is based on previous observations by Padoa-Schioppa and Assad [2006]. In this work, Padoa-Schioppa and Assad recorded the activity in neurons of two monkeys. While doing so, they presented monkeys with choices between two liquid treats and varied both the flavors and the amounts offered. By training monkeys to use eye movements to indicate their choices, they could record the neuronal activity while monkeys participated in the experiment.

This experiment was related to economic setups (see also section 1.3 for a primer). Indeed, monkeys had to perform a choice between two options. For example, a thirsty monkey may have preferred apple juice over water, but would he rather have chosen one drop of apple juice over five drops of water? The choice could be resolved with a utility function. Monkeys could compute the utility of both choices and pick the choice with

[1]This work was a collaboration with Hannah Bayer, Peter Bossaerts, Bernd Figner, Jack Grinband, and Elke Weber.

the highest utility. Padoa-Schioppa and Assad repeatedly offered these choices to the two monkeys and used their revealed preferences to compute the utility of each choice. Then, they recorded the activity of neurons and related it to the revealed utility.

They discovered four main types of neurons (for simplicity, we restrict ourselves to the choice of water vs apple juice and we do not add other types of liquid treats):

- some neurons fired proportionally to the utility of the amount of water offered
- some neurons fired proportionally to the utility of the amount of apple juice offered
- some neurons fired proportionally to the utility of the chosen treat (maximum of the two utilities above)
- some neurons fired in a binary way depending on whether monkeys chose water or apple juice

It is worth noting one important design detail. Padoa-Schioppa and Assad randomly alternated the side where one treat was presented. For example, they sometimes presented the apple juice to the left or to the right of the monkeys. This way, they ruled out a classification by position and proved that the flavor of the treat was the categorization criterion.

We would like to generalize the results in the case of humans confronted with monetary gambles. Two main questions arise.

First, how do humans index gambles? In the case of liquid treats, flavor is the obvious idiosyncratic parameter used for indexing. But when it comes to choosing between two gambles, it is not obvious what index the brain uses. While it is possible that humans use position to index gambles, it seems unlikely given that monkeys do not use it to index liquid treats. We hypothesized that either expected reward (ER) or risk could be used as an index. Indeed, these are the two main parameters that describe gamble (section 1.3) and they have been found to be encoded by the brain (Preuschoff et al. [2006]).

Second, how do we investigate our hypothesis by recording fMRI data? While direct neuronal recording has a high spatial resolution, a single fMRI voxel could encode the overlapping activations of the four types of neurons we described above. We will describe an extension of current analysis methods that allows us to discover the indexing method in the case of overlapping activations.

3.2 Experiments

For each trial, we simultaneously presented subjects with a choice between two gambles. Gamble A paid $\$x_a$ with probability p_a and $\$y_a$ with probability $1 - p_a$. Gamble B paid $\$x_b$ with probability p_b and $\$y_b$ with probability $1 - p_b$. We varied amounts between \$0.10 and \$78 and probabilities between 0.10 and 0.90. We asked subjects to play three sessions with about 55 trials per session.

We drew subjects from a restricted pool in order to screen for risk attitude and understanding of the game. We asked subjects to participate in a prior behavioral experiment where we presented them with the same stimulus set as described above and inferred from their revealed preferences whether they were risk averse, risk neutral, or risk seeking. We chose 18 subjects to represent a varied sample of risk attitudes and we asked some of them to return for an fMRI session.

We excluded one person from the 18 fMRI subjects because he/she did not understand the game. We had introduced *dominated* gambles that were either $\max(x_a; y_a) < \min(x_b; y_b)$ or $\max(x_b; y_b) < \min(x_a; y_a)$. In the first case, subjects should always choose gamble B, in the second case, they should always choose gamble A. One subject systematically chose the other gamble; we decided to exclude him/her and we restricted our analysis to the remaining 17 subjects.

3.3 Methods

3.3.1 Estimation of utility

We first estimated the utility function of each subject. We hypothesized that subjects displayed a constant absolute risk-aversion modeled by an exponential utility function (Holt and Laury [2002]). Since the maximum offered amount was \$78, we divided each amount by $x_{max} = 78$ and then computed the utility function:

$$U(x) = \frac{1 - \exp(-\beta x / x_{max})}{1 - \exp(-\beta)} - \frac{1}{2}. \tag{3.1}$$

The expression above was in the range $[-0.5; 0.5]$ no matter what β we choose. For numerical stability, we modified the above expression to:

$$U(x) = \frac{\exp(-\beta x/2x_{max})}{\exp(-\beta/2)} \frac{\sinh(\beta x/2x_{max})}{\sinh(\beta/2)} - \frac{1}{2}. \tag{3.2}$$

We optimally chose β for each subject. The optimal β values minimized the classification error and predicted the outcome of gambles correctly in 74% of the cases overall.

3.3.2 The Hotelling test

How do we investigate fMRI data where the signal of one voxel reflects the activity of several types of neurons? A similar problem appeared when researchers tried to investigate the visual system (Kamitani and Tong [2005]). However, Kamitani and Tong investigated parts of the visual system ("columns") that are highly structured, and we could not use their approach because we did not know of any *a priori* structure.

We decided to use a method based on Hotelling's T^2 test instead (appendix A.7, Hotelling [1931], and Anderson [2003]). Instead of testing the effect of a single predictor, we test several predictors at the same time. Since Hotelling's test is a generalization of the t-test from a scalar to a vector, it is the natural extension of the techniques previously used in fMRI.

The main advantage of the T^2 test we proposed was scale invariance. Indeed, if we had multiplied any predictor by an arbitrary constant, the results would have remained identical. This was particularly useful because we did not know *a priori* in what proportion each types of neuron were in each voxel. Additionally, this repartition was likely to change from voxel to voxel and from subject to subject. Thus we could not have created a composite predictor that reflected the overall activity of the various types of voxels because we would have needed to presuppose weight on each of the types of neurons. Similarly, we could not have used contrast techniques because they also presuppose predefined weights.

3.3.3 Construction of predictors

We constructed predictors with the following behavioral considerations. Subjects were free to take as much time as they needed to decide between the two gambles and most of them usually took from 1 s to 2 s to answer. However, in some trials, subjects took less than 0.5 s or more than 10 s; we labeled these trials as "bad." Additionally, we also labeled as bad the trials where subjects picked a dominated gamble.

All the predictors were boxcar functions of various heights with an onset at the presentation of the gamble and an offset at the subjects' decision time:

- We created a predictor modulated by a constant during good trials. We used it to capture the average brain activation during the good trials.
- We created a similar predictor for the bad trials.
- We created two separate predictors modulated by the revealed utility of each of the two options in the good trials. We tested three different rules to index gambles. First, we indexed by position. One of the two predictors captured the utility of the gamble presented to the left while the other captured the utility of the gamble presented to the right. Second, we indexed by ER. One of the two predictors captured the utility of the gamble with the lowest ER while the other captured the utility of the gamble with the highest ER. Third, we indexed by risk. One of the two predictors captured the utility of the gamble with the lowest risk while the other captured the utility of the gamble with the highest risk.
- We created one predictor modulated by either +1 or -1, depending on the subjects' choices. For example, depending on the indexing rule we used, we assigned a modulation of +1 when the subjects picked the gamble to the left, with the lowest ER, or the lowest risk.

3.4 Results

We observed that the mPFC (around (5;37;21)) was activated at the threshold $p(uncorrected) < 0.001$ when we indexed gambles by risk (figure 3.1). When we indexed gambles by ER or position, no such activation appeared (figure 3.2). We did not see any activations in the mPFC when we indexed gambles by ER or risk until we raised the threshold to $p(uncorrected) < 0.01$.

3.5 Discussion

These results strongly suggest that risk indexes gambles in the human mPFC. As predicted, the position of gambles is not important. The brain could have used ER to index gambles,

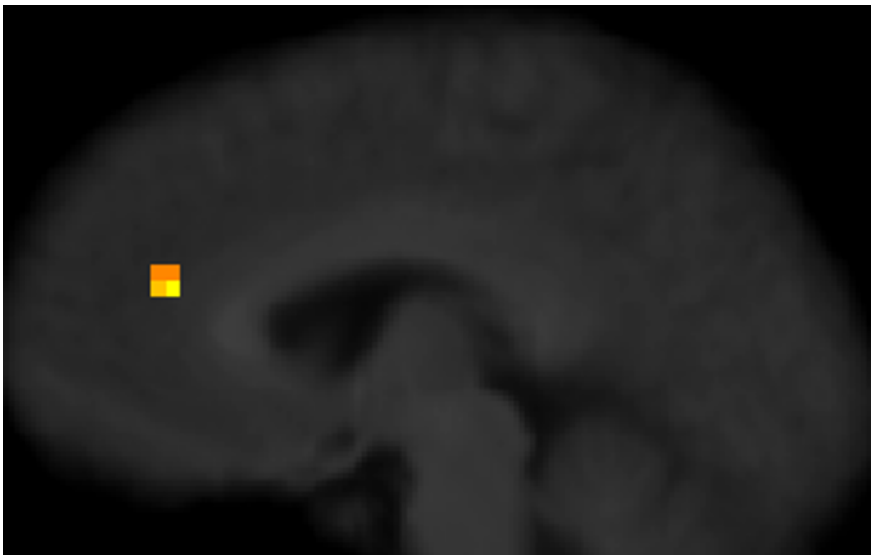


Figure 3.1: Activation of the mPFC when using risk as an indexing method. We show here the brain activation when we use risk as an indexing method. A cluster of voxels around the Talairach coordinates (5;37;21) appears in the mPFC. We used random effects analysis, corrected for serial correlation in fMRI data with ARMA models, and used the threshold $p(uncorrected) < 0.001$ (equivalent to $F_{3,14} > 9.70$).

but our results reveal that this is not the case.

How do we relate these results to the ones in chapter 2? While we could not previously find a utility function and instead found a metric that is positive in both ER and risk, here we find a utility metric.

First, the two areas are not overlapping. The activation in figure 3.1 is more dorsal than in figure 2.3 on page 20. Thus, it may be that separate regions specialize in different tasks. As mentioned in section 2.6.1, it may be that the utility is only computed after some delay period. In that sense, timing would be an essential parameter that would dictate which metric is used.

Second, computing the utility as the difference of ER and risk is only an approximation (section 1.3). Thus, neither the brain nor a CCA (chapter 2) can compute the exact utility function directly from the signals in the insula, ventral striatum, and putamen. Our approach also suggests that estimating the utility as a constant risk-aversion function is more appropriate.

We thus conjecture that the separate encoding of ER and risk serves two purposes, depending on timing. Early after the display of a stimulus, the two signals are added to

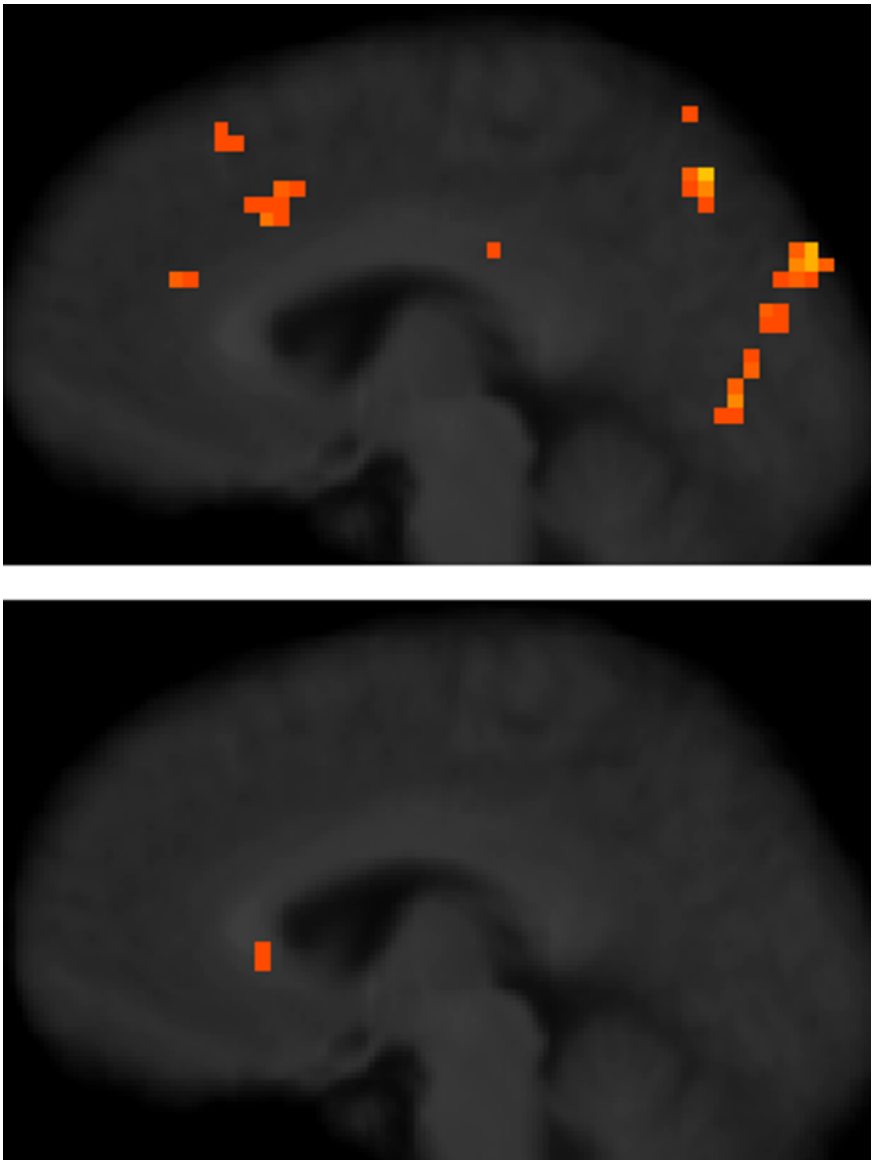


Figure 3.2: Activation of the mPFC when using ER or position as an indexing method. We use here a statistical threshold that is one order of magnitude higher than that in figure 3.1 ($p(uncorrected) < 0.01$). The top half of the figure represents the indexation of gambles by ER, the bottom half by position. In both of these cases, we observe that the mPFC is not activated and that voxels are scattered around the brain, signifying too high a threshold.

form a metric that indicates the saliency of the gamble. Later in the analysis process, a utility function is computed, not by computing the difference of the signals from the areas that encode ER and risk, but rather on its own. The signal of risk is only used as an index.

Why does the brain not use laterality as an index? If the brain indexes gambles by any other way than using laterality, a second region must disentangle the result and map the favored gamble to a specific motor action. In our case, if the gamble with, say, the lowest risk is preferred, then a mapping from a risk index to a laterality index has to be performed in order for the subject to press the correct button. It may be that an index that does not use position is useful because it is more general. Indeed, it is likely that one region specializes in the evaluation of gamble and that another maps the results into a motor action. Thus, the brain will use an idiosyncratic index instead of laterality.

Why does the brain choose risk over ER? We conjecture that risk is chosen because it has a higher discriminative power. Indeed, in most simple situations, the risk can be neglected and the attractiveness of a gamble can be approximated by ER only instead of laterality. In this case, the brain could not use ER as an index. Thus, risk is better suited to distinguish two gambles.

Finally, our novel analysis approach suggests a new investigation of the categorization problem. Ultimately, it could help understand how humans classify other abstract objects.

Chapter 4

Exploring Trader Intuition

In this chapter[1], we investigate an economic experiment much more complex than that in chapter 3: financial markets. The present setup is more complex because of the sheer number of parameters to take into account. To simplify our investigation, we decide not to investigate decisions, as we did in chapter 3, but only examine perception, as we did in chapter 2. We present three consecutive experiments. We use the first experiment with the sole purpose of collecting data. Then, we record fMRI data during the second experiment. Finally, we collect behavioral data in the third experiment.

4.1 Problem statement

How do humans perceive financial markets? This question is vast and there has been little exploration with fMRI. We first need to restrict ourselves to a much narrower question. We investigate here how humans infer knowledge from a financial market.

The *efficient market hypothesis* supposes that markets reflect all the current information available (Fama [1970]). Specifically, this hypothesis supposes that the prices of traded assets accurately reflect all of the publicly available information. Given the intense competition between traders, the prices will quickly converge toward the "right" price.

We would like to understand how the information is transmitted in the market. Indeed, no trader can possess the knowledge of the entire information available at a given time. However, information is transmitted, through the market, from one trader to another. For example, a trader in oil futures may follow more closely the news about the Middle East, while a trader interested in the stocks of U.S. retailers may follow more closely data on

[1]This work was a collaboration with Peter Bossaerts, John Ledyard, Steve Quartz, and William Zame.

domestic consumer spending. However, gasoline prices have an effect on spending. In this over-simplified view, if the market is efficient, the price of barrel of oil will reflect the entire information available on worldwide oil production. The traders of retailers' stocks need not follow current international events, they can infer this information from the price of a barrel of oil (provided the market is efficient). But how do they make this inference?

We would like to use fMRI to begin to understand how humans extract information from financial markets. To do so, we first need to obtain some trading data.

Historical data could not have been used for our experiment. Trading data from real exchanges, such as the New York Stock Exchange (NYSE), is not suited for our need because it is too complex and poorly controlled. Trading on exchanges is extremely rapid and complex. There are too many participants, stocks, and trades. Additionally, it would not be possible for experimenters to know what piece of information influences changes in stock prices. Finally, we would not have any way to control the flow of information. Experiments with fMRI require carefully controlled trials (see section 1.2) and there is no "control state" for the NYSE.

We could not have used simulated data instead. If we had created data by simulating stock prices, we would not have been able to take into account how information is spread. Indeed, this is precisely what we wish to investigate. Since we do not know exactly how stock prices reflect the available information, we cannot create accurate simulated data.

We thus resorted to running a prior data-collection experiment. With this experiment, we could acquire realistic data while at the same time controlling the flow of information. We do not use real-life stocks, but artificial ones. We release information in a controlled way and we record the entire trading activity. We then replay this trading activity in front of new subjects while recording their brain activity.

Our hypothesis is that humans use *Theory of Mind* (ToM) to understand stock markets. Theory of Mind is the ability to attribute internal mental states to others in order to predict other people's behavior and to regulate one's own behavior accordingly (Wimmer and Perner [1983]). Our hypothesis is a generalization of the use of ToM; humans would not only use ToM to understand living entities but also to understand large-scale electronic exchanges. According to this hypothesis, we predict that the brain areas that had been found to participate in ToM will be activated.

However, activation of brain areas associated with ToM is not enough because brain

areas often perform several duties. To corroborate results of an fMRI analysis, we also perform a behavioral experiment. In this experiment, we test subjects' abilities in ToM and predictions of a stock market. Additionally, we also test them using a mathematical quiz.

4.2 Three experiments

4.2.1 Financial market

Researchers collect data with an experimental market where they recruit about 20 subjects to participate in a experiment (Barner et al. [2005]). These subjects trade artificial assets on computers connected to a central server that organizes the trades. After the experiment, the researchers receive data files with all the trading activity and they analyze how subjects behaved. The market is entirely anonymous in the sense that subjects do not know whom they are trading with.

Let us now describe the specific experiment we used. We asked 20 subjects to trade during 13 independent sessions that lasted 5 minutes each. At the beginning of each session, we endowed subjects with varied amounts of cash, bonds, and stocks. Then we told subjects to trade as they saw fit and we paid them at the end of the experiment according to their performance.

The rules of the market were as followed. Subjects could buy the first type of stock ("stock X") for any price that was between 0¢ and 50¢ as long as they had enough cash. They could sell the first type for any price as long as they owned the stock. Each unit of stock X paid a *dividend* that was also between 0¢ and 50¢. However, subjects did not know the dividend at the beginning of each session; they only learned it after all the trading was over. Subjects could not trade the other type of stock ("stock Z") but they received a dividend from this stock too. This dividend was the complementary of stock X's, i.e., the sum of the dividends of both stocks was 50¢. For example, if subjects learned at the end of a session that the dividend of stock X was 42¢, then the dividend of stock Z was 8¢.

We paid subjects according to the following rule. For each session, we computed their earnings according to the dividends, how many of each stocks subjects owned, and how much cash they had. For example, if at the end of a session, a subject owned 8 units of stock X, 4 units of stock Z, and he had 32¢ in cash, we paid him $8 \times 42 + 4 \times (50 - 42) + 32 = 400$¢ (the dividend of stock X was 42¢). We added these \$4 to the subject's earnings and we

paid the total amount to him at the end of the experiment and with real cash. The trading price of the stock did not influence the payoff.

Since the subjects did not know the dividend until the trading session was over, the value of one unit of stock X was 25¢. Indeed, the dividend of stock X was a random number chosen between 0¢ and 50¢, and thus it was on average 25¢. Since subjects did not know what the true value of the dividend was, they could only estimate it.

We predicted that in this setting the price of stock X would converge to 25¢. Whenever the price of stock X dipped below 25¢, a subject should have bought the stock (paid less than 25¢ for something that paid on average 25¢). This purchase tended to push the stock price back up to 25¢. Similar reasoning applied when the price of stock X climbed above 25¢; subject should have sold the stock, lowering its price. In our experimental market, event though we did not detail the above reasoning to subjects, we did observe that the price of stock X converged to the predicted value (Bossaerts et al. [2007]).

The above reasoning was based on the hypothesis that subjects were risk neutral (see section 1.3). While it was not generally true, the non-tradable stock Z cancelled the risk-aversion parameter and our market was not influenced by the risk attitudes.

The setup we just described constituted the baseline of our experiment; we used sessions like the one above as controls. To obtain test sessions, we needed to add some information to be disseminated. We added this information in the form of a number called the *signal*. We chose the signal to be within 10¢ of the true dividend and we only gave this information to a subgroup called the *insiders*. We called the other group the *outsiders* and we randomly assigned subjects to one of the two groups. For example, if the dividend of stock X was 12¢, then at the beginning of the experiment, we would give a signal between 2¢ and 22¢ to the insiders; we would not give any additional information to the outsiders.

At the beginning of each session, subjects from each groups valued stock X differently. Outsiders did not know the signal and valued stock X at 25¢. Insiders knew a more accurate estimate of the dividend and gave stock X a value equal to the signal.

During the trading, both groups had to act strategically. Insiders had an edge when trading with the additional knowledge. However, they had to trade discreetly in order to avoid disseminating information to outsiders. Outsiders had to pay attention to the trading activity and they had to attempt to infer the signal. For example, if the signal was at 12¢ and an insider sold one unit of stock X for 25¢, she would make for sure 3¢, and on average

13¢. Indeed, she knew that stock X would pay a dividend that would be for sure between 2¢ and 22¢. By selling her stock, she would tend to lower the price, thus disseminating some information about the signal.

These two types of trading sessions constituted the two types of conditions. Sessions without insiders constituted the control conditions. Sessions with insiders constituted the test sessions.

4.2.2 Scanner experiment

During the scanner experiment, we replayed the trading data while recording the brain activity of 19 new subjects. The subjects could not trade; the trading had already occurred.

We replayed the 13 sessions with a simplified display (figure 4.1). First, we asked subjects to place a blind bet either onto stock X or stock Z. We designed this bet to give subjects a sense of "randomness" to subjects and to make sure that each played a different game. We paid subjects according to the dividend of stock X, their choice, and their bet: 10 times the dividend of the chosen stock. For example, if a subject had placed a bet on stock X and it paid a dividend of 23¢, we would pay her $2.30 for this session. In that sense, fMRI subjects were outsiders who did not trade. Finally, we told subjects that "insiders" did not refer to illegal insider trading.

After each subject had placed a bet, we replayed the market activity with a simplified display (figure 4.1). We showed the offers to buy (*bids*), sell (*asks*) and all the trades[2]. Since the subjects did not have any trading to perform, we asked them to press a key every time a trade occurred (attention task).

Finally, we revealed the outcome of the blind bet and we continued with the following session.

We built two types of predictors: block predictors and parametric predictors (figure 4.2). Block predictors were boxcar functions modulated by 0 or 1 to indicate whether the session had insiders or not. Parametric predictors were modulated by the distance of the latest trade price to 25¢, i.e., $|price - 25|$. We chose this modulation for two main reasons.

[2]A stock market is like a swap-meet. Traders indicate on the computer that they are willing to buy or sell certain amounts of stocks at certain prices. Many offers to buy and sell from one trader are not matched by another trader and no money and stocks are exchanged, i.e., offers to buy (bids) and offers to sell (asks) remain outstanding. The prices of the asks always higher than the prices of the bids (people want to sell stocks for a high price and buy them for a low price). However, when a bid or an ask is matched, a trade occurs and money and stocks are exchanged.

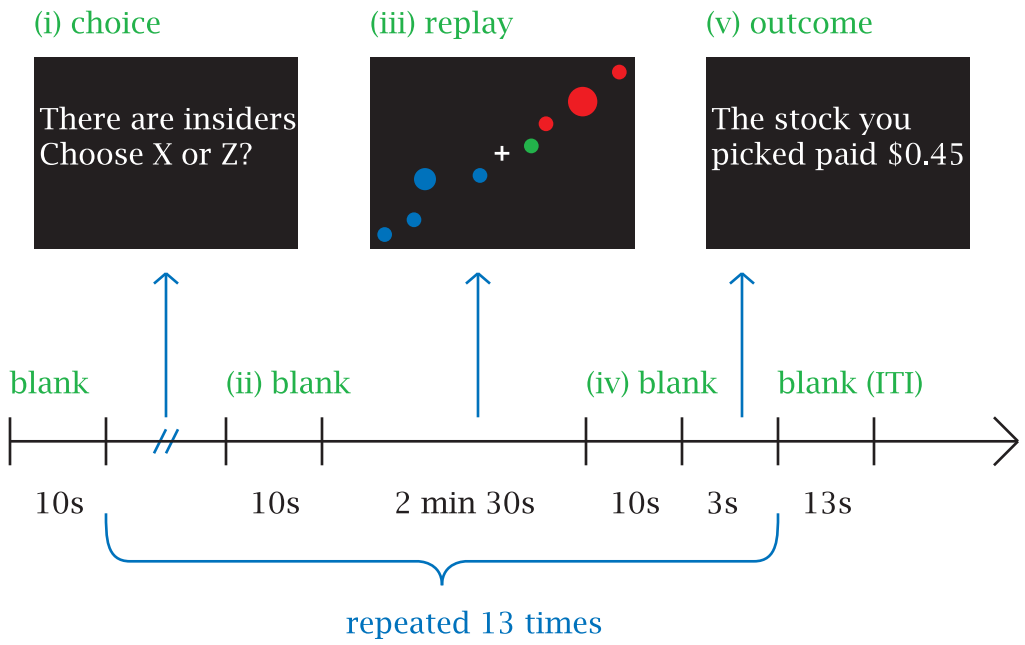


Figure 4.1: Stimulus setup for the fMRI experiment. We replay the 13 trading sessions in a random order. First, we ask subjects to place a blind bet on either stock X or stock Z (section (i)). After a blank screen of random duration (mean of 10 s), we replay the market activity for stock X with a simplified interface (section (iii)). After another blank screen, we inform the subjects of the outcome and repeat the session. We uploaded a video of section (iii) for download at `http://etd.caltech.edu` and `http://www.bruguier.com/pub/stockvideo.html`. In this video, the blue circles indicate the offers to buy (asks) and the red circles indicate the offers to sell (bids). The number inside each circle is the price in cents and the diameter of each circle indicates the number of outstanding offers. The replay is anonymous, i.e., each circle is the aggregate of all the outstanding offers. When a trade occurs, a circle turns green for 500 ms and then returns to its original color. The circles are ordered by increasing price along one diagonal (chosen at random for each period) and the other diagonal is used to space out the circles. The circles grow, shrink, move, and disappear to reflect the trading activity in the market.

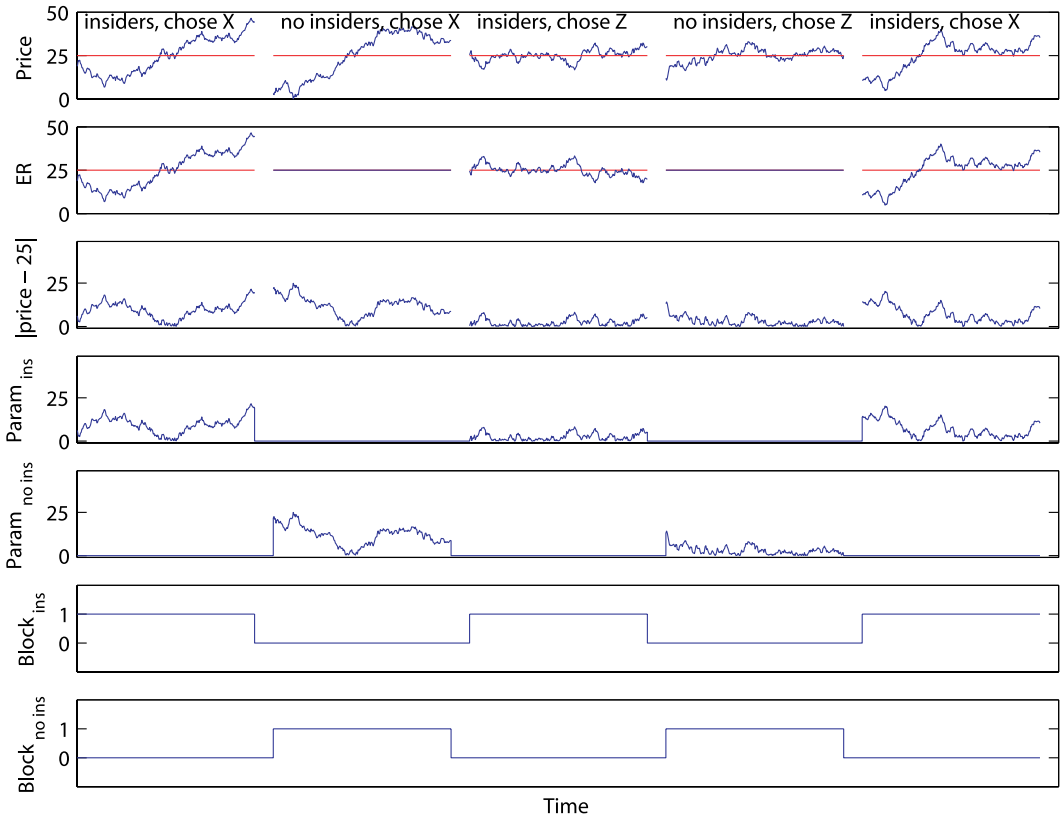


Figure 4.2: Overview of the method we use to create predictors. We present here how we create the two types of predictors (block and parametric). For each of the five fictive sessions (the x axis represents time), we already know the stock price (first row, the y axis represents the price in cents). We also indicate whether insiders are present in the market and whether the fMRI subject have placed a blind bet on stock X or stock Z. We use an horizontal red line to plot the payoff when no insiders are present (25¢). On the second row, we plot the ER for the fMRI subject. If no insiders are present, fMRI subjects expect to earn 25¢; if insiders are present, they expect to earn either the latest stock price or 50 minus the latest stock price. On the third row, we compute $|price - 25|$. The further the trading price is from the uninformed payoff, the more the effect of the insiders is evident. With this predictor, we control for the confounding factor of ER. Indeed, the signal that models the distance between the 25¢ and the trading price is orthogonal to the signal that models ER. To control for other phenomena, we also split the predictors between sessions with insiders and sessions without (fourth and fifth rows); these signals are the parametric predictors. Finally, to investigate the overall effect of insiders, we also compute block predictors (sixth and seventh rows). These predictors are also orthogonal to the ER.

First, the further the price was from 25¢, the more evident the action of insiders was. Many metrics were possible but we chose the absolute value because it was the most simple. Second, this metric was orthogonal to many confounding factors. Since expected reward (ER) was related to the stock price (or 50 minus the stock price, depending on the blind bet), the parametric predictor was orthogonal to ER. Similarly, the parametric predictor is also orthogonal to risk (as measured by the bid-ask spread).

We contrasted the test and control sessions for each type of experiments. Since many other factors may have influenced brain activity, we isolated the specific effect of adding insiders to the market. We had designed the data-acquisition experiment so that the number of trades was roughly the same in both conditions.

4.2.3 Behavioral experiment

We tested 43 new subjects' abilities in three areas: ToM, prediction of a financial markets, and mathematics.

ToM is thought to be an essential human ability. This ability emerges early in childhood. Older children understand that other people may believe that a piece of candy is hidden under a cup even though they themselves know that the candy is inside a drawer. As evidenced by their ability to succesfully pass a *False-Belief* test, children understand that other people may have beliefs different from their own (Wimmer and Perner [1983]). It has been conjectured that autism impairs ToM (Baron-Cohen et al. [2000]). Depending on how functioning patients are, they may or may not be able to pass the False-Belief test.

For the purpose of our experiment, we wish to test healthy adults. While advanced ToM tests exist, such as the *Faux-Pas* test (Stone et al. [1998]), we need a test that is difficult enough for healthy adults and that provides a continuous score, instead of a binary one, as most tests give.

We therefore use the *Eye* test (Baron-Cohen et al. [1997], figure 4.3). The Eye test consists of displaying pictures of the eye and nose areas of humans and asking subjects to pick one word out of four that describes what the person on the picture is thinking or feeling. We use a computer interface and we only give 10 s for the subjects to answer. We reward right answers, do not punish wrong answers, and penalize indecisions. Thus, it is always better to guess than not answer. We obtain a score between 0 and 36 that reflects a subject's ability in ToM. We also created a new test for ToM (appendix B.3).

Figure 4.3: Sample picture of the Eye test. We ask subjects to choose one word among four ("apologetic," "friendly," "uneasy," or "dispirited") that describes what this person is thinking or feeling. The correct answer is "uneasy."

The second test was a stock market prediction task. With this test, we measure the subjects' ability to predict changes in prices. We replay four sessions from the original experiments with insiders. Every 5 s, we stop the display. For half of the pauses, we ask subjects to predict the changes in prices. Specifically, we remind the subjects of the latest trade price and we ask them to predict whether the next trade is going to occur at a higher, lower, or identical price. For the other half of the pauses, we remind the subjects of their prediction and we tell them whether it was correct or not. We reward correct predictions, do not punish wrong ones. However, we only give 5s to subjects to give a prediction and we penalize indecision, thus it is always better to guess.

According to our hypothesis, we predict that ToM and stock market abilities will correlate positively. However, this correlation can be caused by other factors. Indeed, general intelligence or the level of alertness can be confounding factors. In order to control for these possible effects, we introduce a mathematical test.

With our third test, we record the subjects' mathematical abilities (table 4.1). These questions were typical of the ones asked during job interviews for finance positions (Crack [2007]). We ask subjects 7 questions and give them 30 s to read them and type an answer. We do not allow the use of paper or pocket calculators. We obtain a score ranging from 0 to 7.

According to our hypothesis, we predict that we will not observe any significant correlation between the stock market task and the mathematical quiz. In addition, this test will provide a control for the general state of awakeness. Indeed, the more alert the subjects

Consider a game played with a deck of three cards: spades, clubs, and hearts. Your goal is to identify the hearts. The cards are shuffled and displayed in a row, face down. You make your choice. The dealer then turns over one of the two remaining cards, provided it is not hearts. He then offers you the possibility to change your choice and switch to the other card that is left face down. What is the best strategy? Should you switch, stay, or does it not matter? Answer below "switch," "stay," or "either."	switch
Consider a deck of four cards: spades, clubs, hearts, and diamonds. The cards are shuffled and displayed in a row, face down. You choose one card at random and it is discarded. Then the dealer turns over two cards, chosen at random, but provided they are not hearts. Now there is only one card left unturned. If the two cards the dealer turns over are diamonds and clubs, is the probability that the remaining one is hearts more than, less than, or equal to 0.5? Answer below "more," "less," or "same."	less
There are 8 marbles that weigh the same, and 1 marble that is heavier. The marbles are all uniform in size, appearance, and shape. You have a balance with 2 trays. You are asked to identify the heavier marble in at most 2 (two) weightings. How many marbles do you initially have to place on each tray? Input a number below.	3
Divide 100 by 1/2. Is the result more, less than or equal to 100? Answer below "more," "less," or "same."	more
Jenn has half the Beanie Babies that Mollie has. Allison has 3 times as many as Jenn. Together they have 72. Does Mollie have more than, less than, or equal to, 20 Beanie Babies? Answer below "more," "less," or "same."	more
Johnnys mother had three children. The first child was named April. The second child was named May. What was the third child's name? Type the name below.	Johnny
The police rounded up Jim, Bud and Sam yesterday, because one of them was suspected of having robbed the local bank. The three suspects made the following statements under intensive questioning. Jim: I'm innocent. Bud: I'm innocent. Sam: Bud is the guilty one. If only one of these statements turns out to be true, who robbed the bank? Type the name of the robber below.	Jim

Table 4.1: Mathematical quiz. We ask subjects to answer within 30 s by typing their answer on the keyboard. No pencil or paper is allowed and we penalize subjects for failure to answer. We do not penalize wrong answers and reward right answers.

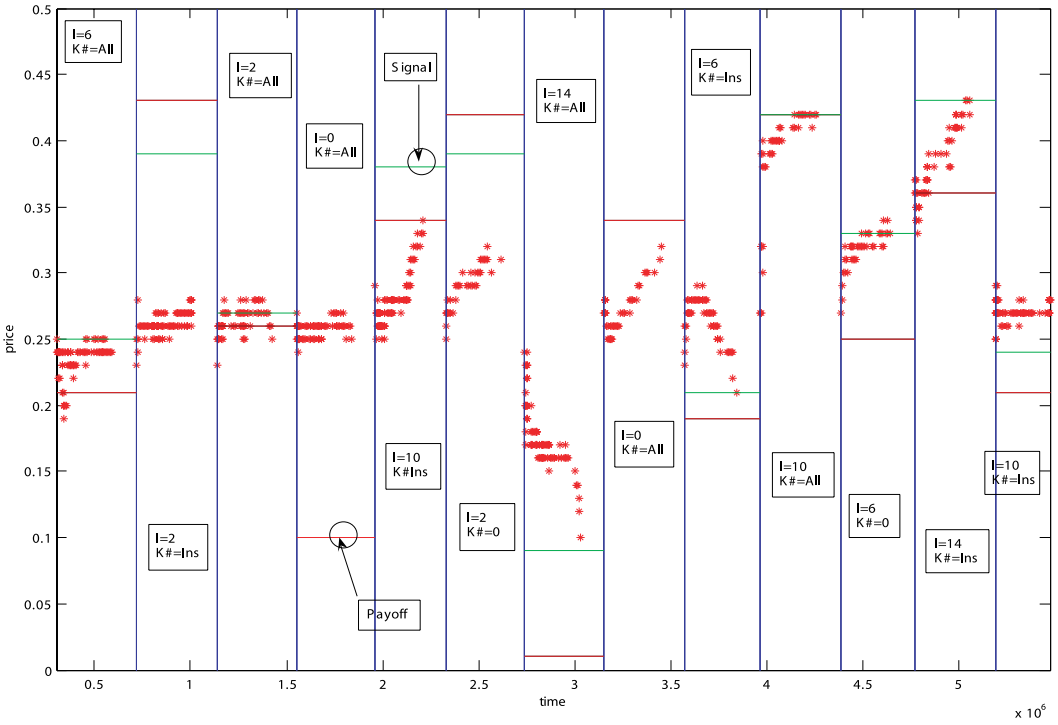


Figure 4.4: Trading activity during the prior data-collection experiment. On the x axis, we display the time in miliseconds. On the y axis, we display the trading price, in dollars. We use vertical blue lines to separate the 13 trading sessions. We display a red star for each trade that occurs. For each session, we indicates with a box the number of insiders (`I=`) and who knows how many insiders there are (`K#=`). Before the start of each session, we give insiders a signal (horizontal green line) that indicates approximately what the dividend (horizontal red line) is going to be; the signal is within 10¢ of the dividend. We can see that during the sessions that contain insiders, the trading price tends to converge toward the signal.

are, the more likely they are to perform well on any of the tasks. If the influence of this counfouding factor is small, then we predict that we will not see any correlation between the abilities to predict stock market prices and the mathematical savvy.

4.3 Results

4.3.1 Data collection experiment

The experiment lasted two hours, including instructions, practice, and trading periods. Trading was brisk throughout; on average, an offer was entered or canceled every 0.7 s and a transaction took place every 3.2 s. Subjects made $55 on average.

x	y	z	cluster size	t_{17}	Area
-30	-7	11	5	4.476	left insula
-14	23	39	5	4.688	frontal part of the anterior cingulate cortex
-9	41	36	22	5.380	paracingulate cortex
-9	32	45	6	4.290	frontal part of anterior cingulate cortex
17	36	43	6	6.322	frontal part of the anterior cingulate cortex
21	-10	-12	5	5.160	right amygdala

Table 4.2: Brain activations in response to the contrast of the parametric predictors. We compute the contrasts of the two parametric predictors (insiders minus no insiders) in a random-effect GLM. Here we report the brain areas that are significantly activated. We use the statistical threshold $p(uncorrected) < 0.001$ and we only report the clusters of more than 5 voxels. In addition, we also indicate the highest t-value of each cluster.

We show a summary of the trades in figure 4.4. We varied both the number of insiders (from 0 to 14) and who knew how many insiders there were in the market (nobody, just the insiders, both insiders and outsiders). In the case when there were insiders, we represent the signal they had been given with a horizontal green line. Whenever we added insiders, the price of the trades tended to converge toward the signal. The speed of the convergence was varied however. For example, in session 6, the price only approached the signal while in session 10, the price reached the signal. When no signal was given, prices stayed around 25¢ (session 4) or irrationally diverged (session 8).

This high variability of the results was useful for the subsequent experiment. Indeed, if we only had observed divergence from 25¢ in sessions with insiders, the comparison of the control and test sessions would have been difficult.

4.3.2 Scanner experiment

We first contrasted the parametric predictors (table 4.2). We found a significant contrast in the paracingulate cortex (PCC, figure 4.5). The area extended for 22 voxels in the Brodmann areas 9/32 (-9; 41; 36). We also found that a smaller region (5 voxels) of the anterior cingulate cortex (ACC) was activated (-14; 23; 39).

Finally, we also discovered significant contrast in the right amygdala (5 voxels around 21; -10; -12; figure 4.6) and the left insula (5 voxels around -30; -7; 11; figure 4.7).

To verify these results, we plotted the averaged activity of the brain in the PCC, amygdala, and insula, as a function of the distance to 25¢ (figure 4.8). We observed that the activity in the three regions followed a linear pattern.

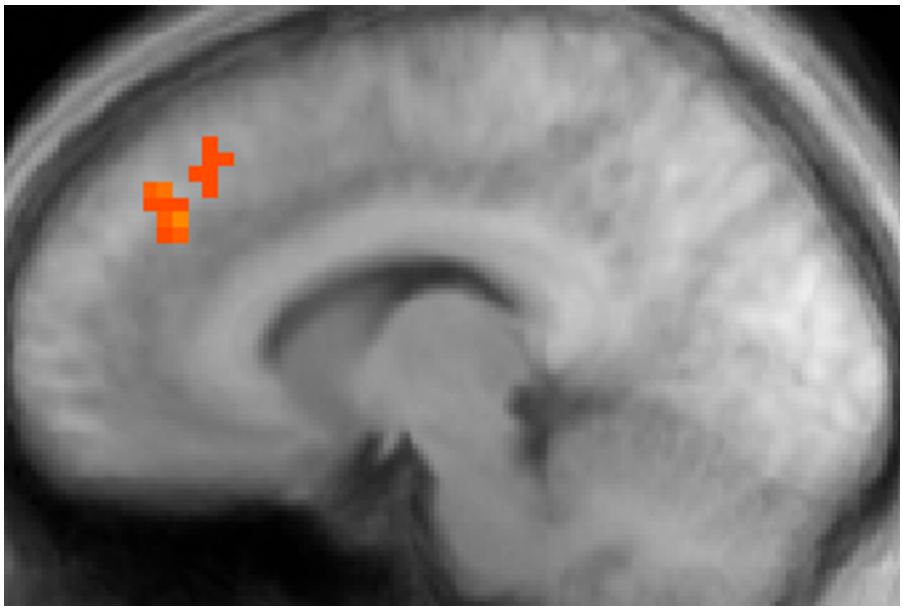


Figure 4.5: Activation pattern when we contrast the parametric predictors (paracingulate cortex). We compute a random effect GLM with a statistical threshold $p(uncorrected) < 0.001$ and we observe that our data significantly explains the activity in the paracingulate cortex.

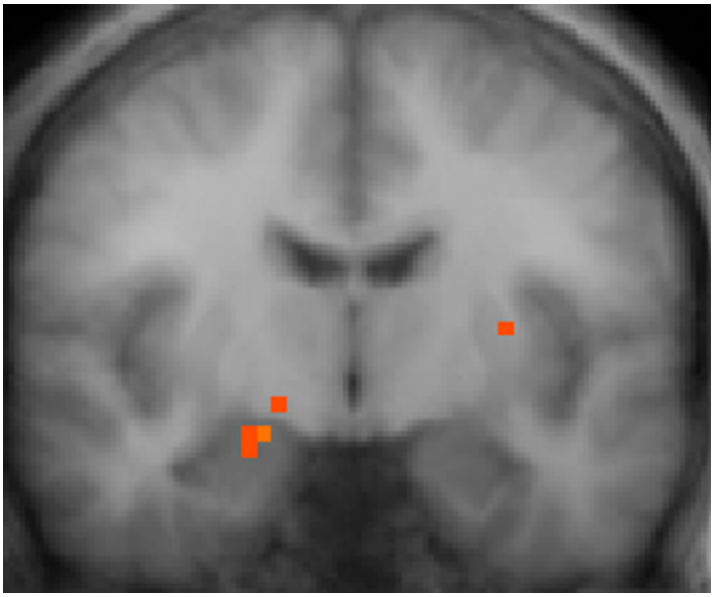


Figure 4.6: Activation pattern when we contrast the parametric predictors (amygdala). We compute a random effect GLM with a statistical threshold $p(uncorrected) < 0.001$ and we observe that our data significantly explains the activity in the amygdala.

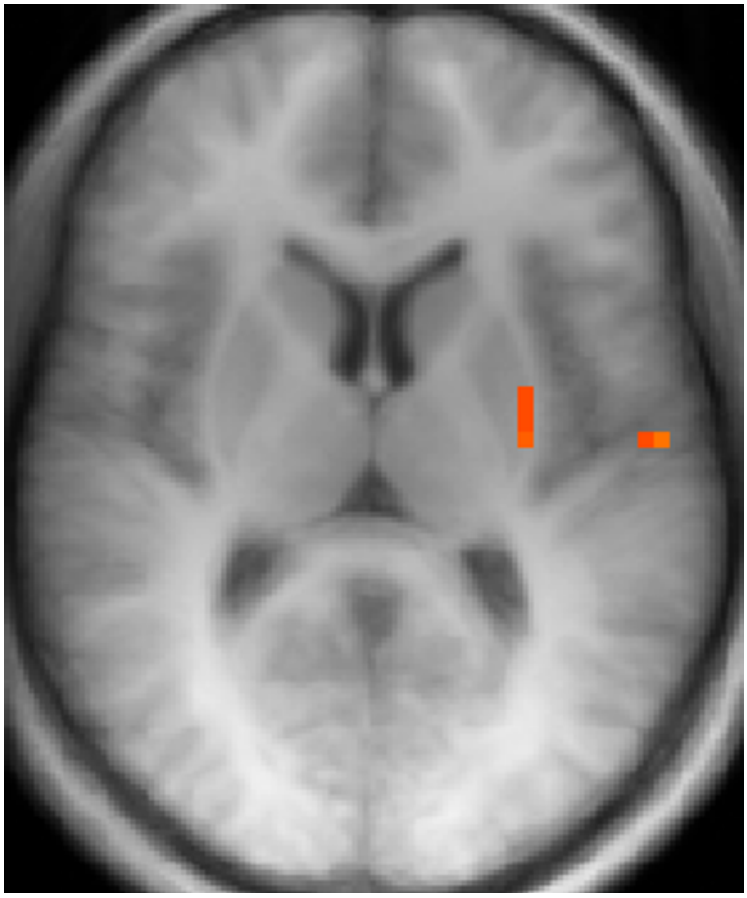


Figure 4.7: Activation pattern when we contrast the parametric predictors (insula). We compute a random effect GLM with a statistical threshold $p(uncorrected) < 0.001$ and we observe that our data significantly explains the activity in the insula.

x	y	z	cluster size	t_{17}	Area
-13	-58	-30	9	4.485	cerebellum
-9	-65	-6	25	4.440	lingual gyrus

Table 4.3: Brain activations in response to the contrast of the block predictors. We compute the contrasts of the two block predictors (insiders minus no insider) in a random effect GLM. Here we report the brain areas that are significantly activated. We use the statistical threshold $p(uncorrected) < 0.001$ and we only report the clusters of more than 5 voxels. In addition, we also indicate the highest t-value of each cluster.

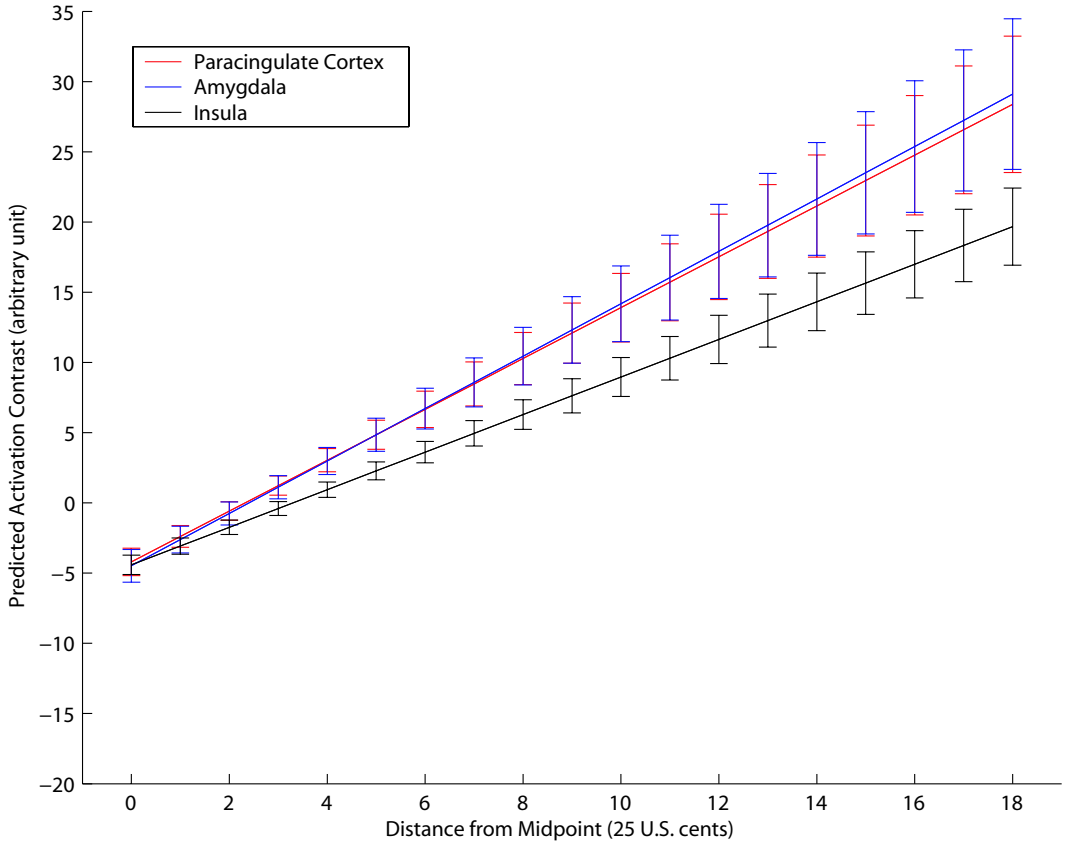


Figure 4.8: Prediction of the activity in the paracingulate cortex, amygdala, and insula. We plot, as a function of the distance to 25¢, the value of the predicted activation along with ± 1 standard error.

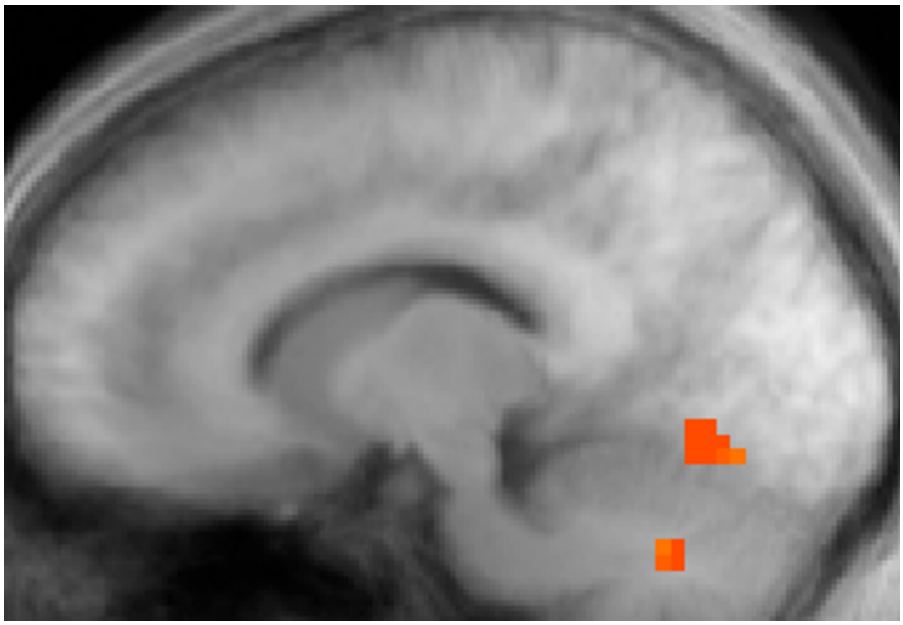


Figure 4.9: Activation pattern when we contrast the block predictors (lingual gyrus). We compute a random effect GLM with a statistical threshold $p(uncorrected) < 0.001$ and we observe that our data significantly explains the activity in the lingual gyrus. We also observe the cerebellum.

We then contrasted the block predictiors (table 4.3, figure 4.9). We observed that the contrast of the predictors was significant in the lingual gyrus (25 voxels around -9; -65; -6) and the cerebellum (9 voxels around -13; -58; -30).

4.3.3 Behavioral experiment

The behavioral experiment confirmed our predictions. We observed a significant correlation ($p = 0.048$) between the subjects' score on the stock market prediction task and the Eye test (figure 4.10). We did not observe any significant correlation ($p > 0.20$) with the mathematical score.

4.4 Discussion

4.4.1 Use of ToM

There is ample evidence that the aptitude to use ToM is crucial to normal social interaction. Our study is different in the sense that we do not investigate the interaction with another human being but the one with an anonymous, electronic financial market.

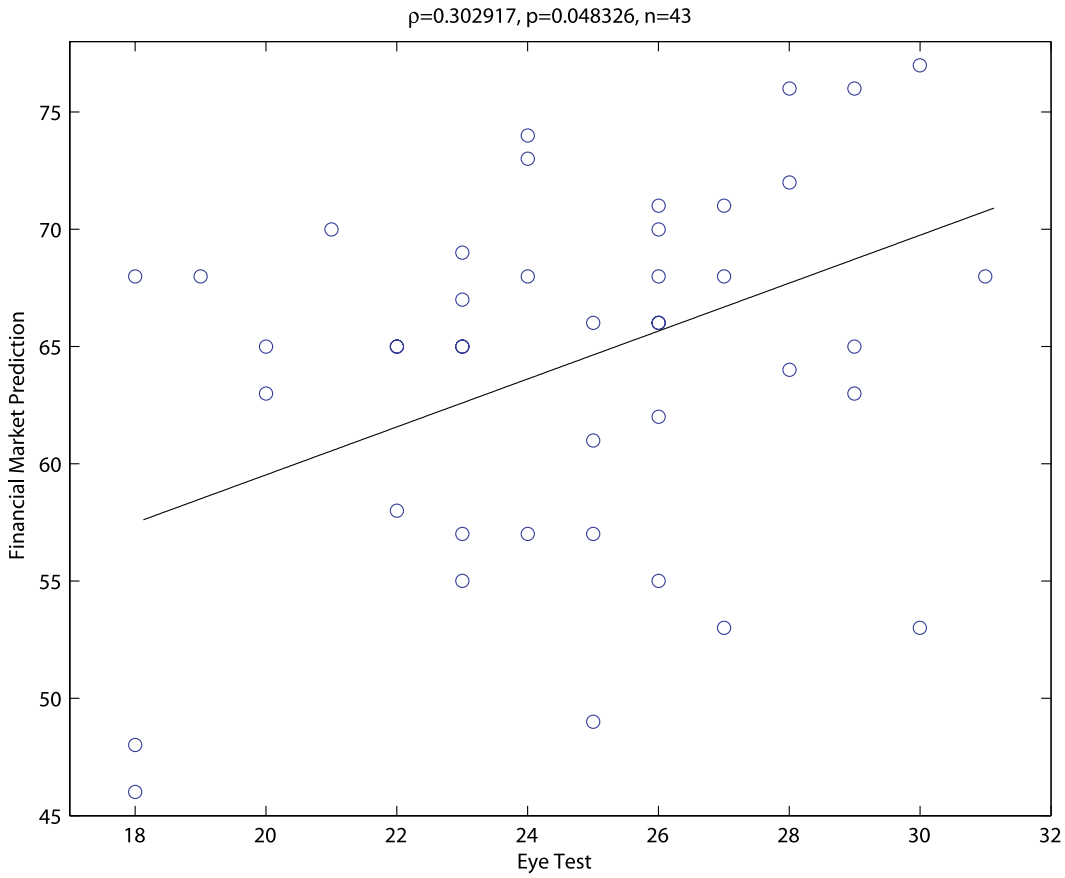


Figure 4.10: Performance on the financial market prediction task as a function of the performance on the eye test. We observe a positive and significant correlation between the abilities to predict the changes in a stock market and the performance on the ToM Eye test.

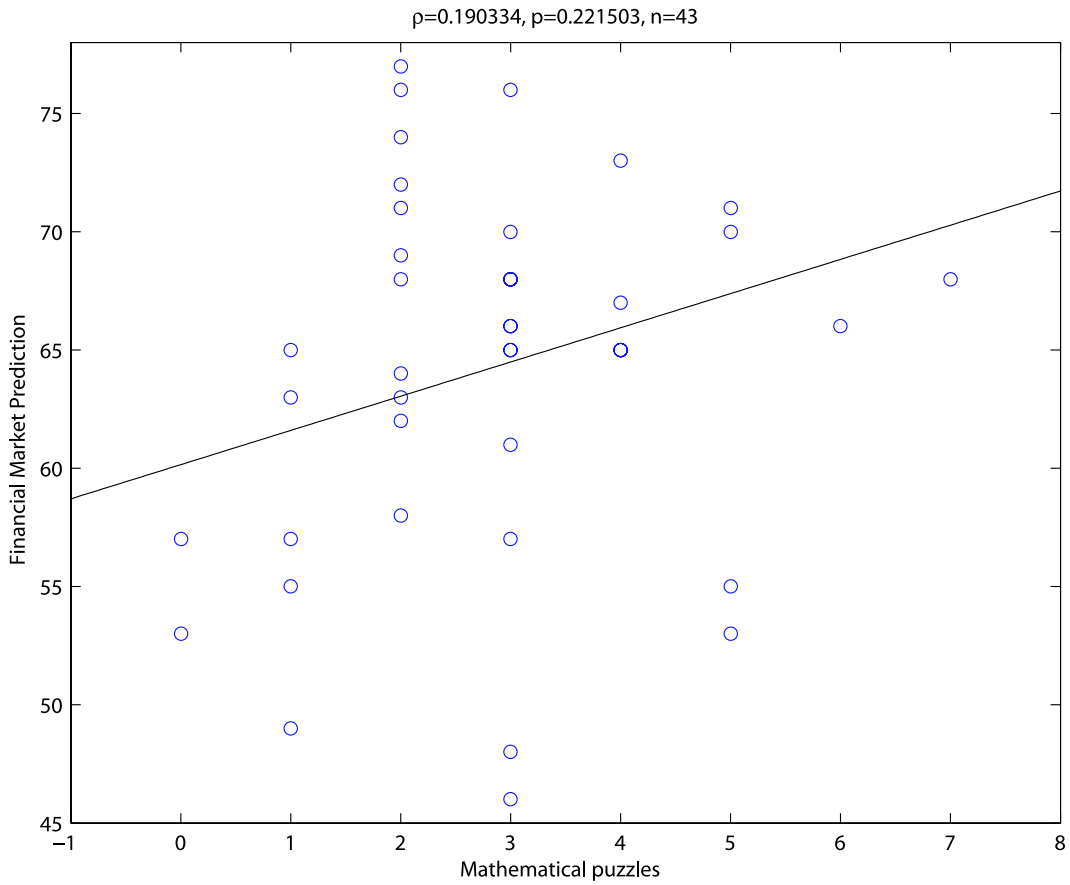


Figure 4.11: Performance on the financial market prediction task as a function of the performance on the mathematical test. We do not observe any significant correlation between the abilities to predict changes in a stock market and the performance on the mathematical test.

By carefully controlling for other parameters, such as ER, risk, conflict, and decision-making, our fMRI experiment reveals that areas previously reported in ToM experiments also activate in our case.

Indeed, the PCC has been reported in tasks that directly involve ToM (Gallagher and Frith [2003], Walter et al. [2004]). Additionally, when humans are playing a strategic game, the activation of the PCC is higher when they think that the opponent is human than when they think that the opponent is a computer (Gallagher et al. [2002], McCabe et al. [2001]). The PCC is also reported in choice vs belief during a strategic game with humans (Bhatt and Camerer [2005], Rilling et al. [2004]). Finally, the PCC is reported in tasks where subjects have to use ToM to describe the movement of geometric shapes as if they were human, a setup that is similar to our display (Castelli et al. [2000]).

We also find an activation in the frontal part of the ACC. Even though it is slightly more dorsal and posterior than other reports (Gallagher et al. [2002], McCabe et al. [2001]), this also suggests a use of ToM.

The activations of the amygdala and insula are also consistent with the use of ToM. These two regions are occasionally reported in ToM tasks (Decety et al. [2004], Baron-Cohen et al. [1999], Critchley et al. [2001]) and learning tasks (Rutishauser et al. [2006], Rolls et al. [2007]). Contrary to the report of the PCC, the activations of the insula and amygdala suggest that the attribution of mental state to a financial market are not purely rational but that there is an emotional compoenent. Indeed, they suggest an empathic response (Singer et al. [2004], Völlm et al. [2006]). Finally, the activation of these two areas corroborates the evidence that market volatility correlates with increased activation of the somatic circuitry (Lo and Repin [2002]).

Moreover, the report of lingual gyrus agrees with our interpretation. A recent study shows that this area is involved in the perception of biological motion (Servos et al. [2002]).

We did not see any activation of areas related to mathematics. We did not see any activation of the frontal and parietal lobes (Newman et al. [2003], Acuna et al. [2002]), nor areas related to the computations of probabilities (Parsons and Osherson [2001]) or arithmetics (Dehaene et al. [1999]). Thus it does not seem that, at least in our experiment, humans used a mathematical approach to understand the changes of prices.

Finally, the behavioral experiments confirm the use of ToM. Despite a careful design and numerous studies confirming the role of the PCC, amygdala, and insula in ToM, a

verification is required because brain areas often perform several duties. We observe that when it comes to predicting price changes in our market, ToM aptitude is a better indicator of success than mathematical abilities.

4.4.2 Implications for finance

Traders disseminate information among themselves without any form of verbal communication and by only using an anonymous, electronic financial platform to interact (Plott and Sunder [1988]). This dissemination of information is a fundamental element of the Efficient Markets Hypothesis (Fama [1970]) because without it, prices will not reflect all the information available.

Finance views financial markets as computers, acting in a pre-programmed way. In the context of the present experiment, finance models the relationship between prices and insider information mechanically, in accordance with the Rational Expectations Theory (Admati [1985], Glosten and Milgrom [1985], Grossman and Stiglitz [1980]). Indeed, the standard theories assume that markets are non-intentional. In this approach, traders only need to find an optimal strategy in the form of a mechanical rule based on a detailed knowledge of the economy ("eduction," Guesnerie [1992]) or based on statistical analysis based on the trading activity (Marcet and Sargent [1989], Easley and O'Hara [1992], Glosten and Milgrom [1985]). However, none of these approaches fully describe human inference and our alternative hypothesis is that traders use ToM.

The use of ToM goes against the traditional approach of finance. Even though financial markets are often anthropomorphized in speech (Oberlechner [2004]), finance rejects this view and sees it as a simple metaphor. Instead, we show here that ToM may be the actual neuronal mechanism used by humans to understand markets. In essence, we show that humans understand markets in a fundamentally different way that current financial theory supposes. Our approach is the first step toward the formalization of "trader's intuition."

Why do humans use ToM instead of a mathematical model? Even though financial markets are very recent evolutionary speaking, they need the same basic skills: interacting with a large group of humans. The standard approach to these problems, a Bayesian analysis, may be energetically demanding for the brain. In effect, when confronted with too complex problems, humans may revert to ToM. Additionally, when the dimensionality of the parameter space of a problem is too high, the Bayesian approach is theoretically impossible

(Diaconis and Freedman [1986]), thus making the ToM approach a more appropriate choice. In general, we conjecture that ToM may be used in other situations where human interaction does not literarily apply but where a mathematical approach would be too demanding.

However, this approach may not always be optimal. It may be that, depending on the situation, the use of ToM is a detrimental cognitive bias. The illusion of control (Langer and Roth [1975]) has been observed in experimental situations (Plott and Sunder [1988], Kagel and Roth [1997]). In our case, while the use of ToM appears to be help traders predict changes in prices, it also suggest that humans treat anonymous, electronic markets as intentional. In this view, they could change and influence the trades, which goes against the rational expectation theory. Indeed, prices cannot be influenced, unlike a human opponent, and traders should only apply a probabilistic approach.

In the light of the present behavioral results however, it remains an open question under what circumstances the use of ToM gives traders an edge. Contrary to the traditional approach, we show that the higher a subjects' ToM abilities, the better their predictions. However, it may be that under other circumstances, the ToM abilities would negatively (or not at all) correlate with the performance on some financial tasks.

Chapter 5

Summary

5.1 Three experiments

We presented three types of tasks of increasing complexity. In the first task (chapter 2), we used previous data to investigate how the two measures of a gamble, expected reward (ER) and risk, are recombined in the brain. In the second task (chapter 3), we used two gambles instead of one. We also asked subjects to perform a choice between the two options. We asked the questions of how the brain indexes the metrics that relate to each gamble. In the third task (chapter 4), we explored a much more complicated task: a financial market. In effect, we went from the discovery that the basic metrics of a gamble are encoded in the brain to more advanced description.

Specifically, in chapter 2, we started with a simple yet powerful discovery: ER and risk are encoded in separate areas of the brain (Preuschoff et al. [2006]). Yet, to evaluate the attractiveness of a gamble accurately, the brain needs to integrate these two dimensions of a gamble into a single one. By using a specific hypothesis on the role of correlation, we created a new technique based on Canonical Correlation Analysis (CCA) and we discovered that the brain adds ER and risk to form a new metric that significantly correlates with the medial prefrontal cortex.

In chapter 3, we introduced a new analysis technique based on Hotelling's T^2 test. This extension of traditional fMRI analysis methods allowed us to understand how several predictors may activate a single voxel. In the context of an experiment with two gambles, we discovered that the cortex uses risk as an index for the utility of the two gambles.

In chapter 4, we investigated how humans perceive financial markets. Contrary to what finance theory predicts, humans do not use a mathematical model to make predictions

but instead use the Theory of Mind (ToM) circuitry. We observed that the paracingulate cortex, a region that had previously been observed in ToM tasks, was also activated in our experiment. According to these results, we correctly predicted that ToM abilities positively correlated with the abilities to predict changes in prices. Mathematical and analytical savvy did not seem to be a significant contributor to investor's success. Hence, we took the first steps into formalizing "trader intuition."

5.2 Implications for economics

It remains an open question how and when the brain switches between the various metrics. We observed a metric that is additive in ER and risk, a metric of utility, and a metric that correlates with ToM. The boundaries of usage of these metrics are unclear at this point. We do not know how the timing and the type of stimuli elicit a brain response that uses a certain metric. Eventually, the switches between the various measures may help explain why humans sometimes behave "irrationally" when they are confronted with economic gambles.

There is more to be asked beyond and between these three tasks. An obvious generalization is to ask about more than two gambles. Neuroeconomics also needs to investigate how amounts of money larger than a few dollars are perceived. For example, neuroeconomics also need to inquire about subjective probabilities (Savage [1972], Anscombe and Aumann [1963]) and state-dependent utility (Arrow [1953], Debreu [1972]). In addition, context, emotional states, and timing all influence a person's perception and decisions when confronted with an economic situation.

Economic theory is particularly well suited for fMRI investigation. Indeed, neuroeconomists already have mathematical measures for their experiments. Instead of using loosely defined notions, they have specific metrics that they can correlate with brain activity.

5.3 Implication for neuroscience

For two of our experiments, we created new analysis methods. With CCA, we based our approach on a hypothesis on the role of correlation (Salinas and Sejnowski [2001]) later observed with direct neuronal recording (Romo et al. [2003]). In this view, the brain directs

integration with correlation. With a CCA, we reversed engineered the biological mechanism. By translating the hypothesis or correlation from the neuronal to the fMRI level, we created a new method that could be used by other researchers. Indeed, the problem of integration of several signals in the brain is general (Santia and Grodzinsky [2007], Sestieri et al. [2006]).

With the adaptation of Hotelling's T^2 statistics to fMRI data, we created a new investigation method. Given the poor spacial resolution of fMRI recordings, future research is bound to be compelled to investigate overlapping activity in a single voxel. In addition, we asked how several percepts are categorized in the brain. Indeed, the human brain is constantly confronted with the percept of several objects and it has to classify them. Using Hotelling's statistics is one approach in understanding such general problems.

Investigation with fMRI has been criticized as being "modern-day phrenology" (Cooter [1985], Lagopoulos [2007]). However, fMRI is just an measurement method and what matters is how we use it. Investigation with fMRI sometimes is restricted to finding the location of "hot spots." We tried instead to answer the question of "how" in addition to "where."

We attempted to explain biological phenomena in chapters 2 and 3. We ground our methods in observation with direct neuronal recording. Our CCA approach is based on a hypothesis by Salinas and Sejnowski [2001] later confirmed by Romo et al. [2003]. Our approach for the Hotelling's T^2 statistics is based on monkey recordings by Padoa-Schioppa and Assad [2006]. Thus, by keeping a strong link with neuronal activity, we tested falsifiable hypotheses with fMRI.

In addition, with the help of fMRI data, we made predictions about human behavior. In chapter 4, after observing ToM areas, we correctly predicted that the more able humans were at ToM, the more able they are at predicting prices in the stock market experiment we designed.

5.4 Conclusion

We tested falsifiable hypotheses that may very well be proven by future work in neuroeconomics. It may well be that the inherent limitations of fMRI will preclude any advanced investigation. It may be that researchers will have to resort to better recording techniques.

Appendix A

Mathematical Proofs

A.1 Simple proof of the partition theorem and proof of the inference method for GLM

A.1.1 Partition theorem

While proofs of the partition theorem already exist (Brownlee [1984], Hald [1952]), we offer a much simpler one here. The purpose of this section is to find the distribution of:

$$Q = \sum_{i=1}^{n} (u_i - \bar{u})^2. \tag{A.1}$$

In the expression above, the u_i are independent normal random variables with mean 0 and variance 1 and $\bar{u}$ is the sample mean defined by:

$$\overline{u} = \frac{1}{n} \sum_{i=1}^{n} u_i. \tag{A.2}$$

By defining the $l_i = u_i - \bar{u}$ and building two vectors L and U such that:

$$L = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix} U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}. \tag{A.3}$$

We can see that $L = AU$ where:

$$A = \begin{bmatrix} \frac{n-1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \dots & -\frac{1}{n} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & \frac{n-1}{n} \end{bmatrix}. \tag{A.4}$$

With all these definitions:

$$Q = \sum l_i^2 = L'L = U'A'AU. \tag{A.5}$$

We want to diagonalize $A'A$. Let us first notice that $A = I - \frac{1}{n}F$ where I is the identity matrix and F is a matrix full of ones (i.e., $\forall i \forall j F_{ij} = 1$). Simple computation gives:

$$A'A = (I - \frac{1}{n}F)'(I - \frac{1}{n}F) = I - \frac{2}{n}F + \frac{1}{n^2}F^2 = I - \frac{2}{n}F + \frac{1}{n^2}nF = I - \frac{1}{n}F. \tag{A.6}$$

Let us call λ_j the eigenvalues of $A'A$ and v_j the corresponding eigenvectors:

$$A'Av_j = \lambda_j v_j. \tag{A.7}$$

The expression above is equivalent to:

$$Fv_j = n(1 - \lambda_j)v_j. \tag{A.8}$$

The rank of F is 1, so there are $n-1$ eigenvectors v_j with eigenvalues $n(1-\lambda_j) = 0$ for $j = 2..n$. For the remaining eigenvalue, using the fact that $\mathrm{Tr}(F) = n = \sum_{j=1}^{n} n(1-\lambda_j) = n(1-\lambda_1)$, we get that $\lambda_1 = 0$. The other eigenvalues of $A'A$ are $\lambda_2 = \lambda_3 = \ldots = \lambda_n = 1$.

Since $A'A$ is symmetric, it can be diagonalized with an orthogonal matrix V containing the orthogonal eigenvectors v_j found above so that:

$$A'A = V'\Sigma V. \tag{A.9}$$

The matrix Σ is diagonal and with values $\{0, 1, \ldots, 1\}$. With all this, we can write Q in

a new way:

$$Q = U'V'\Sigma VU. \tag{A.10}$$

But since V is an orthogonal matrix and U is a vector containing independent normal variables with unit variance, the vector $W = VU$ also contains independent normal variables with unit variance and:

$$Q = W'\Sigma W = \sum_{j=1}^{n} \lambda_j w_j^2 = \sum_{j=2}^{n} w_j^2. \tag{A.11}$$

Thus, Q is χ^2-distributed with $n-1$ degrees of freedom.

A.1.2 Inference in linear regression

The maximum likelihood estimator is (Johnson and Wichern [2002], Weisberg [2005]):

$$\widehat{\beta} = \left(X'X\right)^{-1} X'y. \tag{A.12}$$

By using the fact that $y = X\beta + \epsilon$, where ϵ is a collection of i.i.d. Gaussian random variables with variance σ^2, it follows that $\widehat{\beta}$ is normally distributed with mean β and covariance matrix:

$$\Sigma = \sigma^2 \left(X'X\right)^{-1}. \tag{A.13}$$

The residual sum of squares is computed with:

$$RSS = \widehat{\epsilon}'\widehat{\epsilon} = y' \left(I - X \left(X'X\right)^{-1} X'\right) y. \tag{A.14}$$

We define the error as:

$$\widehat{\epsilon} = y - X\widehat{\beta}. \tag{A.15}$$

The variance of the error is estimated with:

$$\widehat{\sigma}^2 = \frac{RSS}{r}. \tag{A.16}$$

In the above expression, r is the rank of $\left(I - X\left(X'X\right)^{-1} X'\right)$ by similar reasoning as in section A.1.1. By observing that matrix corresponds to a projection, we have $r = n - p$.

The matrix has p eigenvalues equal to zero, and $n - p$ eigenvalues equal to one.

The variable

$$t_i = \frac{\widehat{\beta}_i}{\sqrt{\widehat{\sigma}^2 \left(X'X\right)^{-1}_{ii}}}. \tag{A.17}$$

The variable t_i follows a Student's t-distribution with r degrees of freedom.

A.2 Link between neuronal and synaptic activities and fMRI data

Let $z^{(1)}$ and $z^{(2)}$ be two signals at the neuronal level. fMRI data does not reflect $z^{(1)}$ and $z^{(2)}$ but a version smeared by the HRF. Let h_n be the impulse response that is convolved with the signals at the neuronal level to obtain the fMRI signals:

$$y_n^{(j)} = \sum_i z_i^{(j)} h_{n-i}. \tag{A.18}$$

For simplicity, let $z^{(1)}$ and $z^{(2)}$ have zero mean. Our study is interested in the correlation of these two signals, ρ_{neuron}, defined by:

$$\rho_{neuron}(m,n) = \frac{\mathbb{E}\left[z_n^{(1)} z_m^{(2)}\right]}{\sqrt{\mathbb{E}\left[z_n^{(1)} z_n^{(1)}\right]}\sqrt{\mathbb{E}\left[z_m^{(2)} z_m^{(2)}\right]}}. \tag{A.19}$$

However, we do not have access to this value, because we cannot record the signals at the neuronal level. Is it possible to recover this value with fMRI data? We can compute the fMRI equivalent of equation A.19 for fMRI data:

$$\rho_{fMRI}(m,n) = \frac{\mathbb{E}\left[y_n^{(1)} y_m^{(2)}\right]}{\sqrt{\mathbb{E}\left[y_n^{(1)} y_n^{(1)}\right]}\sqrt{\mathbb{E}\left[y_m^{(2)} y_m^{(2)}\right]}}. \tag{A.20}$$

We will prove that under certain conditions, equation A.18 implies $\rho_{neuron}(m,n) = \rho_{fMRI}(m,n)$ for $m = n$.

For simplicity, let us suppose that, at the neuronal level, samples from two areas are

correlated only if they were recorded at the same time[1]. Mathematically:

$$\rho_{neuron}(m,n) = \rho_0 \delta_{n-m}. \tag{A.21}$$

We defined δ as a discrete-time Dirac function.

$$\rho_{fMRI}(m,n) = \frac{\mathbb{E}\left[\left(\sum_i z_i^{(1)} h_{n-i}\right)\left(\sum_j z_j^{(2)} h_{m-j}\right)\right]}{\sqrt{\mathbb{E}\left[\left(\sum_i z_i^{(1)} h_{n-i}\right)^2\right]}\sqrt{\mathbb{E}\left[\left(\sum_j z_j^{(2)} h_{m-j}\right)^2\right]}}. \tag{A.22}$$

Let us simplify this expression in the case when $m = n$. We will also use the fact that $\mathbb{E}\left[z_i^{(1)} z_j^{(2)}\right] = 0$ if $i \neq j$ by similar assumption as equation A.21. We have:

$$\mathbb{E}\left[\left(\sum_i z_i^{(1)} h_{n-i}\right)\left(\sum_j z_j^{(2)} h_{n-j}\right)\right] = \sum_i \sum_j h_{n-i} h_{n-j} \mathbb{E}\left[z_i^{(1)} z_j^{(2)}\right] = \mathbb{E}\left[z_n^{(1)} z_n^{(2)}\right] \sum_k h_k^2. \tag{A.23}$$

In a similar way:

$$\mathbb{E}\left[\left(\sum_i z_i^{(1)} h_{n-i}\right)^2\right] = \sum_i \sum_j h_{n-i} h_{n-j} \mathbb{E}\left[z_i^{(1)} z_j^{(1)}\right] = \mathbb{E}\left[z_n^{(1)} z_n^{(1)}\right] \sum_k h_k^2 \tag{A.24}$$

$$\mathbb{E}\left[\left(\sum_j z_j^{(2)} h_{n-j}\right)^2\right] = \sum_i \sum_j h_{n-i} h_{n-j} \mathbb{E}\left[z_i^{(2)} z_j^{(2)}\right] = \mathbb{E}\left[z_n^{(2)} z_n^{(2)}\right] \sum_k h_k^2. \tag{A.25}$$

We then have:

$$\rho_{fMRI}(n,n) = \frac{\mathbb{E}\left[z_n^{(1)} z_n^{(2)}\right] \sum_k h_k^2}{\sqrt{\mathbb{E}\left[z_n^{(1)} z_n^{(1)}\right] \sum_k h_k^2}\sqrt{\mathbb{E}\left[z_n^{(2)} z_n^{(2)}\right] \sum_k h_k^2}} = \rho_0 = \rho_{neuro}(n,n). \tag{A.26}$$

[1]This hypothesis is not as restrictive as it seems. Indeed, one could model the correlation at the neuronal level as a Dirac function convolved by a fixed arbitrary function (see, for example, Salinas and Sejnowski [2002], Shadlen and Movshon [1999]) that is the same for both areas, g_n. The proof would be the identical in every respect, except that we would replace the HRF function, h_n, by its convolution with g_n, $h_n \star g_n$.

A.3 Canonical correlation analysis and inference

Though we could compute overall p-values for each row of the CCA with Wilk's lambda method (Anderson [2003], Johnson and Wichern [2002]), the present investigation required p-values on the individual loadings. Resampling methods, such as the bootstrap (Efron and Tibshirani [1993]), were impractical because of the sign ambiguity when computing eigenvectors. Instead, an equivalence between CCA and multiple linear regressions provides the basis for computation of approximate p-values on the loadings[2].

Let X be the (T, n) matrix corresponding to the n predictors and Y be a (T, p) matrix corresponding to the fMRI signals recorded in p regions of interest. Time indexes each row of the matrix (T time samples). Let Σ_{11} and Σ_{22} be the sample covariance matrices of X and Y, respectively, and let $\Sigma_{12} = \Sigma_{21}'$ be the sample covariance matrix between X and Y. For simplicity, we suppose that both predictors and fMRI signals have zero mean.

CCA finds the linear combinations of the column of X and Y that has the largest correlation; i.e., it finds the weight vectors (loadings) a and b that maximize:

$$\rho = \frac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}b}}. \tag{A.27}$$

We follow the derivations of Johnson and Wichern [2002] and we do a change of basis: $c = \Sigma_{11}^{1/2}a$ and $d = \Sigma_{22}^{1/2}b$.

$$\rho = \frac{c'\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}d}{\sqrt{c'c}\sqrt{d'd}} \tag{A.28}$$

By the Cauchy-Schwartz inequality (Strang [2003]):

$$\rho \leq \frac{\sqrt{c'\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}c}\sqrt{d'd}}{\sqrt{c'c}\sqrt{d'd}} = \sqrt{\frac{c'\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}c}{c'c}}. \tag{A.29}$$

The inequality above is an equality when $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}c$ and d are collinear. The right hand side of the expression above is a Rayleigh quotient and it is maximum when c is the eigenvector corresponding to the largest eingenvalue of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$ (we obtain the other rows by using the other eigenvalues in decreasing magnitude). This results if the

[2]It seems that Stata (`http://www.stata.com`) uses this method. However, Stata Corp. does not provide any detail on how p-values are computed and we did not find a suitable reference in the literature.

basis of the CCA. We can compute the two canonical variables: $U_1 = Xa$ and $V_1 = Yb$.

Now, if U_1 is known, then b can be found by a multiple linear regression onto Y. Then, we can apply the methods to compute the p-values for multiple linear regressions (Weisberg [2005] or section A.1.2) and obtain p-values for the CCA. Thus, what remains to be proven is that the coefficients of the multiple linear regression $\widetilde{b} = (Y'Y)^{-1}(Y'U_1)$ are equal to b and then use the inference of the multiple linear regression to obtain p-values for b.

$$\widetilde{b} = \left(Y'Y\right)^{-1}\left(Y'U_1\right) \tag{A.30}$$

$$\widetilde{b} = \left(Y'Y\right)^{-1}\left(Y'X\right)a \tag{A.31}$$

We substitute the definitions of the sample covariance matrices and get:

$$\widetilde{b} = \Sigma_{22}^{-1}\Sigma_{21}a \tag{A.32}$$

$$\widetilde{b} = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}c \tag{A.33}$$

$$\widetilde{b} = \Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}c. \tag{A.34}$$

Since the CCA has found that $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}c$ and d are collinear, we get:

$$\widetilde{b} \propto \Sigma_{22}^{-1/2}d \tag{A.35}$$

$$\widetilde{b} \propto b. \tag{A.36}$$

We do not have equality of $\widetilde{b}$ and b. Fortunately, the inference methods for multiple linear regression are insensitive to scale and will give the same results for both vectors. Indeed, $\widetilde{b}$ and b are perfectly correlated and inference methods need only inspect the correlations between the dependent and independent variables. We can therefore apply the inference methods of linear regressions (section A.1.2).

By symmetry, this method also applies to a. Similarly, these results are easily extended

for the other rows of the CCA.

Since the inference methods for multiple linear regressions suppose that, when testing b, the matrix of predictors X is non-random, the p-values we obtain are only approximations.

A.4 Null hypothesis of zero correlation

We follow the same notation as section A.3 and, for simplicity, we restrict ourselves to the case of $n = 2$ predictors and $p = 2$ fMRI time courses. Let us make explicit the correlation matrices Σ_{11}, Σ_{22}, and Σ_{12}.

Since we designed the predictors of our experiment, ER and risk, to be orthogonal, we have:

$$\Sigma_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{A.37}$$

Similarly, the signals in the brain have a similar correlation matrix with the difference that they may or may not be correlated:

$$\Sigma_{22} = \begin{bmatrix} 1 & \sigma \\ \sigma & 1 \end{bmatrix}. \tag{A.38}$$

The recording noise is supposed independent between voxels and, for simplicity, we normalize all the predictors and signals. We would like to investigate the presence or absence of σ, the correlation of the brain signals. The null hypothesis is that $\sigma = 0$, no correlation exists. Let us not make this simplification for now. Simple computations show:

$$\Sigma_{21} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{A.39}$$

Following the notations of appendix A.3, c is the eigenvector corresponding to the largest eigenvalue of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$. In the case of our experiment, this simplifies to a being the eigenvector of:

$$\frac{1}{1-\sigma^2}\begin{bmatrix} 1 & -\sigma \\ -\sigma & 1 \end{bmatrix}. \tag{A.40}$$

If $\sigma > 0$, the largest eigenvalue is $\frac{1+\sigma}{1-\sigma^2}$, corresponding to an eigenvector:

$$a = \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{A.41}$$

Indeed, if the two signals have positively correlated noise, then subtracting one from the other will cancel out some of the noise. Similarly, if $\sigma < 0$, then the brain should add the two signals and we have the largest eigenvalue $\frac{1-\sigma}{1-\sigma^2}$ with a corresponding eigenvector:

$$a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{A.42}$$

Now, let us return to the null hypothesis; we suppose that no correlation exists ($\sigma = 0$). The recording noise will nevertheless tilt the eigenvector either toward an addition or a subtraction and, by symmetry, each of these events occurs with probability $p = 0.5$. If no correlation exists, we expect the signs to be the same for roughly half the subjects and opposite for the other half.

This provides the basis for hypothesis testing. If we had 100 subjects, and we observed identical signs for 99 of them, then anybody would agree that it strongly suggests that $\sigma < 0$. But where is the threshold? We can use a binomial test and obtain a p-value for any given number of subjects. We will obtain a low p-value (i.e., strongly reject the null hypothesis that $\sigma = 0$) when correlation is likely to be present.

A.5 Statistical power considerations when using a CCA

With a CCA, we computed a new composite predictor that is a linear combination of the ER and risk and we then used this new predictor in a GLM. But a GLM is itself a linear technique that can recreate the composite predictor from expected reward and risk. The use of CCA, however, has two advantages. First, it has greater statistical power. Second, by using a GLM we would not have been able to tell whether the activations were overlapping or recombined. In this section, we will also expose why we need an additional predictor and how to choose it.

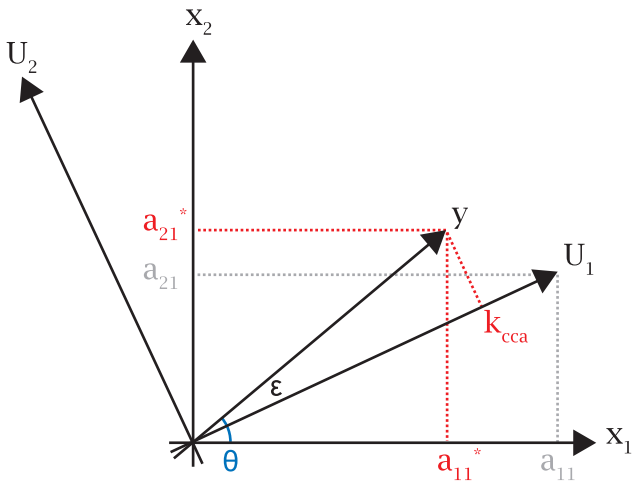


Figure A.1: Illustration of why the CCA has higher power. The coefficient of the projection of y onto U_1, k_{CCA}, is larger than either the coefficient of the projection of y onto x_1, a_{11}^*, or onto x_2, a_{21}^*. The angle between y and U_1 is ϵ and the angle between y and x_1 is θ. We present a general case, but x_1 can be thought of as the ER and x_2 as the risk.

A.5.1 Why CCA has a higher statistical power

We graphically displayed why CCA has higher statistical power on figure A.1. With the CCA step, we found a new composite predictor, U_1 and we propose to project the brain activity, y, onto U_1. If there had been no noise during the recording, we would have found a predictor that is perfectly collinear with the brain activity, i.e., $\epsilon = 0$. Even though there is noise in the recording, ϵ should remain small.

The higher the projection coefficients, the higher the statistical power. Indeed, noise is independent from the experiment and it has the same power in every direction (for simplicity, we normalized the norm of x_1, x_2, and U_1 to 1). Thus, the lower the projection coefficient, the more effect the noise has. The best-case scenario in terms of statistical power for a pure GLM method is when the respective projection coefficients for the ER and the risk, a_{11}^* and a_{21}^*, are equal, i.e., $\theta = \frac{\pi}{4}$. Thus, $a_{11}^* = a_{21}^* = \frac{1}{\sqrt{(2)}}$. But in this case, $k_{CCA} = \cos\epsilon \simeq 1$. Thus, the t-statistics that we use to test the significance of each predictor is $\sqrt{2}$ lower for the pure GLM method, in the best-case scenario. This corresponds to doubling the number of subjects.

The CCA method has a second advantage. If we had not used a CCA and had increased statistical power by using more subjects, we would have observed two overlapping activations and we would not have been able to determine whether the neurons in the area

actually combined the two signals.

A.5.2 Second predictor

In our new setting, we have a new composite predictor, U_1, but we would like to add a second one for two reasons. First, the original study used two predictors, ER and risk, and we would like to compare the two methods fairly by having an equal number of predictors. Second, we would like to have a sanity check. If our method is correct, only one predictor, U_1, is enough to explain the activity in one area. Any additional predictor that we add should not significantly explain the brain activity. If we see a significant activation to a second predictor, then it will cast a doubt on our approach.

How do we choose this second predictor? Instrumental variable estimation (Spanos [1986]) suggests that we use a variable that is correlated with U_1, such as x_1 or x_2. However, we will show that choosing a variable that is orthogonal has better statistical power. U_2 is a perfect candidate, as it is, by construction, orthogonal to U_1.

Let us first investigate the case when we choose U_2 as an additional predictor.

$$U_1 = a_{11}x_1 + a_{21}x_2 \tag{A.43}$$

$$U_2 = a_{21}x_1 - a_{11}x_2 \tag{A.44}$$

$$U = [U_1|U_2] = [a_{11}x_1 + a_{21}x_2|a_{21}x_1 - a_{11}x_2]. \tag{A.45}$$

We do a linear regression with U as a design matrix and estimate k_{CCA}:

$$\left(U'U\right)\left(U'y\right) = \begin{bmatrix} a_{11}a_{11}^* + a_{21}a_{21}^* \\ a_{11}a_{21}^* - a_{21}a_{11}^* \end{bmatrix}. \tag{A.46}$$

By using polar coordinates and a Taylor exponent, we get:

$$k_{CCA} = a_{11}a_{11}^* + a_{21}a_{21}^* = \cos\left(\theta + \epsilon\right)\cos\left(\theta\right) + \sin\left(\theta + \epsilon\right)\sin\left(\theta\right) \tag{A.47}$$

$$k_{CCA} = \cos\left(\epsilon\right) = 1 - \frac{1}{2}\epsilon^2 + O\left(\epsilon^3\right). \tag{A.48}$$

Let us first investigate the case when we choose x_2 as an additional predictor. We have:

$$U = [U_1|x_2] = [a_{11}x_1 + a_{21}x_2|x_2]. \tag{A.49}$$

Similar computations give:

$$k_{CCA} = \frac{a_{11}^*}{a_{11}} = \frac{\cos(\theta)}{\cos(\theta + \epsilon)} \tag{A.50}$$

$$k_{CCA} = 1 + \frac{\sin(\theta)}{\cos(\theta)}\epsilon + O\left(\epsilon^2\right). \tag{A.51}$$

Let us compare equations A.48 and A.51. Choosing U_2 as an additional predictor is superior because the error when estimating k_{CCA} has higher order. Additionally, when $\theta \simeq \frac{\pi}{2}$, the estimate of k_{CCA} becomes numerically unstable. Thus, we should choose U_2 as an additional predictor.

A.6 Random-effects analysis of balanced designs experiments

Random-effects analysis is now the standard method of analyzing fMRI data. The analysis is performed in two steps (chapter 12 of Penny et al. [2003]). First, a linear regression is performed on a subject-by-subject basis. For each subject, we obtain a scalar that follows a Gaussian distribution. In the second step, we perform a t-test that yields a p-value.

The procedure above, known as *summary statistics*, relies on a strong assumption. Each of the matrixes of predictor is the same for every subject. If this assumption is violated, one has to turn to parametric empirical Bayes (PEB) method, a much more complex procedure. Some proof of the validity of summary statistics are available but are somewhat incomplete (chapter 12 of Penny et al. [2003]) or overly complex (Carlin and Louis [2000], Hsiao [2003]) and we propose a proof that is more restricted than the general ones but still adapted to fMRI analysis.

We model the fMRI data as:

$$y_{i,t} = X_i(\mu + z_i) + \epsilon_{i,t}. \tag{A.52}$$

We index subjects with i and time with t. The variable y_{it} is the amplitude recorded

with fMRI data, X_i is the design matrix, μ is the average activity that we wish to estimate, and z_i and $\epsilon_{i,t}$ are noise terms. By hypothesis, the design matrix is the same for every subject and we write $X_i = X_0$. We model z_i as a zero-mean Gaussian random variable with variance $\frac{1}{\alpha}$ and $\epsilon_{i,t}$ as a zero-mean Gaussian random variable with variance $\frac{1}{\beta}$. Samples are independent.

The maximum likelihood estimator of μ is (Weisberg [2005]):

$$\widehat{\mu} = \left(X'C^{-1}X\right)^{-1}\left(X'C^{-1}y\right). \tag{A.53}$$

We use the definitions:

$$y = \begin{bmatrix} y_{1,1} \\ \vdots \\ y_{1,T} \\ y_{2,1} \\ \vdots \\ y_{2,T} \\ \vdots \\ y_{n,T} \end{bmatrix}, X = \begin{bmatrix} X_0 & & & \\ & X_0 & & \\ & & \ddots & \\ & & & X_0 \end{bmatrix}. \tag{A.54}$$

We denote the covariance matrix of X by:

$$C = \frac{1}{\alpha}XX' + \frac{1}{\beta}I. \tag{A.55}$$

Let us simplify the equation A.53. We first do a singular value decomposition (Johnson and Wichern [2002]) of X: $X = U\Sigma V'$. We get:

$$C = \frac{1}{\alpha}U\Sigma\Sigma'U' + \frac{1}{\beta}UU' \tag{A.56}$$

$$C = U\left(\frac{1}{\alpha}\Sigma\Sigma' + \frac{1}{\beta}I\right)U' \tag{A.57}$$

$$C^{-1} = U\left(\frac{1}{\alpha}\Sigma\Sigma' + \frac{1}{\beta}I\right)^{-1}U' \tag{A.58}$$

$$X'C^{-1}X = V\Sigma' \left(\frac{1}{\alpha}\Sigma\Sigma' + \frac{1}{\beta}I\right)^{-1} \Sigma V'. \tag{A.59}$$

By using a short-hand notation for diagonal matrices where the brackets signify the elements of the diagonal indexed by j and denoting by σ_j the diagonal elements of Σ, we have[3]:

$$X'C^{-1}X = V \left\{ \frac{\sigma_j^2}{\frac{1}{\alpha}\sigma_j^2 + \frac{1}{\beta}} \right\} V' \tag{A.60}$$

$$\left(X'C^{-1}X\right)^{-1} = V \left\{ \frac{\frac{1}{\alpha}\sigma_j^2 + \frac{1}{\beta}}{\sigma_j^2} \right\} V'. \tag{A.61}$$

By combining equations A.58 and the SVD decomposition of X, we also get

$$X'C^{-1} = V\Sigma' \left(\frac{1}{\alpha}\Sigma\Sigma' + \frac{1}{\beta}I\right)^{-1} U. \tag{A.62}$$

With the short-hand notation, we have:

$$X'C^{-1} = V\Sigma' \left\{ \frac{1}{\frac{1}{\alpha}\sigma_j^2 + \frac{1}{\beta}} \right\} U. \tag{A.63}$$

By combining equations A.61 and A.63, we get:

$$\left(X'C^{-1}X\right)^{-1} \left(X'C^{-1}\right) = V \left\{ \frac{1}{\sigma_j^2} \right\} \Sigma' U. \tag{A.64}$$

The equation A.64 does not depend on α or β. Even though we do not know the true value for these two noise parameters, their knowledge would not improve the quality of our estimation. Indeed, we can choose $\frac{1}{\alpha} = 0$ and $\frac{1}{\beta} = 1$ and we simplify equation A.53 to:

$$\widehat{\mu} = \left(X'X\right)^{-1} \left(X'y\right) \tag{A.65}$$

With the help of this result, we can now use the special form of equation A.54 and we get:

$$\widehat{\mu} = \left(\{X_0'X_0\}\right)^{-1} \left(\{X_0'y_i\}\right) \tag{A.66}$$

[3]We never have a division by zero problem thorough this proof. The matrix X is designed to have full-rank and we assume that $\beta > 0$

$$\widehat{\mu} = \left\{ \left(X_0' X_0 \right)^{-1} \right\} \left\{ X_0' y_i \right\} \tag{A.67}$$

$$\widehat{\mu} = \left\{ \left(X_0' X_0 \right)^{-1} \left(X_0' y_i \right) \right\}. \tag{A.68}$$

We showed that our new estimate is separable and that we can estimate a separate $\widehat{\mu}$ for each subject and then compute an average.

A.7 Hotelling's T-squared test

In this section, we will present, without proof, the mathematical basis for Hotelling's T^2 test (Hotelling [1931], Anderson [2003]). This test is an generalization of Student's t-test to multiple variables.

Let X be a (n, p) matrix containing n samples of p Gaussian variables. These variables may be correlated (with unknown correlation matrix) and may have non-zero (unknown) mean.

We denote x_{ij} the elements of X and we define the $(1, p)$ vector μ as the sample mean:

$$(\forall j = 1..p) \left(\mu_j = \frac{1}{n} \sum_i x_{ij} \right). \tag{A.69}$$

We define the (p, p) matrix S as the sample covariance:

$$S = \left(X - \mathbb{1}_{(n,1)} \mu \right)' \left(X - \mathbb{1}_{(n,1)} \mu \right). \tag{A.70}$$

In the expression above, $\mathbb{1}_{(n,1)}$ is a $(n, 1)$ vector containing ones.

We define the scalar random variable T^2 as:

$$T^2 = n \mu S^{-1} \mu'. \tag{A.71}$$

Hotelling proved that $F = \frac{n-p}{p(n-1)} T^2$ follows an $F_{p,n-p}$ with p and $n - p$ degrees of freedom. We can use this result to test several Gaussian variables at the same time.

Appendix B

Additional results

B.1 How to recombine risk and expected reward

The CCA gave us a new predictor that was positive in expected reward and risk: $U_1 = 32\,\mathrm{ER} + 65\,\mathrm{Risk}$. Instead of directly correlating this predictor with brain activity, we decided to compute a new predictor. A review of the role of the anterior cingulate cortex (ACC) reveals that this area may respond to conflict (Botvinick et al. [2004]). In this view, the ACC would only respond to changes and not absolute values.

In our experiment, the baseline is when there is a probability of winning equal to 0.5. Indeed, before the first card is shown, the subjects have a probability of winning of 0.5. Thus, at that time, our combined measure should be:

$$U_1(p = 0.5) = 32\,\mathrm{ER}(p = 0.5) + 65\,\mathrm{Risk}(p = 0.5). \tag{B.1}$$

In accordance with the role of the ACC, we predicted that this structure would respond to:

$$\left|U_1(p) - U_1(p = 0.5)\right|. \tag{B.2}$$

We ran a random-effect GLM with a single set of weights and observed that in accordance with our prediction the ACC was activated (figure B.1).

How does the ACC create such a signal? Given the results of the CCA, it is unlikely that the ACC directly combines the signals from the putamen, insula, and ventral striatum. Indeed, that would imply that the correlation changes according to the probability of winning. Instead, we conjecture that the mPFC directly sends its signal to the ACC and that the ACC computes the absolute value.

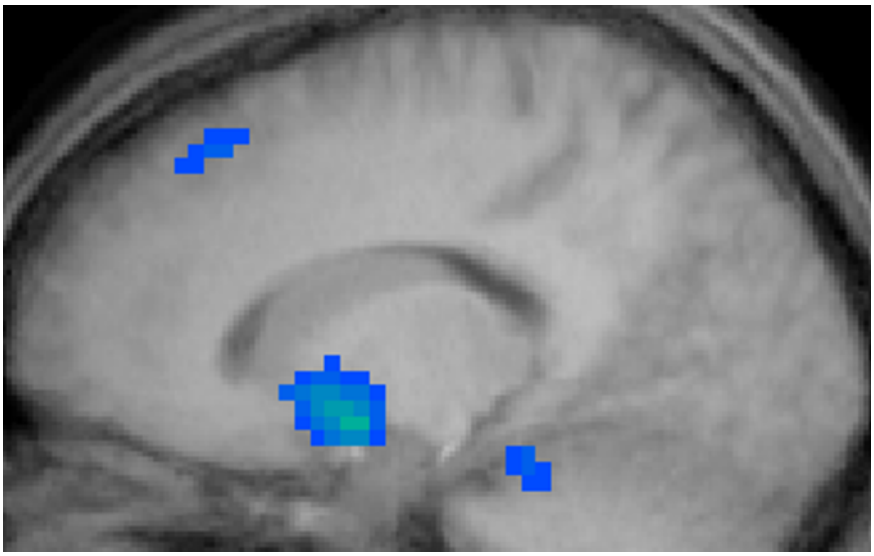


Figure B.1: Activation to the absolute value metric when we use a single set of weights. The area extends for 24 voxels around the center (-11; 30;47) in Talairach coordinates. We use the statistical threshold $p(uncorrected) < 0.001$.

B.2 How utility is indexed in the brain

While we do not observe any activation in the frontal areas when we index by position, we observe an activation in the supplemental motor areas (figure B.2). The activation of this area is in accordance with its role in motor planning (Tanaka et al. [2001]) and is located in the left hemisphere because of a majority of right-handed subjects.

B.3 Exploring trader intuition

In addition to the three tests we administered, we also created a test based on movies by Heider and Simmel [1944] (figure B.3). These movies show geometric shapes moving on a screen. To healthy subjects, these shapes appear to be moving as if they were animals or young children playing during recess.

Without telling subjects about this possible anthropomorphization, we replayed the movie via a computer interface. We paused the replay every 5 s. For half of the pauses, we asked the subjects to predict whether two shapes on the screen would move closer, stay at the same distance, or move farther from each other. For the other half of the pauses, we reminded subjects of their predictions and told them whether their prediction was correct.

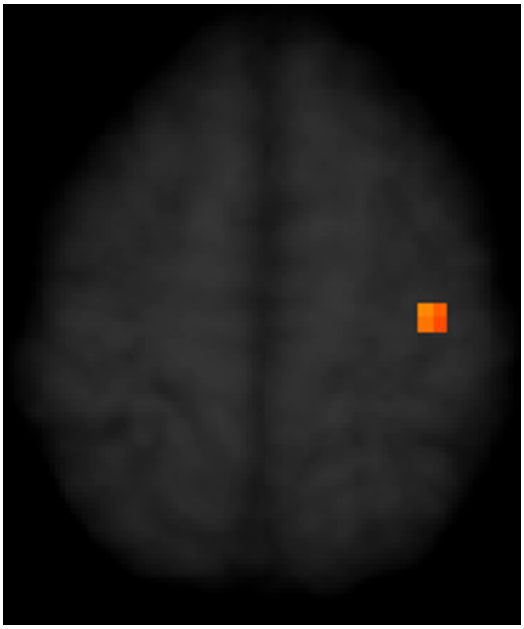


Figure B.2: Activation of the Supplemental Motor Area (SMA) when using position as an indexing method. We show here the brain activation when we use position as an indexing method. A cluster of voxels around the Talairach coordinates (-33;-19;60) appears in the SMA. We used random effects analysis, corrected for serial correlation in fMRI data with ARMA models, and used the threshold $p(uncorrected) < 0.001$ (equivalent to $F_{3,14} > 9.70$).

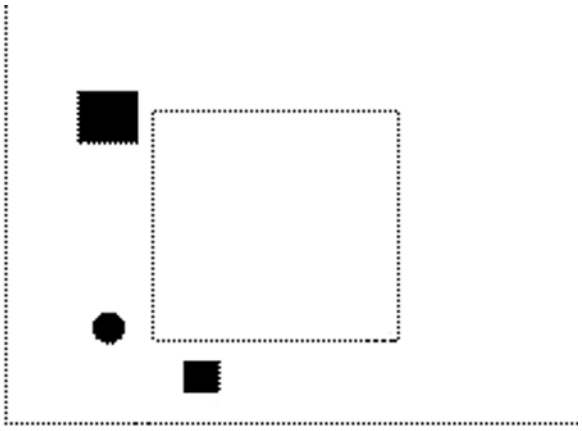


Figure B.3: Sample frame of the Heider test. We display three geometric shapes (two squares and one circle) that move around the screen. We ask subjects to predict whether the two squares will move closer, farther, or stay at the same distance. We reward correct answers, do not punish worng answers, but we penalize indecision. Thus, it is always better to guess.

We rewarded correct predictions, did not punish wrong ones, and punished lacks of answer (subjects had 5 s to make a prediction).

While it is a non-standard ToM test, we nevertheless observed a significant positive correlation between the score on the Heider test and the prediction of the stock market (figure B.4). We did not observe any significant correlation between the Heider test and the Eye test (figure B.5). This lack of correlation suggests that the two tests, while both testing ToM, these tests record different facets of ToM abilities.

Finally, we also asked subjects to answer a questionnaire at the end of the behavioral experiment (table B.1). We see no obvious use of ToM verbalization, suggesting that anthropomorphization is unconscious.

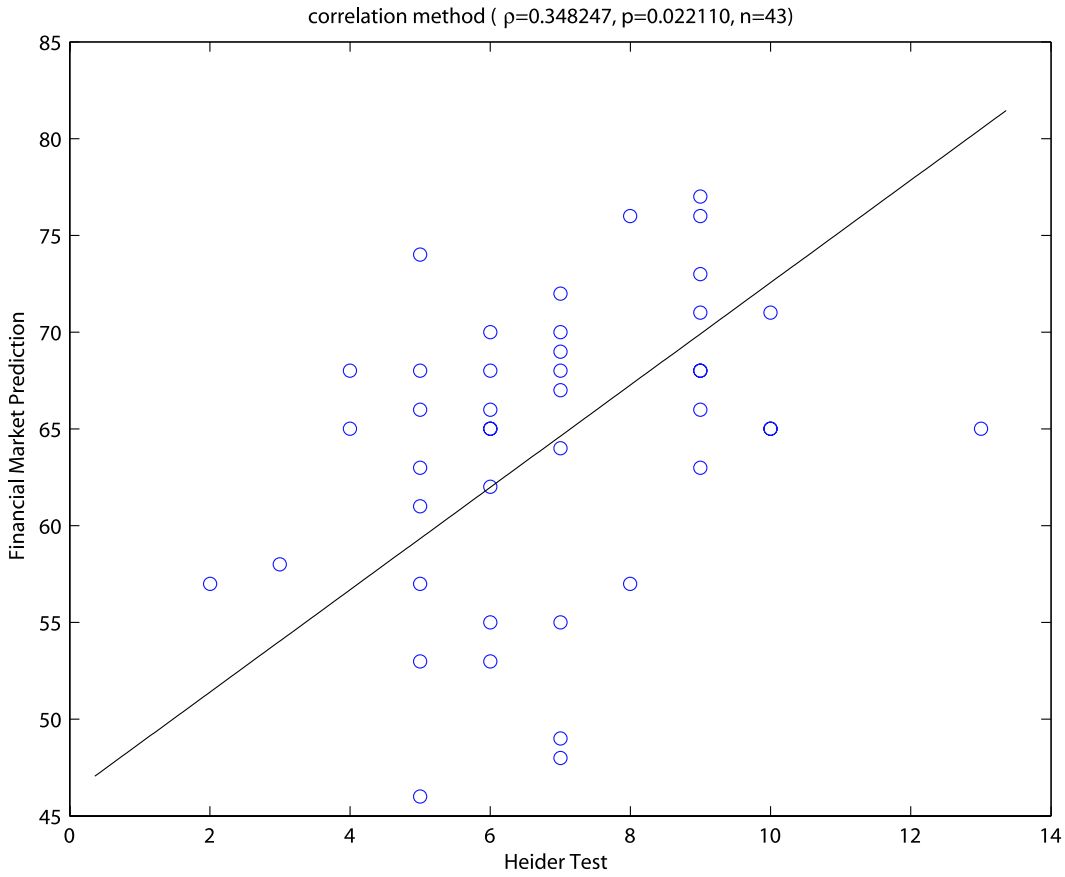


Figure B.4: Performance on the financial market prediction task as a function of the performance on the Heider test. We observe a positive and significant correlation.

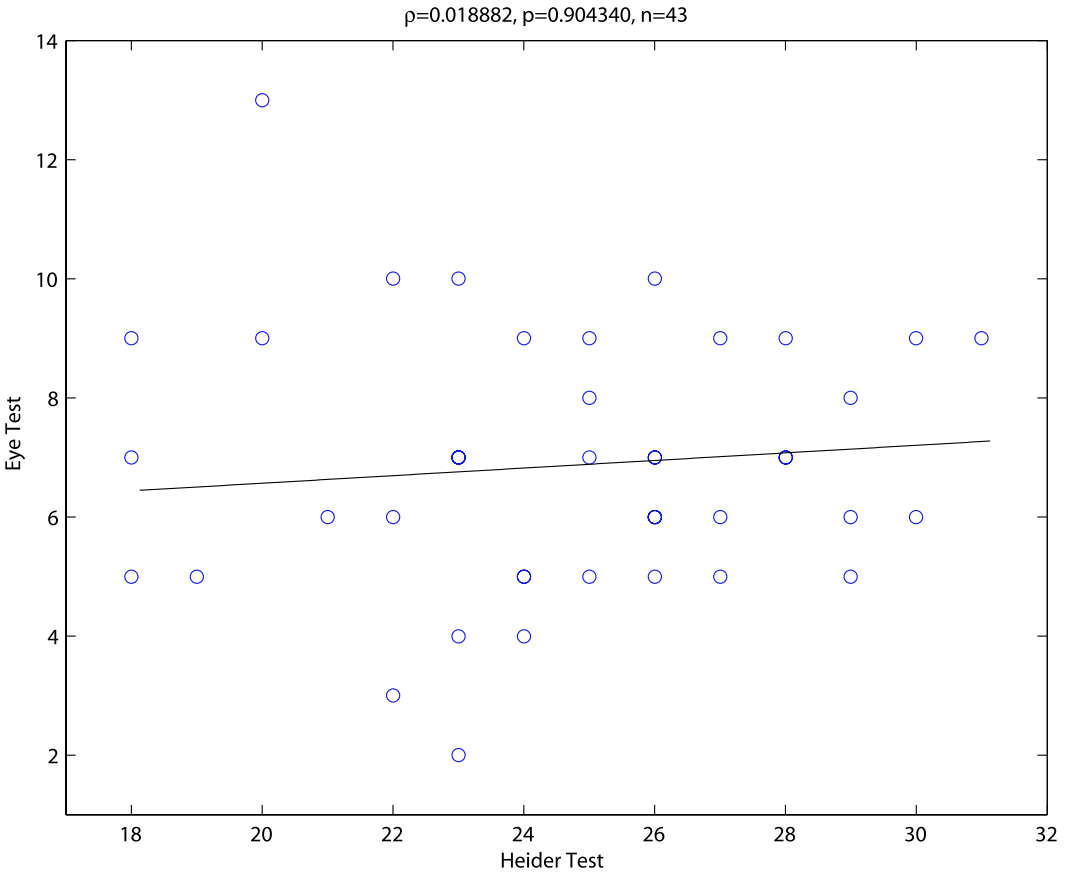


Figure B.5: Performance on the eye test as a function of the performance on the Heider test. We observe no significant correlation.

I predicted them by looking at the different sizes of the prices and previous actions. Usually, things stayed the same
Watching the trade price and focusing on the outer most blue and 2nd red bubble
If there was no movement I would pick same. Also if two #'s were equidistant from the center I would pick the same. Overall I tried different things, from distance to the center, to movement of buyers and sellers
I looked at which bubbles grew quickly - quick growth was a cut that insiders were biased toward that stock value, so I would compare that popular value to the current trade price
I looked at the size of the bubbles and whether it was likely there would be a green flash (transaction made) depending on the supply and demand near current price and the trend of the price change thus far
used the buying & selling pressure to figure out. When there was big diff. in buying & selling price I predicted it to be the same
Looked at the circles & numbers closest to the "+" on the screen
There weren't many changes, so I predicted SAME most of the time. I just watched the circles & tried to get a feel for the overall trend.
I felt that I predicted the stock market game quite well. The most difficult part was predicting gradual increases/decreases at the end
I looked at the trend - if the prize [sic] was going up or down or there was no change for long time.
Decently
I predicted by looking at what the smallest circle closer to the cross was. The closer they are to a big circle, the more likely they are sold
I paid attention to what was the biggest price that people were willing to buy the stock and he lowest that they were willing to sell.
I tried to look for what the insiders were doing as well as any consistencies within the prices
Look at current price and compare sizes of bubbles. The side w. smaller bubble is the side more likely to be approached
Based on the average of prices close to center
I looked at the size of the bubble of the last sale in comparison to the bubbles immediately above and below it
I watched the movement and studied the prices
According to the unit price and number of unit for transaction, I can predict the price. If there is a big gap between selling price and buying price the price would probably stay the same for quite a long time.
I guessed
I watched the total prices, and weighted that value against the most recent trade, and watched for rallies
I noticed the price stayed the same more often than not, but I predicted that the price would decrease after it increased twice
I thought I predicted the prices well. I averaged he highest "willing to buy" with the lowest "willing to sell." 90% of the time this strategy worked
My prediction were based on the smallest amount of units for the last price that would occur in a regular pattern
For the most part I was just making random guesses because I don't have experience in it, which is why I'm unaware if there is a systematic way of predicting the price.
I figured it was safer to mostly click "same" I would figure it usually stays the same. Sometimes I'd guess up or down depending on stock activity
I looked at the trends, which was that it mostly stayed the same. Also, I looked at the circles. If the prices are too different, then it'll probably stay the same. If they're really close, then the price might change
I guessed same after the trade price changed & I would guess lower or higher if a large quantity, at a certain price began moving
I looked for a pattern. If I didn't see a pattern, or if I didn't know I choose "stay the same." I noticed that at the beginning of each phase, the prices were volatile, but became more stagnant as the game went on
I focused on the space increases in each case (red stock buying and the blue being sold). It was a bit tricky considering the numbers would stay the same.
I looked at the size of the circle of the most recent trade
The change of bubble and the size of bubbles. The smaller size bubble, the better price to be predicted
By looking at the trends and noticing that when everyone was offering to sell or buy at a particular price, most likely that would not be the next price. I also realized in the beginning of each round the price was going down to the range of 25 to 35
There are a lot of fluctuations initially, until the market stagnates and nothing happens. Then the prices fall usually one more time
Most of the time, the trading price will remain either the same or going between 2 consecutive numbers. I would choose same, unless there was a very large shift in the "equilibrium," in which price would favor the side w/ less [sic] units
By seeing how many people wanted to share for a higher/lower price

Table B.1: Answers to questionnaire. After the behavioral experiment, we asked subject to answer the question: "How did you predict the changes in the price?" We do not see any evidence of anthropomorphization of the stock market.

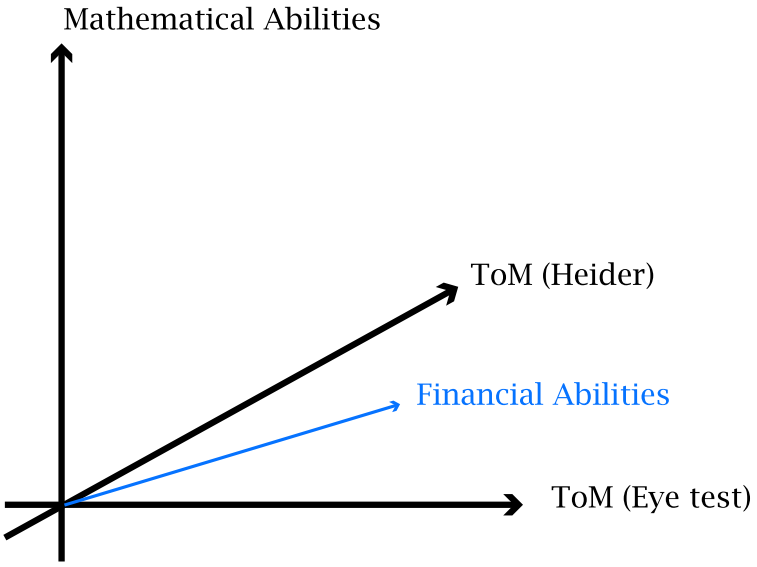


Figure B.6: Graphical representation of why the two ToM are not correlated. We symbolize here the results of our behavioral test. Three facets of human abilities are orthogonal to each other and we represent them with an orthogonal basis (black arrows). The two ToM tests are not collinear but rather define a plane. The financial abilities are not significantly correlated with the mathematical test. They are however correlated with both ToM tests; i.e., they lie in the ToM plane.

References

L. Abbott and P. Dayan. The effect of correlated variablity on the accurancy of a population code. *Neural Computation*, 11:91–101, 1999.

B. Acuna, J. Eliassen, J. Donoghue, and J. Sanes. Frontal and parietal lobe activation during transitive inference in humans. *Cerebral Cortex*, 12:1312–1321, 2002.

A. Admati. A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica*, 53:629–657, 1985.

M. Allais. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21:503–546, 1953.

T. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Hoboken, NJ, 3rd edition, 2003.

F. Anscombe and R. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

K. Arrow. The role of securities in the optimal allocation of risk-bearing. *The Review of Economic Studies*, 31:91–96, 1953.

B. Averbeck, P. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7:358–366, 2006.

M. Barner, F. Feri, and C. Plott. On the microstructure of price determination and information aggregation with sequential and asymmetric information arrival in an experimental asset market. *Annals of Finance*, 1:73–107, 2005.

S. Baron-Cohen, T. Jolliffe, C. Mortimore, and M. Robertson. Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry*, 38:813–822, 1997.

S. Baron-Cohen, H. Ring, S. Wheelwright, and E. Bullmore. Social intelligence in the normal and autistic brain: an fMRI study. *The European Journal of Neuroscience*, 11: 1891–1898, 1999.

S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen. *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. Oxford University Press, New York, NY, 2nd edition, 2000.

M. Bhatt and C. Camerer. Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior*, 52:424–459, 2005.

P. Bossaerts, C. Plott, and W. Zame. Prices and portfolio choices in financial markets: Theory, econometrics, experiment. *Econometrica*, 75:993–1038, 2007.

M. Botvinick, J. Cohen, and C. Carter. Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, 8:539–546, 2004.

K. Brodmann. Beiträge zur histologischen lokalisation der grosshirnrinde. dritte mitteilung: Die rindenfelder der niederen affen. *Journal für Psychologie und Neurologie*, 4:177–226, 1905.

K. Brownlee. *Statistical Theory and Methodology: In Science and Engineering*. Krieger Publication Company, Melbourne, FL, 2nd edition, 1984.

M. Burock and A. Dale. Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. *Human Brain Mapping*, 11:249–260, 2000.

R. Buxton and E. Wong. Dynamic of blood flow and oxygenation changes during brain activations: the balloon model. *Magnetic Resonance in Medicine*, 39:855–864, 1998.

B. Carlin and T. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, Toronto, 2nd edition, 2000.

F. Castelli, F. Happé, U. Frith, and C. Frith. Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, 12:314–325, 2000.

R. Cooter. *The Cultural Meaning of Popular Science: Phrenology and the Organization of Consent in Nineteenth-Century Britain*. Cambridge University Press, New York, NY, 1985.

T. Crack. *Heard on the Street: Quantitative Questions from Wall Street Job Interviews*. Timothy Crack, Dunedin, New Zealand, 10th edition, 2007.

H. Critchley, C. Mathias, and R. Dolan. Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron*, 29:537–545, 2001.

G. Debreu. Continuity properties of paretian utility. *International Economic Review*, 5: 285–293, 1964.

G. Debreu. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Yale University Press, New Haven, CT, 1972.

J. Decety, P. Jackson, J. Sommerville, T. Chaminade, and A. Meltzoff. The neural bases of cooperation and competition: an fMRI investigation. *NeuroImage*, 23:744–751, 2004.

S. Dehaene, E. Spelke, P. Pinel, R. Stanescu, and S. Tsivkin. Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284:970–974, 1999.

P. Diaconis and D. Freedman. On the consistency of Bayes estimates. *The Annals of Statistics*, 14:1–26, 1986.

D. Easley and M. O'Hara. Time and the process of security price adjustment. *The Journal of Finance*, 47:577–605, 1992.

B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, Toronto, 1993.

J. Esteban and D. Ray. Conflict and distribution. *Journal of Economic Theory*, 87:379–415, 1999.

E. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.

K. Friston. Bayesian estimation of dynamical systems: an application with fMRI. *NeuroImage*, 16:513–530, 2002.

K. Friston, C. Frith, R. Frackowiak, and R. Turner. Characterizing dynamic brain responses with fMRI: A multivariate approach. *NeuroImage*, 2:166–172, 1995.

K. Friston, C. Buechel, G. Fink, J. Morris, E. Rolls, and R. Dolan. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6:218–229, 1997.

H. Gallagher and C. Frith. Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, 7:77–83, 2003.

H. Gallagher, A. Jack, A. Roepstorff, and C. Frith. Imaging the intentional stance in a competitive game. *NeuroImage*, 16:814–821, 2002.

M. Gazzaniga, R. Ivry, and G. Mangun. *Cognitive Neuroscience*. W. W. Norton, New York, NY, 2nd edition, 2002.

C. Genovese, N. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15:870–878, 2002.

J. Gillespie and A. Jackson. *MRI and CT of the Brain*. Arnold Publishers, London, UK, 2000.

L. Glosten and P. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14:71–100, 1985.

M. Goncalves and D. Hull. Connectivity analysis with structural equation modelling: an example of the effects of voxel selection. *NeuroImage*, 20:1455–1467, 2003.

S. Grossman and J. Stiglitz. On the impossibility of informationally efficient markets. *The American Economic Review*, 70:393–408, 1980.

R. Guesnerie. An exploration of the eductive justifications of the rational-expectations hypothesis. *American Economic Review*, 82:1254–1278, 1992.

A. Hald. *Statistical Theory with Engineering Applications*. John Wiley and Sons, Hoboken, NJ, 1952.

F. Heider and M. Simmel. An experimental study of apparent behaviour. *American Journal of Psychology*, 57:243–259, 1944.

C. Holt and S. Laury. Risk aversion and incentive effects. *The American Economic Review*, 92:1644–1655, 2002.

H. Hotelling. Relation between two sets of variants. *Biometrika*, 28:312–377, 1936.

H. Hotelling. The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2: 360–378, 1931.

C. Hsiao. *Analysis of Panel Data*. Cambridge University Press, New York, 2nd edition, 2003.

S. Huettel, A. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, Sunderland, MA, 2004.

S. Huettel, A. Song, and G. McCarthy. Decisions under uncertainty: Probabilistic context influences activation of prefrontal and parietal cortices. *Journal of Neuroscience*, 25: 3304–3311, 2005.

P. Jezzard, P. Matthews, and S. Smith. *Functional MRI, an introduction to methods*. Oxford University Press, Oxford, UK, 2001.

R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, NJ, 5th edition, 2002.

J. Kagel and A. Roth. *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ, 1997.

Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8:679–685, 2005.

B. Knutson, C. Adams, G. Fong, and D. Hommer. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience*, 21:RC159 1–5, 2001.

J. Lagopoulos. Functional MRI: an overview. *Acta Neuropsychiatrica*, 19:64–65, 2007.

E. Langer and J. Roth. Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, 34:191–198, 1975.

A. Lo and D. Repin. The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience*, 14:323–339, 2002.

N. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157, 2001.

R. Lucas. Expectations and the neutrality of money. *Journal of Economic Theory*, 4: 103–124, 1972.

A. Marcet and T. Sargent. Convergence of least-squares learning in environments with hidden state variables and private information. *Journal of Political Economy*, 97:1306–1322, 1989.

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

K. McCabe, D. Houser, L. Ryan, V. Smith, and T. Trouard. A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, 98:11832–11835, 2001.

M. McKeown, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6:160–188, 1998.

R. Mukamel, H. Gelbard, A. Arieli, U. Hasson, I. Fried, and R. Malach. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science*, 309:951–954, 2005.

J. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29: 315–335, 1961.

J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1953.

S. Newman, P. Carpenter, S. Varma, and M. Just. Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia*, 41:1668–1682, 2003.

J. Niessing, B. Ebisch, K. Schmidt, M. Niessing, W. Singer, and R. Galuske. Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science*, 309:948–951, 2005.

T. Oberlechner. Understanding the foreign exchange market through metaphors. *British Journal of Social Psychology*, 43:133–156, 2004.

C. Padoa-Schioppa and J. Assad. Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441:223–226, 2006.

L. Parsons and D. Osherson. New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cerebral Cortex*, 11:954–965, 2001.

W. Penny, A. Holmes, and K. Friston. *Human Brain Function*. Academic Press, New York, 2nd edition, 2003.

W. Penny, K. Stephan, A. Mechello, and K. Friston. Modelling functional integration: a comparison of structural equation and dynamic causal models. *NeuroImage*, 23:S264–S274, 2004.

C. Plott and S. Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56:1085–1118, 1988.

K. Preuschoff, P. Bossaerts, and S. Quartz. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, 51:381–390, 2006.

J. Rilling, A. Sanfey, J. Aronson, L. Nystrom, and J. Cohen. The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, 22:1694–1703, 2004.

D. Robbe, S. Montgomery, A. Thome, P. Rueda-Orozco, B. McNaughton, and G. Buzsaki. Cannabinoids reveal importance of spike coordination in hippocampal function. *Nature Neuroscience*, 9:1526–1533, 2006.

E. Rodriguez, N. George, J. Lachaux, J. Martinerie, B. Renault, and F. Varela. Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397:430–433, 1999.

E. Rolls, C. McCabe, and J. Redoute. Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cerebral Cortex*, *in print*, 2007.

R. Romo, A. Hernandez, A. Zainos, and E. Salinas. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron*, 38:649–657, 2003.

U. Rutishauser, A. Mamelak, and E. Schuman. Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron*, 49:805–813, 2006.

E. Salinas and T. Sejnowski. Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, 2:539–550, 2001.

E. Salinas and T. Sejnowski. Integrate-and-fire neurons driven by correlated stochastic input. *Neural Computation*, 14:2111–2155, 2002.

P. Samuelson. A note on the pure theory of consumers' behaviour. *Economica*, 5:61–71, 1938.

P. Samuelson. *Foundations of Economic Analysis*. Harvard University Press, Cambridge, MA, enlarged edition, 1983.

A. Santia and Y. Grodzinsky. Working memory and syntax interact in broca's area. *NeuroImage*, 37:8–17, 2007.

L. Savage. *The Foundations of Statistics*. Dover Publications, Mineola, NY, 2nd edition, 1972.

E. Schneidman, M. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–1012, 2006.

P. Servos, R. Osu, A. Santi, and M. Kawato. The neural substrates of biological motion perception: an fMRI study. *Cerebral Cortex*, 12:772–782, 2002.

C. Sestieri, R. Di Matteo, A. Ferretti, C. Del Gratta, M. Caulo, A. Tartaro, M. Olivetti-Belardinelli, and G. Romani. "What" versus "Where" in the audiovisual domain: An fMRI study. *NeuroImage*, 33:672–680, 2006.

M. Shadlen and J. Movshon. Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron*, 24:67–77, 1999.

M. Shadlen and W. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4:569–579, 1994.

F. Shellock. *Magnetic Resonance Procedures: Health Effects and Safety*. CRC Press, Boca Raton, FL, 2001a.

F. Shellock. *Pocket Guide to MR Procedures and Metallic Objects*. Lippincott Williams and Wilkins, New York, NY, 2001b.

F. Shellock. *Reference Manual for Magnetic Resonance Safety, Implants, and Devices*. Biomedical Research Publishing Group, Los Angeles, CA, 2005.

T. Singer, B. Seymour, J. O'Doherty, H. Kaube, R. Dolan, and C. Frith. Empathy for pain involves the affective but not sensory components of pain. *Science*, 303:1157–1162, 2004.

A. Spanos. *Statistical Foundations of Econometric Modeling*. Cambridge University Press, Cambridge, UK, 1986.

P. Steinmetz, A. Roy, P. Fitzgerald, S. Hsiao, K. Johnson, and E. Niebur. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404: 187–190, 2000.

V. Stone, S. Baron-Cohen, and R. Knight. Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10:640–656, 1998.

M. Stopfer, S. Bhagavan, B. Smith, and G. Laurent. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature*, 390:70–74, 1997.

G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, MA, 3rd edition, 2003.

J. Talairach and P. Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System – an Approach to Cerebral Imaging*. Thieme Medical Publishers, New York, 1988.

S. Tanaka, T. Inui, S. Iwaki, J. Konishi, and T. Nakai. Neural substrates involved in imitating finger configurations: an fMRI study. *NeuroReport*, 12:1171–1174, 2001.

B. Völlm, A. Taylor, P. Richardson, A. Corcoran, J. Stirling, S. McKie, J. Deakin, and R. Elliott. Neuronal correlates of theory of mind and empathy: a functional magentic resonance imaging study in a nonverbal task. *NeuroImage*, 29:90–98, 2006.

H. Walter, M. Adenzato, A. Ciaramidaro, I. Enrici, L. Pia, and B. Bara. Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, 16:1854–1863, 2004.

S. Weisberg. *Applied Linear Regression*. John Wiley and Sons, New York, 3rd edition, 2005.

H. Wimmer and J. Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13:103–128, 1983.