A Common Architecture for Human and Artificial Cognition Explains Brain Activity Across Domains

Authors: Andrea Stocco[1*], Zoe Steine-Hanson[2,†], Natalie Koh[1‡], John E. Laird[3], Christian J. Lebiere[4], and Paul S. Rosenbloom[5]

Affiliations:

[1]University of Washington, Seattle, WA 98195.

[2]Oregon State University, Corvallis, OR 97331.

[3]University of Michigan, Ann Arbor, MI 48109.

[4]Carnegie Mellon University, Pittsburgh, PA 15213.

[5]University of Southern California, Los Angeles, CA 90089.

[†]Now at University of Washington, Seattle, WA 98195.

[‡]Now at Northwestern University, Chicago, IL 60208.

*Correspondence to: stocco@uw.edu.

Abstract: The Common Model of Cognition (CMC) is a consensus architecture for human and human-like artificial cognition. We hypothesized that, because of its generality, the CMC could be a candidate model of the large-scale functional architecture of the human brain. To this end, we analyzed neuroimaging from *N*=200 participants across seven tasks that cover the broad range of cognitive domains. The CMC framework was translated into a model of neural connectivity between brain regions homologous to CMC components. After the model was implemented and fitted using Dynamic Causal Modeling, its performance was compared against four alternative large-scale brain architectures that had been previously proposed in the field of neuroscience. The results show that the CMC outperforms the other four architectures within and across all domains. These findings suggest that a common, functional computational blueprint for human-like intelligence also captures the neural architecture that underpins human cognition.

One Sentence Summary: A consensus computational architecture for modeling both human and human-like intelligence also best explains human neuroimaging activity across cognitive domains.

Main Text

The fundamental organizational principle of a complex system is often referred to as its "architecture," and represents an important conceptual tool to make sense of the relationship between a system's function and structure. For instance, the von Neumann architecture describes the organizing principle of modern digital computers; it can be used both to describe a computer at a functional level of abstraction (ignoring the specific wiring of its motherboard) and, conversely, to conduct diagnostics on an exceedingly complicated piece of hardware (properly identifying the components and pathways on a motherboard and the function of their wiring).

The stunning complexity of the human brain has inspired a search for a similar "brain architecture" that, akin to von Neumann's, could relate its components to its functional properties. Succeeding in this quest would lead to a more fundamental understanding of brain function and dysfunction and, possibly, to new principles that could further the development of artificial intelligence (*1*).

Most attempts in this direction have been "bottom-up," that is, driven by the application of dimensionality-reduction and machine-learning methods to large amounts of connectivity data, with the goal of identifying clusters of functionally connected areas (*2–4*). Although these models can be used to predict task-related activity, they rely on large-scale connectivity and are fundamentally agnostic as to the function of each node. The results of such approaches are also dependent on the type of data and the methods applied. For instance, one researcher might focus on purely functional measures, such as task-based fMRI and the co-occurrence of activity across brain regions and domains; a second researcher, instead, might focus on spontaneous, resting-state activity and slow-frequency time series correlations.

As recently pointed out (*5*), none of these methods is guaranteed to converge and provide a functional explanation *from* the data. However, the same methods can be successfully used to test a model *against* the data via a "top-down" approach (*5*). That is, given a candidate functional model of the brain, traditional connectivity methods can provide reliable answers as to its degree of fidelity to the empirical data and its performance compared to other models. A top-down approach, however, critically depends on having a likely and theoretically-motivated functional proposal for a brain architecture.

The Common Model of Cognition

A promising candidate architecture is the Common Model of Cognition (CMC) (*6*). The CMC is an architecture for general intelligence that reflects the historical convergence of multiple computational frameworks (developed over the course of five decades in the fields of cognitive psychology, artificial intelligence, and robotics) that agree on a common set of organizing principles. The CMC assumes that agents exhibiting human-like intelligence share five functional components: A feature-based, declarative *long-term memory*, a buffer-based *working memory*, a reinforcement-learning-based set of state-action patterns represented in *procedural memory*, and dedicated *perception* and *action* systems. Working memory acts as the hub through which all of the other components communicate, with one additional connection

between perception and action (Fig. 1A). The CMC also includes a set of constraints on the mechanisms and representations that characterize each component's functional properties.

The CMC's components and assumptions distill lessons learned over the last fifty years in the development of computational cognitive models and artificial agents with general human-like abilities. Surprisingly, these lessons seem to cut across specific application domains. For instance, the cognitive architecture Soar (*7*) is predominantly used in designing autonomous artificial agents and robots, while the cognitive architecture ACT-R (*8*) is predominantly used to simulate psychological experiments and predict human behavior (*9*); yet, they separately converged on the CMC assumptions (*6*). Similarly, the SPAUN large-scale brain model (*10*) and the Leabra neural architecture (*11*) are independently designed to simulate brain function through artificial neurons, and yet they also make use of similar components and functional assumptions. Even recent AIs that are made possible by advances in artificial neural networks employ, at some level, the same components. DeepMind's AlphaGo, for example, includes a Monte Carlo search tree component for look-ahead search and planning (working memory) and a policy network (i.e., procedural memory), in addition to dedicated systems for perception and action (*12*). Similarly, the Differentiable Neural Computer uses supervised methods to learn optimal policies (procedural memory) to access an external memory (symbolic long-term memory) (*13*).

Because the CMC reflects the general organization of systems explicitly designed to achieve human-like flexibility and intelligence, the CMC should also apply to the human brain. Therefore, it provides an ideal candidate for a top-down examination of possible brain architecture.

Assuming that the CMC is a valid candidate, how can its viability as a model of the human brain architecture be assessed? Operationally, a candidate model should successfully satisfy two criteria. The first is the *generality* criterion: The same cognitive architecture should account for brain activity data across a wide spectrum of domains and tasks. The second is the *comparative superiority* criterion: An ideal architecture should provide a superior fit to experimental brain data compared to competing architectures of similar complexity.

To test the CMC against these two criteria, we conducted a comprehensive analysis of task-related neuroimaging data from 200 young adult participants in the Human Connectome Project (HCP), the largest existing repository of high-quality human neuroimaging data. Although the HCP project contains both fMRI and MEG data, fMRI was chosen because it allows for unambiguous identification of subcortical sources of brain activity, which is crucial to the CMC and problematic for MEG analysis. The HCP includes functional neuroimages collected while participants performed seven psychological tasks. These tasks were taken or adapted from previously published influential neuroimaging studies and explicitly selected to cover the range of human cognition (*14*), therefore making it an ideal testbed for the *generality* criterion. Specifically, the tasks examine language processing and mathematical cognition (*15*), working memory, incentive processing and decision making (*16*), emotion processing (*17*), social cognition (*18*), and relational reasoning (*19*). The seven tasks were collected from six different paradigms (Table S1; language processing and mathematical cognition were tested in the same paradigm).

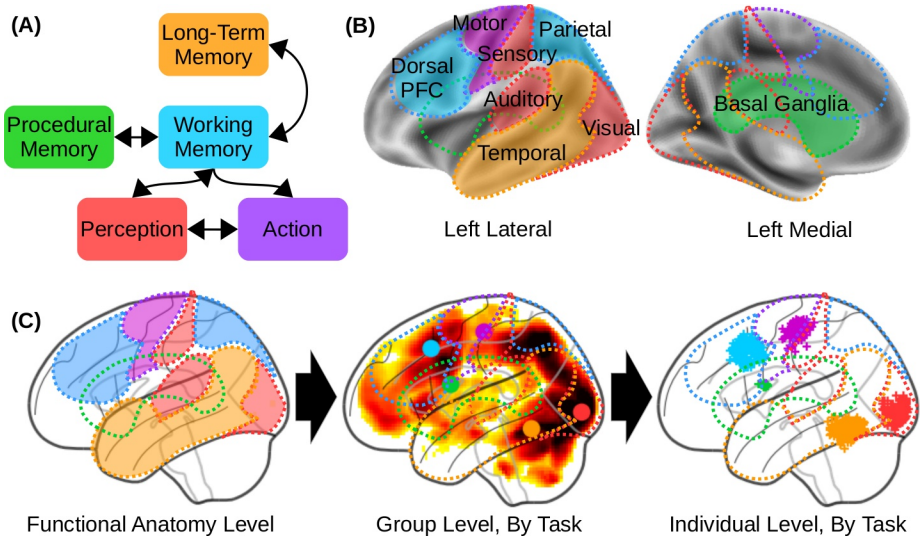

Fig. 1. The components of the Common Model of Cognition and their homologous brain regions. (A) Architecture of the Common Model of Cognition, as described in *(6)*. (B) Theoretical mapping between CMC components and homologous cortical and subcortical regions, as used in this study's pipeline to identify the equivalent Regions of Interest (ROIs). (C) Progressive approximation of the ROIs, from high-level functional mappings (left) to task-level group results (middle, with group-level centroid coordinated marked by a color circle) to the individual functional centroids of the regions in our sample (right; each individual centroid represented by a "+" marker). Group-level and individual-level data come from the Relational Reasoning task (see Fig. 2D).

To properly translate the CMC into a brain network architecture, its five components need to be identified with an equal number of spatially-localized but functionally homologous Regions of Interest (ROIs). To objectively define these ROIs for each task and participant, a processing pipeline was set up (See Supplementary Materials for details). The starting point of the pipeline was a-priori, theoretical identification of each CMC component with large-scale neuroanatomical distinctions. This initial identification was based on well established findings in the literature, and is also consistent with the function-to-structure mappings that had been proposed in other neurocognitive architectures, such as the mappings suggested for ACT-R's module-specific buffers (*8*, *20*, *21*) or for SPAUN's component neural circuits (*10*).

At this level, the working memory (WM) component was identified with the fronto-parietal network comprising the dorsolateral prefrontal cortex (PFC) and posterior parietal cortex; the long-term memory (LTM) component with regions in the middle, anterior, and superior temporal lobe; the procedural knowledge component with the basal ganglia; the action component with the premotor and primary motor cortex; and perception with sensory regions, including the primary and secondary sensory and auditory cortices, and the entire ventral visual pathway, the latter of which comprises the occipital and inferior temporal lobes (Fig. 1B).

Beginning with these macro-level associations, the pipeline progressively refined the exact ROI for each component through two consecutive approximations. Fig 1C provides a visual illustration of this procedure using the data from the relational reasoning task. First, data from each task was analyzed at a group level through canonical General Linear Model (GLM) analysis, and the most functionally active focus within each macro-region was identified (Fig. 1C, middle panel; Fig. S1). This step provided reasonable neural correlates of each component for each task. Note that allowing task-based variation in the localization of each ROI implicitly makes it harder for each model to achieve the generality criterion.

The coordinates of each focus were then used as the starting point to search in 3D space for the closest active peak within the individual statistical parameter maps obtained from GLM models of each participant (see Fig. 1C, right panel). Finally, the individualized ROI coordinates were then used as the center of a spherical ROI (Table S2). Fig. 2 illustrates the distribution of the individual ROI centroids for each task, overlaid over a corresponding group-level statistical map of task-related activity. For each ROI of every participant in every task, a representative time course of neural activity was extracted as the first principal component of the time series of all of the voxels within the sphere.

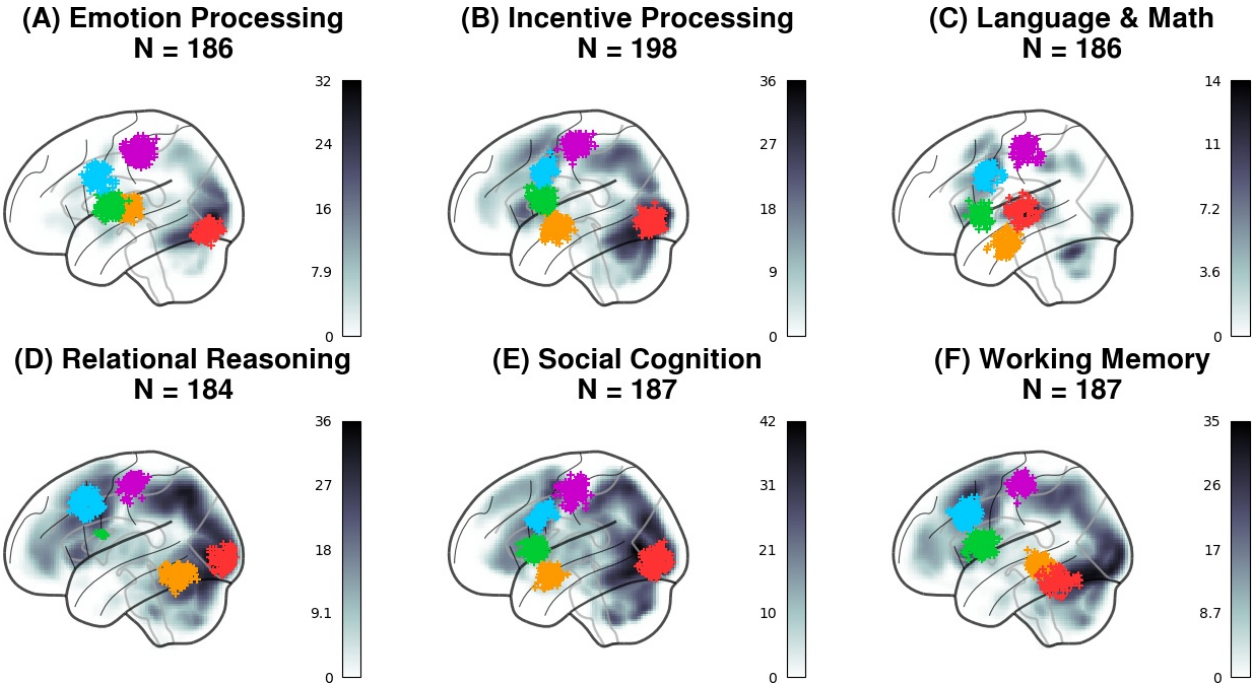

Fig. 2. Lateral view of the distribution of the ROI centroids across individual participants and tasks. Each "+" marker represents the centroid of an ROI for one participant. Colors represent the components, following the conventions of Fig 1A-C. The background represents the statistical group analysis used to identify the seed coordinate for each ROI (i.e., Step 2 in Fig. 1C).

All the ROIs were located in the left hemisphere; this simplifying approach was preferred to possible alternatives, such as including homologous regions in the right hemisphere (which would have required introducing additional assumptions about inter-hemispheric connectivity) or creating bi-lateral ROIs (which would have reduced the amount of variance captured in each ROI). Because all tasks show stronger activation in the left hemisphere than in the right, our results are still representative of brain activity in these domains. Finally, a network was created by connecting all the individually-defined ROIs according to the specifications of the CMC (Fig. 3A). It should be noted that synaptic pathways exist that connect every pair of components; thus, this network model is designed to capture the fundamental layout of a brain architecture in terms of functionally necessary connections, rather than anatomical details.

The link between the network of ROIs and their neural activity was provided through Dynamic Causal Modeling (DCM) (*22*), a neuronal-mass mathematical modeling technique that approximates the time-course of brain activity in a set of brain regions as a dynamic system that responds to a series of external drives. Specifically, the time course of the underlying neural activity *y* of a set of regions is controlled by the bilinear state change equation:

$$dy/dt = \mathbf{A}y + \sum_i x_i \mathbf{B}^i y + \mathbf{C}x$$

where *x* represents the event vectors (i.e., the equivalent of a design matrix in traditional GLM analysis), **A** defines intrinsic connectivity between ROIs, **C** defines the ROI-specific effects of task events, and **B** defines the modulatory effects that task conditions have on the connectivity between regions. For simplicity, the modulatory effects in **B** were set to zero, reducing the equation to the form $\mathbf{A}y + \mathbf{C}x$. A predicted time course of BOLD signal was then generated by applying a biologically-plausible model (the ballon model: (*23*, *24*)) of neurovascular coupling to the simulated neural activity *y*. The parameters of the full DCM model were estimated by applying the expectation-maximization procedure (*22*) to reduce the difference between the predicted and observed time course of the BOLD signal in each ROI.

Our preference for this technique was motivated by the existence of an integrated framework to design, fit, and evaluate models; by its ability to estimate the directional effects within a network (as opposed to traditional functional connectivity analysis); and by its underlying distinction between the modeling of network dynamics and the modeling of recorded imaging signals (as opposed to Granger causality), which makes it possible to apply the same neural models to different modalities (e.g., M/EEG data) in future work.

Alternative Architectures

To address our second criterion of *comparative superiority*, the CMC dynamic model was compared against other DCM models that implement alternative brain architectures. Because the space of possible models is large, we concentrated on four models that are representative of theoretical neural architectures previously suggested in the neuroscientific literature (Fig. 3). These four models can be divided into two families. In the “Hierarchical” family, brain connectivity implements hierarchical levels of processing that initiate with Perception and

culminate with Action. In this family, the brain can be abstracted as a feedforward neural network model with large-scale gradients of abstraction (*3*).

Within this hierarchical structure, each ROI represents a different level and projects both forward to the next level's ROI and backwards to the preceding level's ROI. In the *open-loop* variant, Perception gives rise to two distinct branches, one feeding the LTM component (representing the abstraction of perception into memory) and one ascending to procedural memory, WM and, eventually, Action components. The second, *closed-loop* hierarchical model also incorporates bidirectional connections between LTM and WM, thus reconnecting the two pathways into a loop (Fig. 2B).

In the "Hub and Spoke" family (Fig. 2C), a single ROI is singled out as the network's "Hub" and receives bidirectional connections from all the other ROI (the "Spokes"). With the exception of the Hub, no ROI is mutually connected to any other one. In the first variant, the role of the Hub is played by the WM component. Because, in our mapping, the WM component corresponds to the lateral PFC, this model captures the increasingly popular and supported view of PFC as a flexible hub for control (*2*, *25*). In the second variant, the role of the Hub is played by the Procedural Memory component, which reflects the centrality of procedural control in many production-system-based cognitive architectures (*7*, *8*, *26*). Because, in our mapping, Procedural Memory is identified with the basal ganglia, this architecture also reflects the centrality of these nuclei in action selection and in coordinating cortical activity (*10*, *27*, *28*).

Like the CMC, these architectures are representative of how the five components could be organized in a large-scale conceptual blueprint for the brain architecture; they simply make different choices as to which connections between components are more fundamental and better reflect the underlying neural organization.In addition to representing plausible alternative architectures, these alternative models differ minimally from the CMC and can be easily generated by replacing at most five connections from the CMC architecture (dashed lines and red lines, Fig. 2B-C). Thus, any resulting differences in fit are unlikely to arise because of differences in network complexity.

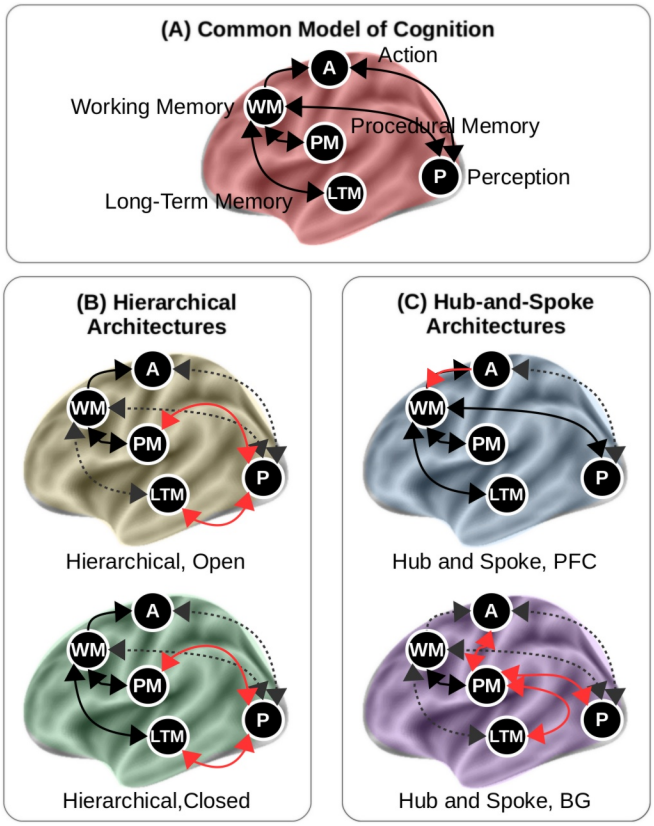

Fig 3. The five large-scale brain architectures tested in this study. (A) The CMC; (B) The "Open" and "Closed" versions of the hierarchical family; (C) The prefrontal (PFC) and basal ganglia (BG) versions of the hub-and-spoke architecture. In (B) and (C), pathways that are common to the CMC are shown in black; pathways that are present in the CMC but not included in the alternative models are shown in grey dashed arrows; and pathways that are present in the alternative models but not in the CMC are shown in red.

Once the five DCM models were separately fitted to the functional neuroimaging data, they were compared against each other using a Bayesian random-effects procedure (*29*). Like many other model comparison procedures, this approach provides a way to balance the complexity of a model (as the number of free parameters) versus its capacity to fit the data. Compared to popular log-likelihood-based measures (e.g., Akaike's information criterion (*30*)), this procedure is more robust in the face of outlier subjects, and thus better suited for studies that, like the present one, include a large number of participants and introduce considerable inter-individual variability (*29*, *31*). (See Supplementary Materials).

Specifically, the probability r_k that a model k would fit a random individual in a sample of participants is drawn from a Dirichlet distribution Dir(α_1, α_2,… α_K), and the distributions of probabilities of architectures 1, 2… k across n individuals are then drawn from multinomial distributions (see Fig. S4). The result of this modeling effort is a distribution of probabilities r_k for each model. These distributions can then be compared in terms of their relative *expected* and *exceedance* probabilities, that is, the mean probability of each model's r_k across the sample and the probability that r_k is larger than the competing models.

This procedure ultimately yields five probability distributions (one for each model), each corresponding to a specific Dirichlet value of the Dirichlet parameter α. These distributions are visualized for each task in Fig 4A-F. Furthermore they can be characterized with two metrics: the distribution's *expected* probability (i.e., the mean value of the distribution) and its *exceedance* probability (i.e., the probability that a value sampled from a distribution will exceed any value sampled from any other distribution) (*29*). The expected probabilities are represented as the colored vertical lines in Fig. 4A-F, while the exceedance probabilities are summarized as colored bars in Fig 4H; their exact values are listed in Table S3.

Both metrics provide evidence in favor of the CMC. As shown in Fig. 4A-F, the CMC provides a better fit to the data than any alternative architecture, and its exceedance probabilities range from 0.75 to 1.0 (Fig. 3H). Thus, the CMC uniquely satisfies both the generality and comparative superiority criteria. By contrast, all of the other architectures are consistently outperformed by the CMC in every domain (violating comparative superiority) and their relative rankings change from task to task (violating generality).

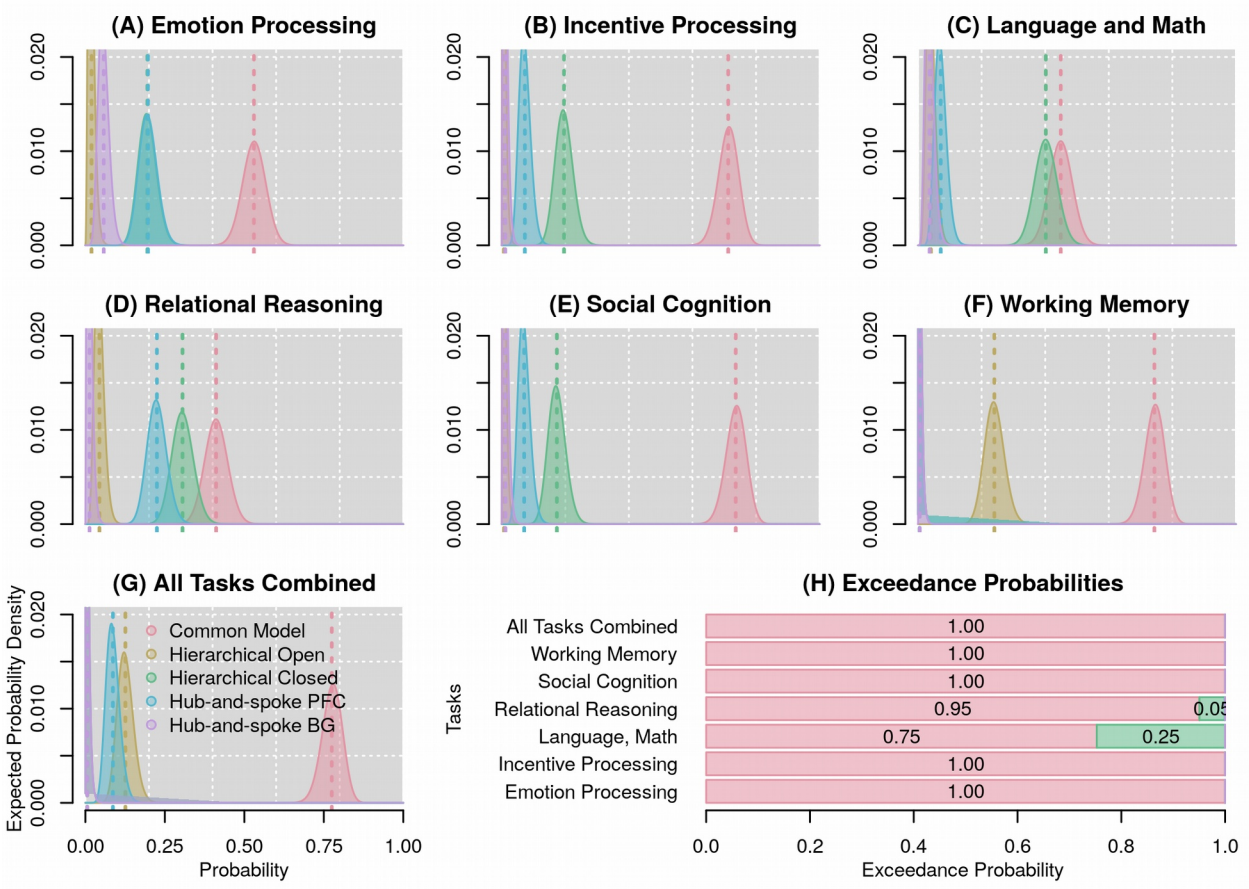

Fig 4. Results of the Bayesian model comparisons within and across tasks. In all plots, different colors represent different architectures.(A-G) Probability distributions that each of the five architectures is true, given the data within each task and across all tasks combined. Vertical dotted lines represent the mean of each distribution, i. e. the expected probability of each model. (H) Corresponding exceedance probabilities, within and across tasks, represented as stacked horizontal bars.

It is interesting to note that, across tasks, the relative ranking of the architectures does not reflect their relative similarity, measured as the number of different connections. For instance, the PFC variant of the "Hub-and-Spoke" family (Fig. 3C) is the one model most similar to the CMC, but is consistently outperformed by other models that, across all tasks, come closer to the CMC distributions (Fig. 4A-F). This fact further suggests that the CMC superiority is due to the holistic nature of its connectivity (i.e., how the components "go together") rather than to the sum of its specific connectivity elements.

The only task in which another model comes close to the CMC in terms of fit was the Language and Mathematical cognition paradigm, in which the Hierarchical Closed model had a 0.25 exceedance probability against the CMC (Fig. 3H). This paradigm was unique because it included two entirely different tasks of comparable difficulty, instead of a single task with two conditions of different difficulty, as was the case in all other tasks. This peculiarity raises a potential concern that the CMC's superiority could be an artifact of modeling each task in isolation, and that in conditions where multiple tasks were modeled simultaneously, a different model could potentially provide a superior fit. To examine this possibility, a second analysis was carried out, which included only the 168 participants for whom data for all seven tasks was available. In this analysis, the data from each of the six paradigms performed by the same individual is modeled as a different run from a "meta-task" performed by that individual. When such an analysis was performed, the CMC maintained its superiority, all other models having a combined exceedance probability $< 1.0 \times 10^{-10}$ (Fig. 4G-H).(*32*)

As noted earlier, although the competing architectures were chosen to represent current alternatives views, we cannot entirely rule out the existence of alternative architectures that explain the data better than the CMC. It is possible, however, to decide whether all of the connections in the CMC are necessary, or whether a simpler model could potentially fit the data equally well. To this end, a Bayesian parameter averaging procedure (*33*) was conducted to generate the posterior distributions of parameter values across participants for each task. Fig. 4 depicts the mean value (as the square color) and the associated posterior probability (as the overlaid number) for each CMC connection in each task. As the figure shows, the parameter values change significantly from task to task, implying that the CMC architecture is adaptively leveraged to meet the specific requirements of each paradigm. Nonetheless, virtually all parameters have a posterior probability $p \approx 1.0$ of being different than zero (with just two parameters having smaller probabilities, of $p = 0.75$ and $p = 0.98$), suggesting that all the components and their functional connections remain necessary across all domains.

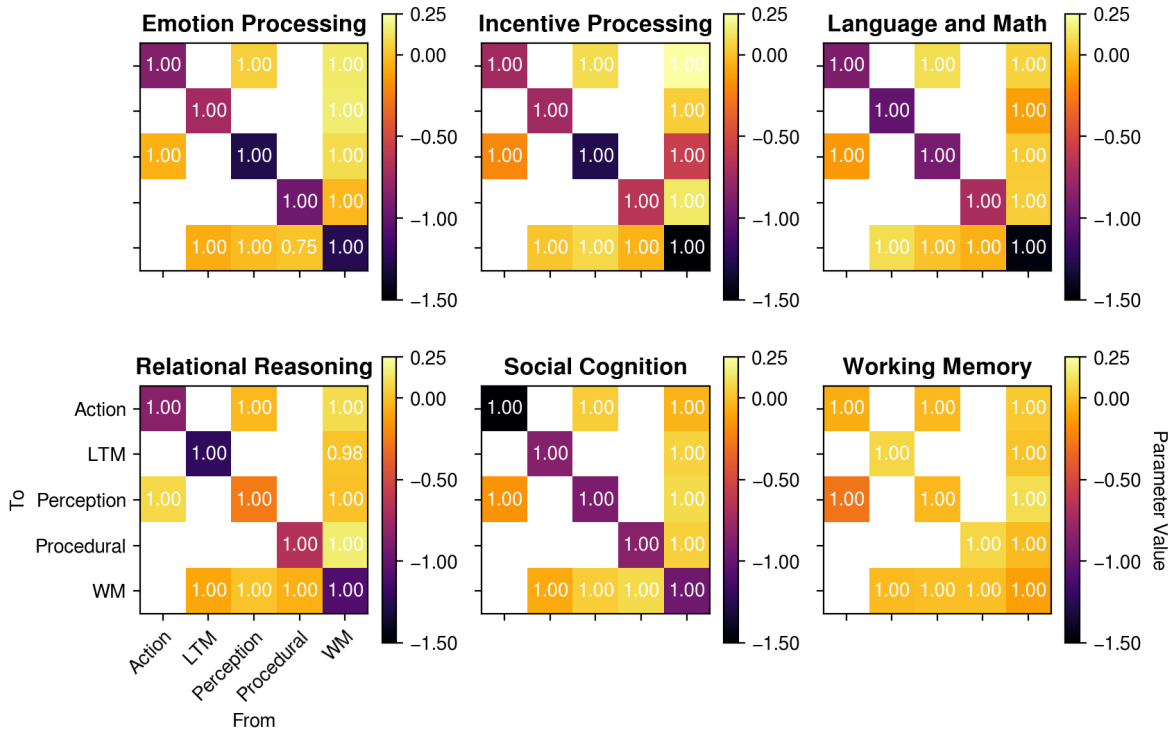

Fig. 5. Estimated DCM parameters for the CMC model cross tasks. In each plot, the color indicates the parameter value, and the white text indicates the posterior probability that the parameter value is significantly different than zero. White squares represent non-existing connections (see Fig. 1A).

Discussion

In summary, these results provide overwhelming and converging evidence for the CMC as a high-level blueprint of the human brain's architecture, potentially providing the missing unifying framework to relate brain structure and function for research and clinical purposes. Although surprisingly robust, these results should be considered in light of three potential limitations. First, our conclusions are based on an analysis of task-related brain activity. Despite being established in the literature, the HCP tasks remain artificial, laboratory tasks, and their ecological validity is unknown. In contrast, many prominent studies have focused on task-free, resting-state paradigms. Thus, although the use of the task-related activity provides the most natural test for the generality criterion, the extent to which the CMC applies to resting-state fMRI remains to be explored. Second, as noted above, our selection of alternative models was representative but not exhaustive. Although most distinct architectures that can be generated using just the CMC components are likely to be unreasonable from a functional standpoint, some of them could outperform the CMC. Finally, it can be argued that our approach does not take full advantage of the possibilities of DCM, which makes it possible to accommodate non-linear, modulatory effects in the dynamic model. For example, the strategic role of procedural knowledge in the CMC (Assumption B3 of the original paper) is compatible with a "modulatory" view of the basal ganglia, which has also been proposed (*27*) and observed (*34*).

These limitations notwithstanding, the fact that the CMC, which draws inspiration from high-level models of human cognition and *artificial* intelligent systems, also accounts for the neural activity of the human brain, which is a low-level *biological* intelligent system, is worthy of consideration. A mundane explanation could simply be that humans reproduce the fundamental architecture of their intelligence when designing human-like models and intelligent systems. A more radical explanation is that the architectural space for general (or, at least, human-like) intelligence is inherently constrained and possibly independent of its physical realization, whether organic or artificial. Both hypotheses are worth exploring in future research.

References and Notes

1. D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick, Neuroscience-Inspired Artificial Intelligence. *Neuron*. **95**, 245–258 (2017).
2. M. W. Cole *et al.*, Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* **16**, 1348–1355 (2013).
3. J. M. Huntenburg, P.-L. Bazin, D. S. Margulies, Large-Scale Gradients in Human Cortical Organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).
4. K. J. Gorgolewski *et al.*, A correspondence between individual differences in the brain's intrinsic functional architecture and the content and form of self-generated thoughts. *PLoS One*. **9**, e97176 (2014).
5. E. Jonas, K. P. Kording, Could a Neuroscientist Understand a Microprocessor? *PLoS Comput. Biol.* **13**, e1005268 (2017).

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

7. J. E. Laird, *The Soar Cognitive Architecture* (MIT Press, 2012).

8. J. R. Anderson, *How Can the Mind Occur in the Physical Universe?* (Oxford University Press, 2007).

9. I. Kotseruba, J. K. Tsotsos, 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, **40**, 1–78 (2018).

10. C. Eliasmith *et al.*, A large-scale model of the functioning brain. *Science*. **338**, 1202–1205 (2012).

11. R. C. O'Reilly, T. E. Hazy, S. A. Herd, The Leabra Cognitive Architecture: How to Play 20 Principles with Nature. *The Oxford handbook of cognitive science*, 91 (2016).

12. D. Silver *et al.*, Mastering the game of Go with deep neural networks and tree search. *Nature*. **529**, 484–489 (2016).

13. A. Graves *et al.*, Hybrid computing using a neural network with dynamic external memory. *Nature*. **538**, 471–476 (2016).

14. D. C. Van Essen *et al.*, The WU-Minn Human Connectome Project: An overview. *Neuroimage*. **80**, 62–79 (2013).

15. J. R. Binder *et al.*, Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. *Neuroimage*. **54**, 1465–1475 (2011).

16. M. R. Delgado, L. E. Nystrom, C. Fissell, D. C. Noll, J. A. Fiez, Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.* **84**, 3072–3077 (2000).

17. A. R. Hariri, A. Tessitore, V. S. Mattay, F. Fera, D. R. Weinberger, The amygdala response to emotional stimuli: a comparison of faces and scenes. *Neuroimage*. **17**, 317–323 (2002).

18. T. Wheatley, S. C. Milleville, A. Martin, Understanding animate agents: distinct roles for the social network and mirror system. *Psychol. Sci.* **18**, 469–474 (2007).

19. R. Smith, K. Keramatian, K. Christoff, Localizing the rostrolateral prefrontal cortex at the individual level. *Neuroimage*. **36**, 1387–1396 (2007).

20. J. P. Borst, M. Nijboer, N. A. Taatgen, H. van Rijn, J. R. Anderson, Using data-driven model-brain mappings to constrain formal models of cognition. *PLoS One*. **10**, e0119673 (2015).

21. J. P. Borst, J. R. Anderson, Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1628–1633 (2013).

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

23. K. J. Friston, A. Mechelli, R. Turner, C. J. Price, Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage*. **12**, 466–477 (2000).

24. R. B. Buxton, E. C. Wong, L. R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* **39**, 855–864 (1998).

25. M. W. Cole, T. Yarkoni, G. Repovs, A. Anticevic, T. S. Braver, Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *J. Neurosci.* **32**, 8988–8999 (2012).

26. D. E. Kieras, D. E. Meyer, An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction*. **12**, 391–438 (1997).

27. A. Stocco, C. Lebiere, J. R. Anderson, Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychol. Rev.* **117**, 540–574 (2010).

28. T. E. Hazy, M. J. Frank, R. C. O'reilly, Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 1601–1613 (2007).

29. K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, K. J. Friston, Bayesian model selection for group studies. *Neuroimage*. **46**, 1004–1017 (2009).

30. H. Akaike, in *Selected Papers of Hirotugu Akaike* (Springer, 1974), pp. 215–222.

31. K. E. Stephan *et al.*, Ten simple rules for dynamic causal modeling. *Neuroimage*. **49**, 3099–3109 (2010).

32. C. H. Kasess *et al.*, Multi-subject analyses with dynamic causal modeling. *Neuroimage*. **49**, 3065–3074 (2010).

33. C. S. Prat, A. Stocco, E. Neuhaus, N. M. Kleinhans, Basal ganglia impairments in autism spectrum disorder are related to abnormal signal gating to prefrontal cortex. *Neuropsychologia*. **91**, 268–281 (2016).

34. D. M. Barch *et al.*, Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage*. **80**, 169–189 (2013).

35. R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, B. T. T. Yeo, The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–2345 (2011).

36. F. Castelli, F. Happé, U. Frith, C. Frith, Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*. **12**, 314–325 (2000).

37. W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (Academic Press, 2011).

38. J. Ashburner *et al.*, SPM12 manual. Available at: http://www.fil.ion.ucl.ac.uk/spm/doc/spm12 manual.pdf

39. G. Schwarz, Estimating the Dimension of a Model. *Ann. Stat.* **6**, 461–464 (1978).

Acknowledgments

The authors would like to thank Jim Treyens and John R. Anderson for their comments on early drafts of this manuscript. **Funding:** This effort has been sponsored by award FA9550-19-1-0299 from the Air Force Office of Scientific Research (AFOSR) to AS, by award FA9550-19-0180 from AFOSR to JL, by award W911NF-14-D-0005 from the U. S. Army to PR, and by an award from the Defense Advanced Research Projects Agency (DARPA) to CL. Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred. **Author contributions:** Conceptualization: AS, CL, JL, and PR; Data curation; AS, NK, and ZSH; Formal analysis: AS, NK, and ZSH; Funding acquisition AS, CL, JL, and PR; Investigation: AS, CL, JL, and PR; Methodology: AS; Project administration: AS; Resources AS, CL, JL, and PR; Software: AS, NK and ZSH; Visualization: AS; Writing: AS, CL, JL, NK, PR and ZSH. **Competing interests:** The authors declare no competing interests; and **Data and materials availability:** All the raw imaging data is available through the Human Connectome Project (http://www.humanconnectome.org). All the pipeline scripts designed to preprocess the data, generate and fit the DCM models, extract and interpret the results, and generate the figures reported in this paper are available as an open-source code repository on https://github.com/UWCCDL/CMC-DCM.

Supplementary Materials:

Materials and Methods

Figures S1-S4

Tables S1-S3

References (None).

Science
AAAS

Supplementary Materials for

A Common Architecture for Human and Human-Like Artificial Cognition Explains Brain Activity Across Domains

Andrea Stocco, Zoe Steine-Hanson, Natalie Koh, John E. Laird,
Christian J. Lebiere, Paul S. Rosenbloom

Correspondence to: stocco@uw.edu

This PDF file includes:

Materials and Methods

The study presented herein consists of an extensive analysis of a large sample (*N*=200) of neuroimaging data from the Human Connectome Project, the largest existing repository of young adult neuroimaging data. The analysis was restricted to the task fMRI subset, thus excluding both the resting state fMRI data, the diffusion imaging data, and all of the M/EEG data. The task fMRI data consisted of two sessions of each of seven paradigms, designed to span different domains.

Tasks 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 Motor Mapping task was not included because it would have required the creation of multiple ROIs in the motor cortex, one for each effector (arm, leg, voice), thus making this model intrinsically different from the others. A full description of these tasks and the rationale for their selection can be found in the original HCP papers (*14*, *34*). This section provides a brief description of the paradigms, while Table S1 provides an overview.

Task (Representative Reference)	**Relevant Conditions (for GLM analysis)**	**Included in DCM analysis?**
Motor Mapping (35)	*Hand, arm, foot, leg, voice responses*	*No*
Emotion Processing (*17*)	Neutral shapes vs. **Fearful and angry faces**.	Yes
Incentive Processing (*16*)	"Winning" vs. **"Losing"** blocks of choices	Yes
Language and Mathematical Processing (*15*)	Listening vs. **Answering questions** (in both Language and Math blocks)	Yes
Relational Reasoning (*19*)	Control Arrays vs. **Relational arrays**	Yes
Social Cognition (*18*)	Randomly moving shapes vs. **Socially interacting shapes**	Yes
Working Memory	0-Back vs. **2-Back blocks** of faces, places, tools, and body parts.	Yes

Table S1: Overview of the seven task-fMRI paradigms used in the HCP dataset. Italics indicate tasks and conditions that were not included in our analysis; bold typeface marks experimental conditions that were selected as "Critical" (as opposed to "Baseline") in the design of the experimental matrices (see below, "DCM-specific GLM analysis" section)

Emotion Processing Task. Participants are presented with 12 blocks of six consecutive trials. During each trial, they are asked to decide either which of two visual stimuli presented on the bottom of the screen match the stimulus at the top of the screen. In six of the blocks, all of the visual stimuli are emotional faces, with either angry or fearful expressions. In the remaining six blocks, all of the stimuli are neutral shapes. Each stimulus is presented for 2 s, with a 1 s inter-trial interval (ITI). Each block is preceded by a 3 s task cue ("shape" or "face"), so that each block is 21 s including the cue.

Incentive Processing Task. The task consists of four blocks of eight consecutive decision-making trials. During each trial, participants are asked to guess whether the number underneath a "mystery card" (visually represented by the question mark symbol "?") is larger or smaller than 5 by pressing one of two buttons on the response box within the allotted time. After each choice, the number is revealed; participants receive a monetary reward (+$1.00) for correctly guessed trials; a monetary loss (-$0.50) for incorrectly guessed trials; and receive no money if the number is exactly 5. Unbeknownst to participants, blocks are pre-designed to lead to either high rewards (6 reward trials, 2 neutral trials) or high losses (6 loss trials, 2 neutral trials), independent of their actual choices. Two blocks are designated as high-reward, and two as high-loss blocks. Each stimulus has a duration of up to 1.5 s, followed by a 1 s feedback, with a 1 s ITI, so that each block lasts 27 s.

Language and Mathematical Processing Task. The task consists of 4 "story" blocks interleaved with 4 "math" blocks. The two types of blocks are matched for duration, and adhere to the same internal structure in which a verbal stimulus is first presented auditorily, and a two-alternative question is subsequently presented. Participants need to respond to the question by pressing one of two buttons with the right hand. In the story blocks, the stimuli are brief, adapted Aesop stories (between 5 and 9 sentences), and the question concerns the story's topic (e.g., "Was the story about *revenge* or *reciprocity*?"). In the math blocks, stimuli are addition or subtraction problems (e.g., "Fourteen plus twelve") and the question provides two possible alternative answers (e.g., "*Twenty-nine* or *twenty-six*?"). The math task is adaptive to maintain a similar level of difficulty across the participants.

Relational Processing Task. The task consists of six "Relational" blocks alternated with six "Control" blocks. In relational blocks, stimuli consist of two pairs of figures, one displayed horizontally at the top of the screen and one pair displayed at the bottom. Figures consist of one of six possible shapes filled with one of six possible textures, for a total of 36 possible figures. Both pairs of figures differ along one dimension, either shape or texture; participants are asked to indicate through a button press if the top figures differ on the same dimension as the bottom figures (e.g., they both differ in shape). In the control blocks, the stimuli consist of one pair of figures displayed horizontally at the top of the screen, a third figure displayed centrally at the bottom of the screen, and a word displayed at the center of the screen. The central word specifies a stimulus dimension (either "shape" or "texture") and participants are asked to indicate whether the bottom figure matches either of the two top figures along the dimension specified by the word. Both relational and control blocks have a total duration of 16 s, but they vary in the number of stimuli. Specifically, relational blocks contain four stimuli, presented for 3.5 s with a 500 ms ITI, while control blocks contain five stimuli presented for 2.8 s with a 400 ms ITI.

Social Cognition Task. The tasks consists of 10 videoclips of moving shapes (circles, squares, and triangles). The clips were either obtained or modified from previously published studies (*18*, *36*). In five of the clips, the shapes are moving randomly, while in the other five the shapes' movement reflects a form of social interaction. After viewing each clip, participants

press one of three buttons to indicate whether they believed the shapes were interacting, not interacting, or whether they were unsure. All clips have a fixed duration of 20 s with an ITI of 15 s.

Working Memory. The task consists of eight 2-back blocks and eight 0-back blocks, with each block containing 10 trials. Each trial presents the picture of a single object, centered on the screen, and participants have to press one of two button to indicate whether the object is a target or not. In the 2-back blocks, a target is defined as the same object that had been seen two trials before, so that participants have to maintain and update a "moving window" of the past two objects to perform the task correctly. In the 0-back blocks, a target is defined as a specific object, presented at the very beginning of the block, so that participants have to only maintain a single object in working memory throughout the block. The stimuli belong to one of four possible categories: faces, places, tools, and body parts. The category of the objects being used as stimuli changes from block to block, but is consistent within one block, so that there is an even number of face, place, tool, and body part blocks for each condition. Each block begins with a 2.5 s cue that informs the participant about the upcoming block type (2-back or 0-back). Each stimulus is presented for 2 s with a 500 ms ITI, for a total duration of 27.5 s per block.

Data Processing and Analysis

Imaging Acquisition Parameters As reported in (*34*), functional neuroimages were acquired with a 32-channel head coil on a 3T Siemens Skyra with TR = 720 ms, TE = 33.1 ms, FA = 52°, FOV = 208 × 180 mm. Each image consisted of 72 2.0mm oblique slices with 0-mm gap in-between. Each slice had an in-plane resolution of 2.0 x 2.0 mm. Images were acquired with a multi-band acceleration factor of 8x.

Image Preprocessing Images were acquired in the "minimally preprocessed" format (*14*), which includes unwarping to correct for magnetic field distortion, motion realignment, and normalization to the MNI template. The images were then smoothed with an isotropic 8.0 mm FWHM Gaussian kernel.

Canonical GLM Analysis Canonical GLM analysis was conducted on the smoothed minimally preprocessed data using a mass-univariate approach, as implemented in the SPM12 software package (*37*). First-level (i.e., individual-level) models were created for each participant. The model regressors were obtained by convolving a design matrix with a hemodynamic response function; the design matrix replicated the analysis of (*34*), and included regressors for the specific conditions of interests described in Table S1. Second-level (i.e., group-level) models were created using the brain-wise parametric images generated for each participant as input. The results of these second-level models replicated previous findings *(34)*, and are illustrated in Figure S1.

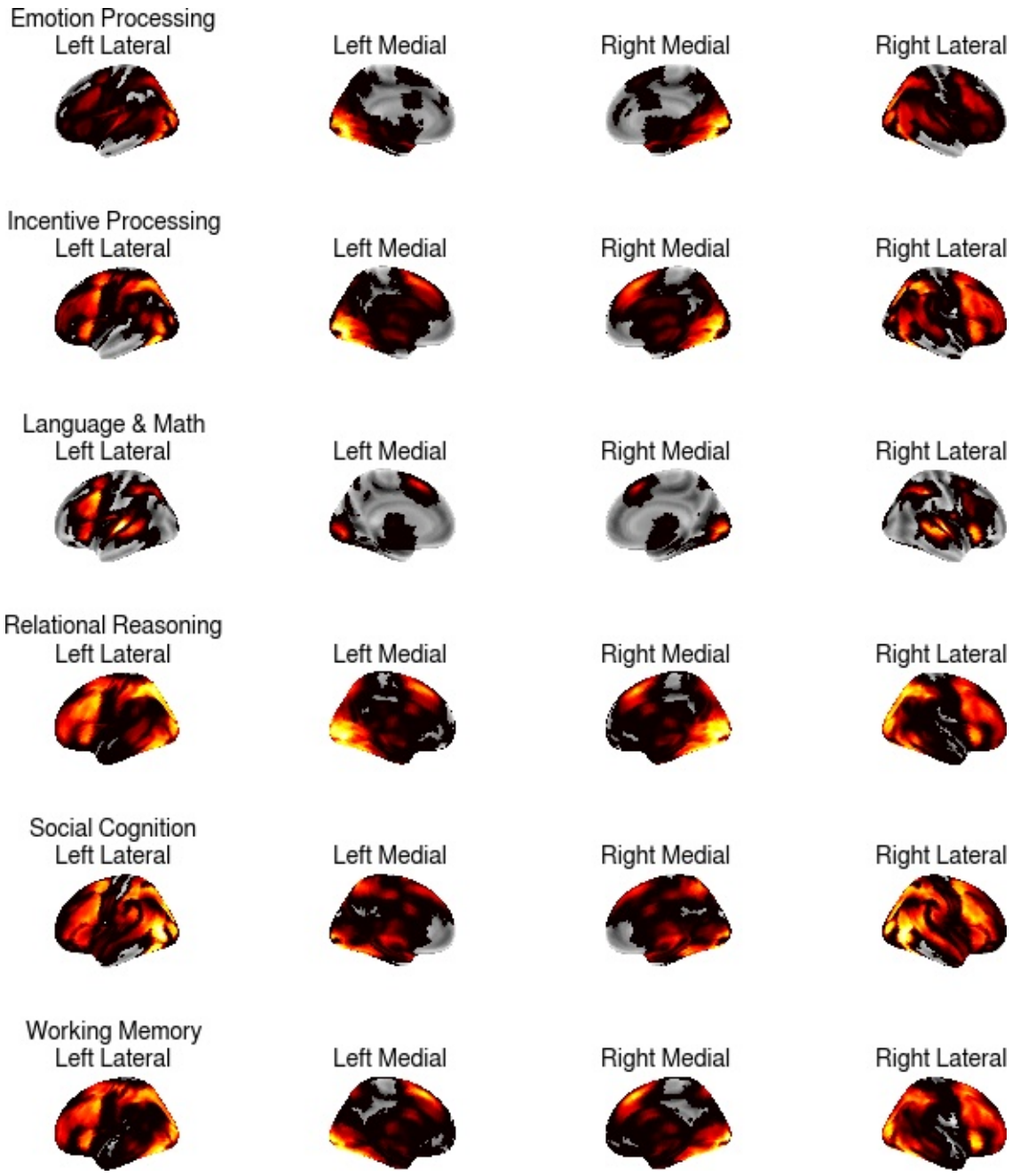

Fig. S1. Results of the group-level GLM analyses for each task.

DCM-specific GLM Analysis In parallel with the canonical GLM analysis, a second GLM analysis was carried out as a an aid to the DCM analysis. The purpose of this analysis was two-fold. First, it was needed to define the even matrix that is used in the DCM equation to measure the parameter matrix **C**. Second, it provided a way to define the omnibus *F*-test that is used in the ROI definition (see below). Because these models are not used to perform data analysis, the experimental events and conditions are allowed to be collinear. Because of the nature of cognitive neuroscience paradigms, all of our tasks include at least two different conditions under which stimuli must be processed in different ways. In all cases, the difference between conditions can be framed in terms of a more demanding, "critical" condition and an easier, "control" condition, with the more demanding events associated with greater mental elaboration of the stimuli. The critical condition of each task is highlighted in Table S1. As is common in DCM analysis, task conditions were modeled in a layered, rather than orthogonal fashion. The difference is illustrated in Figure S2: While in traditional GLM analysis the two conditions are modeled as non-overlapping events in the design matrix, in the DCM-specific definition of the matrix all trials belong to the same "baseline" condition, which represents the basic processing of the stimulus across all trials. Stimuli from the critical condition form a subset of all stimuli presented in the baseline condition. The critical condition is therefore appended to the baseline condition in the design matrix to model the additional processes that are specifically related to it. In DCM, each condition can affect one or more ROIs independently. In our analysis, the association between conditions and ROIs was kept constant across all tasks. Specifically, the baseline conditions selectively affected the perceptual ROI, while the critical condition selectively affected the WM ROI (Fig. S2C). This choice reflects the greater mental effort that is common to all critical conditions, and is confirmed by the greater PFC activity found in all of the GLM analyses of the critical conditions (*34*).

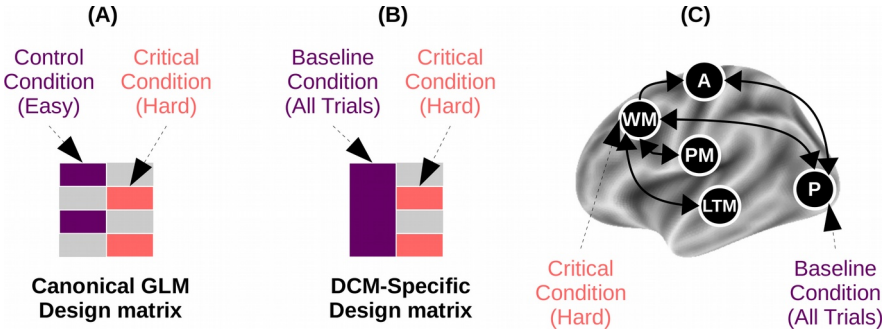

Fig. S2. Difference between design matrix used for canonical GLM (A) and for DCM analysis (B). In the DCMs, the Baseline condition drives neural activity in perceptual areas, while in the Critical condition drives neural activity in the Working Memory component (C).

Regions of Interests (ROI) Definition

The Regions of Interest (ROIs) were selected using a two-step procedure, designed to maximize the sensitivity of our analysis by separately accounting for two sources of variability in the spatial distribution of the ROIs.

The first step is designed to account for group-level variability due to the different tasks and stimuli used in the four datasets. This is necessary because, for example, the different complexity of the visual stimuli determine which portion of the visual cortex is most likely to be engaged, and different stimulus and task characteristics would engage different portions of the PFC. These differences were accounted for by conducting a separate group-level GLM analysis of each dataset, and identifying the coordinates of three points that have the highest statistical response within the anatomical boundaries of the primary visual areas (limited to the occipital lobe), the dorso-lateral PFC, and the basal ganglia (limited to the striatum). Table S2 provides a detailed list of the coordinates of these regions across all tasks.

The second step was designed to account for individual-level variability in functional neuroanatomy. To do so, the task-specific group level coordinates from each task were then used as seed points to locate the closest local maximum in each subject's corresponding statistical parameter map. For maximal sensitivity, the map was derived from an omnibus *F*-test that included all the experimental conditions. In practice, this *F*-test was designed to capture any voxel that responded to any experimental condition. The same *F*-contrast was also use to adjust (i.e., mean-correct) each ROI's timeseries (*37*, *38*)

The individual coordinates, thus defined, were then visually inspected and when the coordinates were outside the predefined anatomical boundaries, manually re-adjusted. Across over 1,200 coordinates examined, only 2 required manual adjustment (~ 0.2%).

Finally, the individual coordinates were used as the center of a sphere. All voxels within the sphere whose response was significant at a minimal threshold of $p < 0.5$ were included as part of the ROI. Fig. S3 illustrates the mean number of voxels and standard deviation for each VOI in each task.

Task	Action	LTM	Perception	Procedural	WM
Emotion Processing	-38, -26, 50	-56, -18, 6	-34, -80, -12	-24, -4, 8	-40, 6, 30
Incentive Processing	-40, -22, 56	-38, -6, -8	-40, -80, -4	-18, 6, 14	-46, 2, 34
Language & Math	-38, -26, 50	-50, -10, -22	-52, -22, 4	-22, 10, 2	-46, 2, 30
Relational Reasoning	-38, -22, 56	-48, -56, -14	-12, -92, -2	-16, 2, 18	-44, 16, 44
Social Cognition	-40, -18, 52	-54, 0, -14	-38, -84, -4	-14, 12, 6	-52, 6, 32
Working Memory	-38, -22, 56	-60, -36, -8	-34, -50, -18	-14, 10, 10	-48, 22, 36

Table S2. Group-level coordinates of the centers of the five ROIs (corresponding to the five components of the CMC) across the seven tasks (Language and Math share the same paradigm and the same coordinates). Coordinates are given in *x*, *y*, *z* dimensions in MNI space. Each coordinate was used as the starting point to identify the closest peak for each individual participant's functional maps (see Fig. 2).

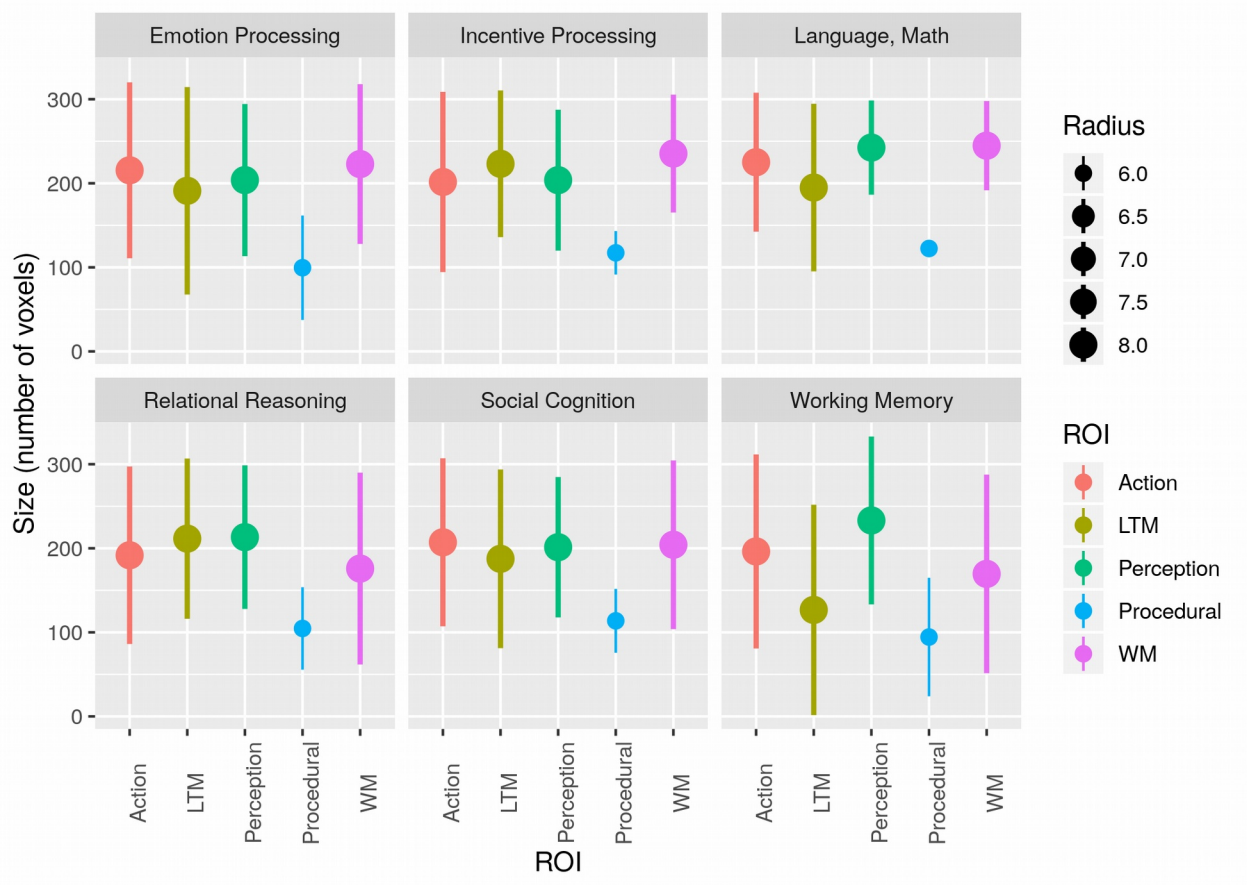

Fig. S3: Mean size (i.e., number of significant voxels included in each ROI spherical volume) of each ROI across participants for each task. Different colors represent different ROIs; dots represent means; dot size represents the ROI radius, and whiskers represent standard deviations.

Bayesian Model Selection

The goal of distinguishing which model provides a better fit to the data is, in turn, a problem of model comparison. In this study, a hierarchical Bayesian approach was employed, using a random effects analysis of the model fits as outlined by (*29*). Compared to fixed-effects methods (such as log-likelihood and log-likelihood-derived methods, Bayesian Information Criterion (*39*) and Akaike Information Criterion (*30*)), a random effects analysis is less sensitive to errors introduced by outlier subjects.

In this analysis, different individuals might potentially be fit by different models (thus allowing a subject-level term), and models are compared on the basis of their relative probabilities of being the best-fitting architecture across the sample of individuals. Statistically, the prevalence of each model k (i.e., the probability r_k that k would fit a random individual) in a sample of participants is drawn from a Dirichlet distribution $\mathrm{Dir}(\alpha_1, \alpha_2, \ldots \alpha_k)$, and the distributions of probabilities of architectures 1, 2… k across n individuals are then drawn from multinomial distributions. The result of this modeling effort is a distribution of probabilities r_k for each model k. These distributions can then be compared in terms of their relative *expected* and *exceedance* probability, that is, the mean probability of each model's r_k across the sample and the probability that r_k is larger than the competing models. Expected probability is calculated as the mean of each distribution; the properties of the Dirichlet distribution guarantee that the sum of the means of all distributions is 1. The Exceedance probability can be calculated by sampling from a multinomial distribution generated from random samples of the original distributions, thus again guaranteeing that all probabilities sum up to 1. Fig. S4, inspired by (*29*), provides a graphical illustration of the procedure. For simplicity, and following the convention of *(29)*, Fig. S4 depicts illustrates exceedance probability in the trivial case in which only two models are present, in which case it reduces to the area of the distribution to the right of $r_k > 0.5$; when more than two models are present, no simply visual interpretation is possible. Table S3 provides a detailed list of model comparison metrics, including the ones derived from the hierarchical Bayesian procedure used in this study (Dirichlet's α, expected, and exceedance probabilities) as well as the group-level log-likelihood of each model.

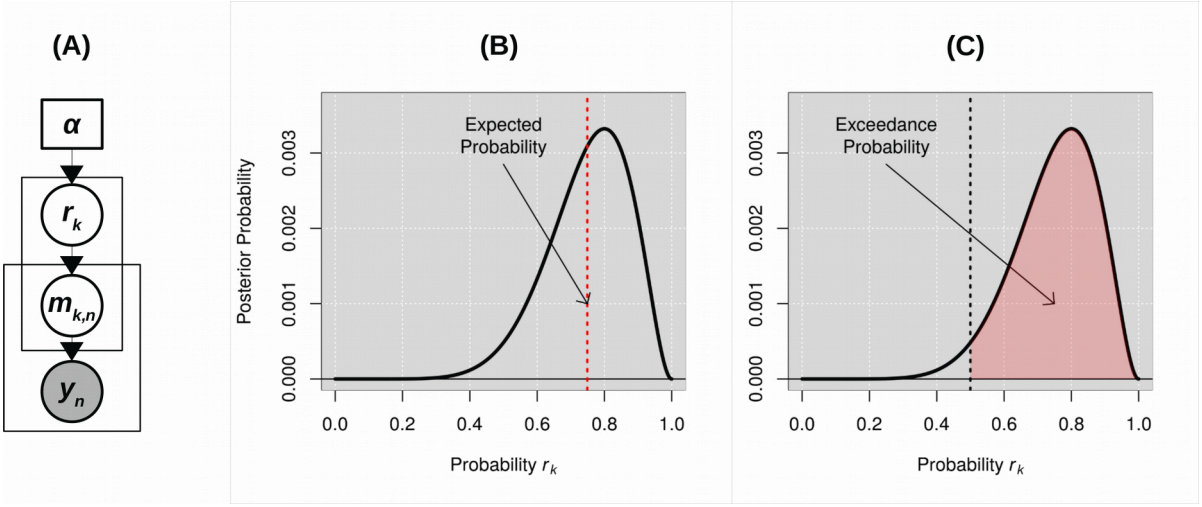

Fig. S4. Hierarchical Bayesian model selection estimation procedure and measures for model comparison (A): Compact visual representation of the hierarchical Bayesian modeling procedure; the procedure yields a distribution of probabilities r_k that each model k would fit any individual n in the sample. (B): Visual representation of a model's posterior probability distribution over r_k (black curve) and its expected probability (red dashed line); the expected probability is the mean expected value of r_k, i.e. $\int_0^1 r_k \times y$; (C) Visual representation of a model's posterior probability distribution over r_k (black curve) and the corresponding exceedance probability (red shaded area) in the hypothetical case of two possible models (i.e., $k = 2$; the exceedance probability, in this case, is simple the area to the right of $r_k = 0.5$). Modified from (29).

Task	Model	Dirichlet α	Expected Probability	Exceedance Probability	Log-Likelihood
All Tasks Combined	Common Model	134.23	0.7759	1.000	-3766837.56
	Hierarchical Closed	21.8	0.1260	0.0000	-3797769.54
	Hierarchical Open	0.99	0.0057	0.0000	-4292199.54
	Hub-and-spoke BG	14.99	0.0866	0.0000	-4250392.03
	Hub-and-spoke PFC	0.99	0.0057	0.0000	-4300912.00
Emotion Processing	Common Model	101.37	0.5307	1.0000	-858282.97
	Hierarchical Closed	3.68	0.0193	0.0000	-865011.12
	Hierarchical Open	37.27	0.1951	0.0000	-860955.74
	Hub-and-spoke BG	37.66	0.1972	0.0000	-861523.46
	Hub-and-spoke PFC	11.02	0.0577	0.0000	-861878.56
Incentive Processing	Common Model	144.77	0.7131	1.0000	-861026.48
	Hierarchical Closed	1.36	0.0067	1.0000	-869976.35
	Hierarchical Open	39.82	0.1962	0.0000	-866148.11
	Hub-and-spoke BG	14.8	0.0729	0.0000	-869471.16
	Hub-and-spoke PFC	2.26	0.0111	0.0000	-866795.76
Language & Math	Common Model	85.83	0.4494	0.7526	-894992.94
	Hierarchical Closed	7.74	0.0405	0.0000	-903067.62
	Hierarchical Open	76.84	0.4023	0.2474	-895740.71
	Hub-and-spoke BG	13.65	0.0715	0.0000	-918300.93
	Hub-and-spoke PFC	6.94	0.0363	0.0000	-902568.43
Relational Reasoning	Common Model	77.8	0.4116	0.9505	-746544.42
	Hierarchical Closed	8.35	0.0442	0.0000	-749922.68
	Hierarchical Open	57.81	0.3059	0.0490	-747385.57

Task	Model	Dirichlet α	Expected Probability	Exceedance Probability	Log-Likelihood
	Hub-and-spoke BG	42.56	0.2252	0.0005	-672161.79
	Hub-and-spoke PFC	2.48	0.0131	0.0000	-750719.3
Social Cognition	Common Model	141.47	0.7368	1.0000	-712975.64
	Hierarchical Closed	1.37	0.0072	0.0000	-720399.27
	Hierarchical Open	33.34	0.1737	0.0000	-716591.85
	Hub-and-spoke BG	13.57	0.0707	0.0000	-719374.15
	Hub-and-spoke PFC	2.24	0.0117	0.0000	-715828.71
Working Memory	Common Model	142.87	0.7441	1.0000	-113247.58
	Hierarchical Closed	46.13	0.2403	0.0000	-112996.33
	Hierarchical Open	1.00	0.0052	0.0000	-680320.81
	Hub-and-spoke BG	0.99	0.0051	0.0000	-680396.45
	Hub-and-spoke PFC	1.01	0.0053	0.0000	-679145.16

Table S3: Results of Bayesian model comparison across tasks and models. The table reports the dirichlet parameters estimated for each model's posterior distribution (see Fig. S4A), as well as the two statistics reported in this paper, expected and exceedance probabilities. For completeness, the last column also reports the corresponding log-likelihood for every model.