

Characterizing Modulatory Interactions Between Areas V1 and V2 in Human Cortex: A New Treatment of Functional MRI Data

K.J. Friston, L.G. Ungerleider, P. Jezzard, and R. Turner

Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom (K.J.F.); Laboratory of Neuropsychology, NIMH (L.G.U.), and Laboratory of Cardiac Energetics, NHLBI (P.J., R.T.), NIH, Bethesda Maryland 20892

Abstract: Reversible cooling experiments in monkey visual cortex have demonstrated that visually driven neuronal activity in V2 depends on feedforward projections from V1, whereas neuronal activity in V1 is modulated by feedback, or reentrant, projections from V2. We present evidence for a homologous asymmetry in reciprocal connections between V1 and V2 in human cortex using physiological measurements obtained with functional MRI. The analysis was based on a nonlinear model of effective connectivity that partitioned the influence that one region exerted over another into an obligatory effect (an effect that depended only on the input) and a modulatory effect (an effect that represented an interaction between input and activity intrinsic to the target region). Using estimates of the modulatory effect we tested two related hypotheses: 1) that V2 would be a major source of modulatory influences on V1; and 2) that the modulatory effects of V2 on V1 would be greater than those of V1 on V2. The first constitutes a hypothesis about the regional or topographic organization of (modulatory) effective connectivity and the second hypothesis directly addresses the functional asymmetry suggested by reversible cooling experiments. The results confirmed that the origins of feedback modulatory effects on V1 were regionally specific and most pronounced in V2. In contrast, feedforward modulatory influences of V1 on V2 were negligible. This apparent asymmetry between feedforward and feedback modulatory interactions was evident in both hemispheres and appears to be a fairly robust feature of nonlinear interactions between striate and extrastriate cortex. © 1995 Wiley-Liss, Inc.

Key words: visual cortex, neuroimaging, nonlinear, effective connectivity, functional magnetic resonance imaging, human brain

INTRODUCTION

The feedforward and feedback connections between striate (V1) and extrastriate areas have distin-

Received for publication July 25, 1994; revision accepted February 6, 1995.

Address reprint requests to either Karl J. Friston, c/o: The MRC Cyclotron Unit, Hammersmith Hospital, DuCane Road, London W12 0HS, U.K., or Leslie G. Ungerleider, Laboratory of Neuropsychology, NIMH, Building 49, Room 1B80, Bethesda, MD 20892.

R. Turner is currently at the Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom.

guishing anatomical and functional characteristics. A common feature of these reciprocal connections is an asymmetry in the anatomy of the anterograde and retrograde elements. One kind of asymmetry is found in the laminar distribution of afferent (incoming) fibers and the distribution of their cells of origin. These laminar distributions have been proposed as the basis for a hierarchical classification of cortical areas [Rockland and Pandya, 1979; Maunsell and Van Essen, 1983]. One area is considered "higher-order" to another if the feedforward projections to that area terminate predominantly in layer 4 and the reciprocal,

feedback projections have a complementary laminar distribution, largely sparing layer 4. Another asymmetry relates to the distributions of projecting cells and feedback [reentrant; Edelman, 1978] afferents to those cells, in that these distributions are not coextensive. For example, in the V1-MT/V5 system, the V1 cells projecting to MT and the returning fibers from MT occupy the same layer in V1 but are not coextensive within it: The terminations of projections from MT are more extensive in the horizontal dimension [Zeki, 1990]. The terms feedforward and feedback are used here to designate an anatomical characteristic and do not imply any particular functional attribution.

Evidence for functional asymmetries is found in the work of Schiller and colleagues on physiological interactions between V1 and V2 [Schiller and Malpeli, 1977; Sandell and Schiller, 1982]. V1 provides a crucial, or obligatory, input to V2, in the sense that visual activation of V2 cells depends on input from V1. This dependency has been demonstrated by reversibly cooling (deactivating) V1 while recording from the retinotopically corresponding region of V2 during visual stimulation [Schiller and Malpeli, 1977; Girard and Bullier, 1989]. In contrast, cooling V2 has a more modulatory effect on V1 unit activity. "Most of these [V1 cells] became less responsive to visual stimulation, while a few became more active during cooling" [Sandell and Schiller, 1982]. The cells in V1 that were most affected by V2 deactivation were in the infragranular layers, suggesting that V2 may use this feedback pathway to modulate the output from V1 [Sandell and Schiller, 1982]. Similar conclusions about the return pathway from MT to V2 were drawn by Girard and Bullier [1989]: Because, in the absence of V1 input, these feedback connections do not constitute an efficient drive to V2 cells, their role is most likely "to modulate the information relayed through area 17" (V1).

Functional magnetic resonance imaging (fMRI) provides a potential means of measuring physiological activity in striate and extrastriate cortex [Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992], and we have used this technique to test the hypothesis that nonlinear interactions between V1 and V2 are asymmetric, with a modulatory effect of V2 on V1, but not of V1 on V2. In order to address this question, we viewed the fMRI data as a set of local neurophysiological measurements that result from stimulus-dependent activation of V1 and the (nonlinear) corticocortical interactions that ensue.

We now describe the theory upon which our analysis was based, the fMRI data used, and the results. The experiments to be described represent a demonstra-

tion study, showing how the theory can be applied in practice. Although the results presented herein are compelling, they should be treated as provisional and as serving to illustrate an idea. We have analyzed several data sets and present results from two independent fMRI time-series obtained from the same individual.

THEORY

A nonlinear model of effective connectivity

To examine the interactions between V1 and V2 we used the concept of effective connectivity [Gerstein and Perkel, 1969] defined (in functional neuroimaging) as the influence one neuronal system exerts over another [Friston et al., 1993a,b]. This influence can be modeled in many ways. The simplest model is a linear one:

$$x_i = \sum C_{ij} x_j \quad (1)$$

where x_i and x_j are the physiological activities (expressed as deviations from their mean) in areas or voxels i and j . C_{ij} represents the effective connectivity which scales, in a linear fashion, the influence x_j has on x_i . The above equality can be extended to include a modulatory interaction (for a discussion of this and other extensions, see Friston et al. [1993b]):

$$x_i = \sum (C_{ij}^O x_j + C_{ij}^M x_j x_i) \quad (2)$$

This model has two terms that allow for the activity in area i to be influenced by the activity in area j . The first represents an effect that depends only on afferent input from area j . This is the activity in j scaled by C_{ij}^O . The coefficient C_{ij}^O will be referred to as an *obligatory* connection strength, in the sense that a change in area j results in an obligatory response in area i . By contrast, the second term reflects a more subtle modulatory influence of area j on area i . The coefficient determining the size of this effect (C_{ij}^M) will be referred to as a *modulatory* connection strength, because the overall effect depends on both the afferent input ($C_{ij}^M x_j$) and intrinsic activity (x_i). In the current model the obligatory effect is linear, although in principle both the obligatory and modulatory terms could be nonlinear. The key distinction between an obligatory and modulatory effect is that modulation includes an interaction between afferent input and intrinsic activity (i.e., the modulatory term includes both x_j and x_i). It is important to note that the terms obligatory and modulatory

are strictly functional and could pertain to either feedforward or feedback connections.

The above equation, or model, can be interpreted from two points of view: 1) by analogy with the nonlinear behavior that characterizes voltage-dependent channels in electrophysiology or 2) in terms of classical pharmacological neuromodulation, where postsynaptic responsiveness is modulated without a direct effect on postsynaptic membrane potential. The voltage-dependent analogy is obtained by considering x_i as postsynaptic potential and x_j as a depolarizing current. According to Equation (2), a high C^M reflects a greater sensitivity to changes in input at higher levels of intrinsic activity. In electrophysiological terms, this translates as a change in postsynaptic depolarization in response to a fixed depolarizing current, which increases with increasing depolarization: This is a characteristic of voltage-dependent interactions [Haberly, 1990].

The neuromodulatory interpretation is based on another way of looking at Equation (2), namely, that the responsiveness of i to its obligatory input increases with modulatory input. This can be demonstrated by differentiating x_i with respect to the total obligatory input. Neuromodulation of this sort is commonly attributed to the action of classical neuromodulatory transmitters (e.g. monoaminergic and catecholaminergic neurotransmitters, such as dopamine and 5-HT; see Friston et al. [1992]). A fuller analysis of the relationship between Equation (2) and the phenomena of voltage-dependent interactions and neuromodulation is presented in the discussion.

Estimating the connection strengths

Examples of the relationship described by Equation (2) are illustrated in Figure 1 for different values of C^M , where, for simplicity, we assume a single input. The sensitivity to input is reflected by the slope of the lines in Figure 1. It can be seen that this sensitivity (slope) increases with intrinsic activity x_i . This activity-dependent effect, determined by the value of C^M , provides an intuitive sense of how we estimated C^O and C^M . This estimation involved measuring the difference in sensitivity (slope) between states with high and low intrinsic activity at the location of interest: Imagine one were able to "fix" the activity in V1 at a low level and measure the connectivity between V2 and V1 assuming a simple linear relationship [Eq. (1)]; a particular value for the sensitivity of V1 to V2 changes could be obtained, say C_{low} . Now, if the procedure were repeated with V1 activity fixed at a high level, a second (linear) estimate would be ob-

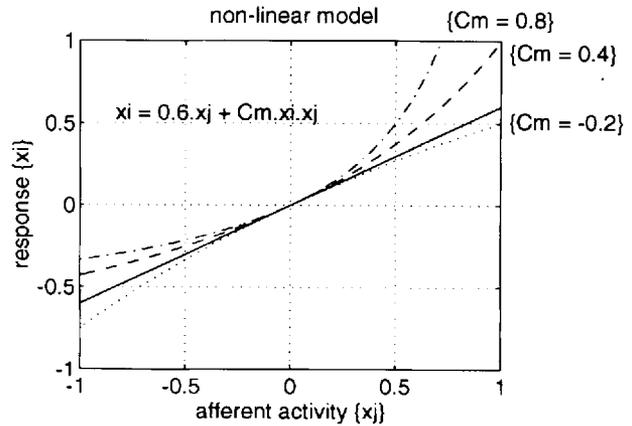


Figure 1.

Graphical illustration of the relationship used to model the nonlinear interaction between an area's response (x_i) and its afferent input (x_j). The value of C^M (the modulatory component of effective connectivity) determines the increased sensitivity to input at higher levels of intrinsic activity. This augmented sensitivity is reflected in the steeper gradient of the curve at higher values of response. --- $C^M = 0.8$, -.- $C^M = 0.4$ and $C^M = -0.2$. The solid line is the linear component ($C^M = 0$).

tained (C_{high}). In the presence of a substantial modulatory interaction between V2 and V1, the second estimate will be higher than the first. This is because the activity intrinsic to V1 is higher for the second estimate and V1 should be more sensitive to inputs from V2. In short, the terms C_{high} and C_{low} provide an estimate of the modulatory influence of V2 on V1 through the first equation of Equation (a2) in the appendix. Similarly, they provide an estimate of the obligatory connection strength through the second equation of Equation (a2) in the appendix (see the appendix for the mathematical details). The activity of V1 can be fixed post hoc by simply selecting a subset of data in which the V1 activity is confined to some small range.

This general approach to characterizing nonlinear systems with a piece-wise series of locally linear models has proved a fruitful strategy in many instances [Tsonis, 1992, p. 221; for a conceptually related approach to multichannel EEG recordings, see Palus et al., 1991]. Although there are direct nonlinear methods, such as the Levenberg-Marquardt method, that could be used to solve Equation (2), the results obtained with that method are very similar to those presented here. Consequently, we report only the method of local linearization as it is aesthetically concise and easier to replicate.

The hypothesis that asymmetrical nonlinear interactions would characterize the interactions between

areas V1 and V2 in human visual cortex can be formulated in terms of C^M : We predicted that 1) the modulatory component of effective connections to V1 would be regionally specific and include V2 ($C_{V1,V2}^M$) would be relatively high compared to other regions, and that 2) the feedforward modulatory influences from V1 to V2 ($C_{V2,V3}^M$) would be smaller than the reciprocal, feedback influences from V2 to V1 ($C_{V1,V2}^M$).

It is important to note that both V1 and V2 are connected with many other visual areas and, accordingly, the present analysis takes this into account. That is, the estimates of modulatory and obligatory effective connectivities (mathematically formalized in the appendix) from a particular reference area (e.g. V2) are obtained in the context of many other possible modulatory and obligatory effects of other areas on the reference area. If the anatomical component of our hypothesis is correct, then the modulatory influences on V1 should be higher from V2 than from elsewhere. Additionally, one should recognize that the method is sensitive to a form of modulatory interaction between the hemodynamics in V2 and V1 which is very specific and would not necessarily be obtained if the response properties in V2 and V1 were simply different in some way (which they almost surely are). This form of interaction is modeled by the second order or nonlinear term in the method. The responsiveness of V1 can be very different from the responsiveness of V2, but if this responsiveness is not dependent on the activity of the other area there will be no modulatory interaction between them.

METHODS

Data

The data were a time-series of 64 gradient-echo EPI coronal slices (5 mm thick, with 64×64 voxels, TE 25 ms, acquisition time 41 ms) through the calcarine fissure (V1) and extrastriate areas. Additional images were acquired in the sagittal plane at an offset of 5 mm left of midline. Images were obtained every 3 s from a normal male subject using a 4.0T whole body MRI system, fitted with a small (27 cm diameter) z-gradient coil. Photoc stimulation (at 16 Hz) was provided by goggles fitted with an array of 16 light emitting diodes (subtending about 24 degrees of visual angle). The stimulation was off for the first 10 scans (30 s), on for the second 10, off for the third, and so on. Images were reconstructed without phase correction. The data were interpolated to 128×128 voxels. Each interpo-

lated voxel thus represented $1.25 \times 1.25 \times 5$ mm of cerebral tissue. Image manipulations and data analyses were performed in MATLAB [MathWorks Inc., Sherborn, MA]. The first four scans were removed to eliminate magnetic saturation effects.

Image preprocessing

The 64 images were translated and rotated to minimize the sum of squares between each of the 64 images and their average (both scaled to the same mean intensity) using the Levenberg-Marquardt method [Press et al., 1988, pp. 542–547]. The rotation and translations were effected using a computationally efficient cubic convolution interpolation method [Keys, 1981]. After the scans were corrected in this fashion for (slight) subject movement, each voxel time-series was corrected to a mean of zero.

ANALYSIS AND RESULTS

Regional specificity of modulatory connections to V1

A reference voxel was chosen in V1, according to the atlas of Talairach and Tournoux [1988], and the effective connection strengths $C_{V1,j}^O$ and $C_{V1,j}^M$ were estimated, as described briefly above and in the appendix. Values for $C_{V1,j}^O$ and $C_{V1,j}^M$ were generated for all voxels, allowing maps of $C_{V1,j}^O$ and $C_{V1,j}^M$ to be constructed. These maps provided a direct test of the first hypothesis concerning the topography and regional specificity of modulatory influences on V1: If the sources of modulatory effects were limited predominantly to V2, then the map of $C_{V1,j}^M$ should highlight V2, and only V2. Figure 2 shows exactly this. The lower figures are maps of $C_{V1,j}^O$ and $C_{V1,j}^M$ and reflect the degree to which the area in question exerts an obligatory (left) or modulatory (right) effect on V1 activity. These maps have been thresholded at 1.64 after normalization to a standard deviation of unity. This value corresponds to an uncorrected threshold of $P = 0.05$. A correction for the many voxels analyzed is not necessary, in this instance, because our hypothesis was anatomically constrained to V2.

The obligatory connections to the reference voxel derive mainly from V1 itself, both ipsilaterally and contralaterally, with a small contribution from contiguous portions of V2. The effective connectivity from contralateral V1 should not be over interpreted given that 1) the source of many afferents to V1 (the lateral geniculate nuclei) were not included in the field of view, and that 2) this finding can be more parsimoniously explained by "common input."

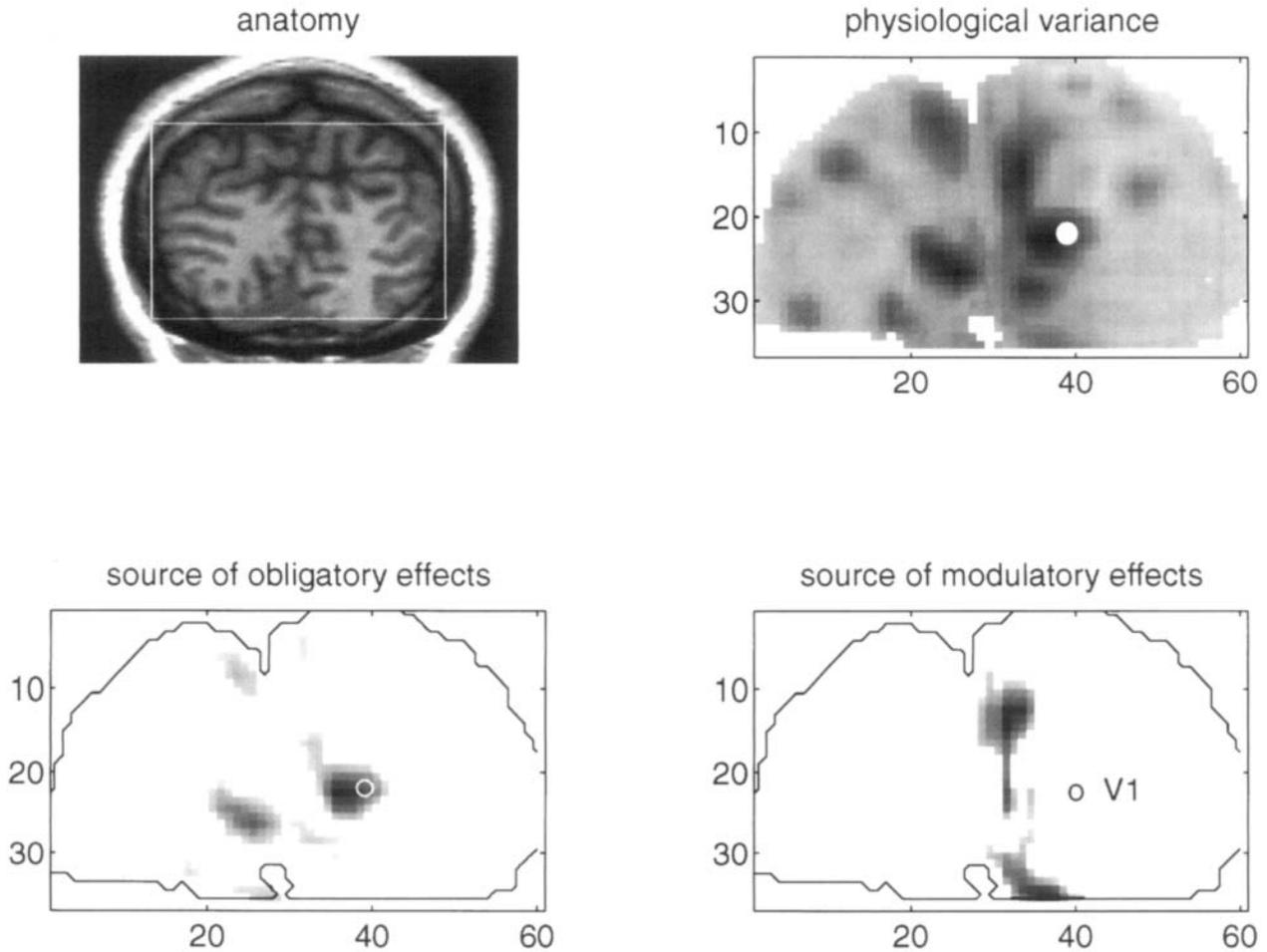


Figure 2.

Maps of the estimates of obligatory and modulatory connection strengths to V1. **Top left:** Anatomical features of the coronal data used. This image is a high resolution anatomical MRI scan (1.5 T) of the subject that corresponds to the fMRI slices. The corresponding coronal section from the atlas of Talairach and Tournoux [1988] is shown in Figure 3A. The box in the anatomical image defines the position of a (36×60 voxel) subpartition of the fMRI time-series selected for analysis. **Top right:** The location of the reference voxel designated as V1 (white dot). This location is

shown on a statistical parametric map [Friston et al., 1991] of physiological variance (calculated for each voxel from the time-series of 60 scans). The image has been scaled to its maximum. **Lower right and left:** Maps of $C_{V1,i}^O$ and $C_{V1,i}^M$. The images have been scaled to unit variance and thresholded at $P = 0.05$ (assuming, under the null hypothesis of no effective connectivity, the estimates have a Gaussian distribution). The reference voxel in V1 is depicted by a circle. Note that V1 is subject to modulatory influences from extensive ipsilateral regions of V2.

As predicted, and with remarkable regional specificity, the modulatory connections with the reference voxel were most marked from ipsilateral V2, dorsal and ventral to the calcarine fissure [Brodmann's area 18, according to the atlas of Talairach and Tournoux, 1988; see Fig. 3A]. Here we infer that the regions involved correspond to V2; however, in the absence of cytoarchitectonic data this inference cannot be substantiated.

To get an idea of the size of the modulatory effect, in biological terms, we compared the sensitivity of the V1

response to obligatory input ($\Sigma C_{ij}^O \cdot x_i$) for the 18 scans with the lowest modulatory input ($\Sigma C_{ij}^M \cdot x_i$) and for the 18 scans with the highest modulatory input. This sensitivity was simply the slope of a plot of V1 activity against obligatory input, as shown in Figure 4. The increase in this slope was about 14%. The difference in slopes or increase in sensitivity due to modulatory effects was statistically significant ($P < 0.001$ $F = 30.75$ $df: 1, 32$). The values for similar analyses (on moving the reference point around in V1) ranged from about 3% to 16%.

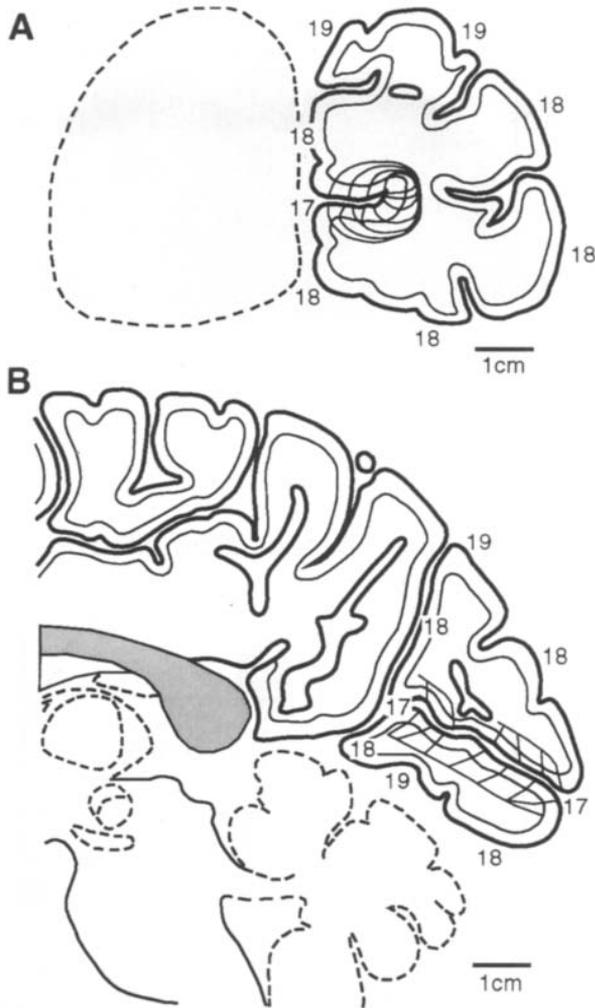


Figure 3.

Coronal (**A**) and sagittal (**B**) sections, adapted from the atlas of Talairach and Tournoux [1988; see their Figures 97 and 44, respectively], corresponding to the fMRI slices selected for analysis. The numbers indicate Brodmann areas. Areas 17 and 18 correspond to V1 and V2, respectively. The thin lines surrounding the cortex of the calcarine fissure indicate the location of the optic radiations.

Asymmetry of functional interactions between V1 and V2

The analysis described in the previous section was repeated for a reference voxel in V2. This revealed obligatory connections from ipsilateral V2 and V1 (Fig. 5, bottom left). In terms of modulatory inputs, those from V1 were strikingly absent (Fig. 5, bottom right). This finding is consistent with our hypothesis. Figures 2 and 5 speak directly to the double dissociation

suggested by reversible cooling experiments in monkey cortex. Feedforward connections from V1 to V2 are, in the context of our model, predominantly obligatory with no evidence of positive modulation. Conversely, feedback connections from V2 to V1 can be characterized as modulatory with relatively less obligatory effects.

To address the functional asymmetry hypothesis more directly, we analyzed all the modulatory connection strengths (or, more accurately, effective connectivities) between two extended regions in ipsilateral V1 and V2 of the left hemisphere. The subset of connections between the two regions was selected to compare the distributions of feedforward and feedback modulatory connection strengths. This was implemented as follows: The values of C_{ij}^M were estimated for every voxel in two (5×5 voxel) squares centered on V1 and V2. The values for each voxel were Euclidean-normalized (sum of squares set to one) over the image. This meant that the estimates reflected a relative regional contribution of modulatory influ-

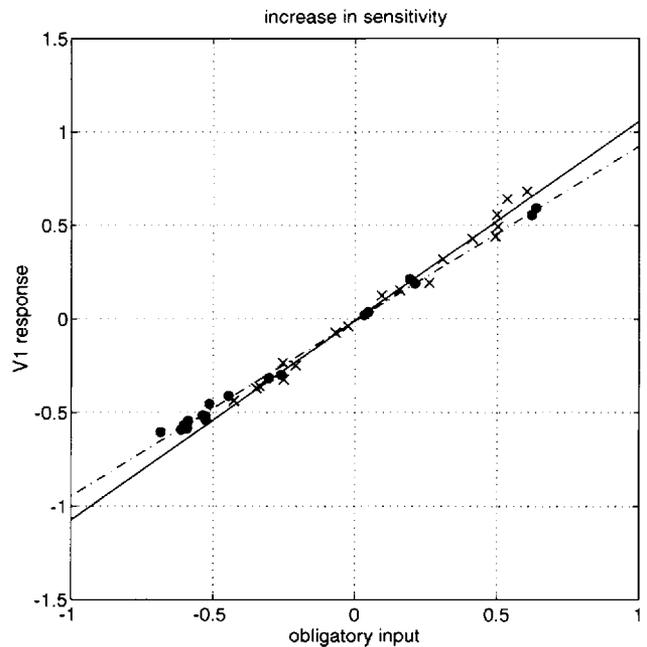


Figure 4.

Regression of the response of a V1 voxel on its obligatory input ($x_i^O = \sum C_{ij}^O \cdot x_j$). The regression has been performed for two subsets of the time-series, corresponding to the 18 scans with minimal modulatory input ($x_i^M = \sum C_{ij}^M \cdot x_j$) (circles) and to the 18 scans with the greatest modulatory input (crosses). Observe that the slope is higher when modulatory input is high. The increase in slope can be interpreted as a modulation-dependent increase in responsiveness or sensitivity. In this example, the 14% increase was statistically significant ($P < 0.001$).

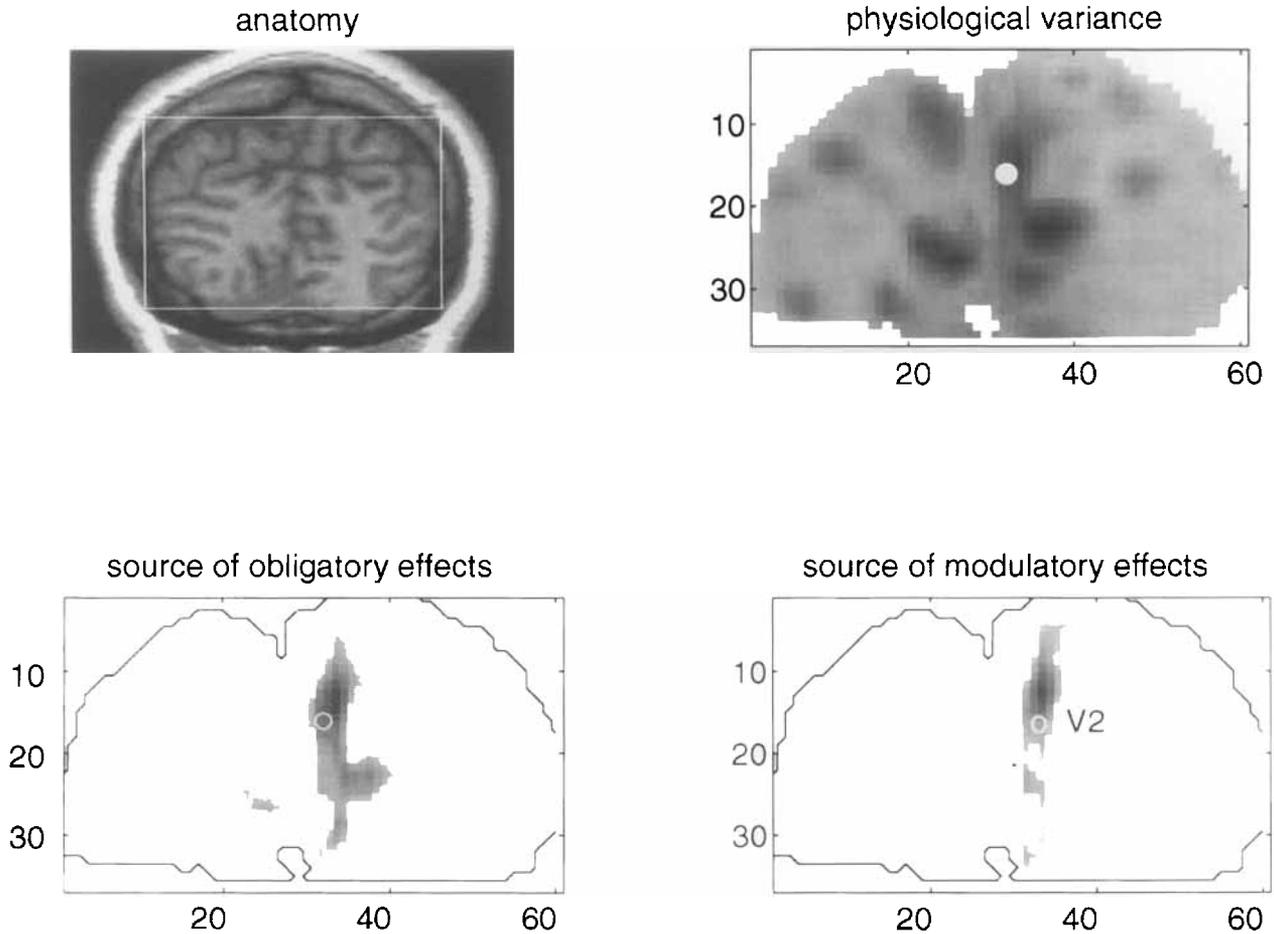


Figure 5.

Maps of the estimates of obligatory and modulatory connection strengths to V2 (Brodmann's area 18; see Figure 3A). Note that although V2 responds in an obligatory way to changes in ipsilateral V2 and V1, it is not subject to modulatory effects from V1. These

findings, taken in conjunction with Figure 2, suggest that V1 is modulated by feedback connections from V2 (in a regionally specific fashion), but this modulation is asymmetric, i.e., it is not reciprocated. All conventions as in Figure 2.

ences. Figure 6 shows the frequency distribution of the 625 estimates for connections from the V1 box to the V2 box (broken line) and the corresponding estimates for connections from V2 to V1 (solid line). There is a remarkable dissociation, with feedback effects (V2 to V1) being much greater than feedforward ones (V1 to V2). This finding can be considered a confirmation of our second hypothesis.

The distributions in Figure 6 suggested that the feedforward modulatory connection strengths are generally negative. It is difficult to comment on the meaning of this unexpected finding. Negative values for the modulatory connection strengths are perfectly permissible and reflect a phenomenon not unlike shunting inhibition, with a reduction in sensitivity at higher levels of activation. However, in subsequent

analyses the distribution of feedforward modulatory connection strengths was centered around zero (see below).

To replicate the asymmetry of feedforward and feedback modulatory effects, an identical analysis was performed for the right hemisphere. If anything, the asymmetry between the distributions was even more striking (Fig. 7). In this instance, the distribution of feedforward modulatory connection strengths was centered on zero, while that corresponding to feedback connections was clearly greater than zero. To ensure that these results were not artifacts of the coronal data set described, we repeated the experiment using the same subject and stimuli, but acquiring data in the sagittal plane (see Fig. 3B). The results are shown in Figure 8 and again point to an unequivocal

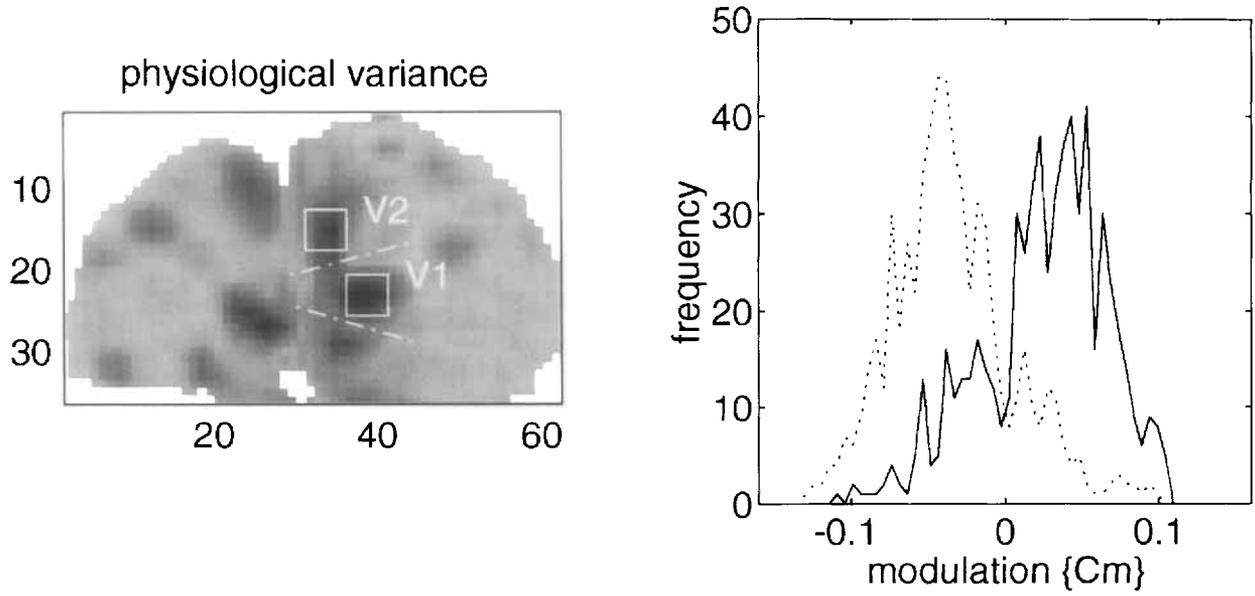


Figure 6.

Graphical presentation of a direct test of the hypothesis concerning the asymmetry between feedforward and feedback V1–V2 interactions. **Left:** Map of physiological variance showing the positions of two boxes defining regions in V1 and V2. The broken lines correspond (roughly) to the position of the V1/V2 border according to the atlas of Talairach and Tournoux [1988; see Figure 3A for

the corresponding coronal section from their atlas]. The value of C_{ij}^M was computed for all voxels in either box and Euclidean normalized to unity over the image. The frequency distribution of C_{ij}^M connecting the two regions is presented on the right. The feedback connections (V2 to V1—solid line) are clearly higher than the corresponding feedforward connections (V1 to V2—broken line).

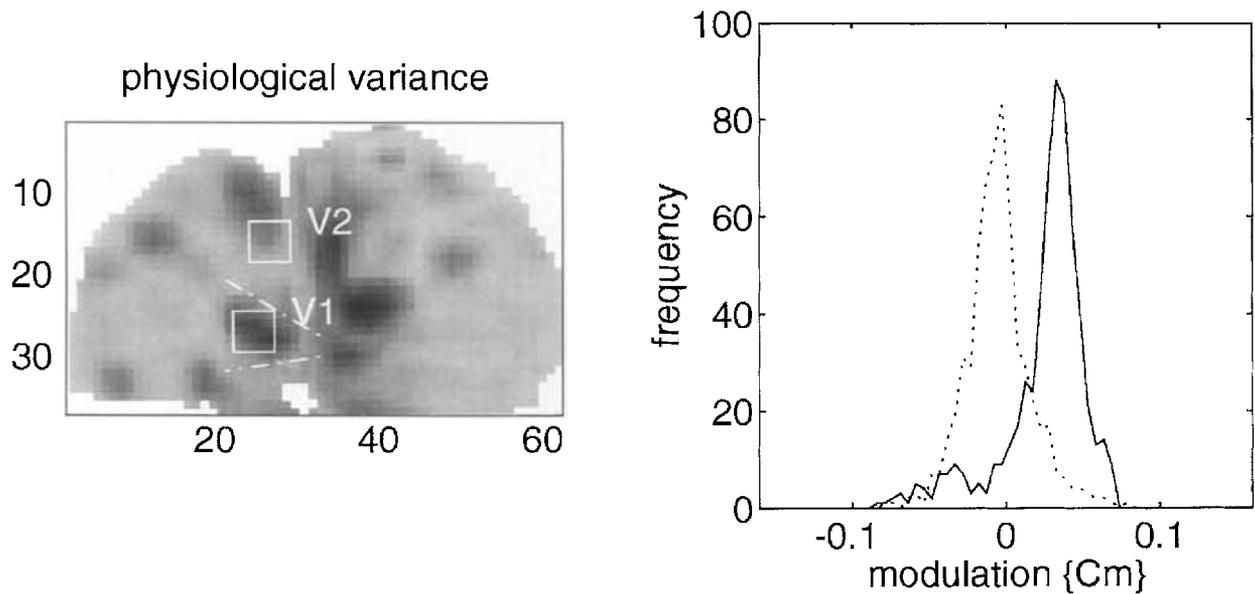


Figure 7.

Identical analysis to that presented in Figure 6, but for boxes placed on V1 and V2 in the opposite hemisphere. Again a clear dissociation between feedforward and feedback modulatory effects is evident.

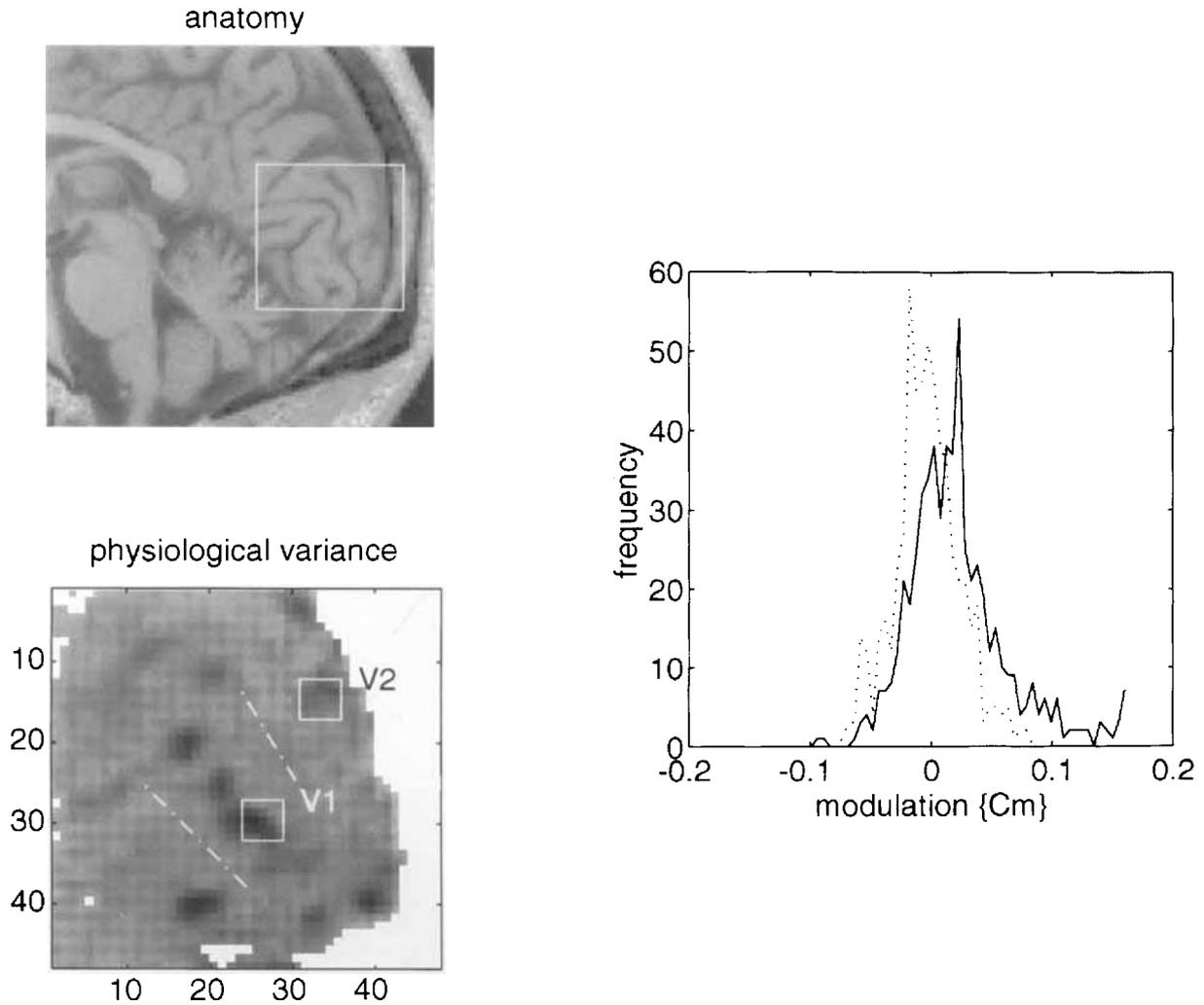


Figure 8.

Identical analysis to that presented in Figures 6 and 7, but for a different data set, which was acquired using the same experimental paradigm but in the sagittal plane. **Top left:** Anatomical high resolution image to illustrate the position of the (48 × 48 voxel) subpartition of the image used in analysis. This subpartition is shown in terms of a map of physiological variance (**lower left**) used to locate the position of the V1 and V2 boxes. The broken lines

correspond (roughly) to the putative V1/V2 border according to the atlas of Talairach and Tournoux [1988; see Figure 3B for the corresponding sagittal section from their atlas]. The placement of the V1 and V2 boxes is at the approximate coronal level depicted in Figures 2 and 5. These results represent a further replication of the initial findings.

asymmetry in modulatory V1–V2 interactions. This effect was seen over a variety of V1 and V2 locations in this sagittal data set. We have replicated these results in a third experiment with coronal sampling at a slightly more posterior level in a different subject (data not shown).

As a final test of the C_{ij}^O and C_{ij}^M estimates we assessed their predictive validity. $C_{V1,j}^O$ and $C_{V1,j}^M$ were estimated on the basis of the first half of the coronal

data set (scans 5–30) and then used to predict the time course of activity in V1 during the second half (scans 31–64). The predicted response of V1 was calculated on the basis of activity everywhere in the images, and the connection strengths according to a rearrangement of Equation (2):

$$x_i = \sum C_{ij}^O \cdot x_j / (1 - \sum C_{ij}^M \cdot x_j). \quad (3)$$

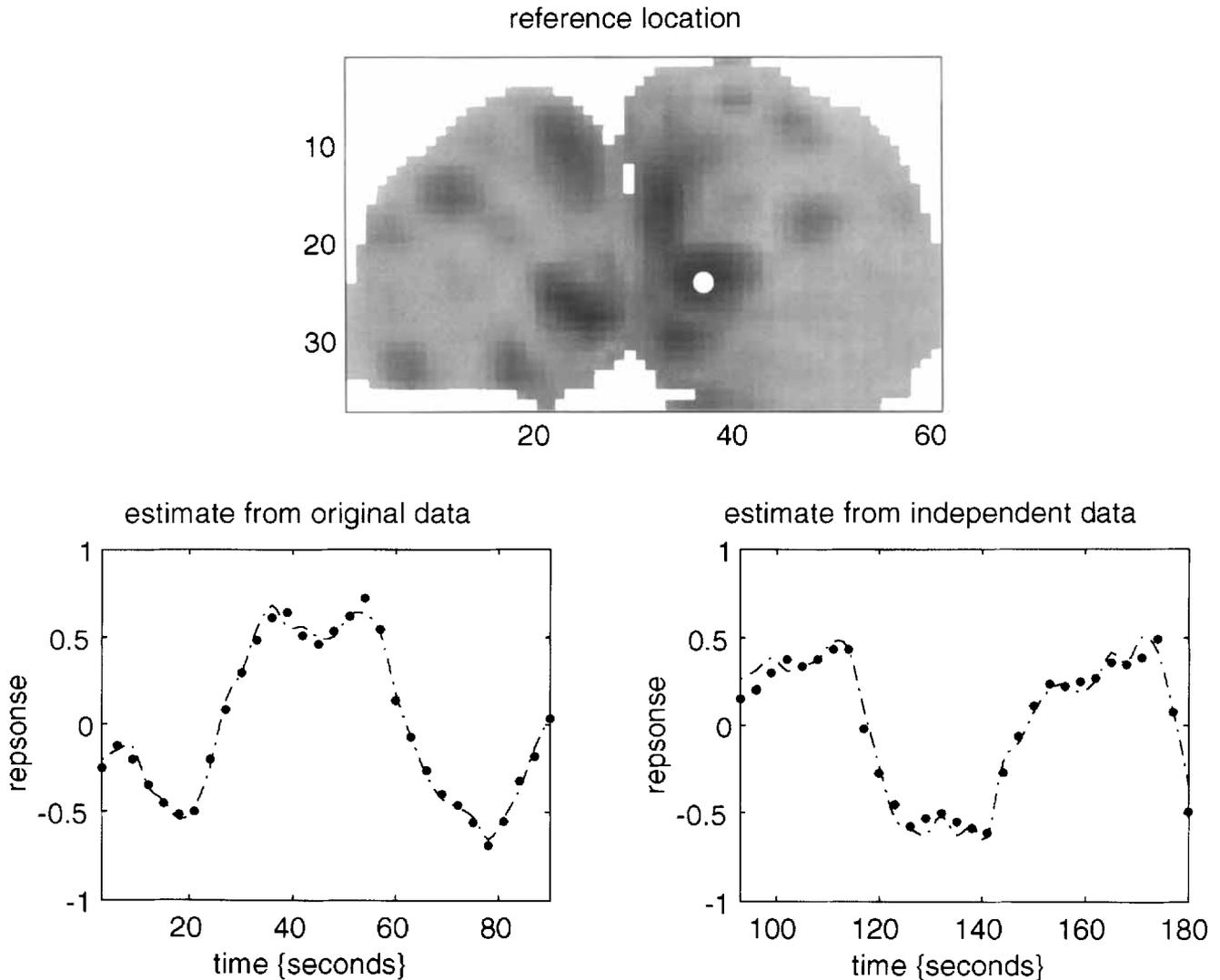


Figure 9.

Validation of the estimates for an arbitrary location in V1. **Top:** Location of reference voxel (white dot) on a map of variance over the 60 scans. **Bottom left:** Actual (dots) and predicted (line) activity obtained by estimating the connections using the first half of the data set. The lines have been interpolated for clarity. **Bottom**

right: Actual (dots) and predicted (line) activity obtained by using the estimates of C_{ij}^O and C_{ij}^M derived from the first half but applied to the second half of the coronal MRI time-series. The predictive validity of these estimates is evident.

This is a weak test of validity, in the sense that the striate/extrastriate system was driven with the same (visual) input in both cases; nevertheless this is a test of the effective connectivity estimates using independent data. The results of this analysis showed the estimates to perform extremely well. Figure 9 illustrates the reference location in V1 (top) and compares the estimated and actual time-dependent activity (bottom) for both halves of the time-series. It may be thought that predicting V1 activity on the basis of data that included V1 activity is somewhat trivial. How-

ever, as described in the appendix, these estimates were obtained using the spatial modes, where there is no "V1 activity."

DISCUSSION

On the basis of reversible cooling studies that have examined the interaction between V1 and V2 cells [Schiller and Malpeli, 1977; Sandell and Schiller, 1982], we predicted that V1 and V2 physiology would show a specific nonlinear and asymmetric relationship. Func-

tional MRI measurements of cerebral activity during visual stimulation in man confirmed this prediction. Using a nonlinear model of effective connectivity [Friston et al., 1993b], we have shown that the feedforward influence of V1 on V2 is predominantly obligatory. By contrast, the feedback influence of V2 on V1 has a substantial modulatory component. This component indicates 1) that the sensitivity of V1 to V2 input increased as a function of intrinsic activity in V1 or, alternatively, 2) that the response of V1 to its total obligatory input was positively modulated by input from V2. Our findings are consistent both with 1) functional asymmetries in reciprocal connections between V1 and V2 and with 2) the notion that feedback (reentrant) connections to lower-order areas in visual cortex modulate physiological activity [Sandell and Schiller, 1982; Girard and Bullier, 1989; Zeki, 1990].

Nonlinear model of effective connectivity

In this section we expand on the interpretation of our effective connectivity model by relating it to voltage-dependent effects and neuromodulation. The modulatory component of effective connectivity (C^M) scales a modulatory interaction between local activity and afferent input. A high C^M reflects a greater sensitivity to input at higher levels of intrinsic activity. This nonlinear interaction is analogous to current-voltage relationships that characterize voltage-sensitive channels. Voltage-dependent effects are revealed by a change in postsynaptic depolarization (in response to a fixed depolarizing current) which increases with increasing depolarization [Haberly, 1990]. In a similar sense, a modulatory influence of area j on area i is manifest as a change in x_i (in response to a fixed change in x_j) that depends on activity in i . This behavior is made obvious on (partial) differentiation of Equation (2) with respect to x_j :

$$\partial x_i / \partial x_j = (C_{ij}^O + C_{ij}^M \cdot x_i) / (1 - \sum C_{ik}^M \cdot x_k) \quad (4)$$

where $\partial x_i / \partial x_j$ represents the sensitivity of changes in x_i to (and only to) changes in x_j . It can be seen that this sensitivity increases with x_i . The activity-dependent effect of inputs from area j is determined by the value of C_{ij}^M . The estimation of the obligatory and modulatory connection strengths appealed to this formulation in the following sense: The sensitivity to afferent input was assessed for low and high levels of intrinsic activity using a linear least squares approach formally identical to multivariate regression. The difference in these sensitivities was then used as an index of the modulatory effect.

There is an alternative way of formulating the model which reveals a close relationship with neuromodulatory effects that are usually attributed to the action of modulatory neurotransmitters. This involves relating the response of a particular region to the total obligatory and modulatory input. Input is simply the activity in remote regions times the appropriate connection strengths (i.e., the activity "seen" by the region in question). It is simple to show that sensitivity to obligatory input is augmented by modulatory input: Let x_i^O denote the total obligatory input (summed over all afferents, i.e., $x_i^O = \sum C_{ij}^O \cdot x_j$) and similarly let x_i^M ($= \sum C_{ij}^M \cdot x_j$) be the total modulatory input. Substituting these expressions into Equation (2) we have:

$$\partial x_i / \partial x_i^O = 1 / (1 - x_i^M) \quad (5)$$

(following partial differentiation with respect to x_i^O and treating x_i^O and x_i^M as independent variables). This equality states that the responsiveness of i to its obligatory input increases as a nonlinear function of modulatory input (provided $x_i^M < 1$). The contribution from area j to the total modulatory input is reflected in the value of C_{ij}^M . This formulation of the model provided for an assessment of the size of the modulatory effect (see Fig. 3) in terms of the increase in sensitivity to obligatory input effected by modulatory input.

It is pleasing that a relationship as simple as Equation (2) can provide a framework that embraces two such disparate aspects of modulation in the brain.

Biological mechanisms of modulatory dynamics

The measurements used in this study were hemodynamic in nature. This limits an interpretation at the level of neuronal interactions. However the analogy between the form of the nonlinear interaction between V1 and V2 activity and voltage-dependent connections is a strong one. It is possible that the modulatory impact of V2 on V1 is mediated by predominantly voltage-dependent connections. The presence of horizontal voltage-dependent connections within V1 has been established in cat striate cortex [Hirsch and Gilbert, 1991]. We know of no direct electrophysiological evidence to suggest that feedback connections from V2 to V1 are voltage-dependent; however, our results are consistent with this. It should be noted that voltage-dependent feedback (reentrant) connections in extrastriate cortex is a strong prediction in theoretical neurobiology: Their potential importance has been demonstrated using

synthetic neural modeling in the context of visual integration and binding [Tononi et al., 1992].

An alternative mechanistic explanation for modulatory effects, which does not necessarily involve voltage-dependent connections, can be found in the work of Aertsen and Preissl [1991]. These authors have investigated the behavior of artificial networks, analytically and using simulations. They concluded that effective connectivity varies strongly with, or is modulated by, pool activity. Pool activity is the product of the number of neurons and their mean firing rate. The mechanism is simple; the efficacy of subthreshold excitatory postsynaptic potentials in establishing dynamic interactions is a function of postsynaptic depolarization, which in turn depends on the tonic background of activity. This clearly relates to the idea that sensitivity to afferent input increases with intrinsic activity. Note that this explanation does not explain the empirical asymmetry we have observed.

Validating the estimates

The validity of the present results depends on the validity of Equation (2) to model comprehensively the interactions of interest. Effective connectivity is not an operational definition. It always depends on some modeling of the influence one neuronal system exerts over another. We have found it useful to consider construct, face, and predictive validity separately. In the present study, construct validity (validity in terms of another construct or framework) has been established with reference to models of voltage-dependent effects and neuromodulation (see above). Face validity (does the model measure what it is supposed to?) was established by the post hoc sensitivity analysis presented in Figure 4. The fact that the increase in sensitivity was significant suggests that the model is “tuned” to the right sort of phenomenon. Finally, we presented a direct assessment of predictive validity in the Results section, which, although not exhaustive, was a step in the right direction. We are currently considering validations that can be effected by a more sophisticated stimulation paradigm.

Some qualifications

Although it is possible to invoke a universe of nonlinear mechanisms that may account for the observed nonlinear effects reported, it is important to discount artifacts. The fMRI measure is a long way removed from neuronal activity. Can spurious nonlinear relationships be introduced by differential delay and dispersion of the signal (i.e., a macrovascular

effect)? Strictly speaking the answer to this question is no. Irrespective of how complicated the delay and dispersion effects, they can be modeled by convolution of the underlying signal [see Friston et al., 1994]. Convolution is a linear operation and, in the context of stationary stochastic processes, linear operators cannot introduce nonlinearities. We made great efforts to ensure the immobility of the MRI time-series by accounting for progressive magnetic saturation effects and movements of the subject’s head. Consistent differences in hemodynamic response profiles (as opposed to the underlying changes in neural activity) in the areas we have designated V1 and V2 would not give rise to the nonlinear asymmetries we have shown. If anything, response differences would tend to obscure any systematic nonlinear features.

The model used in this paper is nonlinear but is very simple when compared to models typically found in computational neurobiology. For example, we have assumed linear response functions. The sorts of models used in estimating effective connectivity should be distinguished from neural network models. The former are designed to allow connection parameters to be estimated, whereas the latter are trying to emulate a physiological process. Effective connectivity models are necessarily simple because they must be sufficiently constrained to permit reasonable parameter estimation. Clearly there is a balance here between practical usefulness and biological validity, and this should always be borne in mind. In addition, the model does not include any explicit time-dependent effects (for example, the activity in V2 is not a function of the activity in V1 100 ms ago). This simplification can be justified by noting that the hemodynamic response function has a time constant of about 8 s. This means that fine-scale temporal relationships can be considered to be averaged out.

The mathematical solution obtained for C^O and C^M are optimum in a least squares sense but derive from a very underdetermined set of equations. This means that the solutions presented are not unique (although they are unique in terms of having a minimum norm—see appendix). While completely reliable, the validity of these estimates is, mathematically, open to question. From one perspective the present findings could be regarded as a validation of the mathematical technique, given that V2–V1 interactions are known to be modulatory and asymmetric [Sandell and Schiller, 1982]. However this is not how the analysis was originally conceived.

It should be noted that the present analysis was based on single coronal and sagittal slice data and that many potential interactions with other brain regions

were not quantified. In this work we considered the time-series from each voxel in a visual area as representative of the activity from other voxels of that area. Clearly these early results will have to be replicated.

Conclusion

Deactivating V1 and V2 by cooling the cortex in monkeys has shown that the influence V1 exerts over V2 activity is not the same as the influence V2 exerts over V1. More exactly, V2 depends in an obligatory way on input from V2, but V1 activity is modulated by (feedback) input from V2. This is a fundamental observation and provides a neurophysiological mechanism whereby higher-order areas can modulate the activity of lower-order ones. Such modulation could modify the impact of sensory input on early visual representations and might be important for the integration of features across extended regions of visual space, attentional effects, perceptual categorization, and, possibly, some illusory phenomena. If the modulatory effect of V2 on V1 exists in human cortex, then the relationship between activity changes in V1 and V2 should show some specific nonlinear behaviors. We have used fMRI to provisionally confirm the presence of this nonlinear relationship in humans. These results speak to a number of important points: 1) concepts like effective connectivity, originated in the field of multiunit electrode recording, find a powerful application in functional neuroimaging; 2) the nature and quality of functional neuroimaging can support some quite ambitious analyses of corticocortical interactions; and 3) the ability of basic neuroscience findings to inform human studies in a concrete and revealing way.

ACKNOWLEDGMENTS

We thank Michelle M. Adams for assistance in preparing the manuscript.

REFERENCES

Aertsen A, Preissl H (1991): Dynamics of activity and connectivity in physiological neuronal networks. In: Schuster HG (ed): *Nonlinear Dynamics and Neuronal Networks*. New York: VCH Publishers Inc, pp. 281–301.

Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS (1992): Time course EPI of human brain function during task activation. *Magn Reson Med* 25:390–397.

Edelman GM (1978): Group selection and phasic reentrant signaling: A theory of higher brain function. In: Edelman GM, Mountcastle VB (eds): *The Mindful Brain: Cortical Organization*

and the Group-Selective Theory of Higher Brain Function. Cambridge, MA: MIT Press, pp. 51–100.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991): Comparing functional (PET) images: The assessment of significant change. *J Cereb Blood Flow Metab* 11:690–699.

Friston KJ, Grasby P, Bench CJ, Frith CD, Cowen PJ, Liddle PF, Frackowiak RSJ, Dolan RJ (1992): Measuring the neuromodulatory effects of drugs in man with positron tomography. *Neurosci Lett* 141:106–110.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1993a): Functional connectivity: The principal component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* 13:5–14.

Friston KJ, Frith CD, Frackowiak RSJ (1993b): Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping* 1:69–79.

Friston KJ, Jezzard P, Turner R (1994): The analysis of functional MRI time-series. *Human Brain Mapping* 1:153–171.

Gerstein GL, Perkel DH (1969): Simultaneously recorded trains of action potentials: Analysis and functional interpretation. *Science* 164:828–830.

Girard P, Bullier J (1989): Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. *J Neurophysiol* 62:1287–1301.

Golub GH, Van Loan CF (1989): *Matrix Computations* (2nd ed). Baltimore and London: The Johns Hopkins University Press, pp. 241–248.

Haberly LB (1990): Olfactory cortex. In: Shepherd GM (ed): *The synaptic organization of the brain*. Oxford: Oxford University Press, pp. 317–345.

Hirsch JA, Gilbert CD (1991): Synaptic physiology of horizontal connections in the cat's visual cortex. *J Neurosci* 11:1800–1809.

Keys RG (1981): Cubic convolution interpolation for digital image processing. *IEEE Trans. on Acoustics, Speech, and Signal Processing* 29:1153–1160.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R, Cheng HM, Brady TJ, Rosen BR (1992): Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci USA* 89:5675–5679.

Maunsell JHR, Van Essen DC (1983): Connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J Neurosci* 3:2563–2586.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992): Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA* 89:5951–5955.

Palus M, Dvorak I, David I (1991): Remarks on spatial and temporal dynamics of EEG. In: Dvorak I, Holden AV (eds): *Mathematical approaches to brain functioning diagnostics*. Manchester and New York: Manchester University Press, pp. 369–385.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988): *Numerical recipes in C: The art of scientific computing*. Cambridge: Cambridge University Press.

Rockland KS, Pandya DN (1979): Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res* 179:3–20.

Sandell JH, Schiller PH (1982): Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J Neurophysiol* 48:38–48.

Schiller PH, Malpeli JG (1977): The effect of striate cortex cooling on area 18 cells in the monkey. *Brain Res* 126:366–369.

Talairach P, Tournoux J (1988): *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: An approach*

to cerebral imaging. Translated by Mark Rayport. New York: Thieme Medical Publishers, Inc.

Tononi G, Sporns O, Edelman GM (1992): Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex* 2:310–335.

Tsonis AA (1992): *Chaos: From theory to applications*. New York and London: Plenum Press.

Zeki S (1990): The motion pathways of the cerebral cortex. In: Blakemore C (ed): *Vision: Coding and efficiency*. Cambridge: Cambridge University Press, pp. 321–345.

APPENDIX

This section presents the mathematical details behind the estimation of obligatory and modulatory connection strengths. For any reference location (i) assume that a subset of the time-series can be selected so that x_i is limited to some small range about its mean ($\langle x_i \rangle$). For this subset Equation (2) can be approximated by:

$$\begin{aligned} x_i &\approx \Sigma C_{ij}^O \cdot x_j + \langle x_i \rangle \Sigma C_{ij}^M \cdot x_j \\ &\approx \mathbf{x} \cdot (\mathbf{C}^O + \langle x_i \rangle \mathbf{C}^M) \\ &\approx \mathbf{x} \cdot \mathbf{C} \end{aligned} \quad (\text{a1})$$

where the second two expressions are in matrix notation (matrices are in bold). Now assume two such subsets are selected, one with a high mean ($\langle x_i \rangle_{\text{high}}$) and one with a low mean ($\langle x_i \rangle_{\text{low}}$) giving two solutions for \mathbf{C}

(\mathbf{C}_{high} and \mathbf{C}_{low}), then:

$$\begin{aligned} \mathbf{C}^M &\approx (\mathbf{C}_{\text{high}} - \mathbf{C}_{\text{low}}) / (\langle x_i \rangle_{\text{high}} - \langle x_i \rangle_{\text{low}}) \\ \mathbf{C}^O &\approx (\mathbf{C}_{\text{high}} - \mathbf{C}_{\text{low}}) \cdot \langle x_i \rangle_{\text{high}} / \langle x_i \rangle_{\text{low}} \\ &\quad / (1 - \langle x_i \rangle_{\text{high}} / \langle x_i \rangle_{\text{low}}) \end{aligned} \quad (\text{a2})$$

The estimation of \mathbf{C}^O and \mathbf{C}^M using this local linearization therefore reduces to solving the rank deficient problem [Eq. (a1)] for two subsets of data. In the present work these subsets were obtained by splitting the data according to whether activity at the reference location was above or below zero. Equation (a1) is underdetermined because the number of voxels (2,304 or 2,160) far exceeds the number of observations (60). We employed a device that has been established in the measurement of effective connectivity using positron emission tomography (PET) data under similar rank deficient conditions. This involves solving Equation (a1) in the space defined by the singular vectors of \mathbf{x} [Friston et al., 1993b]. This represents a standard approach to the rank deficient (underdetermined) least squares problem, where the solutions obtained (\mathbf{C}_{high} and \mathbf{C}_{low}) have the smallest 2-norm of all minimizers [Golub and Van Loan, 1989]. In the above work we used only singular vectors associated with singular values that were greater than unity. Typically there are four to six such vectors which account for over 90% of the observed variance.