CellPress

Neuron
Review

Translational Perspectives for Computational Neuroimaging

Klaas E. Stephan,[1,2,3,*] Sandra Iglesias,[1] Jakob Heinzle,[1] and Andreea O. Diaconescu[1]
[1]Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, 8032 Zurich, Switzerland
[2]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK
[3]Max Planck Institute for Metabolism Research, 50931 Cologne, Germany
*Correspondence: stephan@biomed.ee.ethz.ch
http://dx.doi.org/10.1016/j.neuron.2015.07.008

Functional neuroimaging has made fundamental contributions to our understanding of brain function. It remains challenging, however, to translate these advances into diagnostic tools for psychiatry. Promising new avenues for translation are provided by computational modeling of neuroimaging data. This article reviews contemporary frameworks for computational neuroimaging, with a focus on forward models linking unobservable brain states to measurements. These approaches—biophysical network models, generative models, and model-based fMRI analyses of neuromodulation—strive to move beyond statistical characterizations and toward mechanistic explanations of neuroimaging data. Focusing on schizophrenia as a paradigmatic spectrum disease, we review applications of these models to psychiatric questions, identify methodological challenges, and highlight trends of convergence among computational neuroimaging approaches. We conclude by outlining a translational neuromodeling strategy, highlighting the importance of openly available datasets from prospective patient studies for evaluating the clinical utility of computational models.

Introduction

Non-invasive measurements of human brain activity have been available for almost a century. Following electroencephalography (EEG) in the 1920s, the more recent developments of positron emission tomography (PET), magnetoencephalography (MEG), and fMRI have greatly enriched human neuroscience. Collectively, these methods have enabled major advances in our understanding of brain physiology and cognition.

Neurology and psychiatry have welcomed these techniques enthusiastically, in the hope that non-invasive readouts of brain function might enable more precise diagnoses and better predictions for individual patients. While thousands of functional neuroimaging studies over the past few decades have made important contributions to elucidating pathophysiological processes, the impact on clinical practice has been limited. Success stories where functional neuroimaging has contributed concrete diagnostic tools are restricted to neurology, e.g., presurgical evaluation of epilepsy, differential diagnosis of coma, and brain-computer-interfaces for locked-in patients. By contrast, in psychiatry, functional neuroimaging procedures are yet to be established as diagnostic tools for routine clinical practice.

There are several explanations for this poor translational success rate in psychiatry (Kapur et al., 2012). One issue of interest to this article is that conventional analyses of neuroimaging data—such as statistical parametric mapping or functional connectivity analyses—are essentially descriptive. While they are powerful methods to identify potential nodes and connections of disease-relevant circuits, on their own neither "blobs" (regional activations) nor "networks" (patterns of functional connectivity) provide a mechanistic account of circuit function, i.e., what computations are performed and how they are implemented physiologically.

An alternative are mathematical models that describe putative processes underlying the generation of neuroimaging data. These are forward models that embody a probabilistic mapping from unobservable ("hidden") brain states—cognitive or neurophysiological—to experimental measurements. In other words, these models seek explanations of data, as opposed to statistical characterizations. Importantly, some of these forward models can be inverted, i.e., they allow one to infer hidden brain states from neuroimaging measurements. This opens up the possibility of detecting pathophysiological processes in individual patients ("computational assays") (Stephan and Mathys, 2014) and renders these models attractive candidate techniques for stratifying patients into mechanistically distinct groups.

In this article, we concentrate on three major approaches: (1) biophysical network models, (2) generative models sensu stricto of neuroimaging data, and (3) model-based fMRI analyses of neuromodulation (Figure 1). For simplicity, we will refer to all these models by the umbrella term "computational models," appealing to the multiple meanings of "computation" (e.g., information processing, or algorithmic—as opposed to analytical—mathematical treatments). Furthermore, these models represent different facets of an emerging research program, "Computational Psychiatry" (Deco and Kringelbach, 2014; Friston et al., 2014; Maia and Frank, 2011; Montague et al., 2012; Stephan and Mathys, 2014; Wang and Krystal, 2014).

This article has two major aims. First, it provides an overview of contemporary computational models of neuroimaging data, discussing what mechanistic insights these models may allow for and exploring trends of their convergence. Second, we outline strategies how these models can be applied to clinical questions such that not only novel pathophysiological insights result but, eventually, concrete diagnostic procedures. To ensure

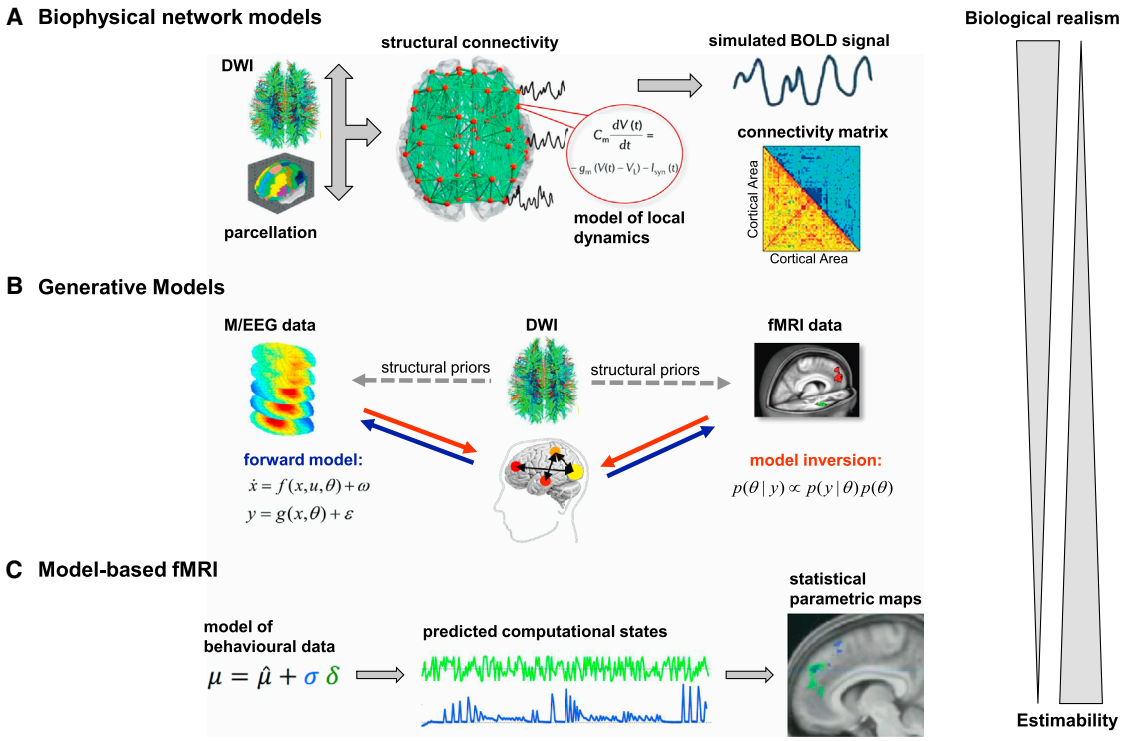

Figure 1. Graphical Overview of Modeling Approaches Discussed in This Paper
The figure contains graphics that has been adapted, with permission, from Deco et al. (2013a) and Chen et al. (2009). See main text for details.

coherence and focus, we concentrate on a single psychiatric spectrum disorder—schizophrenia. This is because pathophysiological theories of schizophrenia highlight themes that feature prominently in existing computational neuroimaging frameworks, i.e., connectivity, synaptic plasticity, neuromodulation, and perceptual inference.

Due to space limitations, this article strictly focuses on "forward modeling" approaches of how measured signals are generated by hidden mechanisms. Other important approaches—e.g., graph-theoretical analyses or analyses of functional connectivity—are covered by existing excellent reviews (Buckner et al., 2013; Bullmore and Sporns, 2009; Fornito et al., 2015).

Why Computational Modeling—And What to Focus on?

Standard classification schemes like the Diagnostic and Statistical Manual of Mental Disorders (DSM) define schizophrenia as a syndrome, i.e., a collection of symptoms and phenomenology over certain periods. The predictive validity of this classification is limited, and patients with the same diagnosis often exhibit markedly different clinical trajectories, outcomes, and treatment responses (Casey et al., 2013; Cuthbert and Insel, 2013; Krystal and State, 2014). This spectrum nature of schizophrenia stems from at least three sources. First, a polygenetic basis, with a large number of genome variants conveying risk (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014); the functional consequences of these genetic variants, however, may converge on only a small set of intracellular signaling cascades and synaptic processes (Krystal and State, 2014; Stephan et al., 2006). Second, environmental factors such as infections, nutrition and stress interact with risk-conveying genes and modulate their expression (gene-by-environment interactions) (van Os et al., 2008). Third, environmental factors can also affect pathophysiological processes directly, e.g., immunological, metabolic, and hormonal factors can alter NMDA receptor (NMDAR) function (Stephan et al., 2009).

Collectively, these considerations imply that different pathophysiological pathways can be affected in different combinations across patients. The ensuing lack of pathophysiological interpretability of the label "schizophrenia" under current diagnostic schemes has major consequences for clinical practice, such as the necessity of resorting to trial-and-error treatment and the difficulties of stratifying patients for clinical studies (Kapur et al., 2012). Only very few physiologically defined subgroups of patients with psychotic symptoms can presently be identified through clinical tests, e.g., patients suffering from neurosyphilis or NMDAR antibodies.

Computational modeling may help addressing this problem by inferring disease mechanisms from non-invasive readouts of circuit function. In analogy to diagnostic procedures in internal medicine, it is the functional status quo of disease-relevant circuits that may prove crucial to assign patients to pathophysiological subgroups and to derive individual treatment predictions. However, what are the most relevant pathophysiological processes and circuits that should inform the development of computational models? Clearly, more than one theory of schizophrenia exists and could provide guidance here. For example, long-standing theories have focused on dopamine

(DA), neurodevelopment, NMDARs, GABA receptors, and excitation-inhibition (E-I) balance, respectively (Gonzalez-Burgos and Lewis, 2012; Howes and Kapur, 2009; Insel, 2010; Lisman et al., 2008; Uhlhaas, 2013). Regardless of their specific propositions, however, a shared perspective among these theories is that schizophrenia is essentially a network disease (Harrison and Weinberger, 2005), where a primary pathology at the level of synapses leads to maladaptive reconfigurations of circuits for learning and perceptual inference (Stephan et al., 2006).

Viewing schizophrenia as a network disease has a long tradition dating back to the early 20th century when Wernicke (1906) and Bleuler (1911) stressed structural and cognitive deficits of functional integration in schizophrenia, respectively. With the advent of neuroimaging, observations of abnormally distributed activity and functionally disconnected areas paved the way for the concept of ''dysconnectivity'' in schizophrenia (Andreasen, 1999; Friston and Frith, 1995). This widely adopted view refers to disturbances of functional integration that manifest as abnormal connectivity and oscillatory activity (Buckholtz and Meyer-Lindenberg, 2012; Bullmore et al., 1997; Friston, 1998; Pettersson-Yeo et al., 2011; Stephan et al., 2006; Uhlhaas, 2013).

Notably, dysconnectivity could result from a variety of synaptic mechanisms. For example, theories differ in their relative emphasis on abnormalities of glutamatergic, GABAergic, dopaminergic, and cholinergic signaling (Gonzalez-Burgos and Lewis, 2012; Lisman et al., 2008; Stephan et al., 2009). This motivates the construction of computational models for clarifying how alterations in different ionotropic and metabotropic receptors impact on functional coupling as assessed by neuroimaging.

In addition to physiology, however, models are required that link neuronal to computational processes and explain how aberrant cognition arises from circuit dysfunction. One theory that provides a framework for constructing such models is the dysconnection hypothesis (Friston, 1998; Stephan et al., 2006, 2009; Adams et al., 2013). It postulates that, physiologically, dysconnectivity in schizophrenia results from abnormal NMDAR-neuromodulator interactions (NNI)—i.e., aberrant regulation of NMDAR-dependent synaptic plasticity by DA or acetylcholine (ACh)—with failures of perceptual inference as a computational consequence. More specifically, the dysconnection hypothesis and conceptually related concepts of psychosis (Corlett et al., 2011; Fletcher and Frith, 2009) build on the ''Bayesian brain'' notion that the brain constructs a model of the world in order to predict its sensory inputs and infer on the environmental causes of its sensations (Dayan et al., 1995; Doya et al., 2011). One particular variant of this Bayesian view is the ''free-energy principle'' (Friston, 2010) that postulates that the brain's central function is to minimize surprise, by updating beliefs and/or choosing actions that lead to expected sensory inputs. Physiologically, Bayesian message passing is typically assumed to rest on glutamatergic signaling of predictions and prediction errors (PEs) via cortical long-range connections, weighted by estimates of precision (inverse uncertainty) that may be encoded by slow changes in release of neuromodulatory transmitters like DA or ACh (Corlett et al., 2010; Friston et al., 2012). Impairments of these processes lead to abnormal perceptual inference that in turn may explain a range of salient symptoms in schizophrenia, e.g., hallucinations or delusions, as discussed below.

This brief overview has outlined target processes for computational models suggested by current pathophysiological theories of schizophrenia. We now turn to different classes of computational models that may prove useful for inferring these putative disease processes from neuroimaging data.

Biophysical Network Models of Neuroimaging Data

Over the last decades, a variety of single-neuron models have been developed that describe the dynamics of ion channel conductances, membrane potential, and firing rate. A straightforward way of constructing neuronal population models are ''direct simulations'' (Omurtag et al., 2000), i.e., simulating a large number of individual neurons and linking them via local synaptic connection rules. The resulting neuronal ensembles can be treated as distinct nodes that are linked by anatomical long-distance connections to yield a large-scale biophysical network model (BNM). Simulated neuronal population activity in each of the regions can then be fed into a forward model that predicts regional fMRI, M/EEG, or PET measurements.

Some BNMs of fMRI data have used the strategy of ''direct simulations,'' usually considering on the order of 10^3 neurons per network node (Deco and Jirsa, 2012). Most present BNMs of neuroimaging data, however, do not pursue a ''direct simulation'' approach. This is not only because of the high computational costs. First, models with large numbers of biophysically detailed single neurons are too complex for parameter estimation; this necessitates fixing model parameters a priori, usually referring to electrophysiological studies in animals. However, many biophysical and morphological parameters show pronounced variability within and across species (Kötter and Feizelmeier, 1998; Marder and Goaillard, 2006). Furthermore, in large-scale models it is difficult to identify the decisive mechanisms underlying a particular empirical measurement: both simulations with systematic exploration of parameter space and analytical treatments become impractical.

For these reasons, most BNMs of neuroimaging data have sought lower-dimensional representations of neuronal mechanisms that strike a balance between biophysical realism and model complexity. This typically rests on ''mean-field'' reduction, a concept from statistical physics that describes system behavior in terms of average effects resulting from the probabilistic interactions of many individual components (e.g., temperature and pressure of a gas). In the context of neuronal population models, instead of accounting for all interactions between individual neurons, the mean-field approach only considers interactions between the statistical moments of neuronal populations (Freeman, 1975). In other words, neurons of one population are only influenced by the mean activity of another population (and possibly higher order moments such as variance). Effectively, this perspective transforms the representation of neuronal dynamics from the microscopic (single neuron) to the mesoscopic (neuronal population) level.

The last decade has seen important advances in mean-field formulations of neuronal population models. A systematic overview and nomenclature can be found in Deco et al. (2008) who adopt a population density perspective, conceptualizing the

temporal evolution of the population's probability density in the form of a flow-diffusion process. Considering the statistical moments of this density corresponds to a "dynamic mean field" approach; considering only the mean activity of each population results in "neural mass" models. Finally, "neural field" models capture the spread of activity across the brain (Jirsa and Haken, 1996; Robinson et al., 2001).

These advances have paved the way for tractable large-scale BNMs of neuroimaging data. The general strategy consists of three steps (Figure 1): (1) representing each network node as a neural mass or mean-field model of local neuronal populations (e.g., excitatory and inhibitory neurons within a cortical area); (2) linking these nodes by long-range connections; and (3) feeding the resulting network activity into an observation model that predicts regional fMRI, M/EEG, or PET data. (It is worth noting that such models possess a likelihood function and allow one to generate synthetic data; this, however, does not yet render them "generative models" in a statistical sense. This distinction will be revisited below.)

For M/EEG, the history of BNMs goes back to neural mass models of event-related potentials (ERPs) (Freeman, 1975; Jansen and Rit, 1995; Valdes et al., 1999). Recent BNMs have covered whole-brain activity, demonstrating, for example, the importance of conduction delays for explaining distributed oscillations in the "resting state" (Nakagawa et al., 2014). Neural field models have also been applied to empirical M/EEG data, elucidating general principles of brain dynamics, such as multistability and scale-invariance (Freyer et al., 2011, 2012), and providing important insights into disorders such as epilepsy (Breakspear et al., 2006).

The importance of conductance delays in M/EEG models and the high spatial resolution of fMRI data highlight the need of BNMs for accurate information on anatomical long-range connections. This information can be obtained from human diffusion-weighted imaging (DWI) or from the CoCoMac database of tract tracing studies in the macaque monkey (for review, see Stephan, 2013). Neither approach is without uncertainty: while DWI data cannot resolve directionality of connections and has limited resolution, CoCoMac rests on mapping procedures that integrate data across different parcellation schemes (and, for human studies, species).

Despite this limitation, BNMs of M/EEG and fMRI data have become important tools for investigating the mechanisms that link microscopic (single neurons), mesoscopic (areas), and macroscopic (networks) levels of description. For fMRI, initial BNMs focused on task-specific networks (Husain et al., 2004); subsequent models have encompassed the entire brain, using parcellations with up to 10^3 regions. These models have typically focused on the "resting state," i.e., unconstrained cognition in the absence of sensory perturbations (Ghosh et al., 2008; Honey et al., 2007). Collectively, these studies provided important insights into how large-scale dynamics are constrained by the anatomical "skeleton" of long-range structural connectivity (Deco et al., 2013a).

Despite all simplifications, the models discussed so far are still too complex for parameter estimation, and the applications described above used simulations under fixed parameters. Inferring subject-specific parameters from neuroimaging data, however, is crucial for future diagnostic applications of BNMs (Woolrich and Stephan, 2013). While this motivates the simpler generative models discussed below, recent BNMs have begun to acquire a limited capacity of estimating parameters from empirical data. In particular, Deco et al. (2013b) derived a linearized simplification of the model by Wong and Wang (2006), allowing for the estimation of a global parameter (that uniformly scales connection strengths across the brain) from empirical fMRI data.

A notable example of the rapid development of large-scale BNMs for neuroimaging data is the "Virtual Brain" project (Jirsa et al., 2010; Sanz-Leon et al., 2015). This open-source software provides a platform for constructing whole-brain models, allowing for the combination of different neural mass and neural field models with different measures of long-range connectivity (CoCoMac or human DWI data). Using realistic head models and different forward models, fMRI and M/EEG signals can be simulated simultaneously from the same underlying neuronal model (http://thevirtualbrain.org).

Applications to Clinical Questions

While BNMs of neuroimaging data have developed rapidly, their application to diseases has only begun recently. In schizophrenia, working memory (WM) is an attractive target for biophysical modeling: it is frequently impaired, and the underlying circuit mechanisms are known in great detail (Brunel and Wang, 2001; Durstewitz et al., 2000; Lisman et al., 1998). One key mechanism concerns dopaminergic regulation of glutamatergic receptor conductances in the dorsolateral prefrontal cortex (DLPFC). In brief, a DA-mediated increase in the conductance of NMDARs, relative to those of AMPARs, is necessary to switch pyramidal cells into a high-frequency firing mode that is critically required for and time-locked to memory maintenance. This is of relevance for schizophrenia, given that the interaction of NMDARs and DA is a central theme in pathophysiological theories of schizophrenia (Laruelle et al., 2003; Stephan et al., 2006).

Recent BNM studies of WM have used the NMDAR antagonist ketamine—an established pharmacological model of schizophrenia symptoms (Corlett et al., 2011)—in healthy volunteers. These studies were particularly interested in disturbances of E-I balance, given that NMDAR antagonism may exert a preferential loss of excitatory drive at GABAergic interneurons, leading to disinhibition of pyramidal cells (Homayoun and Moghaddam, 2007). A first study used a circuit model of spatial WM in DLPFC that suggested that ketamine would reduce lateral inhibition and hence decrease the selectivity of stimulus representations; the ensuing predictions about ketamine effects on WM performance were verified in behavioral experiments (Murray et al., 2014). A second study using fMRI found that ketamine disrupted functional connectivity between fronto-parietal areas and the default mode network (DMN) during WM (Anticevic et al., 2012). Furthermore, ketamine reduced DMN deactivation during the task; a BNM comprising fronto-parietal and DMN modules suggested that this could be explained by local disinhibition and the resulting decrease in sensitivity to long-range inputs.

BNMs have also been used in three recent fMRI patient studies. The first compared adolescents with early-onset

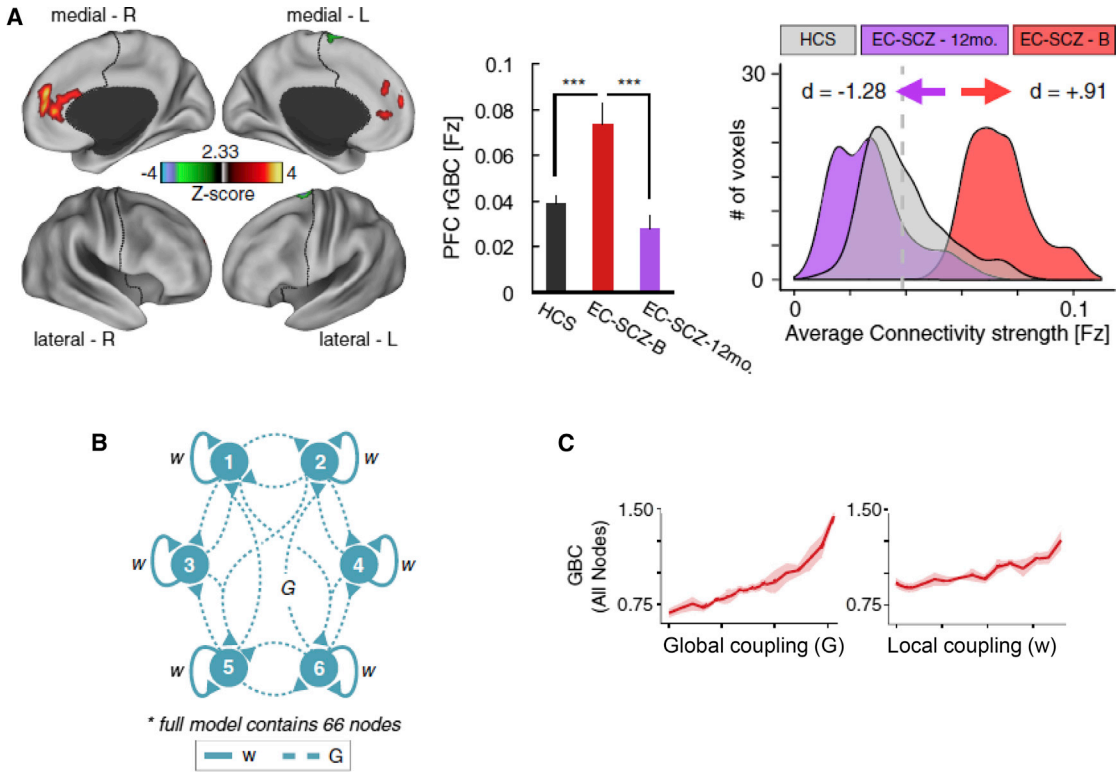

Figure 2. Application of a BNM to fMRI Data from Patients with Schizophrenia

(A) Voxels in medial PFC with enhanced global connectivity at 12-month follow-up compared to baseline. Connection strengths for healthy controls are shown for comparison.

(B) BNM with two key parameters: local coupling (w) within nodes and long-range global coupling (G) between 66 nodes.

(C) Simulations showed enhanced global connectivity when increasing either w or G. Adapted from Anticevic et al. (2015), with permission.

schizophrenia to healthy controls, using a whole-brain model (90 regions) informed by individual DWI data (Cabral et al., 2013). While structural network properties did not differ significantly when correcting for multiple comparisons, applying the BNM to the fMRI data indicated a trend toward reduced global coupling in patients; this would explain lower small-world indices of functional connectivity in schizophrenia (Lynall et al., 2010).

However, a second BNM study of adult patients with schizophrenia (Yang et al., 2014) suggested that reduced functional connectivity may not occur universally across the brain. This study focused on functional connectivity of the DLPFC, finding "hyperconnectivity" in schizophrenia; a result mimicked in the BNM by increasing either within-node or inter-node coupling. Importantly, this study showed that global signal regression during fMRI data preprocessing decisively affected conclusions about global coupling estimates in schizophrenia (cf. Fornito and Bullmore, 2015).

The same BMN was used by a longitudinal fMRI study of 129 patients with early-stage schizophrenia who were scanned prior to medication; 25 patients were followed up after 1 year (Anticevic et al., 2015). Focusing on medial prefrontal (mPFC) cortex, its coupling estimates with the rest of the brain were mostly increased (for other areas, there was a mixed pattern of increased and decreased connectivity). Importantly, mPFC hyperconnectivity normalized after one year and predicted positive (but not negative) symptoms. As Yang et al. (2014), Anticevic et al. (2015) found that increasing either global or local coupling parameters mimicked the observed pattern of prefrontal hyperconnectivity. They interpreted this as altered E-I balance in early-stage schizophrenia that normalized in parallel to symptom improvements over time (Figure 2).

Generative Models of Neuroimaging Data

While BNMs offer a detailed representation of (patho)physiological mechanisms, a central challenge for clinical utility is the difficulty to estimate these mechanisms from subject-specific measurements. This motivates considering a different class of models, so-called "generative models" that represent the joint probability of data and model parameters (Figure 1). They require two things: a prior distribution, indicating the expected range of parameter values and a likelihood function. The latter encodes a probabilistic forward model, quantifying the probability of obtaining a particular observation (e.g., pattern of regional BOLD signals) as a function of the model parameters (e.g., synaptic connection strengths). Once likelihood and prior are specified, it is possible, in principle, to "invert" the model and compute the posterior probability of the parameters given the measured data; this fully characterizes a mechanism (parameter) of interest, providing both its expected value and one's uncertainty about it.

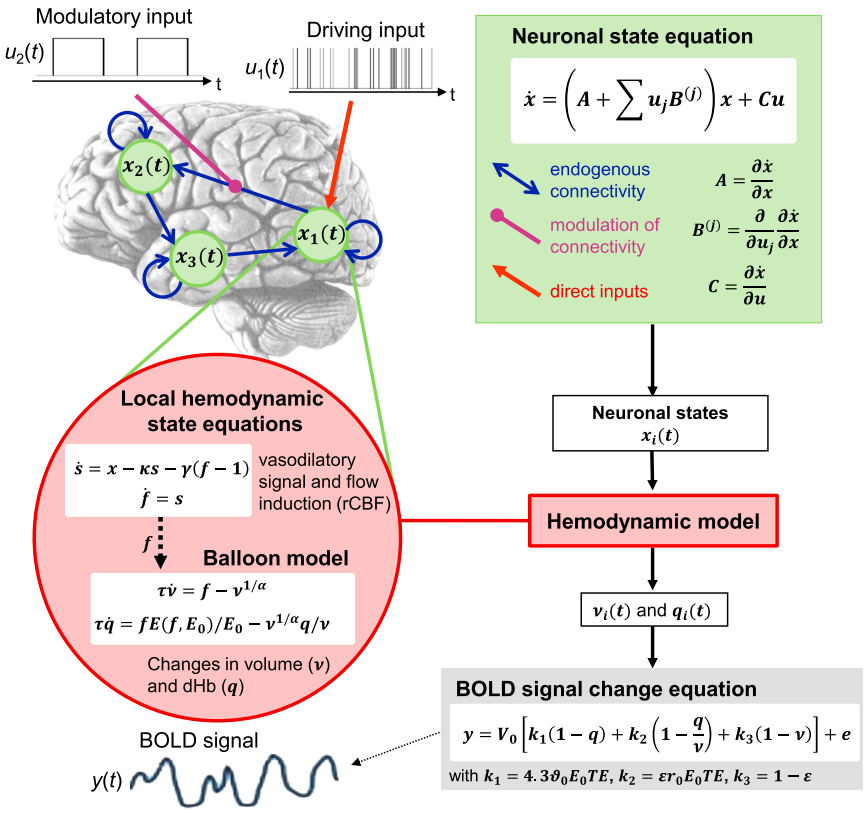

Figure 3. Summary of the Classical Deterministic DCM for fMRI
For mathematical details, see Friston et al. (2003) and Stephan et al. (2007).

In addition, a generative approach also allows for inference on model structure: by integrating out the dependency on model parameters, one obtains the model evidence, a principled index of the trade-off between a model's accuracy and complexity. The model evidence provides a basis for Bayesian model selection and averaging (Stephan et al., 2007; Penny et al., 2010). These procedures allow for comparing and integrating alternative model formulations (e.g., whether a connection exists, or whether a particular form of plasticity is present).

Importantly, in neuroimaging, neuronal activity is not observed directly; instead, the measurements reflect a (potentially complicated) transformation of neuronal activity. This means the likelihood function takes on the hierarchical form of a state-space model and distinguishes between a hidden neuronal level and an observation level (Figure 3). This formulation as a hierarchical generative model is critical for inference on neuronal processes and disambiguating them from potential confounds. For example, in fMRI, regional variations in neurovascular coupling can severely confound inference on neuronal connectivity (David et al., 2008).

Dynamic Causal Modeling

The idea of using a hierarchical generative model to infer on neuronal processes from neuroimaging data was first implemented by dynamic causal modeling (DCM) for fMRI (Friston et al., 2003). The original formulation rests on a low-order Taylor approximation and describes the dynamics of interacting neuronal populations by bilinear differential equations via three types of mechanisms: experimental perturbations (e.g., sensory stimuli) that "inject" activity into the system, fixed synaptic connections by which this activity is conveyed to target populations, and modulatory inputs that invoke contextual changes of connection strengths (e.g., short-term plasticity and neuromodulatory influences) (see Figure 3). Subsequently, the neuronal model in DCM for fMRI has been extended in several ways, including non-linear (Stephan et al., 2008) and stochastic differential equations (Li et al., 2011). The latter can model endogenous fluctuations in neuronal activity and extends the applicability of DCM to "resting state" fMRI data. The predicted neuronal dynamics are linked to region-wise BOLD signals via a nonlinear hemodynamic model (Stephan et al., 2007). Notably, the same hemodynamic model has been incorporated into most BNMs of fMRI data described above.

DCM represented the first complete generative model of BOLD data that spanned both neuronal and hemodynamic levels and was sufficiently simple that it could be inverted. While this is usually done with variational Bayesian techniques, alternative schemes based on Markov chain Monte Carlo (MCMC) sampling or Gaussian processes are currently under development. Compared to the BNMs discussed above, current DCMs are restricted to smaller subgraphs of brain-wide connectivity, typically with up to ten nodes, in order to maintain feasibility of model inversion.

By replacing the hemodynamic forward model with an electromagnetic one, DCM can be generalized from fMRI to electrophysiological data. The rich temporal information in M/EEG measurements allows for constructing DCMs with more detailed neuronal representations and for building bridges to the BNMs discussed above. The first DCM for M/EEG data by (David et al., 2006) was based on a classical neural mass model (Jansen and Rit, 1995). This DCM describes how cortical areas—each represented by a macrocolumn composed of pyramidal cells, excitatory, and inhibitory interneurons—interact through long-range glutamatergic connections whose laminar patterns follow neuroanatomical rules of cortical hierarchies. This allows for considerably more fine-grained physiological inference than DCM for fMRI, e.g., on the relative strength of glutamatergic versus GABAergic transmission (Moran et al., 2011a).

Subsequent developments of DCM for M/EEG have made further strides toward physiological interpretability, for example, using a conductance-based formulation that distinguishes

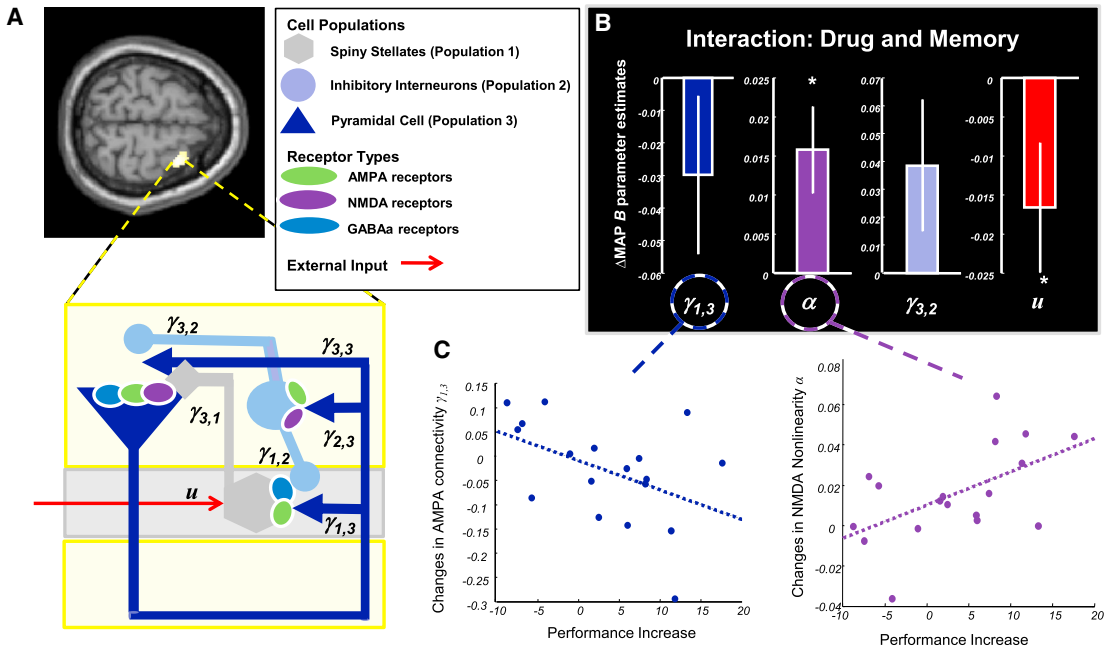

Figure 4. A Prototype Computational Assay for Inferring DA Effects on Glutamate Receptor Conductances
Adapted, with permission, from Moran et al. (2011b).
(A) A prefrontal microcircuit DCM with three cell classes and receptor types was used to model MEG data of healthy subjects performing a WM task twice, under L-Dopa or placebo.
(B) Differences in parameter estimates across drug conditions. In line with previous data, administration of L-Dopa reduced AMPAR conductance ($\gamma_{1,3}$) and enhanced the sensitivity (nonlinearity) of NMDARs (α).
(C) Changes in AMPAR conductance (left) and NMDAR nonlinearity (right) each significantly ($p < 0.05$) correlated with drug-induced change in performance.

ionotropic receptors with sufficiently distinct time constants, i.e., AMPA, NMDA, and $GABA_A$ receptors (Marreiros et al., 2010). Pharmacological validation studies have demonstrated that this model is capable of pathophysiologically relevant inference, for example, the identification of DA-induced changes in NMDAR conductance from MEG data (Moran et al., 2011b) (Figure 4).

Other Generative Models

Beyond DCM, various generative models of neuroimaging data have been proposed in recent years. These include fMRI models of dynamic effective connectivity (Havlicek et al., 2011) or state-space models of M/EEG data that capture neuronal interactions by multivariate autoregressive formulations (Olier et al., 2013; Fukushima et al., 2015). Finally, an entirely different class of generative models aims at explaining trial-by-trial variations in M/EEG responses; we turn to these below.

Application to Clinical Questions

Generative models have been used in numerous studies on schizophrenia, guided by the various pathophysiological concepts described above. Due to space constraints, here we focus exclusively on studies of perceptual inference; this is a theme at the core of the dysconnection hypothesis and related concepts (Corlett et al., 2010; Fletcher and Frith, 2009; Stephan et al., 2006) that provides a bridge to Bayesian brain theories like predictive coding (Rao and Ballard, 1999). In brief, predictive coding posits that the brain constructs a generative model of its sensory inputs and updates its beliefs about the environmental causes of its sensations by inverting this model (Figure 5). Belief updating rests on message passing between hierarchically related neuronal populations, such that each population sends predictions about expected input to the next lower level and, following sensory input, a prediction error (PE) to the next higher level; this PE is then used to update subsequent top-down predictions. This recurrent message passing serves to minimize PE at all levels. Importantly, the impact of PEs is context-dependent and varies with their relative precision (inverse uncertainty). For example, PEs arising from vague (uncertain) predictions signal less necessity for belief adjustment than PEs based on precise predictions. Overall, this suggests a simple classification of computational causes why perceptual inference could break (abnormal computation of PEs, predictions, or precision-weighting) and offers potential explanations for concrete clinical symptoms that can be tested with computational models (Adams et al., 2013; Jardri and Denève, 2013).

Importantly, these computational variables can be linked to physiological processes (Figure 5): PE and prediction signaling is assumed to be mediated by glutamatergic signaling via bottom-up/forward (AMPAR and NMDAR) and top-down/backward (NMDAR) connections in cortical hierarchies (Corlett et al., 2011; Friston, 2005a; Stephan et al., 2006). Precision-weighting might be implemented by the postsynaptic gain of PE-encoding supragranular pyramidal cells, under the influence of slow changes in neuromodulatory transmitter levels such as DA or ACh (Friston et al., 2012; Moran et al., 2013). On the other hand, synaptic

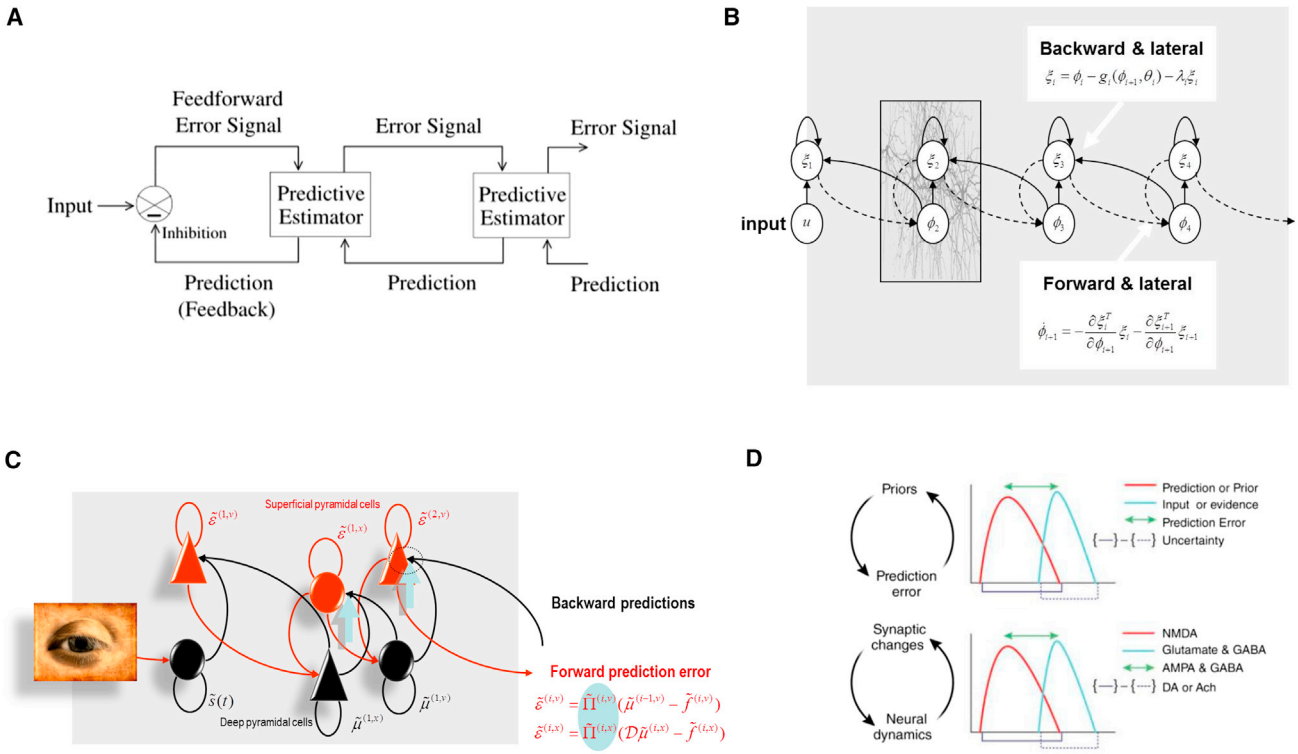

Figure 5. An Overview of Predictive Coding Architectures

Figures are reproduced, with permission from the respective publishers.

(A) Predictive coding (PC) assumes that the brain constructs a hierarchical generative model of its sensory inputs and infers their causes by model inversion. Belief updating rests on message passing between hierarchically related neuronal populations: each population sends predictions about expected input to the next lower level and prediction error (PE) to the next higher level (Rao and Ballard 1999). This recurrent message passing serves to minimize PE at all levels.

(B) A putative neuronal implementation of PC (Friston 2005a): upper and lower circles represent neural units encoding PE (ξ) and the posterior expectation of causes (ϕ), respectively.

(C) A more refined version assigns PEs and expectations to supra- and infragranular pyramidal cells, respectively, and distinguishes causal states (v) and hidden states (x); while the former connect hierarchical levels, the latter represent the temporal dynamics of expectations and endow the model with memory (Friston et al., 2012).

(D) Proposed neurobiological components of PC architectures (Corlett et al., 2011). PEs are assumed to be signaled by fast glutamatergic (and in some circuits GABAergic) transmission, predictions via NMDARs, and precision by neuromodulatory transmitters (DA, ACh).

gain is influenced by NMDAR activation itself and its interactions with GABAergic mechanisms (Adams et al., 2013); this provides a link to theories emphasizing disturbances of E-I balance in schizophrenia (Uhlhaas, 2013).

The idea that abnormal perceptual inference results from deficiencies of the brain's generative model (i.e., aberrant signaling of precision-weighted PEs and predictions via forward and backward connections) can be tested empirically. This rests on using generative models (of neuroimaging data) to determine changes in effective connectivity under experimental variations of the difficulty of perceptual inference. An example is the "hollow mask" illusion where a concave mask of a human face is perceived as a normal convex face. As many other illusions, it can be understood as arising from the biasing influence of a strong prior (here: that faces are convex objects) during the inversion of the brain's model of sensory inputs. Intriguingly, patients with schizophrenia are, on average, considerably less susceptible to this illusion than healthy controls (Dima et al., 2009). Two separate fMRI and EEG studies examined potential mechanisms for this phenomenon: applying structurally analogous DCMs to fMRI and EEG data, these two studies consistently found a strengthening of bottom-up connections and diminished top-down connectivity in patients, consistent with the notion of reduced precision of predictions about facial stimuli (Dima et al., 2009, 2010). Notably, weakening of top-down predictions may explain a range of perceptual alterations in schizophrenia (Notredame et al., 2014) and may also play a role in the initial formation of delusions (Schmack et al., 2013).

Another paradigm that is commonly interpreted from a predictive coding perspective is the mismatch negativity (MMN), an event-related potential (ERP) elicited by unpredicted sensory stimuli ("deviants"). Reduced MMN amplitudes are one of the most consistently found electrophysiological anomalies in schizophrenia (Umbricht and Krljes, 2005). Initial DCM studies on healthy volunteers showed that both forward and backward connection strengths change at the presentation of a deviant, reflecting the bottom-up signaling of PEs and ensuing adaptation of top-down predictions (for review, see Garrido et al., 2009). Subsequent DCM studies in patients with schizophrenia have demonstrated striking alterations. For visual MMN, Fogelson

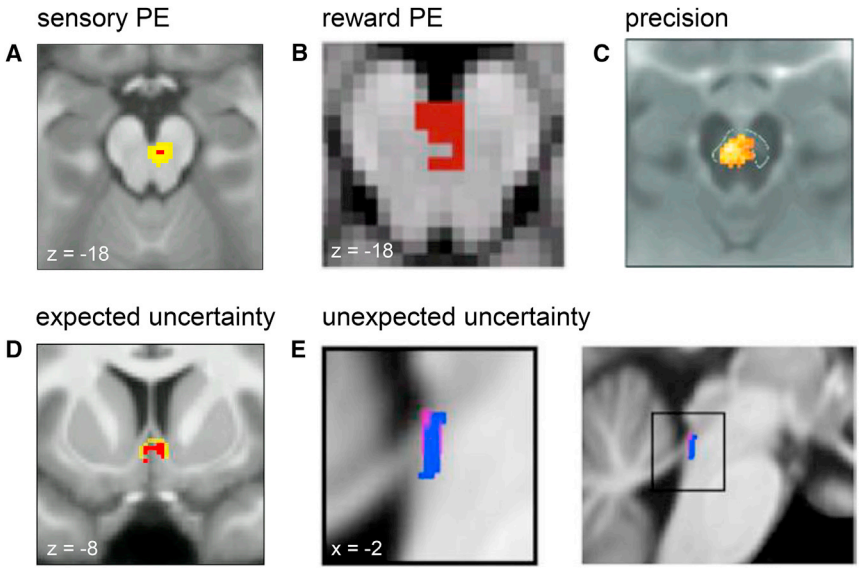

Figure 6. Computational fMRI Studies of Neuromodulation

Figures are reproduced, with permission from the respective publishers. Please see individual papers for significance levels.

(A–C) Midbrain activity reflects precision-weighted sensory PEs (Iglesias et al., 2013) (A), reward PEs (Klein-Flügge et al., 2011) (B), and precision (of beliefs about the value of policies) (Schwartenbeck et al., 2014) (C).

(D and E) Encoding of expected uncertainty (operationalized by probability PEs) in the septal part of the cholinergic basal forebrain (Iglesias et al., 2013) (D) and of unexpected uncertainty in the locus coeruleus (Payzan-LeNestour et al., 2013) (E).

et al. (2014) reported, similar to the hollow mask illusion study by Dima et al. (2010), a marked reduction of top-down connections in the visual hierarchy. For auditory MMN, the DCM results by Dima et al. (2012) suggested that patients exhibited both abnormal forward and backward effective connectivity, accompanied by reduced postsynaptic gain of pyramidal cells in primary auditory cortex.

Pharmacological studies in healthy volunteers have played an important part in elucidating the mechanisms underlying MMN deficits in schizophrenia. This has highlighted the roles of NMDARs and ACh. For example, the NMDAR antagonist ketamine reduces MMN amplitude in healthy volunteers, similar to schizophrenia, and individual MMN amplitude predicts psychosis-proneness under ketamine (Umbricht et al., 2002). Applying DCM to ERPs acquired under ketamine versus placebo, Schmidt et al. (2013) found that ketamine significantly reduced the PE-induced plasticity of forward connections in auditory cortex. A DCM study of MEG data acquired under the acetylcholinesterase inhibitor galantamine suggested that the observed increase in MMN amplitude resulted from ACh enhancing the precision of bottom-up signaling in the auditory hierarchy by increasing the postsynaptic gain of supragranular pyramidal cells (Moran et al., 2013).

An alternative generative modeling framework of M/EEG data is worth mentioning in the context of MMN. This approach concentrates on trial-by-trial responses and views across-trial fluctuations in MMN amplitude as a reflection of Bayesian belief updating. It rests on hierarchical generative models where a hidden layer of computational processes is linked to EEG channels through a linear forward model. Under this framework, competing hypotheses that specific computational variables might be encoded by MMN amplitude can be disambiguated by model comparison. Applying this strategy to a somatosensory MMN paradigm, Ostwald et al. (2012) found that Bayesian surprise (simply speaking, adjustments of model predictions in response to new observations) better explained trial-wise MMN amplitudes than classical interpretations of MMN as change detection. In the auditory domain, Lieder et al. (2013) compared 13 alternative implementations of five major MMN theories. Analogous to Ostwald et al. (2012), they found that Bayesian belief updating proved superior to classical interpretations of MMN in terms of change detection, adaptation or novelty, and adjustments of model predictions provided the best explanation for trial-wise MMN amplitudes.

Model-Based fMRI: Computational Models of Neuromodulation

The previous sections have described computational models of neuroimaging data with decreasing complexity, from whole-brain BNMs to DCMs of circumscribed circuits. This section reduces the spatial scope even further and considers model-based explanations of single-voxel fMRI data. These rest on generative models of behavior (e.g., trial-wise choices or reaction times), yielding trajectories of computational states; these can, in turn, be coupled to a forward (convolution) model and used as regressors in general linear model analyses of voxel-wise fMRI data (Figure 1). This two-step procedure differentiates this approach from the models discussed above and provides a bridge between two types of observations, behavior and brain activity.

This approach—pioneered by O'Doherty et al. (2003) and often referred to as "model-based fMRI"—has been used successfully in many studies, often employing reinforcement learning (RL) or Bayesian models, to determine which circuits implement a particular computational process. Here, we focus on one aspect of particular interest for potential clinical applications. This is the notion that release of modulatory transmitters—e.g., DA from midbrain neurons, norepinephrine from locus coeruleus, or ACh from basal forebrain—may encode the values of specific computational variables (Figure 6).

This idea originated from the seminal observation that the temporal course of phasic DA release resembled the trajectory of reward PEs as predicted by RL models (Schultz et al., 1997). This view nicely connected the role of PEs as "teaching signals" for learning to DA's involvement in modulating synaptic plasticity and triggered the question whether reliable in vivo estimates of neuromodulatory transmitter release in humans could be

obtained through the combination of fMRI and computational models of behavior (D'Ardenne et al., 2008; Düzel et al., 2009).

As key regulators of synaptic plasticity, neuromodulatory transmitters drive (mal)adaptive changes of neuronal circuits and thus play a key role in pathophysiological theories of almost any psychiatric disease (Montague et al., 2012). Another reason for the interest in computational approaches to neuromodulation (Dayan, 2012) is that most drugs used in psychiatry target either synthesis, metabolism, or postsynaptic binding of neuromodulatory transmitters. This suggests that computational assays inferring neuromodulatory processes from neuroimaging data might prove useful for treatment predictions in individual patients (Stephan and Mathys, 2014).

Initial model-based fMRI studies of neuromodulation mainly used conditioning paradigms and RL models to demonstrate that reward PEs explained human fMRI responses in the dopaminergic midbrain or dopaminoceptive regions like the ventral striatum (e.g., D'Ardenne et al., 2008; O'Doherty et al., 2003; but see Klein-Flügge et al., 2011). There are several reasons, however, why DA is unlikely to be restricted to "classical" PEs about primary reward, as examined by Pavlovian and operant conditioning paradigms. First, DA midbrain neurons show pronounced heterogeneity with regard to neurodevelopment, connectivity, and electrophysiology (Roeper, 2013). Furthermore, they contribute to reward-unrelated cognitive processes, such as WM (Matsumoto and Takada, 2013), and may encode PEs about purely sensory events (Iglesias et al., 2013). Third, what constitutes a "reward" is context-dependent and depends on the individual's internal model and the subjective beliefs it rests upon. Finally, DA release fluctuates at different timescales, e.g., tonic versus phasic responses. This suggests that DA neurons may emit a multiplexed signal reflecting several computational variables concurrently (Hiroyuki, 2014).

Other quantities that may be reflected by DA release include novelty (Düzel et al., 2009) and uncertainty or its inverse, precision (Friston et al., 2012). Several studies in animals and humans have provided evidence that uncertainty or precision are encoded by slow fluctuations in DA neuron activity within and across trials (de Lafuente and Romo, 2011; Fiorillo et al., 2003; Schwartenbeck et al., 2014). It has been suggested (Corlett et al., 2011; Friston, 2005b) that DA and other neuromodulators may serve to implement the precision-weighting of PEs that arises from Bayesian models under Gaussian assumptions (Friston et al., 2012; Mathys et al., 2011; Rao and Ballard, 1999) (Figure 5). Empirically, a recent fMRI study demonstrated that trial-by-trial midbrain activity reflected precision-weighted PEs about visual stimuli (Iglesias et al., 2013) (see Figure 6).

Uncertainty has also been a major theme in theories about other neuromodulatory transmitters. An influential proposal by Yu and Dayan (2005) posited that ACh and NE levels may represent "expected uncertainty" (EU; known/estimated unreliability of a prediction) and "unexpected uncertainty" (UU; induced by a change of the environment), respectively. Empirical evidence was obtained by recent fMRI studies (Figure 6) that showed that model-based indices of EU were linked to trial-wise activity in the cholinergic basal forebrain (Iglesias et al., 2013), while UU was found to correlate with activity of the noradrenergic locus coeruleus (Payzan-LeNestour et al., 2013).

Application to Clinical Questions

So far, computational models of neuromodulation in schizophrenia have largely focused on abnormal DA signaling and its potential role in delusion formation. One influential idea is that dysregulated activity of DA neurons might result in PE signals that are ill-timed and/or of abnormal precision, leading to erroneous attribution of importance ("aberrant salience") to random or irrelevant events (Heinz, 2002; Kapur, 2003). This induces ongoing violations of the individual's model of the world such that, eventually, only the compensatory adoption of complicated and seemingly bizarre beliefs can lead to cognitive resolution (Corlett et al., 2010; Roiser et al., 2013). This may rest on an imbalance of precision, where abnormally high precision of PE signals (aberrant salience) at lower levels of the inference hierarchy dominates over relatively low precision of predictions at higher levels; under this perspective, the subsequent adoption of extremely rigid (high precision) beliefs that become visible as delusions may represent a compensatory response (Adams et al., 2013).

A prediction of the "aberrant salience" theory is that patients with schizophrenia should show a diminished difference in PE responses to relevant and neutral stimuli. This hypothesis has been tested by several computational fMRI studies using RL models and conditioning paradigms in patients with schizophrenia (Gradin et al., 2011; Murray et al., 2008; Romaniuk et al., 2010) and individuals at ultra-high risk for psychosis (Roiser et al., 2013). Despite differences in tasks, models, and clinical groups, these studies indeed point to abnormal PE responses in midbrain and ventral striatum (Figure 7): overall, in patients, PEs elicited less activity on both rewarding or aversive trials, but more activity in response to neutral or irrelevant cues. Additionally, midbrain PE responses correlated with psychotic symptom ratings (Gradin et al., 2011; Romaniuk et al., 2010). Finally, abnormal midbrain activity in patients with first-episode schizophrenia has also been reported for other aspects of associative learning (Corlett et al., 2007).

A recent study by Horga et al. (2014) used a PE-dependent learning model to investigate another key symptom of psychosis, auditory hallucinations. Previous fMRI studies demonstrated auditory cortex activation during auditory hallucinations, in the absence of external stimuli (Dierks et al., 1999). From a predictive coding perspective, this points to overly precise and rigid priors that induce misinterpretations of noisy baseline activity in auditory cortex (Fletcher and Frith, 2009; Friston, 2005b). This idea was tested by Horga et al. (2014) in patients with schizophrenia compared to healthy controls. Probabilistically varying the absence and presence of auditory stimuli (voices) and modeling trial-wise fMRI responses as a weighted mixture of predictions and PEs, they found that patients with hallucinations showed reduced PE signals in a voice-sensitive region of secondary auditory cortex and, at the same time, increased activity during silent trials. Both abnormalities varied with the individual severity of hallucinations, but not other symptoms. These findings do not directly prove but are consistent with the notion of rigid priors at higher auditory levels that have become impervious to update requirements signaled by PEs and evoke percepts during silence. It is possible, but presently speculative, that this represents a maladaptive response to initial PE abnormalities, similar as discussed for delusions above.

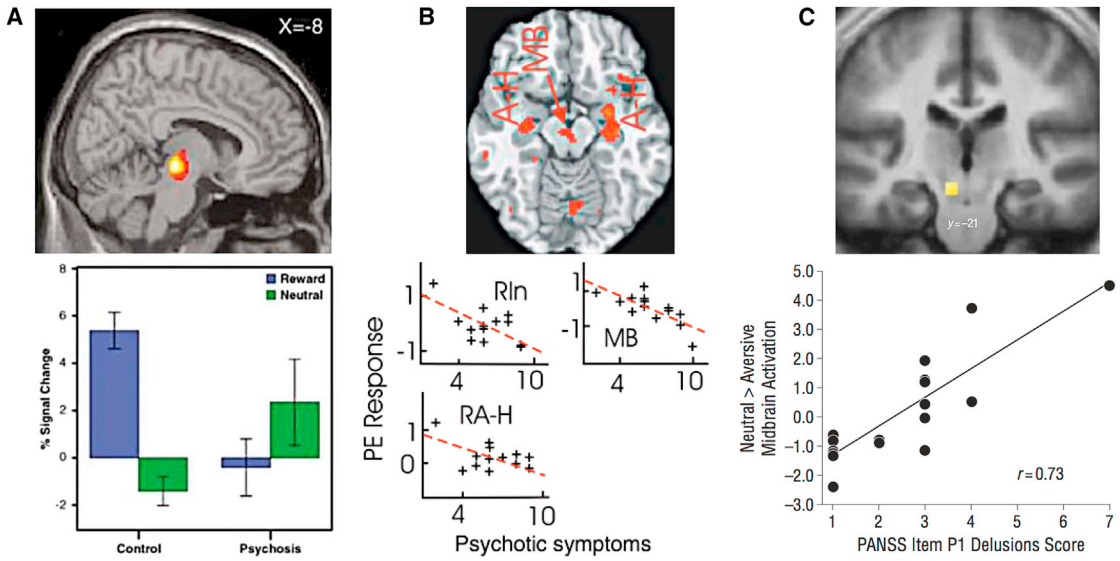

Figure 7. Model-Based fMRI Studies of Neuromodulation in Patients with Schizophrenia versus Controls
Figures are reproduced, with permission from the respective publishers. Please see individual papers for significance levels.
(A and B) During instrumental conditioning, patients showed reduced midbrain PE responses on reward trials (A) and augmented PE responses to neutral cues (Murray et al., 2008) (B). A similar paradigm found a trend to reduced PE activity in the midbrain (MB) of patients. Decreases in PE activity correlated with severity of psychotic symptoms in MB, right insula (RIn) and amygdala-hippocampus (RA-H) (Gradin et al., 2011).
(C) Patients' midbrain activation to neutral stimuli versus fearful stimuli during aversive conditioning correlated significantly with delusional symptoms (Romaniuk et al., 2010).

Strengths, Limitations, and Convergence

Intuitively, the potential diagnostic utility of a model increases the more detailed its representations of biological mechanisms and the more accurately estimates of these mechanisms can be obtained in individual patients. However, these desiderata oppose each other: the more detailed and biologically realistic a model, the greater the challenges of parameter estimation and the danger of overfitting. How do the approaches discussed in this paper fare with regard to this general trade-off?

BNMs have introduced an innovative treatment of whole-brain neuroimaging data, based on detailed representations of neuronal mechanisms. Despite careful abstractions, estimating most of these mechanisms from empirical data is presently not possible, due to various reasons. First, an issue affecting models of fMRI in general is that the low-pass filtering property of neurovascular coupling restricts identifiability to mechanisms that are expressed in a relatively low-frequency domain. Second, in BNMs, the large number of parameters and their ubiquitous inter-dependencies represent difficult numerical challenges for system identification. This requires fixing most parameters to a priori values. Parameters encoding connection strengths represent a particular problem because different configurations of effective connectivity can lead to extremely different dynamic regimes, and there are no simple rules for deducing effective connectivity from anatomical nor functional connectivity. Although the common assumption that synaptic weights are proportional to anatomical connection strengths (as obtained from individual DWI data) is a helpful first approximation, anatomical connectivity only constrains but does not determine effective connectivity. This is because connection strengths change dynamically at short timescales, under the influence of synaptic plasticity and neuromodulation (Stephan et al., 2008). Deco et al. (2013b) introduced a free parameter that scales functional coupling strength globally. This parameter has an indiscriminative effect on all connections; by contrast, cognitive processes invoke selective changes in subsets of long-range connections. A final challenge is the substantial variability in neurovascular coupling across brain regions and across subjects. Assuming fixed hemodynamic parameters across regions may confound identification of neuronal mechanisms from fMRI data (Valdes-Sosa et al., 2011).

Some of these methodological issues are addressed by generative models, such as DCM. These adopt a simpler mathematical characterization of neuronal dynamics and are restricted to smaller circuits. While their parameters can still be numerous, they are constrained by priors; this enables computing the posterior distribution of the parameters (model inversion). This should not be confused with point estimates of parameters as obtained, for example, in BNMs that optimize the correlation between observed and predicted BOLD functional connectivity) (Deco et al., 2013b). Importantly, the regularization afforded by priors avoids overfitting and reduces identifiability issues, enabling one to estimate both neuronal parameters and region-specific hemodynamics in individual subjects. There are, however, non-trivial issues for the inversion of current DCMs, such as local extrema during optimization or the choice of priors; see Daunizeau et al. (2011) for discussion. These methodological challenges have inspired ongoing developments for DCM, such as global optimization schemes and empirical Bayesian procedures for a "data-driven" choice of priors.

The physiological interpretability of model parameters in DCM for fMRI is limited, given the relatively abstract state equations.

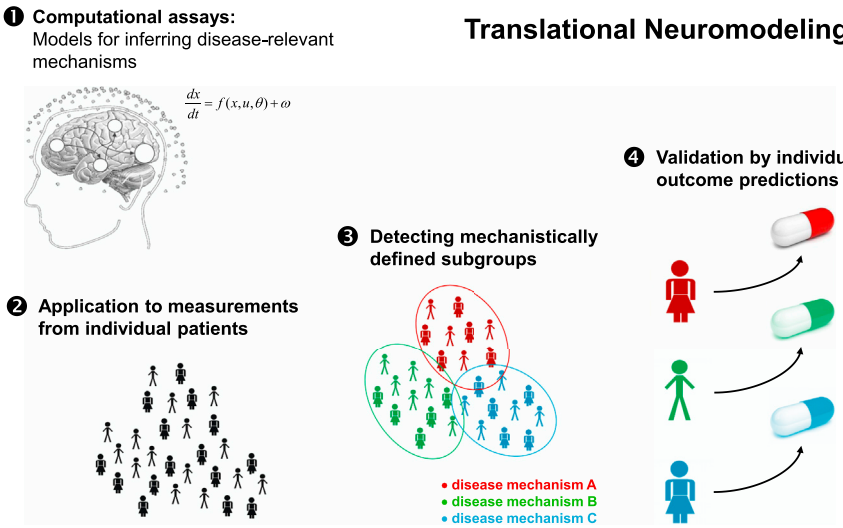

Figure 8. Graphical Summary of the Translational Neuromodeling Strategy Outlined in This Paper

Panel 1 adapted, with permission, from Chen et al. (2009).

For example, there is only a coarse representation of types of synaptic transmission (excitatory versus inhibitory), and synaptic plasticity is only characterized phenomenologically (as a change in connection strength). By contrast, M/EEG data contain much richer temporal information than fMRI, allowing for representation of different types of neurons and postsynaptic receptors. This enables a finer characterization of pathophysiological processes, such as E-I balance or DA modulation of glutamatergic synapses, under suitable pharmacological challenges (Moran et al., 2011a, 2011b). While this indicates promising potential for treatment predictions, a major future challenge for clinical utility will be to develop models that can disambiguate different neuromodulatory mechanisms from M/EEG data in the absence of designed pharmacological perturbations.

Given that both BNMs and DCMs move toward the same goal, but from different ends of the complexity spectrum, it is not surprising to see signs of methodological rapprochement. For example, there are efforts to extend DCMs to larger networks, while preserving the capability of model inversion (Seghier and Friston, 2013). For BNMs, one of the most important future developments concerns parameter estimation; at present, only very few BNMs possess free parameters at all (Deco et al., 2013b; Freyer et al., 2011) and these are of a global and physiologically unspecific nature. As discussed recently (Deco and Kringelbach, 2014; Woolrich and Stephan, 2013), this might be improved by turning BNMs into fully generative models. For example, importing model inversion and comparison procedures from DCM might allow for obtaining connection-specific parameter estimates, for resolving uncertainty about model structure (e.g., connectivity layout) and for detecting overfitting.

Overall, it is likely that we will see a convergence of BNM and DCM in the future. This will need to be complemented by computational perspectives, describing how circuit architecture gives rise to specific computations. For BNMs and DCMs, so far only very simple "neurocomputational" models exist; for example, inferring short-term plasticity from trial-wise PE-dependent changes in effective connectivity (den Ouden et al., 2009). An important goal is to obtain generative models that predict both measured brain activity and behavior from the same underlying neuronal state equations (Rigoux and Daunizeau 2015; Wiecki and Frank 2013).

Translational Neuromodeling

The examples discussed in this article illustrate how computational neuroimaging can contribute to unraveling pathophysiological mechanisms in schizophrenia. However, direct demonstrations of diagnostic utility do not exist so far. One reason is that the vast majority of studies to date have contrasted patients with schizophrenia (as defined by DSM) to controls or other patient groups. This does not address the spectrum nature of schizophrenia with the likely existence of different pathophysiological pathways that underlie the variability in clinical trajectories and treatment responses across patients (Krystal and State, 2014). Establishing models that can distinguish between DSM-defined patient groups only recapitulates the current diagnostic categories—with their lack of predictive validity (Casey et al., 2013; Cuthbert and Insel, 2013)—but using a considerably more expensive approach than conventional clinical interviews. Instead, psychiatry needs approaches that allow for inference on disease mechanisms in individual patients and a stratification of patients according to pathophysiological types with predictive validity (Kapur et al., 2012). In brief: we need tools and strategies for dissecting the spectrum.

Provided the methodological issues discussed above can be addressed, it might become possible to establish computational assays with a similarly important role for differential diagnosis in psychiatry as biochemical assays in internal medicine; cf. Stephan et al. (2006). However, we do not wish to claim that this is a panacea for psychiatry; additionally, various pitfalls exist and will need to be addressed carefully. In this final section, we consider some of these issues and outline strategic aspects for translational neuromodeling (Figure 8). For simplicity, we adopt a categorical perspective and assume distinct pathophysiological subgroups in spectrum diseases. However, model-based differential diagnosis can be equally approached from a dimensional perspective; see Brodersen et al. (2014) for discussion.

Basic Validation Studies

Candidate models for clinical applications should pass basic validation studies. For example, face validity should be examined in initial simulation studies, probing whether the model can recover known system structure and parameter values from noisy data (Havlicek et al., 2011; Stephan et al., 2008). Evaluating predictive validity requires experimental perturbations,

challenging the model to identify a system state induced by a well-defined experimental manipulation, e.g., a selective pharmacological intervention. For example, previous studies have challenged DCM to identify, from fMRI data, the source of epileptiform activity as established by invasive recordings (David et al., 2008), or to detect the consequences of changes in anesthesia levels and application of selective drugs, respectively (Moran et al., 2011a, 2011b). Animal studies play an important role for model validation since they allow for more controlled and specific test scenarios, including highly selective perturbation techniques like optogenetics or "Designer Receptors Exclusively Activated by Designer Drugs" (DREADDS).

Despite their importance, predictive validation studies of computational models are rare. As it is often prohibitively difficult, slow, and expensive to acquire experimental data for model validation from scratch, it will be crucial to establish open datasets with well-defined perturbations of physiological processes. Such datasets would greatly speed up model development and allow one to assess at an early stage the sensitivity and specificity of a candidate computational assay for detecting known physiological states in individual subjects.

Strategies for Identifying Patient Subgroups

Following initial model validation, model-based detection of pathophysiological subgroups in a spectrum disease could proceed in two different ways. If theories predict the nature of subgroups in advance and suitable models of subgroup-specific mechanisms exist, subgroup assignment can be formulated as a model selection problem. That is, in each individual patient, the plausibility of competing models is evaluated (in terms of model evidence), and the patient is assigned to the subgroup associated with the most likely model. A compelling example, albeit in a non-clinical spectrum of synesthetic subjects, is provided by van Leeuwen et al. (2011).

In psychiatric spectrum diseases like schizophrenia, however, we usually do not know how many subgroups exist and what (combination of) mechanisms may define them. Here, subgroups can be identified using an unsupervised variant of "generative embedding" (Brodersen et al., 2011, 2014). This approach applies the same generative model to each subject's data separately and uses the resulting posterior densities for subsequent unsupervised learning (clustering). In other words, the model is used as a theory-guided dimensionality reduction device that creates a de-noised and mechanistically interpretable feature space. Compared to conventional approaches, such as classifying patients on the basis of functional connectivity, generative embedding has shown significantly higher performance in fMRI studies on patients with stroke (Brodersen et al., 2011) and schizophrenia (Brodersen et al., 2014). Notably, the latter study identified three connectivity-defined subgroups in schizophrenia that mapped onto significant differences in clinical symptom scores.

When trying to derive subgroups through generative embedding, one approach is to start with DSM categories in toto, as in Brodersen et al. (2014). Alternatively, one can use behavioral data or symptoms for identifying phenomenologically more homogenous subgroups within these broad spectra and proceed with model-based inference on physiological mechanisms within these subgroups (cf. Hyett et al., 2015). Finally, it is possible to disregard physiological data altogether and define subgroups based on generative modeling of behavioral data alone (Wiecki et al., 2015).

Prospective Patient Studies

Once potential patient subgroups are identified, one needs to verify that the proposed division has predictive validity with regard to individual patients. This requires testing whether a model-based assignment of individual patients to proposed subgroups results in clinically meaningful advances, that is, whether it improves the physician's ability to predict future clinical outcomes and select optimal treatments for individual patients (Figure 8). The importance of this validation in terms of clinical utility cannot be overemphasized: regardless how perfect the physiological validity of a given computational assay, if it does not enable a differential diagnosis that improves predictions and treatment choices for individual patients, it will have no role in clinical practice. Critically, predictive validity of this sort cannot be established in cross-sectional comparisons but requires prospective designs with a focus on individual predictability. In contrast to neuropsychological studies (Barch et al., 2003), so far, prospective computational neuroimaging studies are extremely rare (but see Anticevic et al., 2015; Iyer et al., 2015).

Potential Pitfalls

The success of translational neuromodeling will depend on our ability to develop computational assays (of pathophysiological or psychopathological) states that yield a real improvement in clinical decision-making for individual patients, specifically outcome predictions and treatment selection. In addition to the technical challenges described above, however, several conceptual and practical pitfalls loom. For example, a detected abnormality of circuit parameters in patients may either reflect a primary biological disease mechanism, or simply a different cognitive process (e.g., task strategy) than assumed or instructed (Schlagenhauf et al., 2014). Furthermore, even when the same task strategy is employed, differences observed in patients may arise from unusual beliefs that have formed in response to certain (perhaps unfortunate) experiences, while the neuronal machinery itself is physiologically intact (cf. Mathys et al., 2011). Disambiguating these possibilities is presently rarely done, but can, in principle, be addressed by comparing generative models of behavior that embody veridical versus unusual beliefs about the task and its context.

Another potential problem is that even highly accurate model-based pathophysiological inference may only allow for relatively short-term predictions, given that the model is agnostic about future environmental and biological perturbations (e.g., social stress or infections) that may significantly impact on disease-relevant brain circuits, not only by altering gene expression but also synaptic plasticity directly. One could address this by longitudinal neuroimaging measurements in single patients; this, however, may introduce practical problems and render the cost-benefit ratio unattractive. Alternatively, predictions by a computational assay could be augmented (and updated iteratively) by additional measurements that are more easily and affordably obtained over time than neuroimaging but are

sensitive to perturbations of known relevance. This could include, for example, plasma levels of hormones and markers of inflammation, but also behavioral readouts in individual patients. The latter are particularly attractive for longitudinal measurements in individual patients as they can be obtained as part of games designed for mobile devices (Rutledge et al., 2014) and lend themselves to computational analysis, using identical models (e.g., RL or Bayesian models) as those for neuroimaging data. Overall, this suggests that a more comprehensive (and yet to be developed) modeling framework may become necessary—one where models derive and update clinical predictions by treating computational neuroimaging estimates of an initial pathophysiological state as an "anchor" for subsequent disease dynamics expressed by biochemical and behavioral time series.

Open Datasets and Open Code

Prospective patient studies are essential for establishing the clinical utility of candidate computational assays. However, these studies are expensive, require close cooperation between computational and biomedical scientists, and take years to complete—this represents a serious bottleneck for translation and a career risk for scientists depending on funding renewal. Similar to physiological validation studies described above, there is an urgent need for openly available datasets that can be used to evaluate the potential clinical utility of models at an early development stage and help resolve uncertainty about the most promising directions. While laudable data sharing initiatives have been established in the fMRI community, there is a lack of data from prospective patient studies with clinically relevant targets, such as treatment response, against which the diagnostic utility of candidate models can be benchmarked.

Another important desideratum is the sharing of source code: this accelerates development and reduces errors by providing standard building blocks and facilitates reproduction of results. Following the success of open source packages for "classical" neuroimaging, this development is also gaining ground in computational neuroimaging. Some models discussed in this review are already available as open source code, e.g., in SPM (http://www.fil.ion.ucl.ac.uk/spm), TAPAS (http://www.translationalneuromodeling.org/tapas), and the Virtual Brain platform (http://thevirtualbrain.org).

Conclusions

Choosing schizophrenia as an exemplary spectrum disease, this article has outlined progress in computational neuroimaging over the past decade, with a focus on generative or forward models. While all modeling approaches have made impressive advances and there is a promising trend of convergence, challenging technical and validation problems remain to be addressed in order to establish computational assays as candidate clinical tools. Given successful physiological validation, a translational neuromodeling strategy for psychiatry foresees the use of computational assays for spectrum dissection, where application to patients of a conventionally defined disease (e.g., schizophrenia) yields patient-specific vectors of model parameter estimates and/or log-evidence for alternative disease mechanisms. This quantitative profile could be used to delineate mechanistically distinct patient subgroups. Establishing the validity of these subgroup definitions requires prospective studies with regard to clinically relevant outcome criteria, such as treatment response.

ACKNOWLEDGMENTS

We acknowledge support by the René and Susanne Braginsky Foundation, the University of Zurich, the UZH Clinical Research Priority Programs (CRPP) "Molecular Imaging" and "Multiple Sclerosis," and the Deutsche Forschungsgemeinschaft (TR-SFB 134).

REFERENCES

Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D., and Friston, K.J. (2013). The computational anatomy of psychosis. Front. Psychiatry *4*, 47.

Andreasen, N.C. (1999). A unitary model of schizophrenia: Bleuler's "fragmented phrene" as schizencephaly. Arch. Gen. Psychiatry *56*, 781–787.

Anticevic, A., Gancsos, M., Murray, J.D., Repovs, G., Driesen, N.R., Ennis, D.J., Niciu, M.J., Morgan, P.T., Surti, T.S., Bloch, M.H., et al. (2012). NMDA receptor function in large-scale anticorrelated neural systems with implications for cognition and schizophrenia. Proc. Natl. Acad. Sci. USA *109*, 16720–16725.

Anticevic, A., Hu, X., Xiao, Y., Hu, J., Li, F., Bi, F., Cole, M.W., Savic, A., Yang, G.J., Repovs, G., et al. (2015). Early-course unmedicated schizophrenia patients exhibit elevated prefrontal connectivity associated with longitudinal change. J. Neurosci. *35*, 267–286.

Barch, D.M., Carter, C.S., MacDonald, A.W., 3rd, Braver, T.S., and Cohen, J.D. (2003). Context-processing deficits in schizophrenia: diagnostic specificity, 4-week course, and relationships to clinical symptoms. J. Abnorm. Psychol. *112*, 132–143.

Bleuler, E. (1911). Dementia Praecox oder Gruppe der Schizophrenien. Handbuch der Psychiatrie (Leipzig: Deuticke).

Breakspear, M., Roberts, J.A., Terry, J.R., Rodrigues, S., Mahant, N., and Robinson, P.A. (2006). A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. Cereb. Cortex *16*, 1296–1313.

Brodersen, K.H., Schofield, T.M., Leff, A.P., Ong, C.S., Lomakina, E.I., Buhmann, J.M., and Stephan, K.E. (2011). Generative embedding for model-based classification of fMRI data. PLoS Comput. Biol. *7*, e1002079.

Brodersen, K.H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W.D., Buhmann, J.M., and Stephan, K.E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin. *4*, 98–111.

Brunel, N., and Wang, X.J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. J. Comput. Neurosci. *11*, 63–85.

Buckholtz, J.W., and Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. Neuron *74*, 990–1004.

Buckner, R.L., Krienen, F.M., and Yeo, B.T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. Nat. Neurosci. *16*, 832–837.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. *10*, 186–198.

Bullmore, E.T., Frangou, S., and Murray, R.M. (1997). The dysplastic net hypothesis: an integration of developmental and dysconnectivity theories of schizophrenia. Schizophr. Res. *28*, 143–156.

Cabral, J., Fernandes, H.M., Van Hartevelt, T.J., James, A.C., Kringelbach, M.L., and Deco, G. (2013). Structural connectivity in schizophrenia and its impact on the dynamics of spontaneous functional networks. Chaos *23*, 046111.

Casey, B.J., Craddock, N., Cuthbert, B.N., Hyman, S.E., Lee, F.S., and Ressler, K.J. (2013). DSM-5 and RDoC: progress in psychiatry research? Nat. Rev. Neurosci. *14*, 810–814.

Chen, C.C., Henson, R.N., Stephan, K.E., Kilner, J.M., and Friston, K.J. (2009). Forward and backward connections in the brain: a DCM study of functional asymmetries. Neuroimage *45*, 453–462.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Corlett, P.R., Murray, G.K., Honey, G.D., Aitken, M.R., Shanks, D.R., Robbins, T.W., Bullmore, E.T., Dickinson, A., and Fletcher, P.C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. Brain *130*, 2387–2400.

Corlett, P.R., Taylor, J.R., Wang, X.J., Fletcher, P.C., and Krystal, J.H. (2010). Toward a neurobiology of delusions. Prog. Neurobiol. *92*, 345–369.

Corlett, P.R., Honey, G.D., Krystal, J.H., and Fletcher, P.C. (2011). Glutamatergic model psychoses: prediction error, learning, and inference. Neuropsychopharmacology *36*, 294–315.

Cuthbert, B.N., and Insel, T.R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. *11*, 126.

D'Ardenne, K., McClure, S.M., Nystrom, L.E., and Cohen, J.D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science *319*, 1264–1267.

Daunizeau, J., David, O., and Stephan, K.E. (2011). Dynamic causal modelling: a critical review of the biophysical and statistical foundations. Neuroimage *58*, 312–322.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., and Friston, K.J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. Neuroimage *30*, 1255–1272.

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., and Depaulis, A. (2008). Identifying neural drivers with functional MRI: an electrophysiological validation. PLoS Biol. *6*, 2683–2697.

Dayan, P. (2012). Twenty-five lessons from computational neuromodulation. Neuron *76*, 240–256.

Dayan, P., Hinton, G.E., Neal, R.M., and Zemel, R.S. (1995). The Helmholtz machine. Neural Comput. *7*, 889–904.

de Lafuente, V., and Romo, R. (2011). Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. Proc. Natl. Acad. Sci. USA *108*, 19767–19771.

Deco, G., and Jirsa, V.K. (2012). Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. J. Neurosci. *32*, 3366–3375.

Deco, G., and Kringelbach, M.L. (2014). Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. Neuron *84*, 892–905.

Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M., and Friston, K. (2008). The dynamic brain: from spiking neurons to neural masses and cortical fields. PLoS Comput. Biol. *4*, e1000092.

Deco, G., Jirsa, V.K., and McIntosh, A.R. (2013a). Resting brains never rest: computational insights into potential cognitive architectures. Trends Neurosci. *36*, 268–274.

Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G.L., Hagmann, P., and Corbetta, M. (2013b). Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. J. Neurosci. *33*, 11239–11252.

den Ouden, H.E., Friston, K.J., Daw, N.D., McIntosh, A.R., and Stephan, K.E. (2009). A dual role for prediction error in associative learning. Cereb. Cortex *19*, 1175–1185.

Dierks, T., Linden, D.E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., and Singer, W. (1999). Activation of Heschl's gyrus during auditory hallucinations. Neuron *22*, 615–621.

Dima, D., Roiser, J.P., Dietrich, D.E., Bonnemann, C., Lanfermann, H., Emrich, H.M., and Dillo, W. (2009). Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. Neuroimage *46*, 1180–1186.

Dima, D., Dietrich, D.E., Dillo, W., and Emrich, H.M. (2010). Impaired top-down processes in schizophrenia: a DCM study of ERPs. Neuroimage *52*, 824–832.

Dima, D., Frangou, S., Burge, L., Braeutigam, S., and James, A.C. (2012). Abnormal intrinsic and extrinsic connectivity within the magnetic mismatch negativity brain network in schizophrenia: a preliminary study. Schizophr. Res. *135*, 23–27.

Doya, K., Ishii, S., Pouget, A., and Rao, R.P. (2011). Bayesian Brain: Probabilistic Approaches to Neural Coding (MIT Press).

Durstewitz, D., Seamans, J.K., and Sejnowski, T.J. (2000). Neurocomputational models of working memory. Nat. Neurosci. *3* (*Suppl*), 1184–1191.

Düzel, E., Bunzeck, N., Guitart-Masip, M., Wittmann, B., Schott, B.H., and Tobler, P.N. (2009). Functional imaging of the human dopaminergic midbrain. Trends Neurosci. *32*, 321–328.

Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science *299*, 1898–1902.

Fletcher, P.C., and Frith, C.D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat. Rev. Neurosci. *10*, 48–58.

Fogelson, N., Litvak, V., Peled, A., Fernandez-del-Olmo, M., and Friston, K. (2014). The functional anatomy of schizophrenia: A dynamic causal modeling study of predictive coding. Schizophr. Res. *158*, 204–212.

Fornito, A., and Bullmore, E.T. (2015). Reconciling abnormalities of brain network structure and function in schizophrenia. Curr. Opin. Neurobiol. *30*, 44–50.

Fornito, A., Zalesky, A., and Breakspear, M. (2015). The connectomics of brain disorders. Nat. Rev. Neurosci. *16*, 159–172.

Freeman, W.J. (1975). Mass Action in the Nervous System ((Academic Press).

Freyer, F., Roberts, J.A., Becker, R., Robinson, P.A., Ritter, P., and Breakspear, M. (2011). Biophysical mechanisms of multistability in resting-state cortical rhythms. J. Neurosci. *31*, 6353–6361.

Freyer, F., Roberts, J.A., Ritter, P., and Breakspear, M. (2012). A canonical model of multistability and scale-invariance in biological systems. PLoS Comput. Biol. *8*, e1002634.

Friston, K.J. (1998). The disconnection hypothesis. Schizophr. Res. *30*, 115–125.

Friston, K. (2005a). A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 815–836.

Friston, K.J. (2005b). Hallucinations and perceptual inference. Behav. Brain Sci. *28*, 764–766.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. *11*, 127–138.

Friston, K.J., and Frith, C.D. (1995). Schizophrenia: a disconnection syndrome? Clin. Neurosci. *3*, 89–97.

Friston, K.J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. Neuroimage *19*, 1273–1302.

Friston, K.J., Shiner, T., FitzGerald, T., Galea, J.M., Adams, R., Brown, H., Dolan, R.J., Moran, R., Stephan, K.E., and Bestmann, S. (2012). Dopamine, affordance and active inference. PLoS Comput. Biol. *8*, e1002327.

Friston, K.J., Stephan, K.E., Montague, R., and Dolan, R.J. (2014). Computational psychiatry: the brain as a phantastic organ. Lancet Psychiatry *1*, 148–158.

Fukushima, M., Yamashita, O., Knösche, T.R., and Sato, M.A. (2015). MEG source reconstruction based on identification of directed source interactions on whole-brain anatomical networks. Neuroimage *105*, 408–427.

Garrido, M.I., Kilner, J.M., Stephan, K.E., and Friston, K.J. (2009). The mismatch negativity: a review of underlying mechanisms. Clin. Neurophysiol. *120*, 453–463.

Ghosh, A., Rho, Y., McIntosh, A.R., Kötter, R., and Jirsa, V.K. (2008). Noise during rest enables the exploration of the brain's dynamic repertoire. PLoS Comput. Biol. *4*, e1000196.

Gonzalez-Burgos, G., and Lewis, D.A. (2012). NMDA receptor hypofunction, parvalbumin-positive neurons, and cortical gamma oscillations in schizophrenia. Schizophr. Bull. *38*, 950–957.

Gradin, V.B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., and Steele, J.D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. Brain *134*, 1751–1764.

Harrison, P.J., and Weinberger, D.R. (2005). Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Mol. Psychiatry *10*, 40–68, 5.

Havlicek, M., Friston, K.J., Jan, J., Brazdil, M., and Calhoun, V.D. (2011). Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. Neuroimage *56*, 2109–2128.

Heinz, A. (2002). Dopaminergic dysfunction in alcoholism and schizophrenia–psychopathological and behavioral correlates. Eur. Psychiatry *17*, 9–16.

Hiroyuki, N. (2014). Multiplexing signals in reinforcement learning with internal models and dopamine. Curr. Opin. Neurobiol. *25*, 123–129.

Homayoun, H., and Moghaddam, B. (2007). NMDA receptor hypofunction produces opposite effects on prefrontal cortex interneurons and pyramidal neurons. J. Neurosci. *27*, 11496–11500.

Honey, C.J., Kötter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc. Natl. Acad. Sci. USA *104*, 10240–10245.

Horga, G., Schatz, K.C., Abi-Dargham, A., and Peterson, B.S. (2014). Deficits in predictive coding underlie hallucinations in schizophrenia. J. Neurosci. *34*, 8072–8082.

Howes, O.D., and Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III–the final common pathway. Schizophr. Bull. *35*, 549–562.

Husain, F.T., Tagamets, M.A., Fromm, S.J., Braun, A.R., and Horwitz, B. (2004). Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. Neuroimage *21*, 1701–1720.

Hyett, M.P., Breakspear, M.J., Friston, K.J., Guo, C.C., and Parker, G.B. (2015). Disrupted effective connectivity of cortical systems supporting attention and interoception in melancholia. JAMA Psychiatry *72*, 350–358.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E., and Stephan, K.E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron *80*, 519–530.

Insel, T.R. (2010). Rethinking schizophrenia. Nature *468*, 187–193.

Iyer, K.K., Roberts, J.A., Hellström-Westas, L., Wikström, S., Hansen Pupp, I., Ley, D., Vanhatalo, S., and Breakspear, M. (2015). Cortical burst dynamics predict clinical outcome early in extremely preterm infants. Brain. Published online May 22, 2015. pii: awv129.

Jansen, B.H., and Rit, V.G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol. Cybern. *73*, 357–366.

Jardri, R., and Denève, S. (2013). Circular inferences in schizophrenia. Brain *136*, 3227–3241.

Jirsa, V.K., and Haken, H. (1996). Field theory of electromagnetic brain activity. Phys. Rev. Lett. *77*, 960–963.

Jirsa, V.K., Sporns, O., Breakspear, M., Deco, G., and McIntosh, A.R. (2010). Towards the virtual brain: network modeling of the intact and the damaged brain. Arch. Ital. Biol. *148*, 189–205.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am. J. Psychiatry *160*, 13–23.

Kapur, S., Phillips, A.G., and Insel, T.R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol. Psychiatry *17*, 1174–1179.

Klein-Flügge, M.C., Hunt, L.T., Bach, D.R., Dolan, R.J., and Behrens, T.E. (2011). Dissociable reward and timing signals in human midbrain and ventral striatum. Neuron *72*, 654–664.

Kötter, R., and Feizelmeier, M. (1998). Species-dependence and relationship of morphological and electrophysiological properties in nigral compacta neurons. Prog. Neurobiol. *54*, 619–632.

Krystal, J.H., and State, M.W. (2014). Psychiatric disorders: diagnosis to therapy. Cell *157*, 201–214.

Laruelle, M., Kegeles, L.S., and Abi-Dargham, A. (2003). Glutamate, dopamine, and schizophrenia: from pathophysiology to treatment. Ann. N Y Acad. Sci. *1003*, 138–158.

Li, B., Daunizeau, J., Stephan, K.E., Penny, W., Hu, D., and Friston, K. (2011). Generalised filtering and stochastic DCM for fMRI. Neuroimage *58*, 442–457.

Lieder, F., Daunizeau, J., Garrido, M.I., Friston, K.J., and Stephan, K.E. (2013). Modelling trial-by-trial changes in the mismatch negativity. PLoS Comput. Biol. *9*, e1002911.

Lisman, J.E., Fellous, J.M., and Wang, X.J. (1998). A role for NMDA-receptor channels in working memory. Nat. Neurosci. *1*, 273–275.

Lisman, J.E., Coyle, J.T., Green, R.W., Javitt, D.C., Benes, F.M., Heckers, S., and Grace, A.A. (2008). Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. Trends Neurosci. *31*, 234–242.

Lynall, M.E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., and Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. J. Neurosci. *30*, 9477–9487.

Maia, T.V., and Frank, M.J. (2011). From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. *14*, 154–162.

Marder, E., and Goaillard, J.M. (2006). Variability, compensation and homeostasis in neuron and network function. Nat. Rev. Neurosci. *7*, 563–574.

Marreiros, A.C., Kiebel, S.J., and Friston, K.J. (2010). A dynamic causal model study of neuronal population dynamics. Neuroimage *51*, 91–101.

Mathys, C., Daunizeau, J., Friston, K.J., and Stephan, K.E. (2011). A bayesian foundation for individual learning under uncertainty. Front. Hum. Neurosci. *5*, 39.

Matsumoto, M., and Takada, M. (2013). Distinct representations of cognitive and motivational signals in midbrain dopamine neurons. Neuron *79*, 1011–1024.

Montague, P.R., Dolan, R.J., Friston, K.J., and Dayan, P. (2012). Computational psychiatry. Trends Cogn. Sci. *16*, 72–80.

Moran, R.J., Jung, F., Kumagai, T., Endepols, H., Graf, R., Dolan, R.J., Friston, K.J., Stephan, K.E., and Tittgemeyer, M. (2011a). Dynamic causal models and physiological inference: a validation study using isoflurane anaesthesia in rodents. PLoS ONE *6*, e22790.

Moran, R.J., Symmonds, M., Stephan, K.E., Friston, K.J., and Dolan, R.J. (2011b). An in vivo assay of synaptic function mediating human cognition. Curr. Biol. *21*, 1320–1325.

Moran, R.J., Campo, P., Symmonds, M., Stephan, K.E., Dolan, R.J., and Friston, K.J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. J. Neurosci. *33*, 8227–8236.

Murray, G.K., Corlett, P.R., Clark, L., Pessiglione, M., Blackwell, A.D., Honey, G., Jones, P.B., Bullmore, E.T., Robbins, T.W., and Fletcher, P.C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol. Psychiatry *13*, 239, 267–276.

Murray, J.D., Anticevic, A., Gancsos, M., Ichinose, M., Corlett, P.R., Krystal, J.H., and Wang, X.J. (2014). Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model. Cereb. Cortex *24*, 859–872.

Nakagawa, T.T., Woolrich, M., Luckhoo, H., Joensson, M., Mohseni, H., Kringelbach, M.L., Jirsa, V., and Deco, G. (2014). How delays matter in an oscillatory whole-brain spiking-neuron network model for MEG alpha-rhythms at rest. Neuroimage *87*, 383–394.

Notredame, C.E., Pins, D., Deneve, S., and Jardri, R. (2014). What visual illusions teach us about schizophrenia. Front. Integr. Nuerosci. *8*, 63.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal difference models and reward-related learning in the human brain. Neuron *38*, 329–337.

Olier, I., Trujillo-Barreto, N.J., and El-Deredy, W. (2013). A switching multiscale dynamical network model of EEG/MEG. Neuroimage *83*, 262–287.

Omurtag, A., Knight, B.W., and Sirovich, L. (2000). On the simulation of large populations of neurons. J. Comput. Neurosci. *8*, 51–63.

Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T.T., Kiebel, S.J., and Blankenburg, F. (2012). Evidence for neural encoding of Bayesian surprise in human somatosensation. Neuroimage *62*, 177–188.

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., and O'Doherty, J.P. (2013). The neural representation of unexpected uncertainty during value-based decision making. Neuron *79*, 191–201.

Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., and Leff, A.P. (2010). Comparing families of dynamic causal models. PLoS Comput. Biol. *6*, e1000709.

Pettersson-Yeo, W., Allen, P., Benetti, S., McGuire, P., and Mechelli, A. (2011). Dysconnectivity in schizophrenia: where are we now? Neurosci. Biobehav. Rev. *35*, 1110–1124.

Rao, R.P., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. *2*, 79–87.

Rigoux, L., and Daunizeau, J. (2015). Dynamic causal modelling of brain-behaviour relationships. Neuroimage *117*, 202–221.

Robinson, P.A., Rennie, C.J., Wright, J.J., Bahramali, H., Gordon, E., and Rowe, D.L. (2001). Prediction of electroencephalographic spectra from neurophysiology. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *63*, 021903.

Roeper, J. (2013). Dissecting the diversity of midbrain dopamine neurons. Trends Neurosci. *36*, 336–342.

Roiser, J.P., Howes, O.D., Chaddock, C.A., Joyce, E.M., and McGuire, P. (2013). Neural and behavioral correlates of aberrant salience in individuals at risk for psychosis. Schizophr. Bull. *39*, 1328–1336.

Romaniuk, L., Honey, G.D., King, J.R., Whalley, H.C., McIntosh, A.M., Levita, L., Hughes, M., Johnstone, E.C., Day, M., Lawrie, S.M., and Hall, J. (2010). Midbrain activation during Pavlovian conditioning and delusional symptoms in schizophrenia. Arch. Gen. Psychiatry *67*, 1246–1254.

Rutledge, R.B., Skandali, N., Dayan, P., and Dolan, R.J. (2014). A computational and neural model of momentary subjective well-being. Proc. Natl. Acad. Sci. USA *111*, 12252–12257.

Sanz-Leon, P., Knock, S.A., Spiegler, A., and Jirsa, V.K. (2015). Mathematical framework for large-scale brain network modeling in The Virtual Brain. Neuroimage *111*, 385–430.

Schlagenhauf, F., Huys, Q.J., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.J., Dolan, R., and Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. Neuroimage *89*, 171–180.

Schmack, K., Gòmez-Carrillo de Castro, A., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J.D., Heinz, A., Petrovic, P., and Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. J. Neurosci. *33*, 13701–13712.

Schmidt, A., Diaconescu, A.O., Kometer, M., Friston, K.J., Stephan, K.E., and Vollenweider, F.X. (2013). Modeling ketamine effects on synaptic plasticity during the mismatch negativity. Cereb. Cortex *23*, 2394–2406.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Schwartenbeck, P., FitzGerald, T.H., Mathys, C., Dolan, R., and Friston, K. (2014). The dopaminergic midbrain encodes the expected certainty about desired outcomes. Cereb. Cortex. Published online July 23, 2014. pii: bhu159.

Seghier, M.L., and Friston, K.J. (2013). Network discovery with large DCMs. Neuroimage *68*, 181–191.

Stephan, K.E. (2013). The history of CoCoMac. Neuroimage *80*, 46–52.

Stephan, K.E., and Mathys, C. (2014). Computational approaches to psychiatry. Curr. Opin. Neurobiol. *25*, 85–92.

Stephan, K.E., Baldeweg, T., and Friston, K.J. (2006). Synaptic plasticity and dysconnection in schizophrenia. Biol. Psychiatry *59*, 929–939.

Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., and Friston, K.J. (2007). Comparing hemodynamic models with DCM. Neuroimage *38*, 387–401.

Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E., Breakspear, M., and Friston, K.J. (2008). Nonlinear dynamic causal models for fMRI. Neuroimage *42*, 649–662.

Stephan, K.E., Friston, K.J., and Frith, C.D. (2009). Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. Schizophr. Bull. *35*, 509–527.

Uhlhaas, P.J. (2013). Dysconnectivity, large-scale networks and neuronal dynamics in schizophrenia. Curr. Opin. Neurobiol. *23*, 283–290.

Umbricht, D., and Krljes, S. (2005). Mismatch negativity in schizophrenia: a meta-analysis. Schizophr. Res. *76*, 1–23.

Umbricht, D., Koller, R., Vollenweider, F.X., and Schmid, L. (2002). Mismatch negativity predicts psychotic experiences induced by NMDA receptor antagonist in healthy volunteers. Biol. Psychiatry *51*, 400–406.

Valdes, P.A., Jimenez, J.C., Riera, J., Biscay, R., and Ozaki, T. (1999). Nonlinear EEG analysis based on a neural mass model. Biol. Cybern. *81*, 415–424.

Valdes-Sosa, P.A., Roebroeck, A., Daunizeau, J., and Friston, K. (2011). Effective connectivity: influence, causality and biophysical modeling. Neuroimage *58*, 339–361.

van Leeuwen, T.M., den Ouden, H.E., and Hagoort, P. (2011). Effective connectivity determines the nature of subjective experience in grapheme-color synesthesia. J. Neurosci. *31*, 9879–9884.

van Os, J., Rutten, B.P., and Poulton, R. (2008). Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. Schizophr. Bull. *34*, 1066–1082.

Wang, X.J., and Krystal, J.H. (2014). Computational psychiatry. Neuron *84*, 638–654.

Wernicke, C. (1906). Grundrisse der Psychiatrie (Leipzig: Thieme).

Wiecki, T.V., and Frank, M.J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. Psychol. Rev. *120*, 329–355.

Wiecki, T.V., Poland, J.S., and Frank, M.J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. Clin. Psychol. Sci. *3*, 378–399.

Wong, K.F., and Wang, X.J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. *26*, 1314–1328.

Woolrich, M.W., and Stephan, K.E. (2013). Biophysical network models and the human connectome. Neuroimage *80*, 330–338.

Yang, G.J., Murray, J.D., Repovs, G., Cole, M.W., Savic, A., Glasser, M.F., Pittenger, C., Krystal, J.H., Wang, X.J., Pearlson, G.D., et al. (2014). Altered global brain signal in schizophrenia. Proc. Natl. Acad. Sci. USA *111*, 7438–7443.

Yu, A.J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681–692.