Schizophrenia: A Disconnection Syndrome?

Karl J. Friston and Christopher D. Frith

We review the evidence of pathophysiological changes in the prefrontal and temporal cortices of schizophrenic subjects and of abnormal integration of the physiological dynamics in these two regions. The argument we develop is that some schizophrenic phenomena are best understood in terms of abnormal interactions between different areas, not only at the levels of physiology and functional anatomy, but at the level of cognitive and sensorimotor functioning. We discuss recent functional imaging evidence suggesting abnormal prefronto-temporal interactions in relation to a psychological analysis of experiential symptoms in schizophrenia. Cortico-cortical interactions have been assessed in terms of functional connectivity and eigenimages, using time series of neurophysiological data obtained with positron emission tomography. The results of these analyses suggest that there is a profound disruption of large-scale prefronto-temporal interactions in schizophrenia. These disruptions are particularly relevant if one considers that many positive symptoms of schizophrenia reflect a failure to integrate intrinsically generated behaviour and concurrent perception.

Key Words: schizophrenia, prefrontal, temporal, functional connectivity, eigenimages, integration, neural activity, functional imaging

The relationship between the brain’s cognitive and sensorimotor operations and the underlying neural dynamics has been addressed, broadly speaking, in two ways. Firstly the brain is considered as an ensemble of functionally segregated areas, each subserving a specific function. This view of functional organization has its roots in early lesion-deficit studies and electrical stimulation experiments in animals [e.g., Goltz, 1881; Ferrier, 1875]. The second complementary view of functional organization emphasizes the integration of functionally specialized areas, and considers distributed interactions as the neurophysiological basis for emergent cognitive or psychological function. These two perspectives are deeply enmeshed both in modern neuroscience and historically: For example a key meeting, on the morning of August 4, 1881, to discuss “localization of function in the cortex cerebri,” addressed this issue. Goltz [1881], although accepting the results of electrical stimulation of motor cortex [e.g., Ferrier, 1875], considered the excitation method inconclusive, in that movements elicited might have originated in related pathways or current could have spread to distant centres [Phillips et al., 1984]. This dialectic, functional segregation vs. functional integration, persists today and forms the basis for this discussion of some recent observations using functional neuroimaging in schizophrenia.

The central thesis of this article is that although localized pathophysiology of cortical areas (e.g., the dorsolateral prefrontal cortex or DLPFC) may be a sufficient explanation for some signs of schizophrenia, it does not suffice as a rich or compelling explanation for the symptoms of schizophrenia. The conjecture we review here is that symptoms such as hallucinations and delusions are better understood in terms of abnormal interactions or integration between different cortical areas. This dysfunctional integration is expressed at a physiological level as abnormal functional connectivity, measurable with neuroimaging, and

Karl J. Friston was an exhibitioner at Gonville and Caius College, Cambridge, where he studied Medical Sciences, Physics, and Psychology. He completed his clinical training at King’s College Hospital Medical School, London University. Dr. Friston specialized in Psychiatry in the Department of Psychiatry, Oxford University, and started his research at the MRC Clinical Neuropharmacology Unit, moving to the MRC Cyclotron Unit in London to work in functional imaging. Dr. Friston returned to the U.K. recently, after a Fellowship in Theoretical Neurobiology at the Neurosciences Institute in San Diego, California. Dr. Friston is currently a Wellcome Senior Research Fellow in Clinical Science at the new Cognitive Neurology Department of the Institute of Neurology and Honorary Consultant Psychiatrist at the National Hospital for Neurology and Neurosurgery, London.

Christopher D. Frith studied Natural Sciences, including psychology, at Christ’s College, Cambridge, and gained a Diploma in Abnormal Psychology at the Institute of Psychiatry, University of London. He completed his doctoral and postdoctoral studies with H.J. Eysenck before joining Tim Dow’s Schizophrenia Research Unit at the MRC Clinical Research Centre. For the last few years he has worked in brain imaging in the Neuroscience Section of the MRC Cyclotron Unit, London. He is currently Professor of Neuropsychology at the new Cognitive Neurology Department at the Institute of Neurology, London.

Address reprint requests to Dr. Karl J. Friston, MRC Cyclotron Unit, Hammersmith Hospital, DuCane Rd., London W12 OHS, England.

© 1995 Wiley-Liss, Inc.
at a cognitive level as a failure to integrate perception and action that manifests as clinical symptoms. The distinction between a regionally specific pathology and a pathology of interaction can be seen in terms of a first order effect (e.g., hypofrontality) and a second order effect that only exists in the relationship between activity in the prefrontal cortex and some other (e.g., temporal) region. In a similar way psychological abnormalities can be regarded as first order (e.g., a poverty of intrinsically cued behaviour in psychomotor poverty) or second order (e.g., a failure to integrate intrinsically cued behaviour and perception in reality distortion; see below).

The idea that schizophrenic symptoms are associated with abnormal integration in the brain is supported by an analysis of the neuropsychology of schizophrenia [see Frith, 1992] and some interesting differences in functional connectivity (as measured with positron emission tomography or PET) [Friston et al., 1994] between normal subjects and schizophrenics that are reviewed in this article. The notion that schizophrenia represents a disintegration or fractionation of the psyche is as old as its name, introduced by Bleuler [Bleuler, 1913] to convey a “splitting” of mental faculties. Many of Bleuler’s primary processes, such as “loosening of associations” emphasize a fragmentation and loss of coherent integration. In what follows we suggest this mentalistic “splitting” has a physiological basis, and furthermore that both the mentalistic and physiological disintegration have precise and specific characteristics that can be understood in terms of functional anatomy.

The evidence for structural and functional abnormalities in both the prefrontal cortices and temporal lobes is strong, particularly in the left hemisphere. This evidence ranges from abnormal quantitative cytoarchitecture [Benes et al., 1986] to gross morphological changes evident on MRI scans [Bogerts et al., 1991] and postmortem [Brown et al., 1986]. Functional abnormalities have been demonstrated using PET [Liddle et al., 1992] and electrophysiology [McCarley et al., 1993].

Abnormal integration of prefrontal neural activity and activity in subcortical, limbic, and temporal structures is a common theme, found in many neurobiological accounts of schizophrenia. For example neurodevelopmental models of schizophrenia [e.g., Weinberger, 1987; Murray and Lewis, 1987] refer to the concurrent maturation of the frontal lobes, in terms of myelination, and the emergence of schizophrenic phenomena. In adolescence there is evidence for progressive changes in the nature of cortical interactions (as measured by the EEG), particularly between the left prefrontal and temporal regions [Buchbaum et al., 1992]. The pathophysiological basis of abnormal cognitive processing in schizophrenia has been discussed in terms of abnormal fronto-stra-tal and fronto-temporal integration [e.g., Robbins, 1990; Frith, 1987]. Interpretations of (glutamergic) neurochemical abnormalities, in prefrontal and left temporal regions, refer to the anatomical connections between these regions [e.g., Deakin et al., 1989].

Functional connectivity refers to the observed temporal correlation between two neurophysiological measurements.

Although there is less direct evidence for an abnormal integration of prefrontal and temporal activity there are some intriguing and suggestive observations. For example psychotic symptoms, including complex auditory hallucinations and delusions, are a prominent feature of Metachromatic Leukodystrophy presenting in early life [Hyde et al., 1992]. The pathology of Metachromatic Leukodystrophy is demyelinating affecting many systems but particularly the subcortical white matter. The possibility that the pathophysiology of schizophrenia leads to a similar prefronto-temporal functional “disconnection” suggests that functional connectivity between these two brain systems should be demonstrably abnormal. Our initial results [Friston et al., 1994] suggest that this is indeed the case.

In neuroimaging, functional connectivity is a relatively new concept and furnishes a simple but powerful way of characterizing distributed changes in the brain. A single cortical region is implicated in a distributed system by virtue of its functional interactions with other regions. The identification of distributed systems therefore relies on the characterization of neuronal interactions and correlations. In the past two decades the concepts of functional and effective connectivity have been most thoroughly elaborated in the analysis of multi-unit recordings of separable neuronal spike trains, recorded simultaneously from different brain areas in animals [Gerstein and Perkel, 1969]. Temporal coherence is interpreted as a signature of functional connectivity [Aertsen and Preissl, 1991; Gersten et al., 1989]. In neuroimaging functional connectivity refers to the observed temporal correlation between two neurophysiological measurements from different parts of the brain [Friston et al., 1993a] and can be defined as the temporal correlation between spatially remote neurophysiological events.

In general the interesting aspects of functional connectivity pertain to the patterns of correlations evoked by an experiment. These patterns define distributed systems whose components exhibit coherent activity. One characterisation of these patterns is in terms of the eigenvectors of the functional connectivity matrix. These eigenvectors are obtained using singular value decomposition or SVD to decompose a neuroimaging time-series into a series of independent patterns that embody, in a stepdown fashion, the greatest amounts of functional connectivity. Each eigenvector defines a distributed brain system that, when displayed as an image, is referred to as an eigenimage or spatial mode. Eigenimages have been used to capture the spatio-temporal dynamics of neurophysiological time-series from several modalities including: multi-unit electrode recordings [Mayer-Kress et al., 1991], electro-encephalography [Friedrich et al., 1991], magneto-encephalography [Fuchs et al., 1992], PET [Friston et al., 1993a], and functional magnetic resonance imaging [Friston et al., 1993b]. Many readers will notice that the eigenimages associated with the functional connectivity or covariance matrix are simply the principal components of the time-series [e.g., Friston et al., 1993a]. In what follows eigenimages will play a central role in revealing the important features of functional connectivity in normal subjects and in schizophrenia.
The remainder of this article is presented in three sections and concludes with a discussion. The first section summarizes the nature of the data we will use to illustrate these ideas, the experimental paradigm, and the subjects involved. The second section characterizes and compares functional connectivity using the most prevalent eigenimage or spatial mode from each group studied. The third section presents the eigenimages that embody the functional connectivity expressed by normal subjects but not in schizophrenic patients and conversely those patterns of coherent activity that are seen in schizophrenia but not in normals.

The data were acquired from four groups of six subjects with a PET camera (CTI model 9310B/12, Knoxville TN) using a fast dynamic H2O technique. Each subject was scanned six times during the performance of three-word production tasks. The order of the tasks was balanced for time effects (A B C C B A). Task A was a verbal fluency task, requiring subjects to respond with a word that began with a heard letter. Task B was a semantic categorization task where the subject responded “man-made” or “natural,” depending on the heard noun. Task C was a word shadowing task where the subject simply repeated what was heard. In the current context the detailed nature of the tasks is not very important. They were used to introduce variance and covariance in cerebral activity that could support an analysis of functional connectivity. However, the choice of tasks is crucial in the sense that they engage specific brain systems, and consequently constrain the cortico-cortical interactions that can be examined. The tasks used typically evoke large and widespread regional cerebral blood flow (rCBF) changes in prefrontal, cingulate, and temporal regions [Friston et al., 1991a, b; Friston et al., 1991a].

All the images were stereotactically normalized and mapped into a standard anatomical space [Friston et al., 1991b; Talairach and Tournoux, 1988]. Differences in rCBF due simply to whole brain differences were removed using ANCOVA [Friston et al., 1990]. A mean rCBF estimate for each voxel, for each of the six conditions (scans), for each group, was obtained by averaging over the six subjects in each group. A subset of voxels was selected, for subsequent analysis, if the differences between any of the six scans accounted for a significant amount of variance (ANOVA F > 3.9, P < 0.001, df 5.24) in one or more of the four groups. The result was a large (mean corrected and Euclidean normalized) matrix (M) of rCBF estimates for each of the four groups, comprising six rows (one for each scan) and 4,802 columns (one for each voxel).

The four groups comprised one group of six normal subjects and three groups of six schizophrenic patients [DSMIII-R; American Psychiatric Association, 1987]. The patients were all relatively chronic, medicated, stable, and middle-aged. The schizophrenic groups were categorized according to their performance on a series of verbal fluency tasks. The first group (poverty) produced less than 24 words on a standard (1 minute) FAS verbal fluency task. The other groups all produced more than 24 words. The second schizophrenic group (odd) produced neologisms and/or words not in the semantic category specified and/or five or more unusual words. Unusualness was defined using the Battig and Montague [1969] category norms. The third group was unimpaired according to the above criteria. Although this categorisation is implicitly in terms of performance on psychological tests germane to the activation paradigm employed, the three groups can be loosely identified with the three dimensions commonly found in clinical ratings of schizophrenia [Liddle, 1987; Bilder et al., 1985; Mortimer et al., 1990; Arndt et al., 1991]: (1) psychomotor poverty, characterized by reduced speech, spontaneous movement, and flattened affect; (2) disorganization, associated with inappropriate affect, thought disorder, incoherence, and poverty of content of speech; and (3) reality distortion, with hallucinations and delusions but less neuropsychological impairment [Liddle and Morris, 1991].

The first eigenimage, based on the data from normal subjects, is presented in Figure 1 (top panels). This eigenimage (e1) is found by solving:

\[ C \cdot e_1 = e_1 \cdot \lambda_1 \]

where \( C = M^T \cdot M \) is the functional connectivity matrix. The positive and negative aspects are presented at the left and right respectively. This pattern involves primarily the left DLPPC (positive) and the superior temporal regions, bilaterally, in auditory and periauditory cortex, with left posterior infero-temporal contributions (negative). The corresponding eigenvalue (\( \lambda_1 \)) is seen on the lower left and is more than twice the eigenvalue of the second eigenimage. Each eigenimage has an associated eigenvalue that reflects the amount of variance-covariance that eigenimage accounts for. The first two eigenimages account for nearly all the variance-covariance observed. This is not uncommon in the analysis of activation studies. The vector on the lower right reflects the degree to which the first eigenimage was expressed during the six scans. It is obvious that this eigenimage is maximally expressed during the verbal fluency tasks and minimal so in the word-shadowing tasks. In summary, normal subjects evidence profound negative prefronto-superior temporal functional interactions associated with intrinsic word generation. This is almost an exact replication of previous work [see Friston et al., 1993a].

The positive parts of first eigenimages from the three schizophrenic groups are presented in Figure 2. There are several important observations to be made. First the three patterns are remarkably similar. This is pleasing, considering they were based on independent data. Secondly they are very different from the normal pattern. Although frontal-temporal interactions are the most prominent features of these eigenimages (1) the correlations are positive, (2) are most evident between left prefrontal and left infero-temporal and tempor-insular regions, and (3) the temporal component of this system is not symmetrical. The negative parts of the eigenimages are not shown because they were very scant. In summary all three groups of schizophrenic patients exhibited very similar patterns of func-
tional connectivity, involving positive interactions between left DLPFC and left inferotemporal and middle temporal regions.

Below we present a more direct analysis of the differences between the patterns of correlated activity between normal subjects and schizophrenic patients. Here we identify the eigenimage that reflects the functional connectivity in the normal subjects not expressed in a schizophrenic group (d1). This eigenimage is obtained by using a generalized eigenvector solution:

\[ C_p^{-1} \cdot C_n \cdot d_1 = d_1 \cdot \lambda_1 \]
\[ C_n \cdot d_1 = C_p d_1 \cdot \lambda_1 \]

where \( C_p \) and \( C_n \) are the poverty schizophrenic and normal functional connectivity matrices, respectively.

The results of this analysis are presented in Figure 3. It is of no surprise to see that the pattern which best captures the differences is very similar to the pattern that is most prevalent in the normal subjects, namely negative correlations between left DLPFC and bilateral temporal regions (Fig. 3, upper panels). The amount to which this pattern was expressed in each individual group is shown in the lower panel of Figure 3 using the appropriate 2-norm \( \| d_1 \cdot C \cdot d_1 \| \). This measure simply reflects the amount to which a pattern \( d_1 \) contributes to the variance-covariance structure C (compare with the eigenvalue). It is seen that this eigenimage, whilst prevalent in the normal subjects, is virtually absent in all the three schizophrenic groups. Equivalently the substantial prefronto-temporal functional connectivity within the system portrayed in Figure 3 is not found in the schizophrenics. It is important to note that only one of the schizophrenic groups (the poverty group) was used to define this eigenimage and yet it is absent in the remaining schizophrenic groups.

Figure 4 shows the complementary analysis where the eigenimage was chosen to show functional connectivity prevalent in the poverty schizophrenic group that was minimally expressed in the normal group. The main feature of this functional connectivity pattern was a positive left prefronto-left temporal interaction. Again although the data from the odd and unimpaired group were not used to constrain the eigenvector solution, this pattern is also found in these groups.

In this article we have reviewed an analysis of cortical interactions in normal subjects and schizophrenic subjects using the concept of functional connectivity. Compared to normal subjects the schizophrenic groups showed a very different pattern of
distributed cerebral interactions. The nature of this difference was remarkably consistent across the groups of schizophrenia analyzed. The main differences between normal subjects and patients was a dissociation in terms of regionally specific prefronto-temporal functional connectivity: (1) Profound left prefronto-bilateral superior temporal negative interactions were seen in normal subjects but not in schizophrenia and (2) marked left prefronto-left temporal positive correlations were marked in the schizophrenic groups but not in normals. These results indicate not only regionally specific and consistent differences in functional connectivity, but a complete reversal in the nature of the large scale prefronto-temporal interactions. This reversal can be regarded as a failure of prefrontal cortex to suppress activity in the temporal lobes (or vice versa).

Underlying our thinking about the nature of fronto-temporal interactions is an assumption that the prefrontal cortices are necessary for intrinsically generated behaviour and that the temporal cortices (in the context of word generation) are necessary for extrinsically generated percepts. The evidence for the former association derives from studies of patients with neurological problems, unit recording studies in behaving primates, and functional imaging studies during cognitive activation. Psychomotor poverty in schizophrenia is closely related to psychomotor retardation [Benson, 1990] which includes decreased spontaneous movement, decreased communication, flatness of vocal inflection, unchanging facial expression, and social withdrawal. Among the most common causes of psychomotor retardation is damage to the frontal lobes and Parkinson's disease. In particular, patients with damage to the anterior cingulate and/or supplementary motor area (SMA) tend to become mute and show decreased spontaneous movement [Damasio and Van Hoesen, 1983]. Passingham et al. [1989] have demonstrated that SMA and anterior cingulate lesions in monkeys impair responses when the movement is self initiated, but have little effect when the movement is extrinsically cued. Largely akinetic patients, with Parkinson's disease, can show a dissociation between movements which are self-initiated and extrinsically cued (e.g., paradoxical kinesis) [Marsden et al., 1982]. In functional mapping it is now well established that the DLPFC is concerned with intrinsically generated behaviour that is implicitly mnemonic [e.g., Frith et al., 1991a; Friston et al., 1992].

On the basis of this sort of evidence one might propose that regionally specific pathophysiological change in the prefrontal cortex is a sufficient explanation for the negative signs of schizophrenia. There is however another possibility; that normal fronto-striatal integration is necessary for intrinsically generated behaviour. Robbins [1990] has reviewed the case for fronto-striatal dysfunction in schizophrenia. He comes down in favour of such a hypothesis and concludes that the pathophysiological
basis of schizophrenia is unlikely to be found in a single area but is more likely to be associated with dysfunctional integration in the cortico-striatal loops. While these observations address negative signs (psychomotor poverty), they do not directly relate to the positive signs (disorganization syndrome) and experiential symptoms (reality distortion) of schizophrenia. We now discuss the role of abnormal fronto-limbic and fronto-temporal interactions in the genesis of positive symptoms.

The generation of intrinsically cured behaviour is, in normal circumstances, one aspect of a coherent integration of perception and action, dictated by the history of the individual (the history may be recent and reflected in neural dynamics or embedded in associative changes in neural connectivity). The point is that intrinsically generated action cannot be divorced from perception. Intrinsically generated behaviour corresponds to an adaptive response to concurrent stimuli, informed by past events. In responding the individual operates on the environment and causes perceived changes, if only at the level of proprioception. The coherent temporal succession of self-initiated action and perception depends on a continuous dialogue between neural systems responsible for executing motor behaviour and sensory systems that register the effect. This dialogue may be mediated by connections between prefrontal cortex and appropriate sensory systems to integrate the sensed and expected consequences of acting. An extremely useful metaphor for this sort of neuronal interaction is found in the oculomotor system [see also Feinberg, 1978].

Helmholtz [1866] pointed out that when we move our eyes the image slips across the retina and yet we perceive the world as stationary. This phenomenon can be accounted for by corollary discharge or re-afference copy [Sperry, 1950; von Holst and Mittelstaedt, 1950]. Robinson and Wurtz [1976] identified cells in the superficial layers of the superior colliculus that respond to moving stimuli but do not respond when the eye is moved across a stationary target. At that time they tentatively identified the frontal eye fields (in the prefrontal cortex) as the source of modulating corollary discharge. Similar selectivity of unit responses for extrinsically and intrinsically caused sensory changes is found in the auditory system: Muller-Preuss and Jurgens [1976] identified cells in the auditory cortex of squirrel monkeys that respond to extrinsically...
generated sounds, but not to self-generated vocalization. Ploog [1979] concludes that the inhibition of these cells is caused by corollary discharges associated with vocalization, possibly from the anterior cingulate (in the prefrontal cortex). The anterior cingulate projects not only to Broca’s area but also auditory areas, including Brodmann’s area 22.

The picture which emerges is of modulatory interactions between prefrontal cortex and posterior sensory cortex that serve to integrate perception and action. A failure of this integration, or at a neurophysiological level dysfunctional prefronto-temporal connectivity, may compromise (1) intrinsically generated action, secondary to disintegration between (prefrontal and premotor) intentional and (discriminative and proprioceptive) sensory systems, and (2) aberrant perception resulting from misattribution of a self-induced sensory change to the external agencies. Many of the symptoms of schizophrenia can be sensibly understood in terms of the second impairment. For example:

Consider the effects of a failure to modulate neuronal activity associated with intrinsically generated semantic and syntactical representations in superior temporal regions. This might result in the perception of formed utterances and words that are experienced as extrinsic. The activation of semantic representations (in the absence of properly functioning afferents from prefrontal and cingulate cortices) is a likely concomitant of neural activity in premotor speech areas (e.g., Broca’s area or BA 44) associated with sub- or pre-vocal speech (profound changes in neural dynamics are observed many hundreds of milliseconds before actual movement in other motor systems).

A similar disintegration can be envisaged in terms of (putative) infero-temporal [Perret et al., 1986; McCarthy and Warrington, 1990] unit activity evoked by particular facial expressions. A failure to integrate responses elicited in others, with the social behaviour eliciting that response, might result in the experience of inappropriate expressions on the part of the listener. Not only would this lend itself to a malignant interpretation but would disrupt any coherent discourse. In this example the sensory representation evoked by an intrinsically generated behaviour has been mediated by the patient acting on the world. In the example above the semantic representations were mediated directly by neural interactions. In both cases they were not properly integrated and were experienced as being caused by another.
Although not easily framed in terms of fronto-temporal interactions, delusions of control and “made acts” are related to the above by noting that proprioceptive input could be subject to the same dysfunctional reentrant modulation as neural activity in cortex concerned with the discriminative senses. In this case failure to suppress proprioceptive input directly resulting from volitional acts might be wrongly attributed to an extrinsic cause.

In conclusion, intrinsically generated behaviour and the integration of that behaviour into the perceptual domain depends on coherent interactions between prefrontal cortices and those devoted to perceptual representations. Of the many interactions among these systems we have focused on efferents from the prefrontal cortex to sensory systems and their role is modulating target activity. This modulatory role may be as simple as suppressing responsiveness to self-induced sensory changes (for which electrophysiological evidence exists) [Muller-Press and Jurgens, 1976] or they may reflect dynamic interactions that are less easy to characterize. Many signs and experimental symptoms of schizophrenia impinge on the relationship of self to others as mediated by language and expression (or absence of it) (e.g., poverty of speech, inappropriate or flat affect, incoherent speech, decreased content of speech, auditory hallucinations, and paranoid delusions). The functional anatomy of language-related perceptual representations centres on the temporal regions. It follows that there might be something quite specific about fronto-temporal disconnection and schizophrenia. In this regard structural MRI studies of schizophrenic brains have found abnormalities in the superior temporal gyrus and underlying white matter with some consistency [e.g., McCarley et al., 1993].

ACKNOWLEDGMENTS

We would like to thank all our colleagues in the Neurosciences section at the MRC Cyclotron Unit for their contribution to this work at both practical and theoretical levels, in particular Seigrid Herold, Paul Fletcher, and Peter Liddle.

REFERENCES


Robbins TW (1990): The case for frontostriatal dysfunction in schizophrenia. Schizophr Bull 16:391–402.