

Investigating signal integration with canonical correlation analysis of fMRI brain activation data

Antoine Bruguier,^a Kerstin Preuschoff,^b Steven Quartz,^b and Peter Bossaerts^{b,c,*}

^aDivision of Engineering and Applied Science, California Institute of Technology, 1200 E. California Blvd, 136-93, Pasadena, CA 91125, USA

^bComputation and Neural Systems, California Institute of Technology, 1200 E. California Blvd, 216-76, Pasadena, CA 91125, USA

^cEcole Polytechnique Fédérale Lausanne, Station 15, CH-1015 Lausanne, Switzerland

Received 24 July 2007; revised 15 January 2008; accepted 24 January 2008
Available online 15 February 2008

How the brain integrates signals from specific areas has been a longstanding critical question for neurobiologists. Two recent observations suggest a new approach to fMRI data analysis of this question. First, in many instances, the brain analyzes inputs by decomposing the information along several salient dimensions. For example, earlier work demonstrated that the brain splits a monetary gamble in terms of expected reward (ER) and variance of the reward (risk) [Preuschoff, K., Bossaerts, P., Quartz, S., 2006. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51, 381–390]. However, since ER and risk activate separate brain regions, the brain needs to integrate these activations to obtain an overall evaluation of the gamble. Second, recent evidence suggests that the correlation of the activity between neurons may serve a specific organizational purpose [Romo, R., Hernandez, A., Zainos, A., Salinas, E., 2003. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38, 649–657; Salinas, E., Sejnowski, T.J., 2001. Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539]. Specifically, it is hypothesized that correlations allow brain regions to integrate several signals in a way that minimizes noise. Under this hypothesis, we show here that canonical correlation analysis of fMRI data identifies how the signals from several regions are combined. A general linear model then verifies whether the identified combination indeed activates a projection area in the brain. We illustrate the proposed procedure on data recorded while human subjects played a simple card game. We show that the brain adds the signals of ER and risk to form a measure that activates the medial prefrontal cortex, consistent with the role of this brain structure in the evaluation of monetary gambles.

© 2008 Elsevier Inc. All rights reserved.

Introduction

General linear models (GLMs) have been used extensively to analyze functional magnetic resonance imaging data (fMRI) and to map the human brain into regions that perform specific tasks (Frackowiak et al., 2003). The fMRI amplitudes track the changes in hemoglobin oxygenation (blood-oxygen-level dependent, or BOLD). The BOLD signals can be related to the neuronal activity either linearly, with a hemodynamic response function (HRF), or non-linearly, with a Balloon model (Buxton and Wong, 1998; Friston et al., 2000). Additionally, experimental studies verified the relationship between neurons' activity and fMRI data (Mukamel et al., 2005). Experimenters analyze the data with GLMs that project the BOLD signal onto a set of predictors designed after the stimulus set of an experiment.

However, GLMs offer researchers limited insight into how several areas interact. In particular, using GLMs does not reveal how information is transmitted from one area to another or how information from multiple areas is integrated. They can help us understand “where” but not “how”. To explain how areas interact, researchers have designed new methods of analysis.

However, some of the new methods are limited because they make no assumptions regarding underlying biological mechanisms. Primary among these methods are principal and independent component analysis (PCA/ICA) (McKeown et al., 1998). PCA and ICA are further limited because they are exploratory and do not use experimenter-defined predictors to test a priori hypotheses.

To address the limitations of PCA and ICA, structural equation models (SEM) and dynamic causal modeling (DCM) were devised (Friston, 2002; Goncalves and Hull, 2003). Researchers used both methods to investigate effective connectivity (Frackowiak et al., 2003), but with DCM, they specify a model at the neuronal level, and with SEM, they examine data at the level of fMRI (Penny et al., 2004). In addition, researchers uncovered connections between areas of the brain with psychophysiological interactions (PPI) (Friston et al., 1997). A strength of DCM and PPI is that they can

* Corresponding author. Ecole Polytechnique Fédérale Lausanne, Station 15, CH-1015 Lausanne, Switzerland.

E-mail addresses: bruguier@alumni.caltech.edu (A. Bruguier), preuschi@caltech.edu (K. Preuschoff), steve@hss.caltech.edu (S. Quartz), peter.bossaerts@epfl.ch (P. Bossaerts).

Available online on ScienceDirect (www.sciencedirect.com).

help researchers explain nonlinear effects in the brain, while all previously mentioned methods rely on purely linear models.

DCM, PPI and SEM have two main drawbacks. First, DCM, PPI and SEM, unlike GLMs, cannot be directly used to relate predictors to specific experimental conditions. Even though this absence of relationship between predictors and experimental conditions offers more flexibility to explain changes in brain activity, it makes results harder to interpret. Second, DCM, PPI and SEM work with predefined regions. While they enable investigators to explain relationships between previously known areas, they cannot help to map new functions in the brain.

Here we overcome these limitations with a novel application of a proven method of analysis. Our application is based on general biological observations. Specifically, evidence from direct neuronal recording shows that neurons are sensitive to correlations in their inputs (Salinas and Sejnowski, 2000; Shadlen and Movshon, 1999). The hypothesis that these correlations reflect changes in connectivity was later proven experimentally with neuronal recording (Romo et al., 2003). The idea was that neurons judiciously used the correlations to combine afferent signals in a way that minimized noise in the integrated signal. We propose here that one can exploit these correlations with a method based on canonical correlation analysis (CCA) in order to identify how neurons integrate signals (Anderson, 2003; Hotelling, 1936; Johnson and Wichern, 2002).

CCA is a statistical procedure that uses as input several dependent variables (in the form of time-indexed vectors) and several independent variables (also in the form of time-indexed vectors). The output is a collection of weights for both the dependent and independent variables. These weights are chosen so that the correlation between the weighted sum of the dependent variables and the weighted sum of the independent variables is maximized (Appendix B).

In the context of our experiment, the dependent variables are several predictors and the independent variables are recordings of the neuronal activity in several brain regions. The predictors reflect the signals that the neuronal activations are known to carry individually. CCA identifies the particular combination of the signals that correlates maximally with the *joint* activations across the regions. CCA exploits the noise-minimizing correlations in baseline activations to effectively “reverse engineer” the way neurons integrate the signals.

Because it is difficult to record neuronal activities in humans, researchers usually record fMRI data instead. However, it is still theoretically possible to detect correlation in neuronal activity with CCA applied to fMRI data. Recent work has established a link between neuronal activity and fMRI recordings (Logothetis et al., 2001; Mukamel et al., 2005). With simple mathematics (Appendix A) we show that if the neuronal activity is directly related to the amplitudes recorded with fMRI, then the correlations of neuronal activities are also directly related to the correlations of fMRI time courses.

Even though CCA has already been used on fMRI data under the name CVA (Friston et al., 1995) our method is fundamentally different. Friston and colleagues first combined signals from numerous voxels to create a much smaller number of signals (data reduction). Then, they discovered new predictors by applying a CVA; but the data reduction step prevents them from linking predictors to specific brain regions. In contrast, we take a different approach. We base our method on the specific hypothesis that particular brain regions integrate signals in a way that minimizes noise, and we use CCA to exploit and test this hypothesis.

The way we apply CCA has two distinct advantages. First, we use predictors in the same way as we would have with GLMs. This

approach makes the results easy to interpret by directly relating predictors to both the experimental conditions and to specific brain areas, unlike DCM, PPI and SEM. By examining areas and predictors, we can test falsifiable hypotheses on the role of brain areas. Second, with our approach to CCA, we can map new functions in the human brain, unlike DCM, PPI and SEM. As we will show in the Materials and methods section, CCA creates a new “composite” predictor that makes explicit how neurons combine their inputs. By introducing this new predictor in a post-CCA GLM and performing a whole-brain analysis, we can map the composite predictor to a new brain area.

The post-CCA GLM also helps us to validate our analysis. If the composite predictor recovered through CCA does not significantly correlate with the activity of any brain region, then the negative result will cast doubt on the validity of the method we propose. Indeed, if we observe correlations in the CCA step but no area is activated in the GLM step, we should conclude that the correlations do exist but that neurons do not use them to integrate signals. In other words, we should reject the hypothesis on which our method of analysis is based.

Materials and methods

Overview

Our method consists of three steps (Fig. 1). In the first step, we average the fMRI time courses (time-indexed vectors containing the amplitudes of the BOLD signal) of several voxels for each region of interest. These fMRI time courses contain two types of signals—the signals of interest that we wish to investigate and “extraneous” signals that we wish to ignore (such as those caused by instruction text on the screen or an image). To remove these extraneous signals, we compute adjusted time courses, i.e., the residuals of linear regressions containing the predictors that correspond to these extraneous signals. The residuals are orthogonal to the extraneous signals but still contain the signals we wish to investigate (Weisberg, 2005). The residuals are the input of the second step, where we use CCA to test the hypothesis that neurons use correlations to combine the signals of interest. The CCA step yields a new composite predictor, and in the third step, we use this new predictor in a post-CCA GLM to determine the location of a downstream area where the combined signals of the upstream areas can be said to be projected. The post-CCA GLM is a whole-brain exploration of regions whose activations reflect encoding of the combined signal communicated through joint activation of the regions of interest.

To illustrate our method, we analyze the integration of signals related to reward and risks in a simple monetary gamble. We use data from a previous study that found that the two key features of a gamble—its expected reward (ER, measured as the statistical mean of the payoff) and its risk (measured as the statistical variance of the payoff)—activate distinct areas of the brain (Preusschoff et al., 2006). We call these areas upstream regions, and we use a CCA to investigate how the brain integrates the two features of the gamble to generate a single-dimensional metric. We then check with a GLM that the predictor corresponding to this new metric activates a downstream area.

Correlation hypothesis

In this section, we present the biological rationale for the method we use. For simplicity, we limit ourselves to the case where the downstream region recombines only two upstream activities, recorded as fMRI time courses y_1 and y_2 (both vectors of T time samples). We

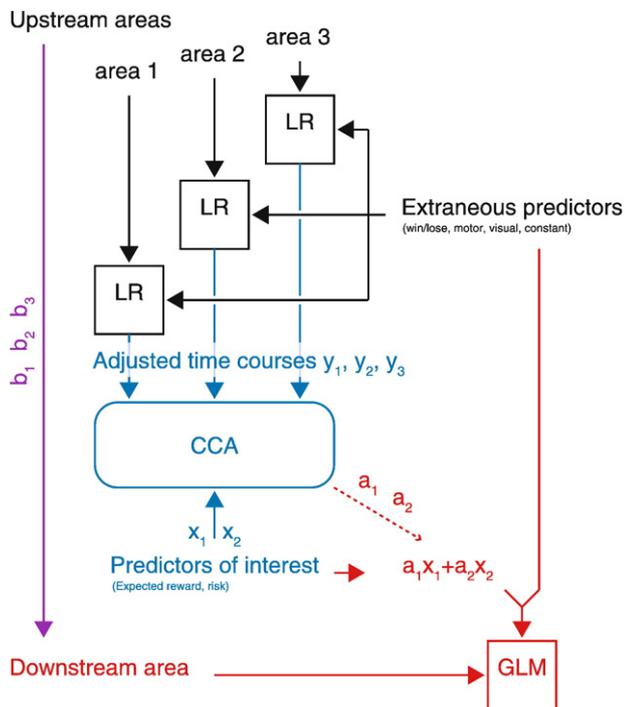


Fig. 1. Overview of the process. With a three-step method, we discover how several “upstream” brain areas may project their combined signal to a “downstream” area (purple arrow). The first step (black section) removes the effect of extraneous predictors (such as visual, or motor) from (in this case) three regions of interest (upstream regions) and generates adjusted time courses (residuals of linear regressions). These adjusted time courses, along with the predictors of interest, are the input of a CCA (second step, blue section; see also Appendix B) that generates a set of weights (a_1 and a_2) that in turn are passed onto the third step (red section). In this last step, we examine every voxel within a candidate downstream region of the brain with a GLM composed of extraneous predictors as well as the newly created predictor ($U_1 = a_1x_1 + a_2x_2$). A predictor U_2 is added to correct for the misestimation of the weights (see Appendix D).

model each y_i ($i = 1, 2$) as the sum of a (predictor) signal induced by the experiment x_i (without loss of generality, we do not assign a weight), a noise term \tilde{y}_i that is possibly correlated with the one in other regions, and a second, Gaussian noise term n_i , assumed to be independent across voxels:

$$y_i = x_i + \tilde{y}_i + n_i.$$

In traditional GLM analysis of fMRI data, the (correlated) noise term \tilde{y}_i and the (uncorrelated) noise term n_i are combined together, and any correlation (across regions) of the \tilde{y}_i components is ignored. Here we do not make this simplification and instead assume a factor structure on noise, with \tilde{y}_i representing the influence of the common factor. We then exploit the correlation of the \tilde{y}_i s across regions.

The correlations of the \tilde{y}_i s are conjectured to ensure maximum signal-to-noise ratio when the signals encoded in the individual activations are combined (Romo et al., 2003; Salinas and Sejnowski, 2001). For example, if there is a negative correlation, the brain will cancel out some of the effect of the \tilde{y}_i s when adding the activations. That is, if the joint activation is to reflect, with minimal noise, the summation of the signals encoded in the separate activations, then the \tilde{y}_i s ought to be negatively correlated. Similarly, if there is a positive correlation between the \tilde{y}_i s, their effect is diminished when the brain subtracts the activations.

Mathematically, we assume that the brain combines activations in a way that minimizes noise. Specifically, given the signal $a_1x_1 + a_2x_2$ that needs to be encoded, the brain implements $\text{corr}(\tilde{y}_1, \tilde{y}_2)$ such that it maximizes the correlation between the combined signal and the combined activations:

$$\max_{\text{corr}(\tilde{y}_1, \tilde{y}_2)} \{ \text{corr}(a_1x_1 + a_2x_2; b_1(x_1 + \tilde{y}_1 + n_1) + b_2(x_2 + \tilde{y}_2 + n_2)) \}.$$

While not made explicit in the mathematical notation, the maximization takes into account physiological constraints.

Here we propose to use CCA to estimate the combined signal $a_1x_1 + a_2x_2$. As researchers, we know neither the weights a_1 and a_2 nor the weights b_1 and b_2 . However, we can exploit the correlation between the \tilde{y}_i s and infer these weights using canonical correlations analysis (CCA).

Use of CCA to discover the weights

CCA (Johnson and Wichern, 2002) is a standard statistical method that, like the conjectured computations in the brain, also maximizes correlation; only, it takes as given the correlation between the \tilde{y}_i s while determining the weights a_1, a_2, b_1 and b_2 , instead of maximizing this correlation given the weights. Specifically, CCA solves the following maximization problem:

$$\rho = \max_{a_1, a_2, b_1, b_2} \{ \text{corr}(a_1x_1 + a_2x_2; b_1(x_1 + \tilde{y}_1 + n_1) + b_2(x_2 + \tilde{y}_2 + n_2)) \}.$$

Standard CCA in fact provides two solutions to this maximization: one solution that maximizes the correlation in an unconstrained fashion and a second solution with a (generically lower) correlation given orthogonality constraints on the signal and activation combinations relative to those obtained in the first solution. When there are n predictors x_i and p fMRI time courses y_i , there will be $\min(n, p)$ solutions. We are only interested in the first solution, i.e., the one that maximizes correlation in an unconstrained fashion. At one point, however, we will use the second solution in order to minimize the effect on inference from noise in estimating the coefficients for the first solution.

Inference techniques have been devised to test the significance of the estimated maximum correlation ρ (Anderson, 2003). We are also interested in testing that each of the weights a_1, a_2, b_1 , and b_2 is significantly different from zero. If any of the weights (in this 2-by-2 situation) is zero, then one cannot meaningfully speak about the brain “combining signals” (if one of the a_i s is zero) or “combining activations” (if one of the b_i s is zero). One of b_1, b_2 will automatically be significant because the signals/predictors were chosen because they activated at least one region of interest (see also Appendix C). The real import is when both b_1 and b_2 are significant. We present a technique (see Appendix B) that tests the null hypotheses that the weights (separately) are zero.

Using GLM to identify projection areas

Once we identify the signal combination $a_1x_1 + a_2x_2$, we can localize a candidate for the projection area by using a GLM that includes $a_1x_1 + a_2x_2$ as a predictor (in addition to the usual extraneous predictors). Obviously, if we do not find any plausible projection, we will reject the principle underlying the prior CCA step. While significance in the canonical correlations analysis would imply that there are correlations, the absence of a projection area suggests that the brain does not use these correlations to direct

projection of the identified signal combination onto some downstream region.

We identify brain regions that activate significantly in the post-CCA GLM as “downstream” areas, to which the regions of interest project or which are activated by the regions of interest. As with any correlation-based procedure, one should be careful with interpretations that imply causality or directionality. However, the predictors (signal combinations) in the post-CCA GLM are identified by joint activation of only the regions of interest. That is, the brain regions eventually uncovered in the post-CCA GLM are not involved in this identification process. As such, the correlations obtained in the post-CCA GLM step are causality (or directionality) in the Granger sense (Granger, 1969).

Why CCA is useful

While the post-CCA GLM is a projection of brain activations onto the recovered signal combination $a_1x_1 + a_2x_2$, it is not equivalent to a GLM where x_1 and x_2 are used as separate predictors. Our post-CCA GLM has higher statistical power, and Appendix D shows that one would need at least twice as many subjects to make up for the loss of power when not using the CCA-based GLM.

However, the advantages of our method go beyond statistical considerations. Our procedure is primarily useful from a neuroscience perspective. Had we improved the statistical power by increasing the number of subjects instead of using CCA, we could not have discovered that the brain integrates activations from separate regions. Instead, with standard GLM and additional subjects, we would only have found overlapping localization, i.e., the activity in the projection area would have correlated significantly with both predictors.

Instrumental variable estimation

In the post-CCA GLM, we use the estimated combination $a_1x_1 + a_2x_2$ as predictor. To mitigate noise in estimation of this predictor, we apply a method similar to, but different from, instrumental variables estimation (Spanos, 1986). In addition to $a_1x_1 + a_2x_2$, we add a second predictor to the post-CCA GLM. We had several options, but we show in Appendix E that the most effective choice is the second combined signal uncovered by the CCA.

Traditional application of CCA actually produces several solutions. Only the first one gives maximum correlation (between combinations of activations and combinations of signals), and only this first solution has been used in our method up to now. The second solution, however, while producing lower correlation, provides a combined signal that is uncorrelated with that of the first solution, yet it is affected by the same estimation error. Appendix E shows that the latter can be exploited to reduce the impact of estimation error in the post-CCA GLM.

Incidentally, adding this second predictor provides an additional check on the validity of the entire procedure. If the signal combination identified in our CCA step truly represents the activity of the downstream region, then only it should correlate significantly with activation in that region; the correlation with the predictor added in the post-CCA GLM to mitigate errors-in-variables ought to be insignificant.

Removal of extraneous predictors

Prior to implementing CCA, we need to remove from the upstream activities the influence of extraneous predictors (predictors that are not

part of the hypothesized signal recombination). Without this removal, these extraneous predictors may spuriously generate erroneous results when using CCA because at least one of these extraneous predictors may correlate with several upstream regions simultaneously.

We remove the influence of the extraneous predictors by using “adjusted time courses” in CCA. Specifically, prior to CCA, we run a linear regression on the activations of each upstream region that contains extraneous predictors that may simultaneously correlate with more than one region of interest. The residuals of these linear regressions provide the adjusted time courses. By construction, they are orthogonal to the extraneous predictors, and hence, they are not correlated across regions because of potential common influence of extraneous predictors.

Results

Experiment

To illustrate implementation of our procedure, we analyzed the data from a previous experiment involving a simple card game (Preuschhoff et al., 2006). In modeling joint brain activation, expected reward (ER) and reward variance (risk) are the two main predictors of interest. They correspond to the variables x_1 and x_2 in the description of our procedure. The experimental paradigm was such that these two predictors were orthogonal. ER had been found to correlate with activation in certain subcortical regions (ventral striatum; putamen) while risk correlated with activation in insula. Our implementation of CCA allowed us to investigate to what extent there was a combined signal in joint activation of ventral striatum, putamen and insula; to identify its nature; and to localize potential projection regions.

We first averaged the signal of 30 voxels in the three regions of interest: ventral striatum (Talairach coordinates $-12; 5; -3$), putamen ($-22; -8; 8$) and insula ($-31; 21; 9$). Subsequently, we removed the influence of extraneous predictors (visual effects, win/loss predictors, decision and motor related effects, and a constant, all constructed as explained in the original study) using a linear regression. We thus obtained three adjusted time courses, which we denoted with y_i before. We then applied CCA to these y_i . Thus, we computed the weights on the adjusted activations y_i s and on the predictors x_i that provided maximum correlation.

We implemented two versions: heterogeneous weights (where the weights varied across subjects) and homogenous weights (where the weights were fixed across subjects).

Finally, we ran random-effects GLM on activation in a large area of the brain that we hypothesized to include potential projection areas of the combined signal inferred through CCA. In this post-CCA GLM, we included the combined signal from the CCA step (and an additional regressor to mitigate errors-in-variables in this predictor, as explained in the Materials and methods section) as well as extraneous predictors.

CCA results

With heterogeneous weights, CCA generated identical (taken as positive) weights on both ER and risk for 14 out of 19 subjects (Table S1). The ER and risk predictors are orthogonal. Consequently, under the null hypothesis that there is no correlation between some combination of ER and risk and some combination of activation in the three regions of interest, the probability of obtaining two positive weights on a single subject is 0.5 (see Appendix C). A binomial test then predicts that the probability of finding two positive weights for 14

Table 1
Results of the CCA computation

Overall p -value $< 10^{-7}$					
Weight	p -value	Predictor/adjusted time course	Talairach		
			x	y	z
$a_1 = 32$	$< 10^{-7}$	ER			
$a_2 = 65$	$< 10^{-6}$	Risk			
$b_1 = 0.24$	< 0.01	Putamen	-22	-8	8
$b_2 = 0.45$	$< 10^{-7}$	Ventral striatum	-12	5	-3
$b_3 = -0.22$	< 0.0002	Insula	-31	21	9

CCA weights for predictors and adjusted time courses, homogenous weights across 19 subjects. The overall significance, computed with Wilks' lambda (Anderson, 2003), is $p < 10^{-7}$; each predictor and each adjusted time course are separately significant (see Appendix B for the computation of p -values).

out of 19 subjects under the null hypothesis is less than 0.032. Hence, we reject the null of no correlation, in favor of the alternative of significant correlation with a combined signal where both ER and risk have positive weights.

With the evidence of positive weights on ER and risk, we reran CCA with homogenous weights across subjects (see Table 1). We found the correlation between the combination of the two predictors and the joint activation in the regions of interest to be highly significant ($p < 10^{-7}$). Moreover, we estimated that the following combined signal: 32 ER + 65 risk. We found that the weights on the brain activations were all separately significant ($p < 0.01$); this confirmed that all regions of interest contributed to the signal combination.

Identifying a potential projection region

In order to localize the downstream area where the combined signal of ER and risk may be projected, we performed a random-effects GLM analysis with the combined signal of ER and risk as one

of the predictors. We used the version of this combined signal obtained from CCA, with weights estimated separately for each subject (Table S1). We only used the 13 subjects for whom the weights on ER and risk were of equal sign and significant (based on a p -value for Wilks' lambda for single comparison and single subject < 0.10 ; high threshold reflects only the quality of the signal and is only used to identify subjects to be included in the downstream area activation analysis. We focused on medial prefrontal cortex (mPFC).

mPFC was identified as potential projection area because our combined signal was a metric that integrated ER and risk, and as such, reflected a value index that had been shown in at least one prior study (Tobler et al., 2007) to correlate with mPFC activation. When subjects are risk tolerant, this value index coincides with the expected utility index of economic theory (Von Neumann and Morgenstern, 1944).

Metrics that are increasing in their components have also been proposed in the game theory literature to evaluate the amount of competition that exists between two antagonists (Esteban and Ray, 1999). Thus, the metric that we identified through our application of CCA could alternatively be interpreted as a conflict metric. This interpretation would suggest that the brain values gamble in terms of conflict between their salient features (in this case, expected reward and risk).

Consistent with our conjecture, the post-CCA GLM revealed significant activation to the combined predictor in the mPFC (Fig. 2, cluster of 11 voxels ($3 \times 3 \times 3 \text{ mm}^3$) around the center (1, 51, -3) in Talairach coordinates, threshold at $p(\text{uncorrected}) < 0.001$). The mPFC did not significantly activate to the second predictor or to ER or risk.

Robustness checks

We found qualitatively similar results (see Supplementary data) with fixed CCA weights across subjects and with CCA weights that were estimated after removal of the serial correlation in the ad-

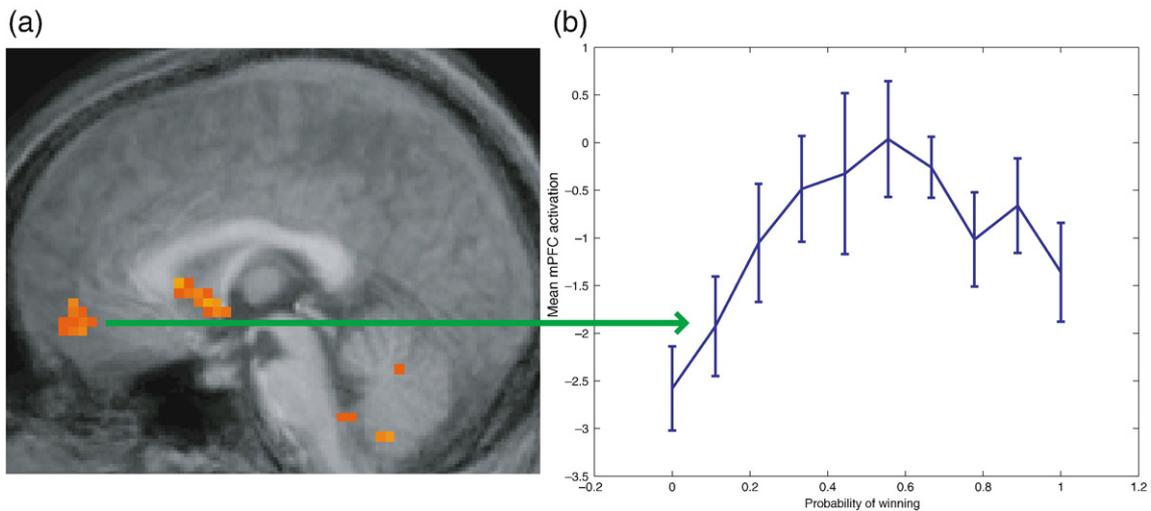


Fig. 2. Activation of the downstream area. (a) Activation that correlates significantly with the combined signal of ER and risk inferred from joint activation of putamen, ventral striatum and insula. The area extends for 11 voxels ($3 \times 3 \times 3 \text{ mm}^3$) around the center (1, 51, -3) in Talairach coordinates (threshold at $p(\text{uncorrected}) < 0.001$). (b) Plot of the mean activation of the region depicted in (a) against the probability of winning. The mPFC encodes both ER and risk with positive sign. The inverted U shape shows that the area encodes risk positively. However, for the activation to solely encode risk, it ought to be symmetrical around $p = 0.5$, while it is not: it is higher for corresponding probabilities of reward above 0.5. The asymmetry reflects the additional influence on activation of ER (which increases monotonically in reward probability). Hence, the activation pattern reflects a weighted sum of ER and risk, as if encoding the utility of the gamble (for a risk-tolerant agent).

justed time courses by means of auto-regressive models (Burock and Dale, 2000). Finally, to check that our findings were not an artifact of correlation induced by the recording system, we removed the first principal component of the adjusted time courses (computed with singular value decomposition) before performing CCA and obtained similar results as well.

Discussion

We introduced here a new method to analyze imaging data based on the hypothesis (Salinas and Sejnowski, 2001), later confirmed (Romo and Schultz, 1990), that neurons exhibit correlated activation in order to direct the integration of signals with minimal noise. We used CCA to reverse engineer this process and to discover how the brain integrated signals. By adding the uncovered combination as predictor in a post-CCA GLM, we identified projection regions and thus verified that the combination of signals was indeed represented in some downstream region.

As an illustration of this approach, we analyzed fMRI data from an experiment involving a simple card game. Prior results indicated that expected reward (ER) and reward variance (risk) were separately encoded in the brain. Our application of CCA indicated that the correlation of the corresponding signals directed the addition of ER and risk to form an integrated signal that activated a third region, namely, mPFC.

Our finding that mPFC activation correlates with the specific combination of ER and risk reflected in joint activation of striatal areas and insula extends our understanding of the role of this cortical region in evaluation of gambles. In a prior study (Tobler et al., 2007), this region had been shown to correlate with both ER and risk, thus establishing its role in overall valuation. In a similar, purely imperative, task, we here show that mPFC correlates with a metric of ER and risk that we extracted from joint activation in ventral striatum, putamen and insula, suggesting that the origin of the value signal in mPFC lies in signal combination of these upstream regions.

In economic theory, value signals presuppose choice. In our paradigm, however, there is no choice in or after the epoch during which we study brain activation. Still, standard economic valuation signals in the brain have been observed before in purely imperative trials (Tobler et al., 2007). Moreover, valuation signals from imperative trials are relevant for choice analysis. In fact, they can provide better predictors of choice in subsequent free-choice test trials than choices in free-choice training trials themselves (Berns et al., *in press*).

CCA complements existing fMRI analysis approaches. Methods such as PCA, ICA, and CVA are exploratory, and like our application of CCA, they allow us to reveal functional connections in the brain. With DCM, SEM, and PPI, nonlinear connectivity can be investigated, unlike with CCA. CCA, however, is based on a known biological mechanism, namely, correlation of activation. In addition, CCA improves on both hypothesis-driven GLM analyses and data-driven techniques such as PCA.

Specifically, our application of CCA generalizes GLMs: instead of investigating the relationship between several predictors and a single voxel, the CCA approach allows one to model the relationship between several predictors and several voxels. At the same time, our approach is a form of “guided” PCA: predictors are included in the analysis in order to obtain meaningful results.

In this study, we only explored implications of across-region correlation with the purpose of improving signal integration. Correlation may play many other roles, both within and between neuronal

populations (Averbeck et al., 2006). Still, even though the wider role of correlation in the brain is being debated, various independent observations support our hypothesis that correlation may direct integration. First, correlation goes hand-in-hand with performance (Hummel and Gerloff, 2004; Stopfer et al., 1997), e.g., reports that correlations are indispensable for efficient sensory perception by honeybees. Second, memory is impaired in rats when correlation is disrupted (Robbe et al., 2006). Third, in normal humans, face perception appears to induce correlations that can be recorded with EEG, but when this correlation is disrupted, the integration of signals no longer works optimally and performance decreases (Rodriguez et al., 1999). Finally, the fact that lower attention level down-modulates correlation (Steinmetz et al., 2000) is consistent with our hypothesis that correlation improves signal integration, and hence, task performance.

Our use of CCA has limitations. First, it cannot exploit correlations within a population. Theoretical boundaries limit the role of intra-population correlation to a group of about 100 neurons (Abbott and Dayan, 1999; Schneidman et al., 2006; Shadlen and Newsome, 1994), too small to be recorded and analyzed with fMRI. As such, we cannot use CCA to investigate, e.g., the role of correlation to improve population coding. Second, CCA imposes a linear model: CCA will discover signal integration only if the downstream area receives a weighted sum of the inputs from the upstream regions and, as with GLMs, the HRF function links the neuronal level to the fMRI amplitude in a linear fashion. Third, because it is based on correlation, CCA can only suggest connectivity, and not prove it. This is, of course, a common drawback of correlation-based approaches (such as GLM). Finally, the brain may use other biological mechanisms besides correlation in activation to direct integration. Our method would fail if this is the case.

The application of our CCA-based procedure to a neuroeconomic experiment confirms its potential to study signal integration. Signal integration is a general problem that the brain confronts, and as such, it is hoped that our method can be successfully applied to problems beyond the present experiment.

Acknowledgments

This work was supported by the NSF Grant SES-0527491, the Swiss Finance Institute and the NCCR FINRISK grant of the Swiss National Science Foundation. We thank John O’Doherty for helpful comments. We thank Steven Youra for help with the redaction.

Appendix A. Effect of the HRF

In this appendix, we show that, if neuronal activities are directly related to the amplitude recorded with fMRI, then correlations of the neuronal activities translate to correlations of fMRI time courses. We do not make a distinction between the correlation in the input, as measured by local field potentials (LFP) (Logothetis et al., 2001), or in the output, as measured by firing rate (Mukamel et al., 2005). In our framework, these correlations are present in both input and output. Moreover, because we average signals over large region of interest, the time courses we record reflect the intracortical processing rather than either the input or the output.

Let $z_1(n)$ and $z_2(n)$ be two signals of average neuronal activity in two regions of interest (where the subscript indexes the region of interest while n denotes time). Time courses in fMRI do not reflect $z_1(n)$

and $z_2(n)$ but represent a version smeared by a hemodynamic response function (HRF) $h(\cdot)$. Given the observations of Mukamel et al., the fMRI signals $y_r(n)$ relate to the corresponding average neuronal activations $z_r(n)$ ($r=1,2$) as follows: $y_r(n) = \sum_i z_r(i)h(n-i)$ (standard convolution).

Without loss of generality, let $z_1(n)$ and $z_2(n)$ be stationary and have zero mean. Our approach focuses on their stationary correlation at times m and n , $\rho_{\text{neuron}}(m,n)$, defined by:

$$\rho_{\text{neuron}}(m,n) = \frac{E[z_1(n)z_2(m)]}{\sqrt{E[z_1(n)z_1(n)]}\sqrt{E[z_2(m)z_2(m)]}}.$$

The issue is: can we recover this correlation with fMRI data recording?

For simplicity, we suppose that this correlation exists only at simultaneous time samples. This hypothesis is not as restrictive as it seems; in case there is some temporal dependency (Salinas and Sejnowski, 2002), one can incorporate it by replacing $h(n)$ with $h(n)$ convolved with a function that models this temporal dependency. That is,

$$\rho_{\text{neuron}}(m,n) = \rho_0 \delta(n-m);$$

where

$$\delta(n-m) = \begin{cases} 1 & \text{if } n=m \\ 0 & \text{otherwise} \end{cases}.$$

Let $\rho_{\text{fMRI}}(m,n)$ be the correlation between observations at times m and n of the corresponding fMRI signals for the two regions of interest, $y_1(n)$, $y_2(n)$. Specifically:

$$\rho_{\text{fMRI}}(m,n) = \frac{E[y_1(n)y_2(m)]}{\sqrt{E[y_1(n)y_1(n)]}\sqrt{E[y_2(m)y_2(m)]}}.$$

We show now that the correlation at the neuronal level, namely, $\rho_{\text{neuron}}(m,n)$, is equal to the correlation readily computed from the fMRI signals, namely, $\rho_{\text{fMRI}}(m,n)$. More specifically, we shall show that $\rho_{\text{fMRI}}(m,n) = \rho_{\text{neuron}}(m,n)$ for $m=n$. First, express $\rho_{\text{fMRI}}(m,n)$ in terms of average neuronal activations:

$$\rho_{\text{fMRI}}(m,n) = \frac{E\left[\left(\sum_i z_1(i)h(n-i)\right)\left(\sum_j z_2(j)h(m-j)\right)\right]}{\sqrt{E\left[\left(\sum_i z_1(i)h(n-i)\right)^2\right]}\sqrt{E\left[\left(\sum_j z_2(j)h(m-j)\right)^2\right]}}.$$

For $m=n$, and by using the fact that $E[z_1(i)z_2(j)] = 0$ if $i \neq j$ we have:

$$\begin{aligned} E\left[\left(\sum_i z_1(i)h(n-i)\right)\left(\sum_j z_2(j)h(n-j)\right)\right] \\ = \sum_i \sum_j h(n-i)h(n-j)E[z_1(i)z_2(j)] \\ = E[z_1(n)z_2(n)] \sum_k h(k)^2 \end{aligned}$$

Similarly:

$$\begin{aligned} E\left[\left(\sum_i z_1(i)h(n-i)\right)^2\right] &= \sum_i \sum_j h(n-i)h(n-j)E[z_1(i)z_1(j)] = E[z_1(n)z_1(n)] \sum_k h(k)^2. \\ E\left[\left(\sum_j z_2(j)h(n-j)\right)^2\right] &= \sum_i \sum_j h(n-i)h(n-j)E[z_2(i)z_2(j)] = E[z_2(n)z_2(n)] \sum_k h(k)^2, \end{aligned}$$

So,

$$\rho_{\text{fMRI}}(n,n) = \frac{E[z_1(n)z_2(n)] \sum_k h(k)^2}{\sqrt{E[z_1(n)z_1(n)] \sum_k h(k)^2} \sqrt{E[z_2(n)z_2(n)] \sum_k h(k)^2}},$$

$$\rho_{\text{fMRI}}(n,n) = \frac{E[z_1(n)z_2(n)]}{\sqrt{E[z_1(n)z_1(n)]}\sqrt{E[z_2(n)z_2(n)]}} = \rho_0.$$

Thus, by computing the correlation at the fMRI level, we can recover the correlation between average neuronal activity:

$$\rho_{\text{fMRI}}(n,n) = \rho_{\text{neuron}}(n,n).$$

Appendix B. CCA and inference

We show here how to obtain p -values for the weight of the CCA. Although the Wilks' lambda method (Anderson, 2003; Johnson and Wichern, 2002) computes a p -value for each row of the CCA (a row is defined as a set of weights for a solution to CCA), to the best of our knowledge, no exact method exists to compute p -values for the weights separately. Moreover, approximation with resampling methods such as the Bootstrap (Efron and Tibshirani, 1993) is impractical because of the sign ambiguity when computing an eigenvector. Instead, we present a method for computing approximate p -values that is based on equivalence between CCA and linear regression.

Let \mathbf{X} be a $(T;n)$ matrix containing T time samples of n predictors and \mathbf{Y} be a $(T;p)$ matrix of T time samples of the time course signal of p regions of interest in the brain. Let Σ_{11} and Σ_{22} be the sample covariance matrices of \mathbf{X} and \mathbf{Y} , respectively and $\Sigma_{12} = \Sigma_{21}$ the covariance matrix between \mathbf{X} and \mathbf{Y} . For simplicity, all vectors have zero mean.

Canonical correlation finds the linear combinations of the column vectors of \mathbf{X} and \mathbf{Y} that maximize their correlation, i.e., we look for vectors \mathbf{a} and \mathbf{b} that give the largest value of:

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

Following the derivations of Johnson and Wichern (2002), we perform a change-of-basis:

$$\mathbf{c} = \Sigma_{11}^{-1/2}\mathbf{a},$$

$$\mathbf{d} = \Sigma_{22}^{-1/2}\mathbf{b};$$

which produces:

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

By the Cauchy–Schwarz inequality:

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

Equality obtains when $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{c}$ and \mathbf{d} are collinear. The rightmost expression of the equation above is the square root of a Rayleigh quotient and is maximized when \mathbf{c} is the eigenvector corresponding to the largest eigenvalue of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$. Wilks' lambda approximation tests the significance of the correlation

(Anderson, 2003). The computed canonical variables are two ($T:1$) vectors:

$$U_1 = \mathbf{X}\mathbf{a},$$

$$U_2 = \mathbf{Y}\mathbf{b}.$$

Closer inspection reveals that the vector \mathbf{b} can be recovered using a linear regression of U_1 onto \mathbf{Y} . Indeed, the ordinary least mean square regression (Johnson and Wichern, 2002) gives:

$$\tilde{\mathbf{b}} = (\mathbf{Y}'\mathbf{Y})^{-1}(\mathbf{Y}'U_1) = (\mathbf{Y}'\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{X})\mathbf{a}.$$

By substituting sample covariance matrices, we get:

$$\tilde{\mathbf{b}} = \Sigma_{22}^{-1}\Sigma_{21}\mathbf{a}, \quad \tilde{\mathbf{b}} = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{c},$$

$$\tilde{\mathbf{b}} = \Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{c}, \quad \tilde{\mathbf{b}} \propto \Sigma_{22}^{-1/2}\mathbf{d}, \quad \tilde{\mathbf{b}} \propto \mathbf{b}.$$

We can use the above result to obtain approximate p -values for the elements of \mathbf{b} . Since for a given \mathbf{a} , a GLM and CCA are equivalent, we can apply the standard test for GLM (Johnson and Wichern, 2002; Weisberg, 2005). It is simple to extend this to the case where the eigenvalue is not the one of maximum magnitude (i.e., in testing the significance of weights in the other rows of the CCA). By symmetry, this method applies to the p -values of \mathbf{a} as well.

In standard tests, the regression matrix \mathbf{Y} is non-random. Hence, we can only infer approximate p -values for \mathbf{a} . The same problem emerges when we use \mathbf{X} to compute p -values for \mathbf{b} . Our approach is useful nevertheless for the analysis of the results of CCA.

Appendix C. Distribution of sign of weights on predictors under the null hypothesis

In this appendix, we follow the same notation as in Appendix B. We restrict ourselves to the case of two predictors and two adjusted time courses. We also restrict ourselves to the design matrices of the experiment in the empirical application, where the predictors were orthogonal. That is, we have $\Sigma_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; for simplicity, we normalize so that the variance of the predictors equals one. Ignoring sampling error, we also have: $\Sigma_{22} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where ρ is the correlation of the \tilde{y}_i 's. Finally, simple computations produce: $\Sigma_{21} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Following the derivations of Appendix B, $\mathbf{c} = \Sigma_{11}^{1/2}\mathbf{a}$ is the eigenvector corresponding to the largest eigenvalue of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$. With the above restrictions, this simplifies to a being an eigenvector of $\frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$.

In case $\rho > 0$, the largest eigenvalue is $1 + \rho/1 - \rho^2$ and the corresponding eigenvector is $[1 - 1]'$. Incidentally, the signs of the elements in this vector are as predicted: under positive correlation between the \tilde{y}_i 's, CCA recovers a signal combination with opposite weights; this reverse engineers the hypothesized brain process: in order to minimize the effect of the \tilde{y}_i when it subtracts two signals, the brain needs the \tilde{y}_i 's to be positively correlated between the \tilde{y}_i . Conversely, if $\rho < 0$, the eigenvalue and eigenvector are $1 - \rho/1 - \rho^2$ and $[1 1]'$. If no correlation exists (null hypothesis; $\rho = 0$), sampling error will nevertheless tilt the results toward either $\mathbf{a} = [1 1]'$ or $\mathbf{a} = [1 - 1]'$; tilting in either direction will happen with a probability of 0.5. This observation provides the basis for hypothesis testing: across subjects, the tilting follows a binomial distribution with mean 0.5. This implies that we expect the weights on the predictors to have the same sign for roughly half the subjects and the opposite sign for the other half.

In the more general case where the predictors are not normalized to 1, CCA will always find a significant correlation, effectively recovering the predictor/region of interest pair that had the highest correlation in the standard region-by-region GLM with which the regions of interest were determined in the first place. The signs of the weights on the other pair will be random. This again implies that we expect the weights on the two predictors to have the same sign for roughly half the subjects and the opposite sign for the other half.

Appendix D. The power of the post-CCA GLM

Our post-CCA GLM has higher statistical power than a standard GLM where the original predictors x_1 and x_2 enter separately. The linear combination of the predictors recovered by CCA, $U_j = \sum_i a_{ij}x_i$, are orthogonal with each other (Anderson, 2003). (For simplicity, we used $a_{i1} = a_i$ in the main text.) We focus on U_1 and U_2 , the combinations corresponding to the first two CCA solutions (i.e., corresponding to the two highest correlations). As depicted in Fig. A1, when we replace the original pair of predictors x_1 and x_2 with U_1 and U_2 , the projection of the signal of interest y onto U_1 has a higher magnitude (k_{CCA}) than the any of the two original projections (a_{11}^* or a_{21}^* onto x_1 or x_2 , respectively). In the best-case scenario (when $\theta = \pi/4$), the coefficient k_{CCA} is $\sqrt{2}$ times higher.

Appendix E. Errors in variables

E.1. Problem

To improve the power of the statistical tests in the post-CCA GLM, we use a method similar to instrumental variable estimation (Spanos, 1986). Specifically, we search for a predictor that is the best

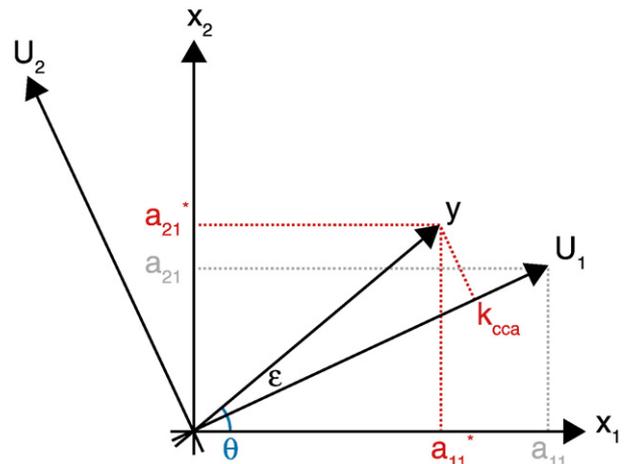


Fig. A1. Why CCA is useful. Graphical representation of the third step of our method (see red section of Fig. 1). Instead of using a GLM to project the activity of a downstream area, y , onto the original basis (given by the variables x_1 and x_2), we use the basis suggested by the CCA (U_1 and U_2). Since the recordings are noisy, the predictor U_1 and the activation y are not exactly collinear but the angle between them is small ($\epsilon \approx 0$). Using the new basis, the projection of y onto U_1 (k_{CCA}) is larger than either of the two original projections (with coefficient a_{11}^* for x_1 and a_{21}^* for x_2), and hence it is less sensitive to noise. To mitigate the effect of misestimating ($\epsilon \neq 0$), we add a predictor. For optimality reasons (Appendix D), we choose the second CCA solution, namely, U_2 , which is orthogonal to U_1 .

choice to mitigate errors-in-variables in the predictor of the post-CCA GLM. In the following, we show that U_2 , the signal combination corresponding to the second CCA solution, is optimal.

The weights estimated through CCA are noisy values of the true weights (Fig. A1). The activity of the downstream region, represented by the vector y , is a linear combination of two, here orthogonal, predictors, represented by the vectors x_1 and x_2 , i.e., $y = a_{11}^* x_1 + a_{21}^* x_2$. For simplicity, we normalize predictors to unit length and use polar coordinates: $a_{11}^* = \cos(\theta)$ and $a_{21}^* = \sin(\theta)$. CCA only provides estimates of the true coordinates a_{11}^* and a_{21}^* to yield the predictor $U_1 = a_{11} x_1 + a_{21} x_2$, which is only approximate. We continue to use polar coordinates ($a_{11} = \cos(\theta + \varepsilon)$ and $a_{21} = \sin(\theta + \varepsilon)$) and we assume that the error from the CCA is small ($\varepsilon \approx 0$).

We first derive an expression for the coefficient in the post-CCA GLM when $U_2 = a_{12} x_1 + a_{22} x_2 = a_{21} x_1 - a_{11} x_2$ is added as a predictor. Let k_{CCA} denote this coefficient (see also Fig. A1). Mathematically, it is the first element of the following vector:

$$(U'U)^{-1}(U'y) = \begin{bmatrix} a_{11} a_{11}^* + a_{21} a_{21}^* \\ a_{11} a_{21}^* - a_{21} a_{11}^* \end{bmatrix},$$

where

$$U = [U_1 \mid U_2] = [a_{11} x_1 + a_{21} x_2 \mid a_{21} x_1 - a_{11} x_2]$$

Switching to polar coordinates and applying a Taylor series expansion, we obtain:

$$\begin{aligned} k_{CCA} &= a_{11} a_{11}^* + a_{21} a_{21}^* = \cos(\theta + \varepsilon) \cdot \cos(\theta) + \sin(\theta + \varepsilon) \cdot \sin(\theta) \\ &= \cos(\varepsilon) = 1 - \frac{1}{2} \varepsilon^2 + O(\varepsilon^3) \end{aligned}$$

Now compare this expression to the one that obtains if any other variable in x_1 and x_2 were to be used as additional predictor. Without loss of generality, we could just choose to use x_2 , in which case the design matrix becomes:

$$U = [U_1 \mid x_2] = [a_{11} x_1 + a_{21} x_2 \mid x_2].$$

With analogous computations as before, the coefficient of the GLM becomes:

$$k_{CCA} = \frac{a_{11}^*}{a_{11}} = \frac{\cos(\theta)}{\cos(\theta + \varepsilon)} = 1 + \frac{\sin(\theta)}{\cos(\theta)} \varepsilon + O(\varepsilon^2).$$

A comparison of the two expressions reveals that the addition of U_2 is better because the influence of the estimation error is of higher order (2nd instead of 1st), and hence, smaller. Intuitively, this result holds because, while U_1 and U_2 are uncorrelated, they are both affected by the same noise.

Inspection of the expression for k_{CCA} also reveals that the choice of U_2 as additional predictor avoids another problem. Indeed, when $\theta \approx \pi/2$, the choice of x_2 as additional predictor makes the estimator k_{CCA} numerically unstable.

Appendix F. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [10.1016/j.neuroimage.2008.01.062](https://doi.org/10.1016/j.neuroimage.2008.01.062).

References

- Abbott, L.F., Dayan, P., 1999. The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11, 91–101.
- Anderson, T.W., 2003. *An Introduction to Multivariate Statistical Analysis*, 3rd ed. John Wiley & Sons, Hoboken, NJ.
- Averbeck, B.B., Latham, P.E., Pouget, A., 2006. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366.
- Berns, G.S., Capra, M., Chappelow, J., Moore, S., C.N., in press. Nonlinear neurobiological probability weighting functions for aversive outcomes. *NeuroImage*.
- Burock, M.A., Dale, A.M., 2000. Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. *Hum. Brain Mapp.* 11, 249–260.
- Buxton, R.B., Wong, E.C., 1998. Dynamic of blood flow and oxygenation changes during brain activations: the balloon model. *Magn. reson. med.* 39, 855–864.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman and Hall, Toronto, ON.
- Esteban, J., Ray, D., 1999. Conflict and distribution. *J. Econ. Theory* 87, 379–415.
- Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Price, C., Zeki, S., Ashburner, J., Penny, W., 2003. *Human Brain Function*, 2nd ed.
- Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application with fMRI. *Neuroimage* 16, 513–530.
- Friston, K.J., Frith, C.D., Frackowiak, R.S.J., Turner, R., 1995. Characterizing dynamic brain responses with fMRI: a multivariate approach. *Neuroimage* 2, 166–172.
- Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6, 218–229.
- Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the balloon model, Volterra kernels, and other hemodynamics. *Neuroimage* 12, 466–477.
- Goncalves, M.S., Hull, D.A., 2003. Connectivity analysis with structural equation modelling: an example of the effects of voxel selection. *Neuroimage* 20, 1455–1467.
- Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral models. *Econometrica* 37, 424–438.
- Hotelling, H., 1936. Relations between two sets of variants. *Biometrika* 28, 312–377.
- Hummel, F., Gerloff, C., 2004. Larger interregional synchrony is associated with greater behavioral success in a complex sensory integration task in humans. *Cereb. Cortex* 15, 670–678.
- Johnson, R.A., Wichern, D.W., 2002. *Applied Multivariate Statistical Analysis*, 5th ed. Prentice Hall, Upper Saddle River, NJ.
- Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157.
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.-P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188.
- Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., Malach, R., 2005. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309, 951–954.
- Penny, W.D., Stephan, K.E., Mechello, A., Friston, K.J., 2004. Modelling functional integration: a comparison of structural equation and dynamic causal models. *Neuroimage* 23, S264–S274.
- Preusschoff, K., Bossaerts, P., Quartz, S., 2006. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51, 381–390.
- Robbe, D., Montgomery, S.M., Thome, A., Rueda-Orozco, P.E., McNaughton, B.L., Buzsaki, G., 2006. Cannabinoids reveal importance of spike coordination in hippocampal function. *Nat. Neurosci.* 9, 1526–1533.

- Rodriguez, E., George, N., Lachaux, J.P., Martinerie, J., Renault, B., Varela, F.J., 1999. Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397, 430–433.
- Romo, R., Schultz, W., 1990. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* 63, 592–606.
- Romo, R., Hernandez, A., Zainos, A., Salinas, E., 2003. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38, 649–657.
- Salinas, E., Sejnowski, T.J., 2000. Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.* 20, 6193–6209.
- Salinas, E., Sejnowski, T.J., 2001. Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539.
- Salinas, E., Sejnowski, T.J., 2002. Integrate-and-fire neurons driven by correlated stochastic input. *Neural Comput.* 14, 2111–2155.
- Schneidman, E., Berry, M.J., Segev, R., Bialek, W., 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012.
- Shadlen, M.N., Movshon, J.A., 1999. Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24, 67–77.
- Shadlen, M.N., Newsome, W.T., 1994. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* 4, 569–579.
- Spanos, A., 1986. *Statistical Foundations of Econometric Modeling*. Cambridge University Press, Cambridge, UK.
- Steinmetz, P.N., Roy, A., Fitzgerald, P.J., Hsiao, S.S., Johnson, K.O., Niebur, E., 2000. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* 404, 187–190.
- Stopfer, M., Bhagavan, S., Smith, B.H., Laurent, G., 1997. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* 390, 70–74.
- Tobler, P.N., O'Doherty, J.P., Dolan, R.J., Schultz, W., 2007. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* 97, 1621–1632.
- Von Neumann, J., Morgenstern, O., 1944. *Theory of Games and Economic Behavior*. Princeton university press, Princeton.
- Weisberg, S., 2005. *Applied Linear Regression*, 3rd ed. John Wiley & Sons, New York, NY.