Analysing connectivity with Granger causality and dynamic causal modelling

Karl Friston[1], Rosalyn Moran[1] and Anil K Seth[2]

This review considers state-of-the-art analyses of functional integration in neuronal macrocircuits. We focus on detecting and estimating *directed connectivity* in neuronal networks using *Granger causality* (GC) and *dynamic causal modelling* (DCM). These approaches are considered in the context of functional segregation and integration and — within functional integration — the distinction between *functional* and *effective* connectivity. We review recent developments that have enjoyed a rapid uptake in the discovery and quantification of functional brain architectures. GC and DCM have distinct and complementary ambitions that are usefully considered in relation to the *detection* of functional connectivity and the *identification* of models of effective connectivity. We highlight the basic ideas upon which they are grounded, provide a comparative evaluation and point to some outstanding issues.

Addresses
[1] The Wellcome Trust Centre for Neuroimaging, University College London, Queen Square, London WC1N 3BG, UK
[2] Sackler Centre for Consciousness Science and Department of Informatics, University of Sussex, Brighton BN1 9QJ, UK

Corresponding author: Friston, Karl (k.friston@ucl.ac.uk)

Introduction

Several dichotomies have proved useful in thinking about analytic approaches to functional brain architectures. Perhaps the most fundamental is the distinction between *functional segregation* and *integration*. Functional segregation refers to the anatomical segregation of functionally specialised cortical and subcortical systems, while functional integration refers to the coordination and coupling of functionally segregated systems [1••]. Within functional integration, two main classes of connectivity have emerged — *functional* and *effective* connectivity. Functional connectivity refers to the statistical dependence or mutual information between two neuronal systems, while effective connectivity refers to the influence that one neural system exerts over another [2•,3]. This distinction is particularly acute when considering the different analyses one might apply to electrophysiological or neuroimaging timeseries.

Functional and effective connectivity

Because functional connectivity is defined in terms of statistical dependencies, it is an operational concept that underlies the detection of (inference about) a functional connection, without any commitment to how that connection was caused. In other words, one tests for dependencies between two or more timeseries, to reject the null hypothesis of statistical independence. This is equivalent to assessing the *mutual information* and testing for significant departures from zero. At its simplest, this involves assessing (patterns of) correlations — of the sort that define intrinsic brain networks. An important distinction — within functional connectivity — rests on whether dependencies are instantaneous or reflect an underlying dynamical process, in which causes precede consequences. This leads to the distinction between analyses of *directed* and *undirected* functional connectivity that do and do not appeal to temporal precedence respectively. Common examples of techniques used to assess undirected functional connectivity (dependencies) include independent components analysis [4] and various measures of synchrony, correlation, or coherence [5]. However, we will focus on analyses of directed functional connectivity — of which the prime example is Granger causality (GC) [6•]. This is because coupling in the brain is both directed and largely reciprocal (producing cyclic graphs or networks with loops that preclude structural causal modelling). As we will see below, GC and related concepts such as transfer entropy (TE) rest on establishing a statistical dependence between a local measurement of neuronal activity and measurements of activity elsewhere *in the past*.

Functional connectivity considers dependencies between measured neurophysiological responses. In contrast, effective connectivity is between hidden neuronal states generating measurements. Crucially, effective connectivity is always directed and rests on an explicit (parameterised) model of causal influences — usually expressed in terms of difference (discrete time) or differential (continuous time) equations. The most popular approach to effective connectivity is dynamic causal modelling (DCM) [7–10,11••,12••]. In this context, causality is inherent in the form of the model, where fluctuations in hidden neuronal states cause changes in others: for example, changes in postsynaptic potentials in one area are caused by inputs from other areas. The

parameters of dynamic causal models correspond to effective connectivity — usually cast as synaptic density or coupling parameters — that are optimised by fitting the model to data. The notion of effective connectivity stems from the pioneering work of Gerstein and Perkel [13] in early attempts to interpret multivariate electrophysiological recordings. At its inception, effective connectivity referred to models; in the sense of the simplest possible circuit diagrams that explain observed responses [14]. In modern parlance, these correspond to dynamic causal models with the greatest evidence: namely, models with the minimum complexity that furnish an accurate explanation for data (see below). In what follows, we review recent developments in the analysis of directed functional connectivity with GC and TE, the analysis of directed effective connectivity with DCM and then consider the approaches in light of each other. Figure 1 provides an overview of recent developments in these techniques.

Granger causality and transfer entropy

The core idea behind GC is that X 'Granger causes' Y if X contains information that helps predict the future of Y better than information already in the past of Y (and in the past of other 'conditioning' variables Z). The most common implementation of GC is via linear vector autoregressive (VAR) modelling of timeseries data, enabling both statistical significance testing and estimation of GC magnitudes [6•,15•,16]. However, GC is not limited to this implementation; it can use nonlinear, time-varying, and non-parametric models [17,18]. In particular, TE [19] represents an information-theoretic generalisation of GC that does not require a parameterised model (is model-free). Specifically, the TE from X to Y is zero if, and only if, Y is conditionally independent of X's past, given its own past. Importantly, for Gaussian data, TE is equivalent to GC [20••], furnishing a useful interpretation of GC in terms of information transfer in 'bits'. Related approaches include partial directed coherence and the directed transfer function; see [21] for a review. Here we focus on the most popular of these techniques, namely GC:

Following its introduction within econometrics [6•,15•], GC has been applied in neuroscience partly because it is simple to estimate, given (stationary stochastic) timeseries. Such data are generated by a wide range of neuroimaging and neurophysiological methods. GC has some useful properties including a decomposition of causal influence by frequency [15•] and formulation in an 'ensemble' form, allowing evaluation of GC between multivariate sets of responses [22]. GC has provided useful descriptions of directed functional connectivity in many electrophysiological studies [23–25]. Recently, Bosman et al. [26••] analysed electrocorticographic data from macaque monkeys to show that 'bottom-up' signals across multiple cortical regions were most prominent in the gamma band, while 'top down' influences dominated at beta frequencies — a finding that is strikingly congru-

ent with neural implementations of predictive coding [27]. GC can also be applied to standard EEG or MEG signals, either at the source or sensor level (following spatial filtering to reduce the impact of volume conduction). For example, Barrett et al. [28•] used source-localised EEG to show that gamma-band GC between posterior and anterior cingulate cortices reliably increased during anaesthetic loss of consciousness, extending previous results obtained using (undirected) phase synchrony [29]. We will turn to this example later in the context of DCM.

The application GC to fMRI is more controversial, given the slow dynamics and regional variability of the haemodynamic response to underlying neuronal activity [30,31]; and see 'Pros and Cons' below. While naïve application of GC to fMRI data is unlikely to be informative, careful consideration of the methodological issues has permitted some useful applications that have produced testable hypotheses. For example, Wen et al. [32•] analysed fMRI data obtained from a cued spatial visual attention task; finding that GC from dorsal to ventral frontoparietal regions predicted enhanced performance, while GC in the reciprocal direction was associated with degraded performance. These findings are consistent with the notion that dorsal attentional regions mediate goal-oriented top-down deployment of attention, while ventral regions mediate stimulus-driven bottom-up reorienting. In a similar paradigm, Bressler et al. [33] found that GC from parietal to occipital areas was predictive of behavioural performance. In a final and unusual example, Schippers et al. [34] used GC of fMRI signals to analyse directed interactions *between* the brains of two subjects engaged in a social game (charades), providing novel evidence for 'mirror neuron system' formulations of social interaction. Another promising application of GC is to intracranial local field potentials, which possess high temporal and spatial resolution and which comprise comparatively few variables (as compared to fMRI voxels or EEG sensors). An early application in this area, Gaillard et al. [35] examined directed functional connectivity during supraliminal as compared to subliminal visual word processing.

Dynamic causal modelling

The basic idea behind DCM is that neural activity propagates through brain networks as in an input-state-output system, where causal interactions are mediated by unobservable (hidden) neuronal dynamics. This multi-input multi-output neuronal model is augmented with a forward, or observation model that describes the mapping from neural activity to observed responses. Together neuronal and observation model comprise a full generative model that takes a particular form depending on the data modality. The key outputs of DCM are the evidence for different models and the posterior parameter estimates of the (best) model, particularly