Investigating a network model of word generation with positron emission tomography

K. J. Friston,1,2 C. D. Frith,3 P. F. Liddle3 and R. S. J. Frackowiak1

1MRC Cyclotron Unit, Hammersmith Hospital, DuCane Road, London W12 0HS, U.K.
2University Department of Psychiatry, Westminster and Charing Cross Hospital, London W6 8RF, U.K.

SUMMARY
By using positron emission tomography (PET) we examined the biological validity of a network model describing changes in cerebral activity associated with intrinsic and extrinsic word generation. The production of words not specified by an extrinsic stimulus constitutes willed or intrinsic generation. Perceiving a heard word is an example of extrinsic generation. The model incorporates three neuronal systems: a pool that stores word representations in a distributed fashion, an afferent system conveying sensory input to the pool and a modulating system that alters the responsivity of neurons in the pool. Simulations based on the model suggested that intrinsic generation would be associated with low activity in the pool, consequent on reduced modulation, and extrinsic generation with high activity. We measured cerebral activity with PET during intrinsic (verbal fluency) and extrinsic (responding to heard words) word generation and found this pattern of changes in the left superior temporal region. We were able to designate this region the site of the distributed word store and implicate the left dorsolateral prefrontal cortex (DLPFC) as the source of modulation. The relation between the superior temporal gyrus and DLPFC was shown by examining the correlation between the two regions in terms of cerebral activity.

We conclude that the left DLPFC is responsible for modulating the responsivity of a neural system in the superior temporal gyrus and is the probable mediator of changes in attentional and intentional states that underlies the intrinsic generation of words.

1. INTRODUCTION
Mesulam (1990) suggests ‘complex behaviour’ is mapped at the level of multifocal neural systems rather than specific anatomical sites, giving rise to brain-behavioural relationships that are both localized and distributed. Each network contains anatomically addressed channels for transferring information content and chemically addressed pathways for modulating behavioural tone.

In our formulation the essence of ‘complex behaviour’ is intrinsic generation of representations and behaviour in the absence of extrinsic cues. This generation can be effected by neurochemical modulation of discrete neural networks. These networks also transform sensory information conveyed by anatomically addressed inputs from sensory cortices. Modulation can bias the network towards perception (extrinsic generation) or generation of stored representations in complex behaviour (intrinsic generation). This bias is mediated by altering the balance of responsiveness of neurons to sensory input and feedback from within the network. Catecholaminergic afferents have been proposed in this modulatory role (Bloom 1988). A recent network model of catecholaminergic effects on signal to noise characteristics concludes: ‘changes in the responsivity of individual elements do not affect their ability to detect signal and ignore noise but the same changes in cell responsivity in a network of such elements do improve the signal detection of the network as a whole’ (Servan-Schreiber et al. 1990). We think that modulation of cortical responsiveness is orchestrated over large cortical networks by association (particularly prefrontal) cortices. This role compares well with the Supervisory Attentional System (Shallice 1988), ‘which modulates rather than dictates the operation of the rest of the system’.

To clarify these modulatory mechanisms we developed a model of the interaction between anatomically addressed sensory input and neurochemically addressed modulation. On the basis of simulations we estimated the optimum levels of modulation a distributed word store would require (i) to activate representations spontaneously in the absence of input and (ii) to discriminate between extrinsically generated words and non-words. We compared these estimates with measured neuronal activity (as indexed by regional cerebral blood flow (rCBF)) during (i) verbal fluency tasks and (ii) discrimination between aurally presented words and non-words.

2. THE MODEL
We follow Servan-Schreiber et al. (1990) in assuming the response of a neuron can be described by a nonlinear, increasing activation function $f_{M}(x)$ of the net afferent input. We make two modifications to their activation function:

$$f_{M}(x) = \frac{1}{1 + e^{-(Mx + 0)}}.$$  \hspace{1cm} (1)

The first is trivial in that $f_{M}(x)$ returns values in the
range \([-1, 1]\). This reflects the use of \(x\) as deviation from mean firing rate \((\mu = 0)\), as opposed to actual rate. Secondly and more importantly the gain \((M)\) has a differential (opposing) affect on input from within the pool \((x_i)\) and that external to the pool \((x_e)\). For the \(j\)th neuron:

\[
    f_{Mj}(x_j) = \frac{2}{[1 + e^{-(1-M)(x_j)-(M-1)x_j+B}]} - 1,
\]

where modulation \((M)\) ranges from 0 to 1 and bias \((B)\) = -1. Input from within the pool \((x_i)\) is calculated by postmultiplying the vector of activities for all neurons with the connectivity matrix \((C)\). The external input \((x_e)\) corresponds to an extrinsically generated pattern \((E)\) plus noise \((\epsilon)\).

\[
    x_i = C f_{Mj}(x)
\]

\[
    x_e = E + \epsilon.
\]

The connectivity matrix \((C)\) was determined in accordance with the following assumptions. The connection structure within the network is featureless (stationary). That is, the relation between connection strength and inter-neuron distance is the same for every part of the network. This relation is Gaussian. The resulting connection matrix is a Gaussian autocorrelation matrix. The matrix used in simulations was 30 x 30 (i.e., 30 neurons in the pool). The width of the Gaussian function was 0.6. These parameters were chosen to allow for 900 possible interactions between neurons and 30 possible eigenvectors (see below). The storage of 30 words in a given category of which only half are readily accessible corresponds, roughly, to human capabilities.

(a) Word representations, activation and information entropy

The pattern of activities corresponding to stored representations are taken to be the eigenvectors of the connection matrix \((C)\). An eigenvector \((e)\) is a pattern of activity which corresponds to a stable state for the network. The prominence \((\pi)\) of an eigenvector is obtained by postmultiplying that eigenvector with the vector that describes the current state \((f_{Mj}(x))\). This is effectively the covariance between the two. For the \(i\)th eigenvector:

\[
    \pi_i = |e_i f_{Mj}(x)|.
\]

For a representation to be active the corresponding eigenvector's prominence must be high relative to competing alternatives. To assess the effects of modulation we used a measure of word activation based on information entropy. The probability that a representation was active was calculated by normalizing the sum of all prominences to unity \((\pi_i - \pi' i)\). The information entropy \((S)\) over all representations was then derived:

\[
    S = \sum \pi' i \ln (\pi' i).
\]

A low information entropy \((S)\) corresponds to a predictable state where a single representation (eigenvector) is clearly dominant.

(b) Relation to observable changes in neuronal activity

The simplifying assumption is made that change in metabolic rate \((R)\) is monotonically related to average displacement of neuronal activity from equilibrium levels. This simplification is based on the rationale that the brain is organized such that energy requirements are at a minimum when all neurons fire at equilibrium levels \((f_{Mj}(x) = 0)\). The average displacement from this equilibrium, over neurons, is given by the standard deviation of activity:

\[
    R = \text{standard deviation} (f_{Mj}(x)).
\]

A maintained increase in metabolic demand \((R > 0)\) can be inferred from increases in rCBF.

3. SIMULATIONS

Simulations were done on a SUN SPARC workstation (SUN Microsystems Europe Inc., Surrey, U.K.) in MATLAB (Mathworks Inc., New York, U.S.A.). Using the above model we investigated the nature of the interaction between extrinsic input and modulation.

(a) The effects of modulation on information entropy in the absence and presence of extrinsic input

\(M\) was increased from 0 to 0.5 in steps of 0.01. For each value of \(M\) the network was updated 100 times and the information entropy \((S)\) calculated for the final stable state. This was done twice, with extrinsic input \((E) = 0\) and with the input to the third eigenvector, or word, of the connection matrix \((e_3)\). On all occasions the network was initialized to the same state \((\frac{3}{4}\) of the second eigenvector and \(\frac{1}{4}\) of the first). Entropy was plotted against \(M\) (using a fourth-order polynomial regression) for the two situations (figure 1a). The standard deviation of noise (over neurons) was 0.5 and that of \(\epsilon_3\) was 20. Metabolic rate \((R)\) was determined for each endstate and plotted against \(M\) (figure 1b).

To illustrate the effects of changing \(M\) the endstates of the network for five values of \(M\) (0, 0.1, 0.2, 0.3, and 0.4) typical of the simulations were plotted on imaginary spheres to depict the relative amounts of the first three eigenvectors contributing to these endstates (figure 2).

(b) In the absence of extrinsic input

As modulation increased, entropy rose and the predictability of the final state of the network fell. This results from noise de-stabilizing the network and preventing ‘relaxation’ into a stable state (figure 1a). Figure 2b illustrates the endstates for increasing values of \(M\).

(c) In the presence of input

The changes in entropy with modulation are more complicated in this case. The entropy increases at first, reaches a maximum \((M = 0.1)\) and then falls to a minimum \((M = 0.4)\). Figure 2a illustrates what is happening. At low \(M\) the effect of input is negligible and the endstate falls near the first eigenvector or attractor. At high values of \(M\) the extrinsic input is sufficiently strong to reinforce the pull of the third eigenvector and the endstate falls into point attractor 3. There is however an intermediate stage where the
natural propensity to drift to attractor 1 is balanced by the extrinsic pull to attractor 3. In this state the network ends up as a mixture of both and in a state of high entropy and confusion.

(d) Changes in metabolic demand ($R$)

Figure 1b shows monotonically increasing values with increasing modulation $M$ for a fixed extrinsic input ($\epsilon3 + \text{noise}$).

(e) Summary

In the presence of extrinsic input, corresponding to a stored pattern (eigenvector), entropy falls with increasing modulation. Therefore modulation increases sensitivity to (eigenvector) input without changing its signal to noise characteristics (modulation changes the relative gain associated with external input and internal feedback, it does not increase input relative to noise). As increased sensitivity holds for, and only for, eigenvectors the ability to discriminate between words and non-words is enhanced by high modulation.

In the absence of extrinsic input entropy increases with modulation. This represents a key interaction between modulation and extrinsic input on word activation. This dissociation suggests that intrinsic word generation depends on increasing the network’s relative sensitivity to internal feedback (decrease in modulation) whereas discrimination depends on the opposite changes (increased modulation).

Figure 2. Graphical illustration of how the network’s state depends on modulation. In each drawing the points correspond to the endstate of the system (over 100 iterations) plotted on the surface of an imaginary sphere. Point attractors on the surface of this sphere correspond to patterns of activity which are eigenvectors or stable states of the network. In this figure the first three eigenvectors are represented. The first eigenvector has the largest eigenvalue and is the ‘strongest’ attractor (point 1 on the right). Point 3 corresponds to the third eigenvector (top) which is also the extrinsic input. To completely represent the state of the network we would need a 30-dimensional hypersphere but the drawings in figure 2 can be thought of as viewing this hypersphere from the side, allowing only 3 dimensions to be seen. (a) In the presence of extrinsic input the endstate is drawn from the strongest attractor (1) to that representing the input (3) with increasing modulation. At an intermediate value of $M (= 0.1$) these opposing forces are balanced and the endstate is ambiguous. This is reflected in a high entropy (figure 1a). (b) In the presence of noise alone increasing $M$ results in endstates that are randomly displaced further away from the first eigenvector. This is reflected by an increasing entropy (figure 1a).
The changes in metabolic demand ($R$) depend on both the level of extrinsic input (= variance of ($x_i$)) and on modulation ($M$). For a fixed level of external input $R$ increases monotonically with modulation. Experimentally, if extrinsic input is controlled for, changes in metabolic demand ($R$) can be used as an index of modulation ($M$).

This suggests intrinsic generation task (verbal fluency) should be associated with low rCBF in the cortical region corresponding to the word store. Extrinsic generation (listening to words) should bring about increases in rCBF. These differences are attributable to decreases and increases in modulation respectively. These (independent) effects should be seen in the same region, probably the superior temporal gyrus.

4. PET STUDY
(a) Subjects and tasks

Four right-handed, normal, male volunteers (age 26–45) underwent six scans whilst performing six different tasks. Task order was balanced using a Latin square design. The tasks were (1) resting, (2) resting, (3) counting out loud, (4) lexical decision with articulated response (‘correct or incorrect’) based on aurally presented words and non words (extrinsic generation), (5) producing names of animals (intrinsic generation) and (6) producing words beginning with the letter ‘a’ (intrinsic generation).

The rate of word production in unpaced verbal fluency tasks starts high and then declines. The critical period from the point of view of scanning is the first 90 s of performance. A pilot study suggested the average rate of production, during this period, was two words per 5 s. This rate was chosen for the lexical decision and counting tasks. We confirmed post hoc that three subjects achieved this rate during the fluency tasks, the fourth subject was somewhat slower.

The lexical decision task was chosen to control for the lexical decision implicit in the verbal fluency tasks. High frequency concrete words were selected from the MRC psycholinguistics data base. All words used began with ‘L’. Non-words were constructed by replacing the first letter with ‘L’ to create an orthographically regular non-word. The number of presented words and non-words was equal. Counting was chosen as a control because the next word produced is uniquely specified by the previous number, consequently there is no generation in the sense addressed by the model.

In all but the rest conditions, heard self-vocalization, articulation and rate of word generation were controlled for. In all conditions eyes were closed.

(b) rCBF measurements

Scanning was done using a PET scanner (CTI model 931-08/12, Knoxville, U.S.A.), the physical characteristics of which have been described (Spinks et al. 1988), according to a standard protocol (Lammersma et al. 1990). Following stereotactic normalization (Friston et al. 1989) each image was smoothed using Gaussian filter 10 pixels wide to accommodate normal variability in gyral anatomy.

(c) Image analysis

Global differences were removed on a pixel basis (Friston et al. 1990) using analysis of covariance (ANCOVA). This analysis generates adjusted pixel means for the six conditions and an error variance required for their comparison using the t statistic. These comparisons generate images of t values ($\text{SPM}(t)$), a procedure called statistical parametric mapping (Friston et al. 1990). The omnibus significance of the $\text{SPM}(t)$s are assessed by comparing the number of expected and observed pixels above a threshold ($p < 0.001$). Two orthogonal comparisons were done to contrast the intrinsic tasks with baseline (counting) and the extrinsic task with baseline. For illustration a third, direct comparison of intrinsic and extrinsic tasks was made.

To show the magnitude of rCBF changes the adjusted mean rCBF for each of the six conditions was displayed graphically for the left superior temporal gyrus.

To test the hypothesis that the left dorsolateral prefrontal cortex (DLPFC) was responsible for modulating superior temporal gyrus activity, maps of the correlation coefficient (SPM($r$)) between the six mean activities across the brain and in the left DLPFC were generated. Regions with a significant ($p < 0.01$) value in the SPM($r$) are areas in which rCBF correlates with that in the left DLPFC.

5. RESULTS

The (one-tailed) SPM($t$) comparing the baseline (counting) and intrinsic tasks was significant in an omnibus sense ($\chi^2 = 199.9, 1$ d.f., $p < 0.0001$). The orthogonal comparison of baseline and extrinsic generation was similarly significant ($\chi^2 = 63.3, 1$ d.f., $p < 0.0001$). The only region that decreased in the first and increased in the second comparison was the left superior temporal region.

The SPM($t$) reflecting changes in cerebral activity on comparing extrinsic and intrinsic tasks directly is seen in figure 3. This (one-tailed) SPM($t$) highlighted two large areas in the superior temporal regions where activity fell during the intrinsic generation tasks. The most significant difference was in the left superior temporal gyrus ($x = -54$ mm, $y = -20$ mm and $z = 4$ mm Talairach & Tournoius (1988)). The most marked activation is seen in the left DLPFC. The omnibus significance of these SPM($t$)s was again high ($p < 0.0001$).

(a) Cerebral activity in the superior temporal cortex

The adjusted mean rCBF (normalized to 50 ml dl$^{-1}$ min$^{-1}$) for all six conditions is given in figure 4. As predicted, the highest rCBF is seen in the lexical decision condition (4 = extrinsic generation). The lowest rCBFs are seen in the verbal fluence conditions (5 and 6 = intrinsic generation). The rCBF in the
(b) Correlations of cerebral activity

The SPM(t) is shown in figure 3. The bright regions correspond to those areas showing significant ($p < 0.01$) negative correlations with the left DLPFC. These areas are mainly limited to the superior temporal regions and correspond to the sites identified in the categorical comparisons.

6. DISCUSSION

The behaviour of a network model of the modulatory effect of distant cortical input on a distributed word store suggested that high activity would be associated with discriminating between extrinsically generated words and non-words and that low activity would characterize intrinsic generation. We interpret findings from a PET activation study as consistent with these model driven observations. rCBF increased in the left superior temporal region when subjects had to distinguish between words and non-words. This same region evidenced rCBF decreases during verbal fluency tasks. These changes were estimated relative to a control task (counting). We extended the analysis to show that rCBF in the left DLPFC correlated significantly and inversely with rCBF in the superior temporal regions. We suggest that this association cortex is the source of a distant modulating input, responsible for mediating intentional states.

We have used correlation SPMs to infer functional connectivity between the DLPFC and bitemporal cortices. Following the terminology of Getting (1989) the present findings suggest that neural networks can be configured in different modes by the modulating effect of functional connectivity. 'A mode is intended to imply a manner in which a network processes information or generates an output pattern...transitions between modes may occur when afferent or modulatory inputs alter the constituent properties of the building blocks'. In this case the two modes correspond to an extrinsic and an intrinsic mode. The extrinsic mode configures the network to act as a sensory filter, allowing extrinsic inputs to affect activity within the network only when the input pattern is an eigenvector of the network. Novel or unfamiliar inputs will fail to impose themselves on the network and not gain access to higher levels of processing. In the second, intrinsic mode the network acts as a distributed store of internal representations which can be activated by shifts in intentional state.

This model has parallels in invertebrate networks. For example the Tritonia escape swim system can operate in two modes according to the prevalent modulation. In one mode each neuron can be activated independently by afferent input and can contribute to the routing of activity to motor neurons that mediate directed reflexive withdrawals (cf. extrinsic generation). In the second mode the network forms a pattern generator that mediates swimming behaviour by producing sequences of alternating bursts (Getting 1989), this corresponds to intrinsic generation of a behavioural state.

These findings are consistent with a study of 39
normal volunteers and nine patients with schizophrenia by Warkentin et al. (1989) who used the Xe inhalation technique to measure rCBF during a verbal fluency activation. The ‘most marked difference was seen in the left prefrontal area with significant increase only in the normal group’. Moreover the normal group showed flow decreases in left central and central-parietal areas. The work of Petersen et al. (1988) on the functional anatomy of single-word processing has identified areas in the left prefrontal cortex activated in a task requiring generation of a ‘use’ for a presented target word.

The fact that the correlation shown between DLFPc and temporal cortex was negative suggests (in terms of the model) that increased trans-synaptic activity in the prefrontal cortices is associated with a reduced excitatory input to the superior temporal regions. This suggests the innervation of the DLFPc neurons that send long association fibres to the temporal cortex is inhibitory. Indeed inhibitory postsynaptic dopamine receptors are recognized in the prefrontal cortex (Glowinski et al. 1984).

There are specific implications the model has for the design of PET activation studies. The modulating effect can be interpreted as mediating selective attention. The prefrontal cortices may orchestrate other cortices to gate information of a specific and relevant nature. If this is the case the activity in lower level cortices will be a function of selective attention and sensory stimulation. The simulation highlights this point, in that a major determinant of changes in regional activity is modulation. Changes in activity can result from changes in modulation even when the stimulus is constant. If we equate modulation and attention the role of attention in experimental design and interpretation becomes supreme. Indeed functional specialization has been inferred using manipulations of attention during PET scans with invariant stimuli (Posner et al. 1989, Corbetta et al. 1990). Roland has synthesized PET findings in terms of recruitment and attentional control of cortical fields (Roland 1985). He notes that in the organization of voluntary behaviour ‘activation of the superior prefrontal cortex was independent... of sensory input and motor output’. The interaction between attention and stimulus is potentially an interesting way in which to use PET to clarify models of complex behaviour.

A few qualifications are required. It has been assumed that measured cerebral activity corresponds to the average deviation from mean firing rate in the model. Given that energy metabolism is largely concerned with the maintenance of resting transmembrane potentials we feel that this is fair. It does of course imply that maintaining a neuron in a state of inhibition is as energy costly as keeping it more active than normal.

Controlling for differences in extrinsic input was not ideal. The increases in rCBF during extrinsic generation may have two components, an increase in modulation and an increase in extrinsic input. The lexical decision task involved hearing words spoken by self and non-self whereas the verbal fluency tasks and counting involved only self vocalization.

Finally we have not addressed the key processes underlying the verbal fluency itself, i.e. how a series of different words are produced or how categorization arises. We felt this extension beyond the scope of the current model but suggest a word string represents a four dimensional pattern of activity in the network whose emergence depends on modulation germane to the current simple model.

K.J.F. was supported by the Wellcome Trust. We thank our colleagues and collaborators at the MRC Cyclotron unit for making the study possible.

REFERENCES


Submitted by S. Zeki; received 11 January 1991; accepted 4 March 1991
Figure 3. Statistical parametric maps (SPMs) of the $t$ statistic which reflects the significance of difference in rCBF between the extrinsic condition and the mean of the two intrinsic tasks. There is a clear relative de-activation in the superior temporal regions on both sides (top right) and activation of the left prefrontal cortex (top left). The brightest points along any of the three lines of view are displayed, all pixels displayed correspond to a significance of $p < 0.001$. The SPM(r) (bottom) is of the correlation of mean rCBFs across conditions with the left dorsolateral prefrontal cortex. Bright areas in the superior temporal gyri correspond to a negative correlation at $p < 0.01$. 