The Cingulate as a Catalyst Region for Global Dysfunction: a Dynamical Modelling Paradigm

Dina M. Kronhaus1,2 and David J. Willshaw2
1Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Cambridge, UK and 2Institute for Adaptive and Neural Computation (ANC), School of Informatics, University of Edinburgh, Edinburgh, UK

The anterior cingulate (AC) often exhibits both structural and functional abnormalities in affective disorders. Neither the cause for this vulnerability nor its effect on behaviour is known. Due to its extensive connectivity, minor output changes from the AC may exert widespread consequences. A causal model describing coupling coefficients (effective connectivity) among several brain regions in healthy subjects performing a memory task inspired our work. This stationary causal analysis provides a theoretical framework for our nonlinear dynamical models. We tested the effects of global and local perturbations upon stability of a systems-level neural network of interconnected brain regions. Interactions between regions, represented by path coefficients, were modelled using connectivity matrices. We found that both characteristic behaviour and response to perturbation differed in networks representing perceptual matching and long-delay conditions. Owing to the highly interconnected character of the networks, activation of a few areas was sufficient to trigger characteristic patterns of behaviour. However, only perturbation of key regions resulted in global dysfunction. Likewise, recovery of function was possible by increasing output from some, but not all, regions. We suggest for this recovery to be context specific, conditional on the task, integrity of other regions and global properties such as neuronal excitability.

Keywords: affective disorders, anterior cingulate, connectivity, neural networks, working memory

Introduction

Neural activation, as assessed at the network level, is subject to both task and contextual dependencies. A recent theory promoted by McIntosh (2004) suggests that certain key regions within a functionally connected network can serve as catalysts, facilitating state transitions (McIntosh, 2004) by virtue of their anatomical connections with several specialized networks. However, these ideas have not been explored in a modelling framework and the functional correlate of integrating a catalyst region into a systems-level model are unknown.

Discrete structural deficits have been identified in various neurological and neuropsychiatric disorders. These suggest that acute damage (neuronal or glial atrophy) to specific areas such as the anterior cingulate (AC) (Drevets et al., 1997; Cotter et al., 2001), prefrontal cortex (PFC) (Rajkowska et al., 1999, 2001) and hippocampus (GH) (MacQueen et al., 2003; Duman, 2004) are involved both in illness aetiology and its clinical course. The functional corollary of these structural deficits is unclear, although a link has been argued between localized atrophy and specific cognitive deficits (Landro et al., 2001; Cordovani-Mackowiak et al., 2003; Van der Werf et al., 2003; Wolf et al., 2004). Localized cortical deficits can be compensated for by increased activity elsewhere, either under transient experimental conditions (Drummond et al., 2000) or irreversible decline (Della-Maggiore et al., 2000). Furthermore, a putative cellular mechanism for functional compensation in aged monkeys has recently been reported (Chang et al., 2005). Therefore, we wished to explore the dependencies between structure and function using artificial neural networks. Furthermore, we wanted to study putative compensatory mechanisms, which may explain the frequent absence of a clear association between behaviour and neural activity.

Computational modelling techniques allow an empirical study of the impact of changing discrete parameters, such as initial inputs or connection strengths, on characteristic network behaviour. Several computational modelling techniques are in existence (Hertz et al., 1991; O’Reilly and Munakata, 2000). These neural network techniques have been applied to large-scale models (Tagamets and Horwitz, 1998; Young et al., 2000); recently, Bayesian methods have also been applied to large-scale neural phenomena (Friston et al., 2003). These modelling approaches complement statistical techniques, such as Structural Equation Models (SEM), which combine anatomical information (usually derived from primate labelling studies) with activation data from neuroimaging studies in human subjects (McIntosh and Gonzalez-Lima, 1994). Structural equation modelling identifies coupling coefficients that minimize the difference between covariance implied under a linear model and those observed empirically. In short, they model statistical dependencies under a linear model in terms of regression coefficients. The resulting connectivity matrices indicate stationary covariance between different brain regions during task performance. Here, we develop a novel modelling framework which is conceptually related to large-scale models, yet is driven by the data summarized by SEM.

Our modelling framework allows us to investigate the dynamical properties of well-established neuronal networks. In a systems level modelling paradigm we explore the complex dynamics that arise from activation and perturbation of different brain regions. We aim to uncover the mechanism(s) responsible for both localized and diffuse deficits, as well as the putative mechanisms by which amelioration of symptoms can occur. The empirical constraints on our neural network model were taken from previously reported SEM analyses of effective connectivity using human whole-brain imaging data (McIntosh et al., 1996). These networks represent interactions [measured by positron emission tomography (PET)] among different brain regions, in healthy subjects performing the delayed match to sample task (Haxby et al., 1995). During this task, subjects were asked to identify a previously viewed stimulus, across variable time delays. The areas chosen to represent different delay conditions,
in the effective connectivity models, were selected for further analysis on the basis of a range of theoretical approaches, suggesting their ubiquitous function in the delayed match to sample task (McIntosh et al., 1996). Connectivity matrices describing different delay conditions thus illustrate different arrangements of areas in a neural network, where the strength of coupling between areas is adapted to perform specific tasks. Using neural network modelling techniques we analyse the structural stability of the network representing two delay conditions: perceptual matching and long delay. Path coefficients are interpreted in our model as weighted connections, these positive and negative coefficients are assumed to indicate a functional relationship (i.e. excitatory or inhibitory connections of a given strength) among regions. We assessed the effects of changes in the parameters (i.e. connection strength) on structural stability. An analogous ‘lesion’ approach has been previously employed by others (e.g. Young et al., 2000; Hoffman and McGlashan, 2001), although these studies were not based on functional neuroimaging data. We also simulate a putative mechanism for neurobiological compensation (Della-Maggiore et al., 2002) by increasing activity elsewhere in the network. Compensation was defined as restoration of sustainable activity.

The novelty of our approach rests on a stability analysis of a neuronal network model that is constrained by empirical results from neuroimaging. Our model proposes a new approach to characterizing the structural and dynamic stability of a large-scale neural model of interacting cortical areas. Critically, the connections describing the architecture of this network were taken from empirical analyses of effective connectivity using whole-brain human imaging. These constrained models allowed us to characterize the stability of the ensuing networks in a way that would not be possible empirically. Our analyses point to some important characteristics of cortical networks involved in high-order cognitive functions such as memory and evaluation of similarity.

Materials and Methods

Simulating the Network

In the neural network models, path coefficients in the effective connectivity networks (McIntosh et al., 1996) are interpreted as connection strengths between different nodes of the network. Throughout the manuscript we refer to brain regions by their Brodmann area (BA) and name (e.g. PFC).

The perceptual matching and the long-delay networks depicted in Figure 1ac were simulated and analysed by programs written in Matlab, using the following components:

Activation Vector \( \mathbf{a} \)

A row vector of 22 units representing 11 areas in the left hemisphere (LH) and 11 areas in the RH corresponding to the area identified by McIntosh (Fig. 1). The value of each element in vector \( \mathbf{a} \) changes over time and reflects the changing activity in the corresponding brain region.

The input vector \( \mathbf{a}_0 \) is initialized with binary values, set by the user. Setting just one element of \( \mathbf{a}_0 \) to 1 was sufficient to trigger sustainable activity in the network.

Connectivity Matrix \( \mathbf{W} \)

\( \mathbf{W} \) is a 22 \( \times \) 22 matrix that represents the pattern of connections among the 22 areas (Fig. 1bd shows the perceptual-matching and long-delay networks, respectively). Rows represent the source and columns represent the target of each connection. This was derived by studying the path-coefficients presented in figures 2-6 of McIntosh et al. (1996). The arrows were mostly in the lower (medium, weak or zero) value-ranges. Conforming to McIntosh et al. (1996), weak connections are given values of 0.1 or 0.35; likewise, medium-strength connections are either 0.36 or 0.65 (lower values did not produce sustainable activity). Inhibitory connections are denoted by negative values.

Update Rule

The value of the internal activations at time \( t \), \( i_k \), is defined as the sum of all activations \( a_k \) weighted by the appropriate entries in \( \mathbf{W} \). In matrix form \( i_k \) is calculated as

\[
i_k = a_k \cdot \mathbf{W}
\]

(1)

The internal activation is then passed through a sigmoidal function, \( f \), to get the activation at the next time step \( a_{k+1} \):

\[
a_{k+1} = f(i_k)
\]

(2)

Note that time is measured in arbitrary units.

Sigmoidal Function

A sigmoidal function is used to compress a wide range of internal activation into a limited range of output. Thus, activity remains bounded. Furthermore, the specific form of the sigmoidal function used in this work limited the output to exclusively positive values to prevent oscillations between positive and negative values.

The sigmoidal function \( f \) takes the form:

\[
f(a_k) = \frac{1}{1 + \exp(-k(a_k - \xi))}
\]

(3)

The value of \( f(a_k) \), and hence each element of \( a_k \), will be bounded in the range [0,1]. The constant \( k \) gives the transition of \( f \) from 0 to 1. \( \xi \) specifies the value of \( a_k \) at which \( f(a_k) \) has half its maximum value. \( k \) was given the value of 10, which defines a fairly sharp transition of \( f \) while \( \xi \) was set at 0.5.

Starting from an initial activation vector, \( a_0 \), the network activity is updated using the above equation for a pre-determined number (1000) of iterations. Activation of the 22 units was plotted against time (Fig. 2), with activation of each area (Fig. 1bd) represented in a separate plot (numbered 1–22). In summary, we constructed simple neural network model whose dynamics were described by difference equations. These neuronal networks had one state variable per region. This state variable can be thought of as representing mean synaptic activity. This renders our model a simple neural mass model of interacting cortical areas. Our objective were to characterize the stability of these neural mass models in terms of perturbations to regional activity and structural perturbations to the coupling among areas. This corresponds to an analysis of their dynamical and structural stability respectively.

Interpretation of Results

In our analyses, we characterized stability in terms of changes in the systems attractor following perturbation. In the absence of perturbations our systems had a quiescent fixed-point attractor. After some perturbations our networks expressed sustained activity with a quasi-periodic attractor. We were particularly interested in which areas and which connections were necessary to support this sustained activity and implicit meta-stability. Investigation of the long-delay network was guided by findings in the matching network. This enabled comparison between their characteristic behaviours, allowing us to assess the effect of perturbation and compensation upon their respective activity patterns.

After running the simulations (for 1000 iterations) we qualitatively evaluate network dynamics (behaviour). We noted the degree to which sustainable, robust oscillations or other integrated behaviours (such as initial bursts or convergence at the upper or lower limits of the sigmoidal function) can occur. Oscillations in a subset of network regions suggested that constituents of this sub-network were closely coupled. Perturbations and restoration of this pattern characterized the stability of the network highlighting potential mechanisms for functional damage and repair. The absence of activation (i.e. if the activation of any unit remained or decayed to 0) was not considered to suggest a pattern of co-activation.
Specification of Experiment 1: Characterizing Network Behaviour

Characterizing network behaviours in the matching (and subsequently the long-delay) network was performed by activation of one unit at a time. This was done by setting the appropriate value of \( a_i \) to 1. This highlighted a functional grouping of the matching network into functional sub-networks, or clusters, of co-activated units. Once a sub-network was identified, we determined which units were necessary for this pattern by individually silencing the activity of units that did not appear to participate in this pattern of behaviour. For unit \( i \) this was implemented by enforcing \( a_i(i) = 0 \). This stepwise elimination of units facilitated an assessment of network stability and an examination of the emergent properties of the network without introducing an \( a \) priori bias. It allowed us to determine which units were necessary and sufficient for generating and maintaining the characteristic patterns of activity. Furthermore, this allowed us to assess the consequence of

Figure 1. Connectivity diagrams depicting the cortico-cortical and cortico-limbic connections of left (LH) and right (RH) hemispheres in (a, b) the perceptual matching and (c, d) the long-delay conditions (McIntosh et al., 1996). (b, d) Weight matrices corresponding to the connectivity diagrams (a, c) denote LH and RH units (1–11, 12–22). Bilateral Brodmann areas (BA) are indicated in the key (bottom left). Discrete values were ascribed to weak and medium weights (positive and negative), to characterize network behaviour. Excitatory connections are depicted in shades of red and inhibitory connections are in blue.
removing certain units that are necessary for maintenance of this pattern without taking an obvious part in its maintenance. To gain insight into the nature of this activity, the activation values of the different units were noted and compared for different values, varying both inputs and weights.

Selection Criteria for Initial Activity
Given the different nature of the connectivity in the perceptual matching and long-delay networks, it was sometimes difficult to find a common ground on which to pursue a comparative study of network behaviour. For example, activation through the visual areas appears to be the most sensible way of emulating the experimental procedure (Haxby et al., 1995; McIntosh et al., 1996). In the matching network, an initial input from left (L) BA18v (fusiform) was sufficient to produce enduring activity in the network. We therefore chose these values as defaults for the remainder of this paper, unless otherwise specified.

By contrast, an initial input from the same unit in the long-delay network produced only transient activity in several units without spreading to the dominant sub-network. However, to compare the activity of these networks, we thought it would be helpful for the activation to be initiated by a visual unit. Therefore, in the long-delay network initial input from L BA18v and BA37. A combination of these two inputs triggers a similar pattern to input from L BA37 alone with initial transients, such as those seen in L BA18v, R BA18v and R GH (dotted circles), associated with the visual input only. Coupled oscillations are observed among units L BA47, BA21, BA37, R BA47 and BA21. The pattern of oscillations is more wide-spread in this network extending to other PFC (L BA46 and BA10) and cingulate (L BA24 and BA23).

Figure 2. Plots show simulated activation over time in areas from the model. Here and in subsequent plots, figures show activation in the 22 areas represented by the network; the top half of each plot shows activation of the 11 areas in the left (L) hemisphere; right (R) hemisphere activation is shown at the bottom. The name of the area appears above each plot. Numbers of iterations are plotted on the x-axis; the y-axis of each plot shows the activation of the area at that time. Activity in the matching (a) and long-delay (b) networks. (a) The matching network can be activated by initial input is provided by L BA18v (fusiform; the input unit is marked by x here and in subsequent plots). This causes a short burst of activity in different units of the left (top) and right (bottom) hemispheres. Coupled and robust oscillations are observed among units L BA46, BA10, BA24 and to a lesser extent BA47. (b) Long-delay network, input from area L BA18v and BA37. A combination of both these inputs triggers a similar pattern to input from L BA37 alone with initial transients, such as those seen in L BA18v, R BA18v and R GH (dotted circles), associated with the visual input only. Coupled oscillations are observed among units L BA47, BA21, BA37, R BA47 and BA21. The pattern of oscillations is more wide-spread in this network extending to other PFC (L BA46 and BA10) and cingulate (L BA24 and BA23).

We tested different values within the range provided by McIntosh et al. (1996) for weak, medium and strong connections. Our simulations revealed that at the lower limit given by McIntosh et al. (1996), weight values were insufficient to create sustainable activity in the networks. We tested interim values and discovered that, in most cases, only the upper limit of the values (absolute values of $w = 0.35$ and $m = 0.65$; see Fig. 1 for definition of $w$ and $m$) typically produced sustained activity; hence we adopted these values throughout the study.

Specification of Experiment 2: Physiology of the Networks
The second set of experiments explored the structural stability of the networks, using the following procedures:

1. Changing global values of either excitatory or inhibitory weights.
2. Assessing the relative contribution of specific areas to characteristic activity patterns, by decreasing their weights of connections to all other areas in the network (emulating significant structural or functional deficits, some of which have been associated with memory deficits).

Positive and negative path-coefficients will be discussed as if they represented excitation and inhibition in the network. However, it is unlikely we can assume a direct correspondence between correlation or anti-correlations among units in the network (represented in the effective connectivity networks) and simple polarity. We would like to stress, therefore, that these are not taken as the direct influence of GABA or glutamate deficits in the brain, but rather as coherent co-activation of network constituents.
3. Exploring the ability of specific units to compensate for a discrete reduction in the output of other areas may provide a quantitative underpinning for normal task in clinical populations in the presence of neural dysfunction.

To avoid experimental bias, we investigated how a decrement or increase in the efficacy of every unit (one at a time) affects network activity.

We plotted activation of one area against another to compare the activity of pairs of units. Further, we tested for periodicity in the signal by examining the activation values over time and by plotting the autocorrelation functions of different units in the statistical program R (www.r-project.org).

Results

Characterizing Network Dynamics

Characterizing Activity in the Matching Network

Initial input was provided to the matching network by setting L BA18v to 1. The activity of the network was then characterized by testing which elements are likely to be co-active. The nature of this coupling was captured in oscillations across several units. In the minimal case of co-activated units, the oscillations occur in L BA46 and BA10 (PFC) and L BA24 (AC) (Fig. 2a). This triangular loop could also be excited by input from visual areas by initially activating either L BA18v (fusiform) (marked by \( \ast \); Fig. 2a) or contralateral right (R) BA18v. Despite this triangle of three areas forming an entirely excitatory loop, they exhibit an oscillatory activation pattern, which is a result of the initial activation (from one unit only) and the discrete (stepwise) update function. Nonetheless, we feel that this is a useful paradigm to examine co-activation and dysfunction in task-related networks, simulating dysfunction and repair.

Characterizing Activity in the Long-delay Network

Classically, working-memory deficits are associated with ageing, depression and frontal lobe lesions. The study of the long-delay network is therefore based upon the investigation of the matching network, guided by the neurobiological and the neuropsychiatric questions we wished to ask.

Unlike the perceptual-matching network, where the dominant sub-network was fairly isolated and the pattern of activity was self-sustaining, the long-delay network appeared to be characterized by an activity pattern distributed over more areas and was thus less vulnerable to disruption. Here, the dominant sub-network consisted of L BA47, BA21 and BA37 (Fig. 2b) where oscillations occur as a product of the discrete activation and update function. However, unlike the matching network, the dominant pattern is closely coupled with L BA24 (AC), L BA46 and L BA10 (PFC), as well as L GH, as shown in Figure 3b. To compare the activity in the matching and the long-delay networks and to express the essentially visual paradigm on which this model was based (Haxby et al., 1995), L BA18v (fusiform) is initially activated (identified by \( \ast \); Fig. 2b) and therefore included in the minimal sub-network (Fig. 3b).

Identifying the Minimal Sub-networks

By observing activity in many trials, we have determined a minimal set of regions that appears dominant in establishing characteristic network behaviour (Fig. 3). In the matching network, activity propagating from L BA18v (fusiform) will invariably pass through L BA37 (IT) and L BA24 (AC). We could therefore expect these two areas to be the most crucial links in the activation of this network. Namely, we may expect that change in their activity will have the most significant effect on the characteristic behaviour of the network.

Both the matching and long-delay networks included a number of loops characterized by feedback connections. However, while in the matching network we found a strong coupling among L PFC and BA24, this interaction is more subtle in the long-delay network extending to L GH. In this subset of regions, sustainable activity can be triggered by one of three closely coupled units (L BA47, BA21 and BA37) and maintained while all the other units are silenced. Nonetheless, owing to the distributed pattern of connections in the long-delay network, propagation of activity is more robust to disruption while the relative contribution of each constituent to the overall pattern cannot be predicted by visual inspection alone.

Having identified the minimal sub-networks, we were able delineate which areas were necessary to maintain network activity. Informed by the definition of catalyst areas (McIntosh, 2004), we could expect that either of the areas participating in the minimal sub-network is a potential candidate for this catalyst role, since each appeared to facilitate activity of the minimal set.

GH and PFC in Working Memory Tasks

During working memory tasks, the GH was reported to display unique properties with relation to the PFC (McIntosh, 1999). In the course of isolating the minimal constituents of the long-delay network, the relationship between the L PFC and GH became apparent (especially the correspondence between L BA10 and L GH; Fig. 3c). Maintenance of patterns in working memory was expected to be dependent upon this interaction (McIntosh et al., 1996) and therefore, we wished to test whether this relationship holds for all three prefrontal areas in the network (i.e. BA46, BA10 and BA47). The long-delay network displayed complex periodicity, especially when only the minimal constituents were activated (Fig. 3c). We plotted activation of GH against activation of each of the PFC regions (BA46, BA10 and BA47) in our model. Figure 3(c-d) shows two plots that are seemingly identical [both include only areas in left hemisphere (excluding L BA23) of the network; L BA23 and all right hemisphere areas are always silenced]. However, to generate Figure 3(c-f), while in (c) the visual units (L BA18v, BA19d and BA17/18) were activated, in (d) they were reset (clamped) after every time-step. The activity in L BA10 and L GH appears to have a similar periodicity, yet the oscillations are somewhat irregular. Furthermore, the activity in L BA47 appears to be unrelated to either of those. Visual inspection alone clearly could not distinguish between the two. Our plots (Fig. 3e-f) chart the activity of these three areas against each other. They allowed us to examine the activity of two pairs of units simultaneously, leading to the conclusion that the relationship between L GH and L BA46 mirrored the relationship between L GH and L BA10 (Fig. 3c). Detailed examination of the activation values in L BA46, BA10 and GH revealed a dominant periodicity of 18 (Fig. 4), twice the number of arms in the network.
Figure 3. The minimal set in (a) matching and (b) long-delay networks: units necessary for creating the characteristic activation patterns. The dominant pattern of activity in the matching network (a) consists of left frontal areas L BA46 and BA10; both send and receive excitatory input from L BA24 (AC). This sub-network is excited by input from visual areas (BA18v, BA19d, BA17/18) via ventral visual stream in LH. Here input originates in the BA18v (fusiform; providing sufficient input for sustained activity), proceeding through BA37, BA21, and BA47, and activating the dominant sub-network. Red lines represent excitation (a darker shade and thicker lines depict stronger weight values) and blue lines represent inhibition; double-headed arrows depict recurrent connections, colour-coded in the direction of the arrowhead. The dotted, double-headed arrows among the excitatory sub-network comprising left frontal areas BA46, BA10 and BA24 (circled) indicate the areas that oscillate together. (b) Minimal number of constituents participating in the activation of the long-delay network. Recurrent loops include the basic sub-network (L BA47, BA21 and BA37; oscillation depicted by dotted arrows), extending to fronto-limbic areas (L BA24, BA46 and BA10) as well as LH. Activity in these recurrent loops cannot be triggered exclusively by L BA18v. However, this area participates in the network by providing initial activation, together with one of the minor constituents of the dominant sub-network (L BA37; see Fig. 2b, 3c–f). Minimal constituents of the long-delay network. Only 10 of the 22 areas in the entire network are active. In LH, all areas excluding L BA23 (posterior cingulate) are active, while the rest (L BA23 and all areas in RH) are reset at every time-step. Thus all area names here refer to LH-only. (c) Initial input from L BA37 and BA18v, or (d) by L BA37 alone, where activation of visual areas (C represents clamped/silenced left visual areas, where activity is set to 0) is reset at every time-step. Default weight values are used in these experiments. Coupled oscillations are observed among the dominant sub-network (L BA47, BA21, BA37) while the amplitude of oscillations in L BA24 is higher than in the default (Fig. 2b) pattern. Oscillations in L BA46, BA10 and GH (all circled) are somewhat irregular. (c, d) Although activity in visual areas is clamped (3d), it seems similar to activity in (3c), where visual units are not clamped (Fig. 3e–f). To further examine activity patterns, we plot pairwise activity of three key areas (3e corresponding to 3c, and 3f corresponding to 3d). In both plots, activity of LGH versus BA10 is shown in black and activity of LGH versus BA46 is in blue. The activity plotted here is for iterations 20–1000, ignoring initial transients. From these plots, it is apparent that the clamped units affect network activity.
activation values of these units over time (directly reflecting activation in Fig. 3d). This lack of close synchronization between areas (as shown by the long periodicity) reflects the complex nature of interaction in the long-delay network.

Network Physiology: Altering Network Connectivity

The Matching Network

The matching network (Fig. 5a) was more sensitive to changes in the efficacy of all excitatory connections (termed global excitation) than to changes in the value of inhibitory weights across the network (global inhibition). Arguably, this may be attributed to the number of excitatory and inhibitory connections. Compared with 21 inhibitory connections ($w = 12; m = 9$), there are 46 excitatory connections ($w = 23; m = 23$). Therefore, we expected that any change in the global value of inhibition, will need to be stronger by (at least) a factor of two, compared with any change in global excitation, to exert a similar effect on the whole network. Surprisingly, however, changes in levels of global inhibition had no impact on the characteristic behaviour of the network (Fig. 5b).

Global Excitation (Experiment 2.1). With stronger excitation, the dominant sub-network (Fig. 5a) is more extended, recruiting L GH and L BA23 (posterior cingulate), which are mutually inhibitory, until some units reach a steady state. However, a small decrease (5%) in global excitation, stops the characteristic activity pattern from developing. In the network associated with initial activation of L BA48 (fusiform) (rather than initial recruitment involving one unit of the dominant sub-network), a 2% reduction in excitation prevents activity from propagating.

Global Inhibition (Experiment 2.1). Changing the value of only inhibitory weights in the matching network, e.g. doubling the negative weights, did not appear to alter network behaviour (Fig. 5b). Likewise, decreasing inhibition in this network to one-tenth of its original values did not affect network behaviour.

However, when excitation and inhibition were altered concurrently, varying inhibition subsequently changed the pattern of network behaviour. Furthermore, removing inhibition from the network entirely (negative weights set to 0) appeared to have an even greater effect compared with increasing the inhibitory weight values to −1. For example, with input from L BA46, when all excitatory weights were decreased by 4%, the characteristic oscillations in the sub-network of L BA46, BA10 and BA24 disappeared. However, when inhibition was eliminated while excitation was still decreased ($w = 0.336; m = 0.624$), the usual pattern of oscillations was reinstated. Nonetheless, a different combination of positive weight values $w = 0.12$ and $m = 0.78$ (20% increase from respective baseline), which normally produced oscillations, was unaffected by large changes in values of inhibitory weights (either decrease to negative weight values of −1, or decrease to 0). This configuration, therefore, appears to be even less sensitive to changes in inhibition.

Localized changes (experiment 2.2). Our modelling paradigm allows us to quantify the effect of certain parameter values on the default behaviour of the network. In this instance, we have assumed that structural deficits imply a functional change in the behaviour of a unit, decreasing the output from the unit in relation to the rest of the network. The ability to enumerate the discrete effects of these perturbations can suggest how specific interactions may happen. For example, they may explain why the network is particularly sensitive to deficits in certain units and how these deficits can be corrected (through increased activity in other units).

The matching network is particularly sensitive to decreased weight value of 20% (output) from either L BA37 or L BA24. We suggest, therefore, that these areas are more likely to be considered catalyst regions, since they facilitate activity throughout the network and in the absence of their integrity the network will be impaired. Thus, a relatively small decrement in the output from these areas prevented the characteristic, self-sustaining activity (Fig. 2a), from developing. These results are in opposition to the absence of a notable effect while the weight value from every other source unit (one at a time) was either increased or decreased by 50%. The exception to this rule was L BA21, where halving the weight did not allow the activity to propagate to the dominant sub-network. Nonetheless, the sustainable activity in the network was more sensitive to decrement in the output from L BA24 than from any of the other units.

Compensation. The critical role of L BA37 and BA24 may be accounted for by their pivotal role in the propagation of activity. From an initial input provided by BA18v (fusiform), activity propagates through a prominent input from L BA37 (IT), allowing activity to propagate through temporal L BA21 and frontal L BA47, reaching the dominant sub-network, which includes frontal areas L BA 46 and BA10, through L BA24 (AC).

Unlike L BA24, where a decrement in its efficacy required a relatively large compensatory increment in output of other units in the network (such as an increase of 40% in the weight values from left prefrontal areas L BA46 or BA10), reduction in
activity of L BA37 was relatively easy to correct for (probably because it is excluded from the dominant sub-network or due to the distributed nature of this network). An increase of 10% only in the output of one of several areas (L BA47, L/R BA21, L/R BA17/18 or R BA37) restored the characteristic pattern of oscillations. Likewise, an increase of 10% in the output of purely visual areas (L BA18, BA19d, BA17/18; Fig. 6a) appears to restore the characteristic activity patterns and perhaps even

Figure 5. Changes in global excitation (a, c) and inhibition (b, d) alter the characteristic pattern of activity in the matching (a, b) and long-delay (c, d) networks. (a) Increasing excitatory weights in the network by 20%, to \( w = 0.42, m = 0.78 \), causes transitory oscillations in the dominant sub-network. Thereafter, areas that were previously oscillating, as well as a few other contralateral areas, reach a steady state of persistent activity (L BA46, BA10, BA24, BA47, BA21; R BA47, BA21) or become inactive (bilateral GH). (b) There is no effect on characteristic behaviour (Fig. 2a) in the matching network, even when inhibitory weights are increased by 100% to \( w = -0.7, m = -1.3 \). (c) Decreased excitation (positive weight values were decreased by 20% \( w = 0.315; m = 0.585 \) ) and (d) increased inhibition (negative weight values were increased by 50% to \( w = -0.525; m = -0.975 \) ) in the long-delay network produce changes to characteristic behaviour. The typical pattern of oscillations (c) is slower to develop with decreased global excitation. Increasing global inhibition (d), however, seems to affect the regularity of oscillations in bilateral GH (most significantly). Coupled oscillations are observed among a similar network of areas (c, d), including high amplitude, regular oscillations in L BA47, BA21, BA37, R BA47 and BA21, and lower amplitude oscillations in a wider network comprising L BA46, BA10, BA24, BA23, R BA37 and bilateral GH. Increased inhibition produces increased amplitude of oscillations in all active areas. Negative weight values remain unchanged in (a, c), as do positive weight values in (b, d).
overcompensate, causing widespread oscillations. Nonetheless, some areas fail to compensate for decreased output or weight values from these two areas (L BA37 and BA24). For example, L BA46 cannot compensate for decreased weight values from L BA37, while L BA37 cannot compensate for the decrement in weight output from L BA24 (Kronhaus, 2004).

The Long-delay Network
Owing to the distributed pattern of activity in this network, changing global and local parameters did not create the extreme changes seen in the matching network.

Changing Global Excitation and Inhibition (Experiment 2.1). Increasing global excitation by 20% in the long-delay network increased the amplitude of oscillations or maximally activated several areas. However, these oscillations did not extend to other units that did not participate in the characteristic pattern. Moreover, unlike the matching network, decreasing global excitation did not have a similar catastrophic outcome (compare Fig. 5b with Fig. 2b).

Despite an initial disruption, after time-step 50, normal network activity was reinstated. Increasing global inhibition by 50% seems to affect network behavior by changing the frequency and/or amplitude of the characteristic oscillations (see, for example, L BA37, Fig. 5b), where activity of both L GH and R GH may be particularly vulnerable. Decimating global inhibition (90% decrease) caused the amplitude of oscillations in all participant units to increase to the maximum (Kronhaus, 2004; see basic results in the matching network for comparison).

Localized Deficits and the Efficacy of Compensation (Experiment 2.2). Using a similar testing procedure to the matching network, we tried to disrupt network activity by decreasing the weight values from one unit at a time. However, unlike the matching network, substantially reducing output from any unit by decreasing its output weight values did not prevent the characteristic pattern from developing. The effect of localized disruption (in the presence of global decrease in efficacy) in the matching and long-delay networks may suggest that the compensatory strategy in the matching and long-delay tasks are different. This may be related to the compensatory strategies in depressive illness, where global activity is often suppressed (Shajahan et al., 1999). Moreover, memory deficits in depressed patients, when present, appear to be accompanied by an altered pattern of neural response (MacQueen et al., 2003).

Decreasing positive (with or without the negative) weight values throughout the network by 10% was followed by individually decreasing the weight values from every source unit, following the procedure described in the matching network. This highlighted the sensitivity of this network to decreased weights (representing reduced activation as a result of structural or functional abnormalities) from L BA24 (Fig. 6b) and, surprisingly, L GH (by 10 and 20% respectively). Hence, these areas are potential catalysts in the long-delay network.

The latter was unexpected because unlike the key participants in the dominant sub-network (L BA47, BA21, BA37 and BA24; all linked by excitatory reciprocal connections), L GH is excited by L BA37 yet inhibited by L BA21, two of the constituents of this excitatory network (see Fig. 3b). Furthermore, L GH does not directly excite any of the constituents of this sub-network nor trigger the activity of this sub-network when it is set in a

Nonetheless, an analogous decrement of 20% in weight values from any other source unit could not disrupt network activity.
Specificity of Network Dynamics
Since the dynamics of our networks were not easily tractable (analytically), we aimed to characterize robust behaviours by exploring a range of parameters. Since weight values were determined on the basis of empirical data from McIntosh et al. (1996), we limited our study to the original values. Larger weight values produced a stronger coupling among an extended network while weaker weights were insufficient to produce sustainable activity. The shape of the activation function was important in determining network activity. However, subtle changes did not prevent oscillations from occurring.

We also considered the effect of activating several areas upon network activity, since our simulations proved that robust oscillations could be triggered by a single input. The cumulative effect of more than one input being set to 1 in $a_0$ of the matching network, achieved the expected activation of either one or more sub-networks. Since the dominant force in this network is excitatory, no specific combination of initial activations could exert an inhibitory effect. In the long-delay network the activation of more than one unit changed the pattern of oscillations in the other units. However, this was typically expressed in the increase of either the amplitude or the frequency (or both) of oscillations. Interestingly, however, if L BA18v and R BA18v are co-activated in the matching network, the activity peaks and dies down very quickly (thus there is no chance for the activation of the dominant cluster to develop), since L BA21, which provides the link to L BA46, BA10 and BA24, was important in determining network activity. However, subtle changes did not prevent oscillations from occurring.

Discussion
Dynamical activation of the effective connectivity models suggested by McIntosh et al. (1996) highlighted a critical role for the integrity of activation of BA24 (AC) in both perceptual matching and long-term maintenance of patterns in working memory. Our model revealed fundamental differences between networks characterizing two experimental conditions, with and without a working memory component. We found that the long-delay network becomes particularly vulnerable to perturbations when global excitation is reduced. Compensation, termed as restoration of the typical pattern of oscillations, is possible in both networks, through different mechanisms. Visual areas can readily compensate for disrupted activity in the matching, but not in the long-delay network. Conversely, BA46 and BA10 (PFC) are more effective at compensating for disrupted activity in the long-delay network. Nonetheless, visual areas (BA18v, BA19d, BA17/18) compensate more easily for decreased efficacy of BA24 and BA37 (IT) in the matching network (with a 10% increment) compared with the PFC's ability to compensate for decrement in activity BA24 and GH in the long-delay network (requiring a 40% increment in efficacy). Finally, the minimal constituents in the matching and long-delay networks suggested different connectivity structures, expressing feed-forward and recurrent characteristics respectively. In the basic framework, the latter implied the comparatively robust response of the long-delay network to minor efficacy changes.

A more general question is how long we would expect activity to persist in these networks with presentation of only one input. For example, the matching network is only supposed to perform a comparison with minimal involvement from the PFC, while longer delays lead to greater recruitment of the PFC (Sawaguchi and Yamane, 1999). However, in our matching network, the PFC appears to be recruited almost instantly and the activity in this cluster persists beyond what we would expect visual activation to be associated with. By contrast, the long-delay network is expected to support maintenance and storage of patterns, perhaps by maintaining persistent activity and therefore, the sustained oscillations we observed in this network are not surprising.

Caveats
Without any informed expectations of the types of activity that would result, we defined activity in the network as a measure of co-activation or functional clustering among units. Several model parameters had to be guessed (such as a limiting function) or approximated (such as the choice of weight values). Therefore, some of the behaviours of these networks are a product of imposed decisions.

Modelling inhibitory and excitatory connections in this network was suggested to be emulating neural inhibition and excitation. Since anti-correlations (expressed as negative path coefficients) are not likely to represent direct inhibition and furthermore, the connections between cortical areas are mostly glutamatergic (Somogyi et al., 1998), this assumption may be incorrect. Moreover, the units in this effective connectivity network (McIntosh et al., 1996) are based on previous analysis of PET data (Haxby et al., 1995) and were sufficient to describe it. However, certain latent interactions may be present yet not accounted for by the effective connectivity model. Though not
necessary to describe the PET data, their presence in our models may have changed both the dynamical activity (perhaps modifying the excitatory loop between BA46, BA10 and BA24 in the matching network) as well as the response of the network to changes in local or global efficacy.

**Neural Context**

McIntosh et al. (1996) suggested that finding performance related regions may be achieved through studying their interactions with other brain regions, termed 'neural context'. Catalyst regions (McIntosh, 2004), through their extensive connectivity and contribution to several networks, can transcend the functional hierarchy depending on the contextual framework (i.e. goal and co-activated regions). According to this theory, catalyst regions subserve contingencies between structure and function allowing for rapid responses to a constantly changing environment. This facilitation of different states, rather than participation in any given state per se, identifies these areas as 'behavioural catalysts' (McIntosh, 2004). In the context of our study, catalyst regions can be defined as areas that are crucial for propagation for activity within a given network. Therefore, network dynamics are highly dependent on the structural and/or functional integrity of catalyst areas. The integrity of L BA24 (AC) was crucial to both networks explored here. Participating in both the matching and the long-delay networks, the AC appears to facilitate different state transitions. Furthermore, its crucial role in both within and between state facilitation suggests that this is a catalyst region.

For illustrative purposes, we consider the mood-related neural context of activity in these networks. In this case, the neural context will be set by the clinical state of patients with major depressive disorder. Both exploring activity in structurally vulnerable regions (such as the AC) and investigating the role of these regions in task-associated networks (matching and long-delay) highlighted a key role for BA24 in propagation and maintenance of activity. Our model suggests that global suppression of the network [possibly an analogue of decreased neuronal excitability (Shajahan et al., 1999) in depressive illness] may confound the effects of discrete deficits in BA24 and further compromise the integrity of the network. Thus, our model enables discrete measurement of the stability of the network, simulating the possible dynamics engendered by compromised function.

**Solving the Task**

Although this model was not designed to study how the delayed match to sample task could be solved, this paradigm encodes the connections in two delay conditions (McIntosh et al., 1996) and suggests several differences between them. These, in turn, may show how different networks are able to solve the delayed match to sample task. However, the results can be evaluated only within the limits of the present modelling framework. Creating self-sustaining activity in the matching network, for example, is superfluous to the recognition task for which this network is putatively specialized, yet in the matching network there is no need to maintain this activity over time. Further, it is unreasonable to assume that the activity is a product of only one input. This, along with other imposed parameters, was chosen to enable understanding the minimal interaction in this complex structure.

Nevertheless, if we can accept that activity in the matching and long-delay networks is qualitatively different, we would suggest that the significance of the PFC-AC loop (the dominant sub-network; Fig. 3) may express the dissociation or transition between the straight-forward matching task (the initial burst in both hemispheres; Fig. 2a) and recruitment of frontal areas to account for the increased working-memory load, associated with different delays.

Further, we have suggested that persistent oscillations in the network are a product of the mode of input (setting a single unit 1 in \(a_n \)) and the discrete update function. Nonetheless, one is tempted to relate the persistent activity among the closely coupled L BA47, BA21 and BA37 in the long-delay network (Fig. 2b) to persistent activity recorded in the IT cortex (represented here by L BA37) and PFC in monkeys (Miyashita, 1988; Wilson et al., 1993) performing a similar task. This persistent activity is thought to maintain familiar patterns in memory until a new stimulus is presented. Our dynamical modelling of coupling among areas in the long-delay network has captured the unique activation pattern associated with task performance in non-human primates.

**Inferences**

The exact nature of the interplay between excitation and inhibition in clinical conditions such as depressive illness is unclear. Though there are numerous findings reporting deficits in different neurotransmitters and neuromodulators, there is no lucid delineation between these deficits and behavioural or cognitive dysfunction. We introduced global changes in excitation and inhibition to examine whether the matching and long-delay networks have a similar response to these perturbations and found that the long-delay network was more sensitive to those. Furthermore, we wished to associate these perturbations with deficits, such as decreased cortical excitability in depressed patients (Shajahan et al., 1999). In this sense, we would like to argue that decreased cortical excitability may be functionally expressed in decreased positive correlations among brain areas. Nonetheless, although various perturbations were discussed in the context of the neurobiological deficits, we would suggest that positive and negative correlations (termed excitation and inhibition) may express coherence, or the absence of it, among different regions. Therefore, if we can accept that characteristic activation (oscillatory pattern) can be both task-independent and sustainable, this pattern was differentially affected in the matching and the long-delay network.

Hence, this study suggests a fundamental difference in the manner of neural recruitment in tasks that do (long-delay) or do not (perceptual matching) involve a memory component. During a relatively easy task, it is possible that presence of the positive correlations is instrumental for successful recruitment of task related areas, while changes in negative correlations are insignificant. By contrast, in the long-delay network both positive and negative correlations must be maintained, for successful completion of the task. Furthermore, the long-delay network is relatively robust to low-level localized perturbation. However, this resilience is expeditiously lost when excitation throughout the network is decreased. These results suggest a putative triangulation between the decreased excitability, specific working-memory deficits and cell loss (in the AC), which are all associated with depression. Finally, our modelling framework should be treated primarily as a ‘proof of concept’. Thus, we do not claim that changes in excitatory or inhibitory efficacy are directly linked to deficits in glutamatergic GABAergic neurotransmission. However, the experimental decrease
in global efficacy may be associated with decreased intracortical connectivity as a consequence of white matter lesions in various neurological and neuropsychiatric disorders (e.g. see Kumar and Cook, 2002; Sassi et al., 2003; Schmidt et al., 2004).

Conclusions

In sum, our models have allowed us to observe the effects of network connectivity upon (resulting) network activity. The networks in our model were made more tangible by associating discrete parameters (such as weight values), described in a stationary model (McIntosh et al., 1996), and characterizing its stability in a dynamical modelling environment. Illustrating interactions between different regions, functional deficits and mechanisms of recovery, our models have shed light on systematic differences between networks of brain areas (and hence the processes) representing different cognitive tasks.

Our models demonstrate the potential benefits of using high-level neural network models to simulate neuroimaging data. Large-scale modelling strategies have been previously implemented (Tagamets and Horwitz, 1998; Young et al., 2000) in a different format. Most of the neuroimaging literature comprises of statistical models of activity and/or connectivity under different experimental constraints. The approach we describe in this paper can validate existing accounts of a non-dynamic nature suggesting discrete mechanisms to explain these models. Furthermore, this modelling approach can produce emergent behaviours which have neither been described nor predicted by the connectivity models. Furthermore, the emergent properties arising from these models often suggest further directions for research.

Notes

This work is supported by the Pintsen-Darwin Trust, the Medical Research Council and ANC. We thank Dr S.J. Eglen, Prof. A.R. McIntosh and Prof. E.T. Bullmore for their helpful comments and encouragement.

Address correspondence to Dr. Mark McIntosh, Department of Psychology, c/o Experimental Psychology, University of Cambridge, Downing Site, Downing Street, Cambridge CB2 3EB, UK. Email: dk323@cam.ac.uk.

References


