Moving Between Functional and Effective Connectivity

Anthony R. McIntosh, PhD

Rotman Research Institute
Baycrest Centre for Geriatric Care
University of Toronto
Ontario, Canada
Introduction
One of the current challenges in neuroscience is to understand how brain operations give rise to mental phenomena ranging from sensation and perception to memory and attention. We are getting to know a great deal about how the brain functions in basic sensory and motor systems. For higher mental functions, however, a long scientific battle has been raging as to whether such functions are localizable. A dominant assumption in neuroscience is that certain parts of the brain play unique roles in mental function. This idea of one region/one function comes from early studies that showed some remarkable cognitive deficits due to lesions in specific parts of the brain. Up until the last 15 to 20 years, the tools available to neuroscientists have allowed them to examine only small parts of the brain at a time; their findings, although limited, have reinforced the notion of discrete functions in specific brain regions.

Modern neuroimaging tools allow us to measure how the entire brain reacts as people perform different mental operations. We are finding that many more brain areas “light up” when someone pays attention, thinks, and remembers than we would have expected based on the results from brain lesion studies. However, many researchers in the field, who continue to focus on one or two critical brain regions, overlook this new information.

The brain is made up of individual elements: from cells to neural ensembles. These elements are connected, so their individual actions can be combined through their interactions. The combined responses of small groups of cells give interacting brain areas a rich response repertoire, ranging from simple sensation to consciousness and reason. When neuroimaging data are examined in terms of brain interactions, it is observed that many regions cooperate in our thought processes. Emerging neurobiological theories emphasize the combined actions of interacting brain elements (cells to ensembles to regions) as the link between the brain and human mental function (McIntosh, 2000a,b).

From a network perspective, anything that affects the integrity of a specific brain region will necessarily influence the operation of the entire network or networks in which this region participates. Behavioral deficits following damage, or arising from disease processes, could thus reflect either the abnormal operation of a damaged network, or the formation of a completely different network with a new behavioral repertoire. Thus, much could be learned about brain dysfunction (as well as normal function) by examining network operations in subjects where mental functions are compromised by damage or disease.

If normal brain function and dysfunction result from the action of distributed networks, then analytic approaches tuned to such dynamics would best capture these actions. What follows reviews some of the basic methods that have been used for network analysis and presents the underlying theory for applying and developing a new perspective that serves to unite the understanding of brain function and dysfunction within one framework.

Theoretical Basis and Tools for Network Analysis
Network analysis, as applied to neuroimaging, can be considered a collection of analytic methods: e.g., interregional correlations/covariances or the corresponding measure in the frequency domain, such as coherence. These methods attempt to measure the interdependency among brain areas during different cognitive states. The driving assumption behind the use of these approaches is that the correlations/covariances of activity measure neural interactions. Neural interactions refer, in a general sense, to influences that different elements in the nervous system have on each other via synaptic communication; the term “elements” refers to any constituent of the nervous system, either a single neuron or collections thereof.

Traditional approaches to understanding neural interactions have focused on studying systematic variation in activity with some manipulated parameter. However, activity changes in one neural element usually result from a change in the influence of other connected elements; thus, focusing on activity in one area will cause one to miss the change in afferent influence. Furthermore, it is logically possible for the influences on an element to change without an appreciable change in measured activity. The simplest example is where an afferent influence switches from one source to another, without a change in the strength of the influence. For example, in the feed-forward network depicted in Figure 1, region C may show similar activity patterns when influence from either A or B is strong. Therefore, monitoring regional activity alone would not be able to differentiate the source of the effects, but measures of the relation of activity between elements (e.g., path v versus w) would be able to.

The measurement of neural interactions in neuroimaging has developed under two general approaches. The first emphasizes pairwise
interactions, often in terms of correlations or covariances. The second incorporates additional information, such as anatomical connections, and considers a simultaneous interaction of several neural elements to explicitly quantify the effect one element has on another. These two approaches are known as “functional connectivity” and “effective connectivity,” respectively. Both terms were introduced in the context of electrophysiological recordings from multiple cells (Aertsen et al., 1987) and have been used with reference to neuroimaging data (Friston et al., 1993; Friston, 1994; Horwitz, 2003).

Although the majority of studies of network interactions have focused on the young healthy brain, tremendous gains could be had in studying these operations in other age-groups and across patient populations. From a developmental perspective, the fact that brain structure changes across the entire life span has obvious implications for network operations. Even where overt behavior does not show an age-related change, there may be quite different sets of regional activity and interactivity between age-groups (Grady, et al., 2003).

In clinical populations, the network reorganization may be even more dramatic, depending on the source of pathology. In cases of brain damage, observed network reorganization will likely be two-fold: primary response to damage (degeneration, diaschisis); and secondary responses, as the networks reconfigure in an attempt to adapt to the insult (compensatory mechanisms). An interesting implication is that some of the behavioral deficits may reflect the secondary response. In degenerative disorders, a similar reorganization likely occurs, although over a more protracted time scale. Finally, mental disorders (e.g., schizophrenia, major depression) also will affect the integrity of network operations (Jennings et al., 1998; Welchew et al., 2002; Seminowicz et al., 2004).

Recent modeling work suggests that the patterns of resting state connectivity directly result from the anatomical and functional architecture of the brain (Honey et al., 2007). In a simulated network, as the architecture and dynamics more closely approximate real neural systems, distributed patterns of functionally connected networks emerge. Remarkably, the spatial patterns of these networks resemble those reported in functional magnetic resonance imaging (fMRI) experiments. If the anatomical structure is perturbed or the dynamics changed, then the patterns break down. Such findings suggest that resting state connectivity may indeed index the integrity of a given brain. This premise has been substantiated by empirical observations of resting state correlations in normal aging, showing a reduction in overall functional connectivity.
particularly among frontal and parietal regions. The reduced functional connectivity correlates with a decline in behavioral measures of executive function and overall processing speed (Damoiseaux et al., 2008). Degenerative disorders such as Alzheimer’s disease also show reduced functional connectivity related to disease severity (Stam et al., 2006). Taken together, these empirical and modeling findings suggest that resting state correlations may act as a useful “fingerprint” for the integrity of functional networks.

Functional and effective connectivity can also show task-dependent changes. Horwitz et al. (1992), using positron emission tomography (PET), showed functional connectivity patterns that mapped on to the use of “what” versus “where” cortical visual pathways. Effective connectivity analyses of these data (McIntosh et al., 1994) showed task-dependent switches in prefrontal feedback and strong suppressive interactions between “what” and “where” pathways.

Importantly, effective connectivity can differentiate between top-down versus bottom-up effects. Category-specific responses have been observed frequently in the ventral occipitotemporal lobe (e.g., fusiform gyrus for faces, parahippocampal gyrus for places), which is typically considered a top-down effect. Using dynamic causal modeling (DCM) (Friston et al., 2003) to estimate effective connectivity, Mechelli and colleagues (2003) found that early sensory areas changed their effects on category-specific areas in relation to the stimuli, but higher-order association regions did not show such changes in effects. Thus, category specificity in these data was a bottom-up effect. It is likely that the real story of the neural instantiation of category specificity is an outcome of reciprocal interactions among neural sites. The results from the effective-connectivity analyses have enriched models of cognitive function by moving them beyond strict hierarchical representations and emphasizing the dynamic and interactive nature of neural instantiations.

**Major Steps in Network Analysis**

The progression from data collection to the final stage of a network analysis will depend, largely, on the question one asks of the data. Assuming a comprehensive analysis is planned, the steps can be outlined as follows:

1. **Perform activation analysis.** This is the usual first step in any image analysis. It is reasonable to assume that regions showing similar activity changes between tasks may also be part of the same functional network, though this is not a certainty (Stephan, 2004). While the typical mode of activation analysis uses a univariate approach, multivariate approaches may be preferable when one is attempting to identify cohesive networks. The primary reason is that, where there are dependencies among measured (dependent) variables, multivariate approaches will have greater sensitivity because they explicitly make use of these correlations (Lukic et al., 2002).

2. **Relate brain activity to behavioral measures.** Although activation analysis is the most common approach in neuroimaging, a growing number of investigators are relating activity patterns to either performance measured during the experiment or to demographic measures. In the former instance, the brain-behavior analysis may be considered as completing a “causal chain”; that is, the activation analysis would be most sensitive to the input side of the chain, and behavior analysis to the output. Combining brain-behavior analysis with activation analysis can be seen as getting the most comprehensive coverage of most, if not all, regions that are part of the functional network for a given task. Finally, relating the patterns of functional or effective connectivity provides an anchor for interpretation and confirms that the patterns of interactivity actually “make a difference” in performance.

3. **Analyze functional connectivity.** Once the candidate nodes are collected, the pattern of interactions can be used to examine functional connectivity. Probably the simplest approach to this analysis is calculating pairwise correlations/covariances. Functional connectivity estimates can be compared across tasks or groups to define dependencies on this dimension.

4. **Analyze effective connectivity.** Functional connectivity can be easily assessed across any number of regions of interest, but effective connectivity requires a more focused approach wherein a subsection of regions identified from the previous steps are considered for more detailed models.

Some neuroscientists are concerned about which source of variance, across tasks or across subjects, is best for estimating neural interactions (Friston, 1995; Strother et al., 1995a,b). The issue of which source of variability is “correct” is not unique to neuroscience (Mandler, 1959), and there is no necessity for making a logical connection between covariances computed across tasks within-subjects, and those computed across subjects within-task. However, there is also no justification for preferring one source of variability to another, particularly in cases where both can be examined, as in fMRI or event-related potential

© 2010 McIntosh
(ERP) studies. Within-subjects analysis assesses the direct relation between regions, while across-subjects analysis indicates the stability of that relation. These are complementary, not contradictory, pieces of information.

For illustration, say we chose ten people of varying heights and weight and asked them to pull on a potentiometer by flexing their arm (an arm curl). If you measured muscle activity in the arm of each subject, say through blood flow, and correlated them, you would probably find a strong correlation with the biceps and brachialis muscles. Although each person would differ in the amount of blood flow to the muscles, from the correlation based on this variance, you would conclude that the muscles on the ventral surface of the arm have something to do with flexion. If, instead, you measured muscle activity in a single subject with a progressive increase in the resistance to arm flexion, you would find a correlation between muscle activity in the ventral part of the arm. Replicating the measurement by running different subjects would lead you to the same conclusion you had reached by using the between-subjects covariance. The point here is that computing covariances between or within subjects can lead to complementary conclusions, so long as there are adequate experimental controls and the statistical analysis ensures the answers are reliable.

It should not be taken as a suggestion that all the network analysis steps listed above must be carried across to every data set. Obviously, the choice of analysis (functional connectivity or effective connectivity) depends on the particular question one has to ask of the data. Functional connectivity analyses are likely satisfactory when the goal is in the exploratory/explanatory mode. For example, if a peculiar activation pattern were noted in one group, assessing the functional connectivity of that region with the rest of the brain could help explain the peculiarity in terms of a difference in the pattern of interactions in that group, relative to controls. On the other hand, if the question were phrased in terms of directed influences, then analysis of effective connectivity would be needed. For example, if the question was whether top-down influences from prefrontal to temporal cortices vary between groups, an analysis of effective connectivity must be performed to distinguish top-down from bottom-up effects.

**Taxonomy of Techniques**

One has only to casually flip through an issue of *NeuroImage* or *Human Brain Mapping* to realize that methodological developments in the estimation of functional and effective connectivity are exploding. The sections below briefly characterize the major methods used for estimating connectivity and list their advantages and disadvantages. This is by no means an exhaustive list.

**Functional connectivity**

**Regional correlation**

This is perhaps the simplest and most often used method. Pairwise correlations of regions of interest, or voxels, provide a snapshot of functional connectivity patterns (Horwitz et al., 1984, 1991). This method has the advantage of simplicity and uses a minimal number of assumptions beyond linearity. Where the technique becomes problematic is when the number of correlations grows and one must correct for multiple statistical tests (the same problem as with other univariate measures). Moreover, as the number of correlations grows, easily summarizing the patterns becomes difficult. It is at this point that multivariate methods may be helpful (see below).

**Psychophysiological interactions**

Linear regression methods sometimes appear to lie in a gray area between functional and effective connectivity. For example, the method to estimate psychophysiological interactions (PPIs) (Friston et al., 1997) in the statistical parametric mapping (SPM) package is used to assess task-dependent changes in the degree that one region (Y) predicts or explains the activity of another (X) (McIntosh and Gonzalez-Lima, 1994). However, the PPI approach provides the same statistical result as would be obtained if the roles of X and Y were reversed. Thus, the PPI method is most similar to an estimate of functional connectivity.

**Principal component analysis**

A tried-and-true method, principal component analysis (PCA) has been applied to a number of neuroimaging data sets to summarize complex patterns of interregional correlations. It is a helpful means to follow from the calculation of pairwise correlations. The PCA solutions are always unique for a given data set (compared with those of independent component analysis [ICA]), and the calculation of the principal components is relatively fast. The main drawbacks include:

- **Orthogonality of components**, which may impose artifactual groupings within a component. This effect can be alleviated somewhat by orthogonal or oblique rotation; and
- **The decomposition depends on the rank of matrix**. If there are more regions and observations, the matrix will be rank-deficient, which can obscure the “true” grouping of regions.
Independent component analysis
ICA is a newer method than PCA and has been applied extensively to fMRI and EEG data. It was originally a denoising method but has since been shown to be quite powerful for extracting resting state networks in fMRI data (using a variation of the usual ICA: tensor ICA) (Beckmann and Smith, 2005). ICA has the advantage over PCA of not assuming orthogonality but rather maximal independence. In this case, it has the capacity to separate artifactual components from those of interest. This capacity depends, however, on the flavor of ICA used and the nature of the artifact. The drawbacks of ICA include:
• Nonunique solutions without additional constraints; and
• Computationally expensive for large data sets.

Partial least squares
The partial least squares (PLS) method has been used in neuroimaging for more than a decade and has been applied to PET, fMRI, and EEG (McIntosh et al., 1994; McIntosh and Lobaugh, 2004). It is related to canonical correlation analysis in that it relates the neuroimaging data to the experimental design (e.g., design contrasts); performance measures; or, for functional connectivity, one or more voxels. In the latter case, it can be considered to be a multivariate extension of PPI. PLS has the flexibility to work on combinations of design, behavior, and voxels and has been extended to merge multiple imaging data sets (Martinez-Montes et al., 2004). It has the advantage of creating a flexible framework for direct testing of statistical dependency in neuroimaging data. Its main drawbacks are as follows:
• Orthogonal extraction of components like PCA may obscure the true dependencies. To offset this effect, the extraction can be done with ICA (Lin et al., 2003);
• Interpretation can be complicated in complex designs; and
• Statistical assessment through resampling is computationally expensive.

Effective connectivity
Structural equation modeling
Structural equation modeling (SEM) is a multivariate linear regression tool and has been used primarily for PET and fMRI data (McIntosh and Gonzalez-Lima, 1994; Buchel and Friston, 1997), although its use has been extended to EEG data (Astolfi et al., 2004, 2005). Its primary use has been to identify changes in effective connectivity between tasks or groups within a defined anatomical network (Protzner and McIntosh, 2006). It has also been used to identify likely patterns of effective connectivity in a given data set (Bullmore et al., 2000). It has the advantage of allowing fast and robust computations and can be used for rather complicated models (McIntosh et al., 1996b); more recently, it was validated for use with neuroimaging data based on large-scale simulations (Kim and Horwitz, 2009; Marrelec et al., 2009). It has a long history and, thus, several software packages and numerous algorithmic variations are available. For its application to neuroimaging, the main drawbacks are as follows:
• Absolute assessment of model fit is very dependent on sample size;
• It needs to prespecify connection directions; and
• It cannot deal with fully reciprocal models.

Granger causality
Granger causality (GC) is a general methodological approach for analyzing dependencies in time series. Its most common implementation comes in the form of autoregressive modeling (Goebel et al., 2003). There are also variations that operate in the spectral domain (Kaminski et al., 2001), although they have not been used in fMRI. Methods that generally fall under this label have the advantage of working directly with the time series, allowing inferences on directionality without needing to prespecify the direction (cf. SEM and DCM). Its main drawbacks are as follows:
• Most implementations are pairwise. Multivariate extensions are possible (Deshpande et al., 2009), but with many regions, the solutions may become unstable;
• For fMRI, GC requires relatively short repetition time ($T_R$) to get a robust time series; and
• There has been a recent observation that GC may provide spurious estimates of directional effects in fMRI data (David et al., 2008). However, a series of papers that will appear in the journal NeuroImage will address this observation (Roebroeck, et al., in press).

Dynamic causal modeling
Unlike SEM and GC, DCM was designed specifically for neuroimaging data and has been applied to fMRI and EEG (Friston et al., 2003; Kiebel et al., 2009). Like SEM, DCM has also received some validation through large-scale simulations (Lee et al., 2006). DCM uses a generative model of the measured signal to infer its neural sources. The effective connectivity estimation then proceeds based on the neural source activity rather than the measured signal (e.g., blood oxygen level–dependent [BOLD] or EEG). The model first estimates the intrinsic connections between sources and then the changes in the connections that come about through external perturbation (usually the experimental design). This can be thought of...
in the general linear model (GLM) framework as estimating the grand mean, and the deviations from the mean, from the experimental manipulation. A Bayesian estimation procedure is used to estimate the effective connections and their change, as well as providing evidence for the “best” model.

The advantages of DCM are the tight coupling to biophysical models, which enables an interpretation of the effective connections in terms of neurophysiology. There is a potential for investigating several models for mediation of effective connections using model evidence. Its main drawbacks are as follows:

- It is computationally expensive, and in its present form, cannot handle more than about six regions;
- It cannot easily model intrinsic activity such as resting state networks; and
- Some researchers question the robustness of parameter estimation, given the extensive constraints on the generative model.

**Final Thoughts**

Functional or effective connectivity estimation has benefits for developing theories of brain operation. Brain imaging researchers will often discuss the results from regional activation analysis in terms of “functional networks” without specifically referring to how these networks are formed. By requiring that the networks be expressed through either functional or effective connectivity estimation, the researcher’s assumptions about the network organization are more obvious.

It is also critical to acknowledge, particularly with effective connectivity, that the results are a model. There are decisions that have to be made in the course of estimation, such as the selection of regions to include in the model and how their interactions are mediated (through specification of anatomical connections). Because it is a model, however, it is an approximation of reality and, by definition, false. To paraphrase the statement from statistician George Box that all models are wrong but some are useful, the utility of any model comes from its capacity to explain neural dynamics and cognitive function and to suggest further avenues of research to test and develop the model.

Finally, one of the greatest sins in analyzing neuroimaging data is to assume that there is a single correct method. While one can certainly make mistakes in the application of a method, there is little to be gained from “analytic chauvinism.” The complexity of the data that are extracted from neuroimaging methods dictates that a single analytic approach is insufficient. I strongly concur with the position advocated by others: that a pluralistic approach will provide a much better appreciation of how the brain brings about human mental function (Lange et al., 1999).

**References**


