Inferring a Cognitive Architecture from Multi-Task Neuroimaging Data: A Data-Driven Test of the Common Model of Cognition Using Granger Causality

Holly S. Hake (hakehs@uw.edu)
Department of Psychology, University of Washington, Seattle, WA 98195 USA

Catherine Sibert (sibert@uw.edu)
Department of Psychology, University of Washington, Seattle, WA 98195 USA

Andrea Stocco (stocco@uw.edu)
Department of Psychology, University of Washington, Seattle, WA 98195 USA

Abstract

A common complaint levied at analyses based on cognitive architectures is their lack of connection to observed functional neuroimaging data, particularly for architectural models that rely on high level, theoretical components of cognition. Previous work has connected task-based functional MRI data to the Common Model of Cognition (CMC), using a top-down modeling approach. Here, a bottom-up method, Granger Causality Modeling (GCM), is applied to the same task-based data to infer a network of causal connections between brain regions based on their functional connectivity. The resulting network shares many connections with those proposed by the Common Model, and also suggests important additions to the Common Model, likely related to the role of episodic memory in control.

Keywords: Cognitive Modeling, Cognitive Architecture, Granger Causality, Functional Connectivity

Introduction

In the field of cognitive architectures, an important topic of discussion is the relationship between the components of an architecture and their relationship to the brain. Some architectures, like SPAUN (Eliasmith et al., 2012), LISA (Hummel & Holyoak, 2005), and Leabra (O'Reilly et al., 2016), are designed to mimic the brain's biological circuits and rely on artificial neurons as their building blocks. These systems take a circuit-level approach to cognitive modeling, based on the notion that function arises from form. An alternate, functional approach forms the basis of another class of architectures, such as Soar (Laird, 2019), or ACT-R (Anderson, 2007), whose building blocks are more abstract and high-level cognitive components such as perceptual systems and memory that have been then mapped *post-hoc* to particular brain regions (e.g., Anderson, Fincham, Qin, & Stocco, 2008).

Ultimately, the success of both bottom-up and top-down approaches depends on one fundamental aspect, that is, the exact nature of the functional connections between the assumed components, or the underlying brain's architecture. Surprisingly, the fields of systems-level neuroscience and the fields of cognitive architectures have rarely interacted in this domain. In this paper, we attempt to resolve some of the tensions between the competing methodologies by using Granger Causality Modeling (GCM) of low-level functional brain activity to find causal connections between brain regions associated with high-level cognitive components. The networks produced by these connections are then compared to existing frameworks of theoretical architectures.

Functional Connectivity

Most research aimed at understanding brain architecture has been done through the analysis of *functional connectivity*, a data-driven and bottom-up method of determining the degree of connection between brain regions through statistical dependencies--typically, the Pearson correlation between times series in different brain regions (e.g., Fox et al., 2005). Through this method, network neuroscientists have identified several distinct networks of brain regions, such as the Default Mode Network (Raichle et al., 2001). However, while functional connectivity analysis can detect the presence of such networks, it can be difficult to characterize the specific function or role that they play in higher level cognition. Furthermore, correlation coefficients have no directionality attached to them, which makes it impossible to draw causal conclusions about the role of different regions and the flow of information along a network.

The Common Model of Cognition

A number of recent studies have tried to connect architecture frameworks to functional brain activity in a top-down fashion, by imposing architectural constraints on a network of connected brain regions. In particular, these studies have capitalized on the Common Model of Cognition (CMC), an abstract description of the principles common to multiple architectures (Laird, Lebiere, & Rosenbloom, 2017). The CMC proposes that, at the highest level, cognition arises from the interaction of five cognitive components, corresponding to Perception, Action, Long-Term Memory, Procedural Memory, and Working Memory. These components can be associated with five corresponding large-scale brain circuits, and a network of directional connections can be drawn between them. Most recently, Stocco et al. (2021) showed that the CMC outperforms a selection of six other architectures in fitting data across six different task paradigms spanning seven different domains, suggesting that it provides a reasonably accurate system-level description of the brain's architecture.

Unfortunately, all of the previous tests of the CMC (Stocco et al., 2018; Steine-Hanson et al., 2018; Stocco et al., 2021, have employed a top-down approach, comparing the relative fit of different possible architectures. This approach was partially constrained by the choice of one particular method of the analysis of effective connectivity, Dynamic Causal Modeling (DCM: Friston, Harrison, & Penny, 2003). The authors justified the choice of DCM because it allows for the distinguishing of the directionality of connections, while the most commonly used functional connectivity measures are based on partial correlations and are non-directional. While DCM allows for directional estimates, it relies on the top-down implementation of a plausible architecture, and it also limits the use of a bottom-up, data-driven approach. Because the space of possible architectures, even when only five components are considered, is extremely large, it is possible that a better candidate architecture exists, but was simply not included among those examined by Stocco et al. (2021).

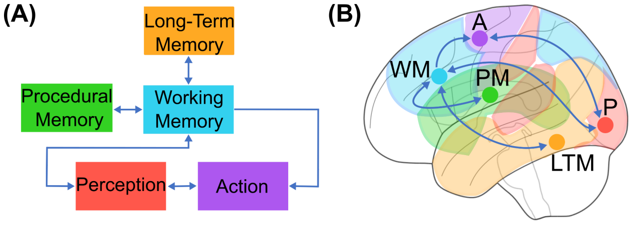

Figure 1: (A) The Common Model of Cognition (CMC); (B) Proposed associations between components and anatomical brain regions.

Granger Causality Modeling

In this paper, we re-analyze the data from the Stocco et al (2021) paper using Granger Causality Modeling. In GCM, the existence of a causal effect between two time series x and y is established by comparing two models (Granger, 1969), one auto-regressive linear model in which the value of y at times t depends only on its past value at time t-1:

$$y(t) = \beta_0 + \beta_1\, y(t-1)$$

and an alternative model that includes the effect of the past state of x:

$$y(t) = \beta_0 + \beta_1\, y(t-1) + \beta_2\, x(t-1)$$

If the second model is significantly better than the first, then x is said to Granger-cause y. Although it was originally developed and applied in the field of economics, Granger causality has been successfully applied to neuroimaging data (Roebreck et al., 2005; Deshpande et al., 2008) and offers similar advantages and comparable performance to DCM (Friston et al., 2012). In this paper, we apply this method to test the existence of all possible causal connections between the five components proposed by the CMC.

Materials and Methods

Participants

The study presented herein consists of an extensive analysis of a large sample ($N = 200$) of neuroimaging data from the Human Connectome Project (HCP), the largest existing repository of healthy young adult neuroimaging data.

Task fMRI Data

The HCP task-fMRI data encompasses seven different paradigms designed to capture a wide range of cognitive capabilities. Of these paradigms, six were included in our analysis (the seventh was a motor localization task). A full description of these tasks and the rationale for their selection can be found in the original HCP papers (Barch et al., 2013; Van Essen et al., 2013).

Data Processing and Analysis

Image Acquisition and Preprocessing. MRI images were acquired and minimally preprocessed according to HCP guidelines (Barch et al., 2013; Van Essen et al., 2013). Scans were taken on a 3T Siemens Skyra using a 32-channel head coil with acquisition parameters set at TR = 720 ms, TE = 33.1 ms, FA = 52°, FOV = 208 × 180 mm. Each image contained 72 2.0mm oblique slices with an in-plane 2.0 x 2.0 mm resolution. Images were acquired with a multi-band acceleration factor of 8X. These raw images then underwent minimal preprocessing including unwarping, motion realignment, and normalization to the standard MNI template. From there, the images were then smoothed with an isotropic 8.0 mm full-width half maximum Gaussian kernel.

Regions of Interest Definition. Regions of Interest (ROIs) for each task and participant were defined using the method described in Stocco et al. (2021) and available on the paper's online repository. For each CMC component, a group-level centroid was first identified by running a canonical GLM analysis that compared the stimuli against their task-specific baseline and then locating the peak of a statistical parametric map within the general areas associated with that CMC component (Figure 1). Because all tasks show stronger activation in the left hemisphere than in the right, all the group-level centroids were located in the left hemisphere.

To account for individual-level variability in functional neuroanatomy, the group-level coordinates were then used as the starting point to search in 3D space for the closest activation peak within each individual statistical parameter map. Figure 2 illustrates the distribution of the individual coordinates of each region for each task, overlaid over a corresponding group-level statistical map of task-related activity (as in Stocco et al., 2021). Each individual coordinate is represented by a point; the ≈200 points for each region form a cloud that captures the spatial variability in the distribution of the individual coordinates for that region. Next, the individualized ROI coordinates were used as the center of a spherical ROI with an 8mm radius. All voxels within the sphere whose response was significant at a minimal threshold of $p < .50$ (that is, a 50% probability of showing a response) were included as part of the ROI.

Finally, for each ROI of every participant in every task, a representative time course of the BOLD signal was extracted as the first principal component of the time series of all of the voxels within the sphere. The resulting time series, one per component, were then entered into a Granger causality model.

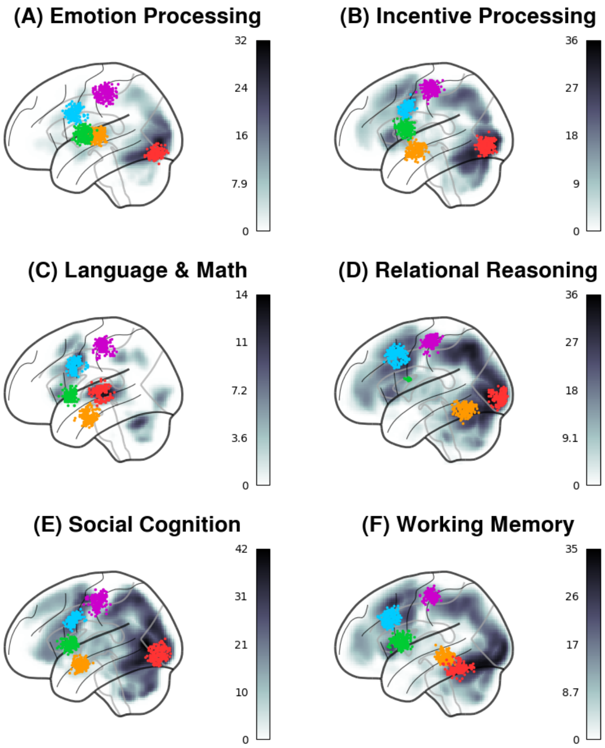

Figure 2: Location of ROI centroids across the six tasks of the Human Connectome Project; variations account for individual differences in functional anatomy.

Granger Causality Model

A multivariate Granger causality model was then set up, in which the BOLD response at time t across all regions, $\mathbf{x}(t)$, was modeled as the contribution of all of the regions (including itself) at lags 1, 2, … k:

$$\mathbf{x}(t) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\, \mathbf{x}(t\text{-}1) + \ldots + \boldsymbol{\beta}_k\, \mathbf{x}(t\text{-}k) \tag{1}$$

To determine the optimal lag value, ten models were created by varying k from $k = 1$ to $k = 10$, and the value of k that gave rise to the model with the lowest Bayesian Information Criterion was selected. Across all participants and tasks, the maximum lag that was observed was $k = 6$, and the modal was $k = 2$. Note that, when $k > 1$, there are multiple different parameter estimates that quantify the directional effect of a region on another region, one for each lag. To reduce the dimensionality of these estimates, only the most significant lag (i.e., the one with the smallest p-value) was selected.

For each participant, a subject-level inferred architecture was then created by discretizing the matrix of connections and maintaining only directed links with $p < .05$. To infer a group-level architecture from the individual-level architectures, the most likely directed links between regions need to be inferred from the frequency of their distribution in the sample of participants. To determine the probability that each directed connection c is part of the group-level architecture, we modeled the probability of it appearing across all participants as a binomial distribution, with a prior probability of $P = .50$.

Results

Group level connection maps for each of the six tasks used in the HCP dataset are shown in Figure 3. The figures show a connectivity matrix representation of the inferred architectures for each task, where the brightness of each matrix cell reflects the probability that the corresponding directional connection should be included in the architecture.

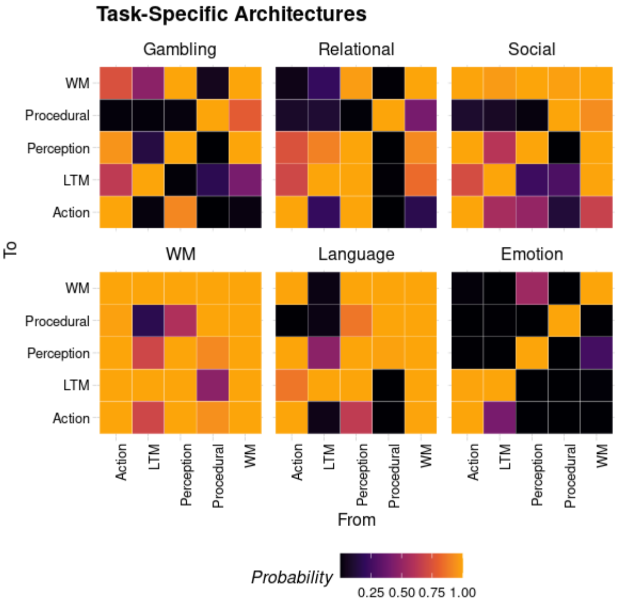

Figure 3: Task-specific connection grids for each of the six HCP task paradigms. Each grid square represents a potential causal connection between two regions, and the brightness of the square reflects the probability of that link at a group level.

The figure highlights the different connections utilized in each task domain, as well as some commonalities shared by all tasks (i.e., the preponderance of connections to and from the WM component, corresponding to the PFC).

As pointed out by Stocco et al. (2021), however, an efficient architecture should be *stable* and maintain its functional characteristics across different tasks. Therefore, to derive a general architecture from these six task-specific ones, we considered each task as an independent experiment to test these connections, and we used Fisher's (1932) method to combine the *p*-values from each task. According to this method, the distribution of the log of *p*-values from independent tests follows a χ^2 distribution with $2N$ degrees of freedom, and the p-value of each connection can be calculated from the χ^2 cumulative distribution function as follows:

$$p_{\text{global}} = p(\chi^2_{2N} > \textstyle\sum_{\text{task}} \log p_{\text{task}})$$

The results are shown in Figure 4, which represents the connectivity matrix of an architecture inferred across participants *and* domains.

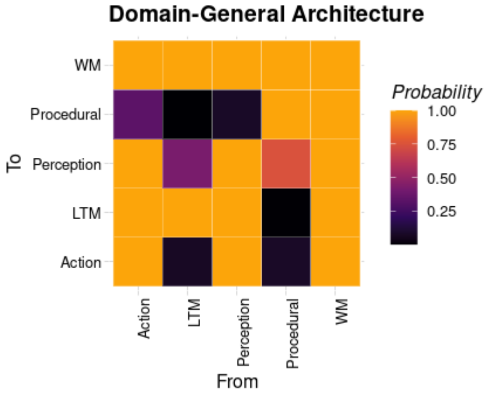

Figure 4: The connection grid for a general architecture incorporating all six HCP tasks. Each grid square represents a potential causal connection between regions, and the brightness of the square reflects the probability of that connection being present.

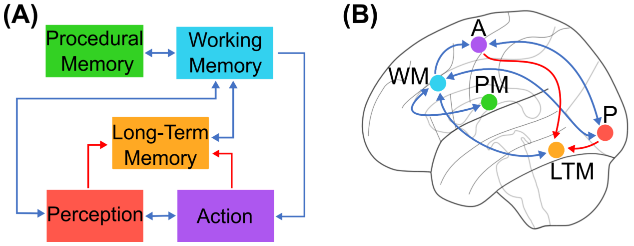

Figure 5: (A) A visual representation of the architecture inferred from the Granger causality model; (B) Proposed associations between components and anatomical brain regions. Arrows: dark blue, connections present in both CMC and GM; red, connections unique to Granger model.

The results support an architecture that is similar, but not exactly identical, to the CMC. If a strict 95% threshold is applied to the map of connections inferred from the GCM analysis, 22 of the possible 25 connections are shared between the human-derived network and the CMC (Fig. 5).

Comparing the Other Architectures

In the previous DCM based analysis of architecture structures, Stocco et al. (2021) were not able to incorporate data-driven inferences about connections. Instead, the plausibility of the CMC was evaluated by comparing its predictions against a set of representative alternative architectures across tasks (Fig. 6). These architectures, divided into two categories, or "families", represent the possible organizational structures of general purpose architectures. All consist of the same five regions or components present in the Common Model, but provide differing accounts of the connections between them.

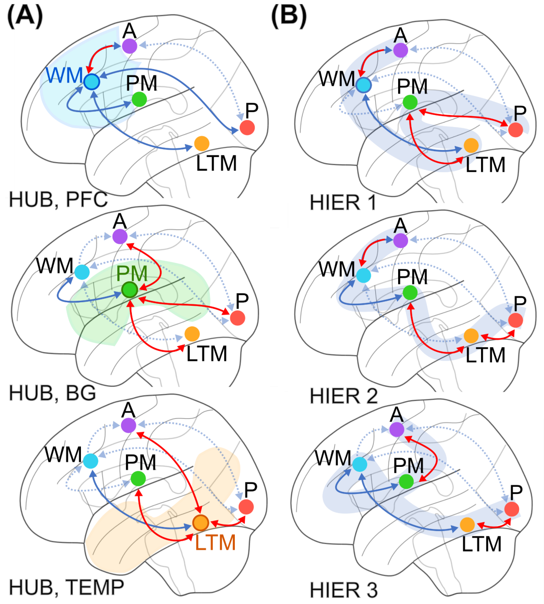

Figure 6: (A) Three variations of Hub-and-Spoke (HUB) models, and of (B) Hierarchical (HIER) models. Arrows: dark blue, connections present in both CMC and alternate models; red, connections unique to alternate models; and dotted, connections present in CMC and absent in alternate models.

The "Hub and Spoke" model family designates a single region as the "Hub" of model activity, with bidirectional connections between it and all other regions. These "Spoke" regions, however, do not connect to one another, and activity passing from one spoke region to another must also pass through the hub. Of the five CMC components, arguments can be made for each of the three memory modules serving in the capacity of a hub: working memory could drive activity from the prefrontal cortex (Hub PFC), long term memory could drive activity from temporal regions (Hub Temporal), or procedural memory could drive activity from the basal ganglia (Hub Procedural).

An alternate account of model structure is posed by members of the "Hierarchical" family. In this account, the architecture serves as a feed-forward system where activity originates in the perception region, travels through the successive memory regions, and culminates in the action region. With the limited number of regions and the fixed position of the perception and action regions, the potential models in this family vary only in the order of the three memory modules. An additional constraint, the assumption that long term memory (LTM) will proceed working memory (WM), leaves the position of procedural memory as the only degree of freedom. It is either the first of the three memory modules (Hierarchical 1), the middle module (Hierarchical 2), or the final module before action (Hierarchical 3).

To test whether the results of our Granger causality model converge with those previously reported with DCM (a test of convergent validity), we performed the same comparison of architectures done by Stocco et. al (2021). To do so, we first derived the theoretical network architectures of the six alternate architectures examined in that study for comparison against the network architecture derived using GCM. These networks are represented in the form of connectivity matrices in Figure 7.

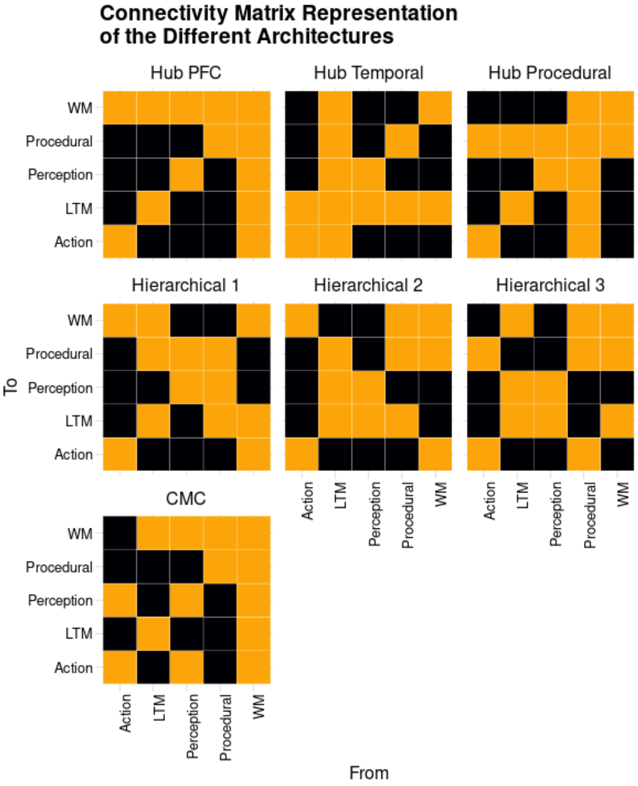

Figure 7: Connectivity matrix representations of the six alternative architectures (Figure 6) and the CMC.

For each alternate architecture, as well as the CMC, we examined the degree of similarity between the network of connections suggested by the GCM analysis and the connections theorized by the architecture. We considered three metrics. The first is the *correlation* between the predicted and observed directed connections in the vector of 25 possible edges in the networks. The second is the proportion of *overlap* between the two vectors, defined as the proportion of exactly matched connectivity predictions or, equivalently, the complement of the proportional Hamming distance between the two vectors of connections. The third and final metric is *likelihood*, defined as the *Z*-scores of predicted vs. expected number of successes in a binomial distribution of 25 connections. The results of each of the three metrics are compared in Figure 8. For all criteria, the CMC reflects the greatest similarity to the network architecture uncovered by the GCM analysis.

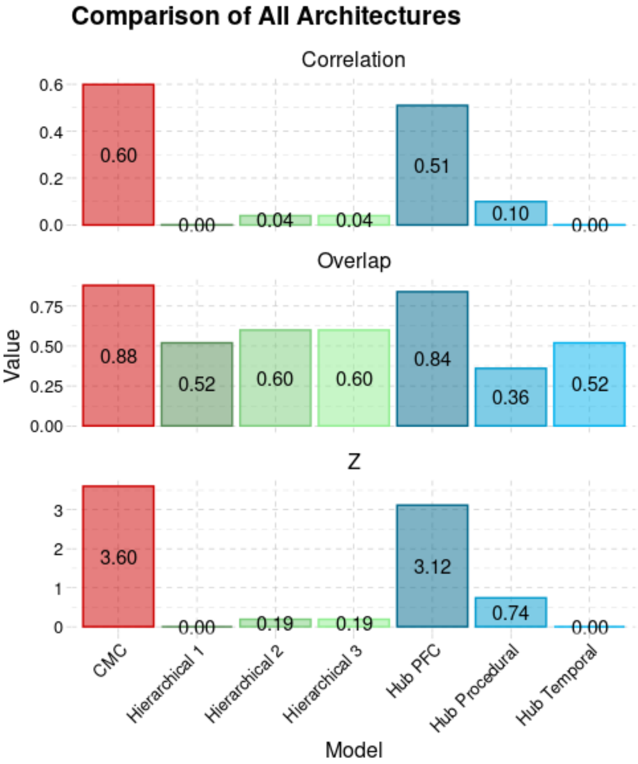

Figure 8: Comparison of the CMC (in red) and six alternate architectures in terms of three measures of similarity to the GCM network.

Discussion

In this paper, we have presented an analysis of Human Connectome Data using the same ROIs as in Stocco et al (2021), but employing Granger causality instead of DCM, to analyze fMRI data. A replication of the original comparison (Stocco et al., 2021) between the Common Model of Cognition and six alternative architectures largely confirmed the previous study's findings, namely, that the CMC performs better than the alternative architectures at explaining effective connectivity within and across all tasks.

Granger causality modeling, however, offers the unique opportunity of deriving a new architecture from data. The new architecture, represented in Figure 5, is a modified version of the CMC with the addition of projections from Perception and Action regions to the LTM component.

It is interesting to note that both connections are unidirectional, i.e., Perception and Action feed to LTM but do not receive projections back. Instead, bidirectional connections exist between LTM and WM. This particular connectivity structure seems to be best adapted to implement a form of instance-based learning, whereby successful episodic memories are formed by encoding previous stimuli, and actions and outcomes are stored to be later retrieved and guide behavior. In this case, direct connections from Perception and Action to LTM would support the encoding of stimuli and actions, respectively, while the connectivity from WM to LTM could support the encoding of evaluation of the outcome (performed by the WM component). Previous episodes could be later retrieved through the directed connection between LTM and WM. Moreover, the existence of additional functional links to LTM suggests that the large-scale brain organization seems to contain multiple hubs of different importance.

Limitations

However, these findings should be considered in light of a number of potential limitations. The first is that, while we ultimately aggregated the results into a single task-independent network, a significant amount of variability exists between the network architectures that can be inferred from the specific tasks. While DCM is intrinsically top-down and limited to examining the fit of specific network models, GCM does not suffer such limitation. Thus, the degree to which an task-independent architecture could be derived from individual tasks is debatable, and reflects the underlying assumption that, at a very high level, brain activity showcases a common invariant architecture. This hypothesis, of course, is not universally accepted and should be examined independently in future studies.

A second limitation is that the estimates of connectivity obtained through Granger causality might change when a larger set of component regions are included; thus, these results cannot be considered stable until the exact number of ROIs is considered canonical. It should be noted that, however, this limitation is also common to DCM and thus similarly affects previous work in this area (e.g., Steine-Hanson et al., 2019; Stocco et al., 2021).

Finally, it should be noted that, although these results do *suggest* that a better architecture (depicted in Figure 5) might outperform the original Common Model architecture and the others tested by Stocco et al. (2021), they do not necessarily imply so. This is because GCM is a different method than DCM, it is entirely possible that the architecture of Figure 5 would not perform as well when its effective connectivity is measured within the DCM framework. Therefore, possible future studies should re-investigate the superiority of this new architecture using the DCM-based comparison, as done in Stocco et al (2021).

Implications for the Common Model of Cognition

These limitations notwithstanding, these results do support credibility to the principles of the CMC. The architecture that was identified through GCM differs only minimally from the CMC, and the CMC remains the architecture that most closely matches our results across the set of potential architectures tested by Stocco et al. (2021). Our new findings, however, suggest important modifications to its structure. We consider these results as an exciting starting point for the future examination of large scale-connectivity of the brain.

Acknowledgement

This research was supported by grant FA9550-19-1-0299 from the Air Force Office of Scientific Research to AS.

References

Anderson, J. R. (2007). How can the human mind occur in the physical universe? *Oxford University Press*.

Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in cognitive sciences*, 12(4), 136–143.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., & WU-Minn HCP Consortium. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189.

Deshpande, G., LaConte, S., James, G. A., Peltier, S., & Hu, X. (2009). Multivariate Granger causality analysis of fMRI data. Human brain mapping, 30(4), 1361-1373.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. science, 338(6111), 1202-1205.

Fisher, R. A. (1932). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Friston, K., Moran, R., & Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current opinion in neurobiology*, *23*(2), 172-178.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27), 9673-9678.

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

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*, 424–438.

Hummel, J. E., & Holyoak, K. J. (2005). Relational reasoning in a neurally plausible cognitive architecture: An overview of the LISA project. Current Directions in Psychological Science, 14(3), 153-157.

Laird, J. E. (2019). *The Soar cognitive architecture*. MIT press.

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*, *38*, 13–26.

O'Reilly, R. C., Hazy, T. E., & Herd, S. A. (2016). The leabra cognitive architecture: How to play 20 principles with nature. The Oxford handbook of cognitive science, 91, 91-116.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*, 676–682.

Roebroeck, A., Formisano, E., & Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage, 25(1), 230-242.

Steine-Hanson, Z., Koh, N., & Stocco, A. (2018). Refining the Common Model of Cognition Through Large Neuroscience Data. Procedia Computer Science, 145, 813–820.

Stocco, A., Laird, J. Lebiere, C., & Rosenbloom, P. (2018). Empirical evidence from neuroimaging data for a Standard Model of the Mind. In C. Kalish, M. Rau, J. Zhou, and T. T. Rogers (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*. Madison, WI, pp. 1094-1099.

Stocco, A., Sibert, C., Steine-Hanson, Z., Koh, N., Laird, J. E., Lebiere, C. J., & Rosenbloom, P. (2021). Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains. *NeuroImage*, *235*, 118035.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79.