The Schizophrenic Brain: A Broken Hermeneutic Circle. Some New Insights and Results

Péter Érdi, Mihály Bányai, Balázs Ujfalussy and Vaibhav Diwadkar

Abstract—Schizophrenia is often regarded as a set of symptoms caused by impairments in the cognitive control in macro-networks of the brain. To investigate this hypothesis, an fMRI study involving an associative learning task was conducted with schizophrenia patients and controls. A set of generative models of the BOLD signal generation were defined to describe the interaction of five brain regions (Primary Visual Cortex, Superior Parietal and Inferior Temporal Cortex, Hippocampus and Dorsal Prefrontal Cortex) and the experimental conditions. The models were fitted to the data using Bayesian model inversion. The comparison of different model connectivity structures lead to the finding that in schizophrenia, there are significant impairments in backward connections from prefrontal cortex to hippocampal and temporal regions in patients. These findings fit very well with the predictions of a neuron network model of encoding-recall switching dynamics we have described previously.

I. INTRODUCTION

The notion that the philosophical approach called hermeneutics could and should be applied to understand the brain was suggested by Tsuda [1]. One of the most important concepts in hermeneutics is the hermeneutic circle. Previous combined behavioral, fMRI and computational comparatives studies of schizophrenic and control patients by an associative learning paradigm suggested that some symptoms of schizophrenia in may be understood as some failures in the hermeneutic circle [2]. More precisely, our model study suggested that the impairment of cognitive control of the prefrontal cortex on hippocampal processes implies uncertainties in the task to be solved and will result in poorer performance in learning and recall processes. Now we re-evaluate the data about the functional macro-networks by adopting Dynamical Causal Modelling (DCM) [3]. We also relate our new findings to the previous results from neural network modelling.

II. DISCONNECTION HYPOTHESES OF SCHIZOPHRENIA

It has been hypothesized that schizophrenia is best understood in terms of pathological interactions between different brain regions. This claim can be formulated on two different levels: on one hand, we can investigate the structural organization of the cortex and look for physical differences between patients and controls, and on the other hand, we can investigate whether there is a difference in the task-dependent functional interaction of cortical areas regardless to the underlying neural structure.

The concept of structural “disconnection syndrome” goes back at least to Wernicke [5], who interpreted psychosis as a result anatomically disrupted connections. It was reintroduced by Geschwind [6], [7], and had a crucial role in behavioral neurology and psychiatry [8]. Newer meta-analysis also supported the hypothesis [9].

Specifically, we are interested in the functional reduction, both the qualitative and quantitative nature of it. Consequently, there are two questions to be answered: (i) what are the differences in the model architectures describing the information processing network of healthy and schizophrenia subjects, and (ii) which connections are significantly impaired during schizophrenia? In technical terms, we analyze effective connectivities, which reflect the causal influence that one brain region exerts over another. Effective connectivity can consist of two components. The first characterizes the intrinsic connectivity of the network, and the second models input-dependent changes in them. (Inputs, however, may have effects for brain dynamics not only by modulating connections, but also via direct or indirect influence on specific regions). Pathological connectivities may appear in both components.

Task-related functional connectivity can be investigated with respect to various functions of the brain, e.g. learning, memory, control, etc. We studied associative learning, since this is a cortical function that requires the integration of multiple sensory, representation and cognitive control pathways, making it a useful approach to grasp disordered functional interaction.

III. METHODS

A. Experimental paradigm

A paired-associate learning paradigm was adopted in which subjects are required to learn arbitrary associations between locations (in space) and objects (with unique identities), see Fig 1. The two kinds of memoranda (“where” and “what”) are processed by two components, i.e. spatial (dorsal) and object (ventral), of the forward visual pathway. These information streams converge in the hippocampus, with potential supervisory inputs from the prefrontal cortex. A macro-network provides a framework to estimate model architectures using DCM.
B. Dynamic causal modelling

DCM provides a complete phenomenological model framework for the analysis of fMRI data. For a detailed description see [3]. The model structure consists of two components: a neural state equation and a hemodynamic model. The neural component describes the time evolution of the neural state variables, \( x \), which refer to the neural activity of the brain areas. This is a bilinear formula of the state variables themselves and the input variables, \( u \), which are the conditions defined by the experiment (Eq. 1). The connectivity parameters of the neural model are the elements of the three matrices, \( \theta_n = \{ A, B, C \} \). \( A \) contains the intrinsic coupling parameters, the causal effects of the areas on each other, \( B \) contains the modulatory parameters, the effects of the inputs on the intrinsic connections, and \( C \) contains the direct effects of the inputs on the areas.

\[
\dot{x} = (A + \sum_{i=1}^{N} u_i B^i)x + C u \tag{1}
\]

\[
y = \lambda(x, \theta_h) \tag{2}
\]

The hemodynamic component describes the nonlinear mapping \( \lambda \) from the neural activity to the fMRI signal, \( y \), actually measured in the \( N \) brain areas (Eq. 2). For the details of the hemodynamic model see [4]. We need to estimate the values of the parameter set, \( \theta = \{ \theta_h, \theta_n \} \) best fitting to measurement data. One possible procedure to do so is the Bayesian maximum a posteriori (MAP) estimation technique defined by Eq. 3, where \( M \) denotes the specific connectivity pattern of the model.

\[
p(\theta \mid y, M) = \frac{p(y \mid \theta, M)p(\theta \mid M)}{p(y \mid M)} \tag{3}
\]

For all probability distributions in 3, we assume that both the prior \( p(\theta \mid M) \) and posterior \( p(\theta \mid y, M) \) distributions are Gaussians, and the MAP estimation is defined as the mean of the posterior distribution. To compare models with different connectivity patterns, we can set the prior probability of having certain connections is a certain model to zero.

C. Comparison of models

We can compare models with different connectivity patterns in a Bayesian way by estimating their model evidence:

\[
p(y \mid M) = \int p(y \mid \theta, M)p(\theta \mid M) d\theta \tag{4}
\]

The model evidence is the probability of obtaining the actual measurement conditioned on the model form, integrated on the whole parameter space of the model. To obtain the expected posterior probabilities of models in a model set, we use a variational Bayesian method, i.e. a variational free energy bound on log evidence for each subject-model pair. These log evidences are then used to compute the relative evidence for different models over the group to provide the expected posterior probability of each model as described in [10].

D. Model definitions

Five regions were used in the modeling studies, specifically Primary Visual Cortex (V1), Superior Parietal (SP), Inferior Temporal Cortex (IT), Hippocampus (HPC) and Dorsal Prefrontal Cortex (PFC). A set of causal models of fMRI signal generation (with the mathematical structure described above) were defined to evaluate connections between five regions material to the task. To model the information processing in the associative learning task, we assumed the presence of two streams connecting the five brain regions. The “forward” or “data” stream propagates sensory information at different levels of processing from the low-level sensory areas towards high-level cognitive areas. The “backward” or “control” stream propagates control signals from the high-level areas towards the lower-level ones. In this paper we examine impairments in cognitive control, so the focus of the investigations is the control stream.

Exogenous inputs corresponded to the box-car stimulus function in Fig 1 and a time-dependent effect that modelled learning or plasticity over blocks. The first effect entered as a driving input to the visual area, whereas the second modulated all connections in any given model.

Multiple models were evaluated by varying hypothesis-related intrinsic connections between regions, while fixing other connections (\( A \) matrix). We included the intrinsic...
connections of the data stream to all models. Based on the hypothetical control stream we defined three additional connections that may extend the basic model in different combinations. The eight possible combinations of these connections constitute the model class. All possible intrinsic connections are visualized in Fig. 3.

Fig. 3. The model space for varying intrinsic connections. Connections marked by dashed line are varied.

IV. RESULTS

A. Model selection

The results show that in the control group there is a clear winner (see Model no 8 in 4 for the intrinsic connection patterns, the model that contains the full control stream. In the SCZ group, there is no clear winner, there are several more probable models, and the differences are smaller between model probabilities. It can also be seen that while the winning model in the HC group contains all the connections defined, while the most probable models in the SCZ group lack more or less connections. This result implies that the information processing network of schizophrenia patients is fundamentally different than the one of controls. However, the model selection does not provide the specific pathways being impaired, so the parameter level analysis is also necessary.

Fig. 4. Model comparison results showing impairment in prefrontal control of memory formation in patients.

B. Effective connectivities

In the next step of the analysis, we give a more detailed quantitative characterization of the results. At the parameter level, we look for significant differences in the effective connectivity in the models fit to the data of the two subject groups, assuming fixed model structure. To do so, we selected a reference model for comparison by running the model selection for all subjects, no distinction by group.

The winning model is the one containing all hypothesized connections. The means and standard deviations of the intrinsic coupling and modulatory parameters are depicted in Fig. 5. To obtain the significance of the differences we applied a two-sample t-test on the parameter values in the two groups.

Fig. 5. Average connectivity parameters for HC and SCZ (dark) groups. The significant differences are in the prefronto-hippocampal and hippocampo-inferior temporal pathways.

The significant differences between groups are in the strength of intrinsic connections between prefrontal cortex and hippocampus and between hippocampus and inferior temporal cortex. All these connections are weakened in the SCZ group, supporting the hypothesis about impaired effective connectivity in the control stream in schizophrenia. Both these connections are playing important roles in the cognitive control of the associative memory formation. Furthermore, we see the reduced effects of Time on these causal links meaning reduced excitatory contextual modulation by learning. This can be seen as reduced task-related plasticity of a pathway in the illness.

Note that we did not apply Bonferroni correction to the T-tests because we hypothesized in advance that backward (controlling) connections from the prefrontal cortex would be impaired and schizophrenia. We report all the T tests to show that this prediction was confirmed and that this was the greatest group difference among all connections.

C. Neural network model

Our previous results demonstrated the schizophrenic patients show a significant impairment in object-location associative learning tasks [11]. Here we developed a neural network model incorporating brain regions involved in paired-associate learning in order to analyze the mechanisms underlying behavioral differences between schizophrenic patients and control subjects.

A feed-forward network creates the representations of the identity of the object and its location in the model of the area IT and SP in the ventral and dorsal visual stream, respectively. The proposed role of the hippocampus is to bind these two representations together so that when cued by the location, the correct object can be recalled. Moreover, in order to model cognitive control, we included a prefrontal region which controls learning and recall processes presumably by modulating the plasticity and the efficiency of hippocampal synapses. We fitted the model’s performance
to the behavioral data (Fig. 6). Our model predicts that the impairment of cognitive control of the prefrontal cortex over hippocampal processes implies inaccurate regulation of hippocampal dynamics and explains the poorer performance of patients in this task.

Fig. 6. Learning curves of schizophrenia patients (red) healthy controls (blue) and simulation results from the neural network model (white) with two parameter settings.

The predictions of the neural network model are in accordance with the findings of the DCM study. This multi-level approach is suitable to integrate the explanatory capabilities of mechanistic neural models with the analytical power of data-driven phenomenological approaches.

V. CONCLUSIONS

A widely studied aspect of schizophrenia is the impairment of cognitive control over information processing cortical circuits. The prefrontal cortex is the area that is consensually considered as the center of cognitive control functions, such as attention, memory and executive functions. Our interpretation is that at least some symptoms of schizophrenia may be understood as some failures in the hermeneutic circle. While the breaking of the circle may lead to schizophrenic symptoms, combined pharmacological psychotherapeutic strategies should act to repair the circle.

REFERENCES