Biological Blueprints for Human Inspired AI

Stanford, April 7 & 9, 2020

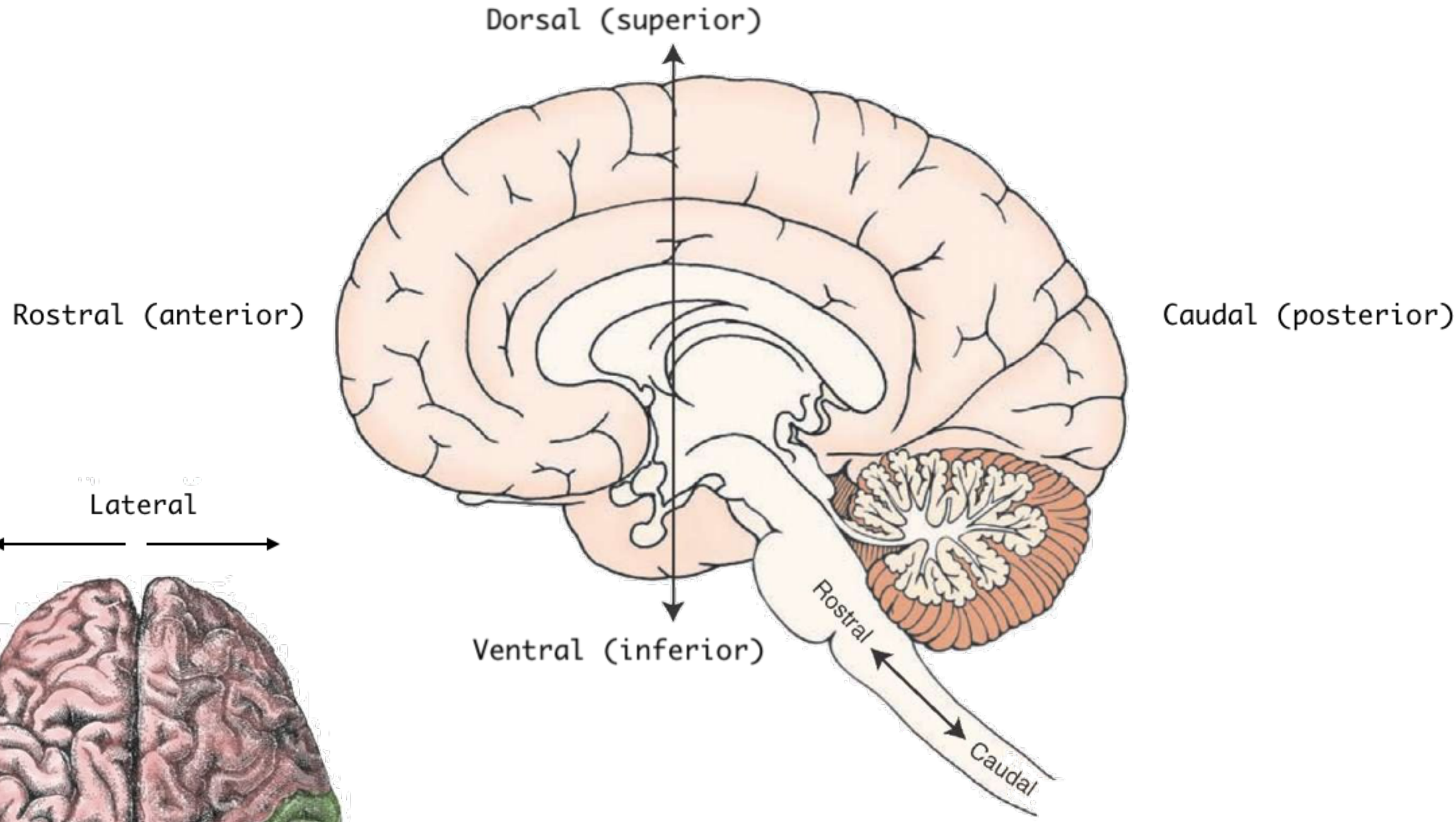
Acknowledgements: Nikhil Bhattasali, Chaofei Fan, Eugene Lewis, Megumi Sano, Yash Savani and Rishabh Singh

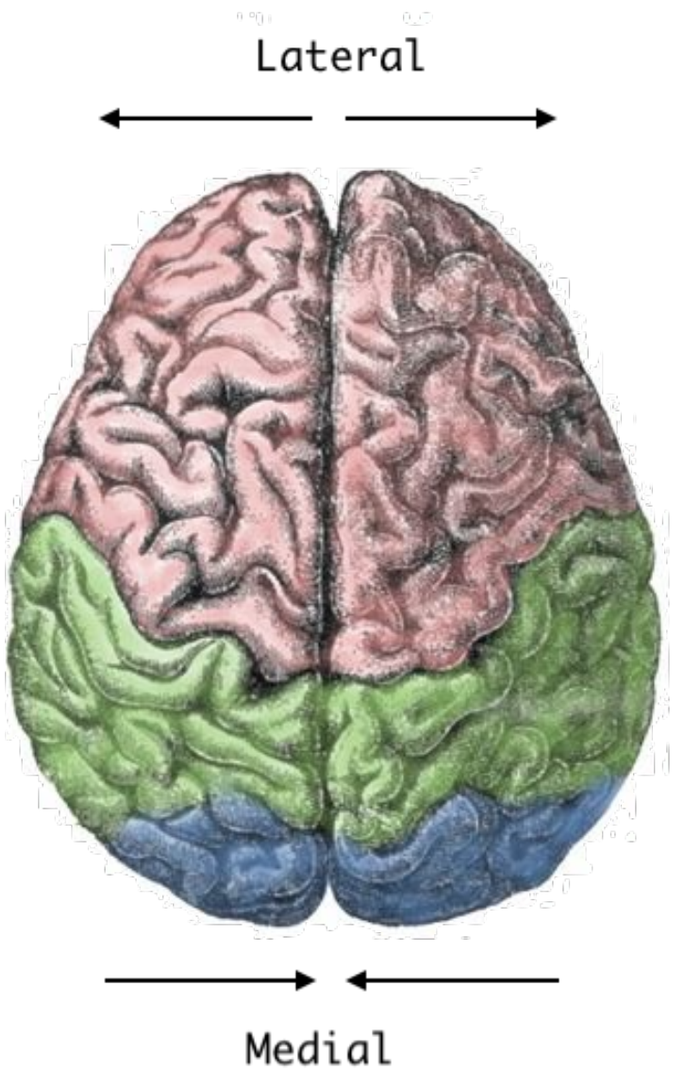
Biological Blueprints for Human Inspired AI[‡]

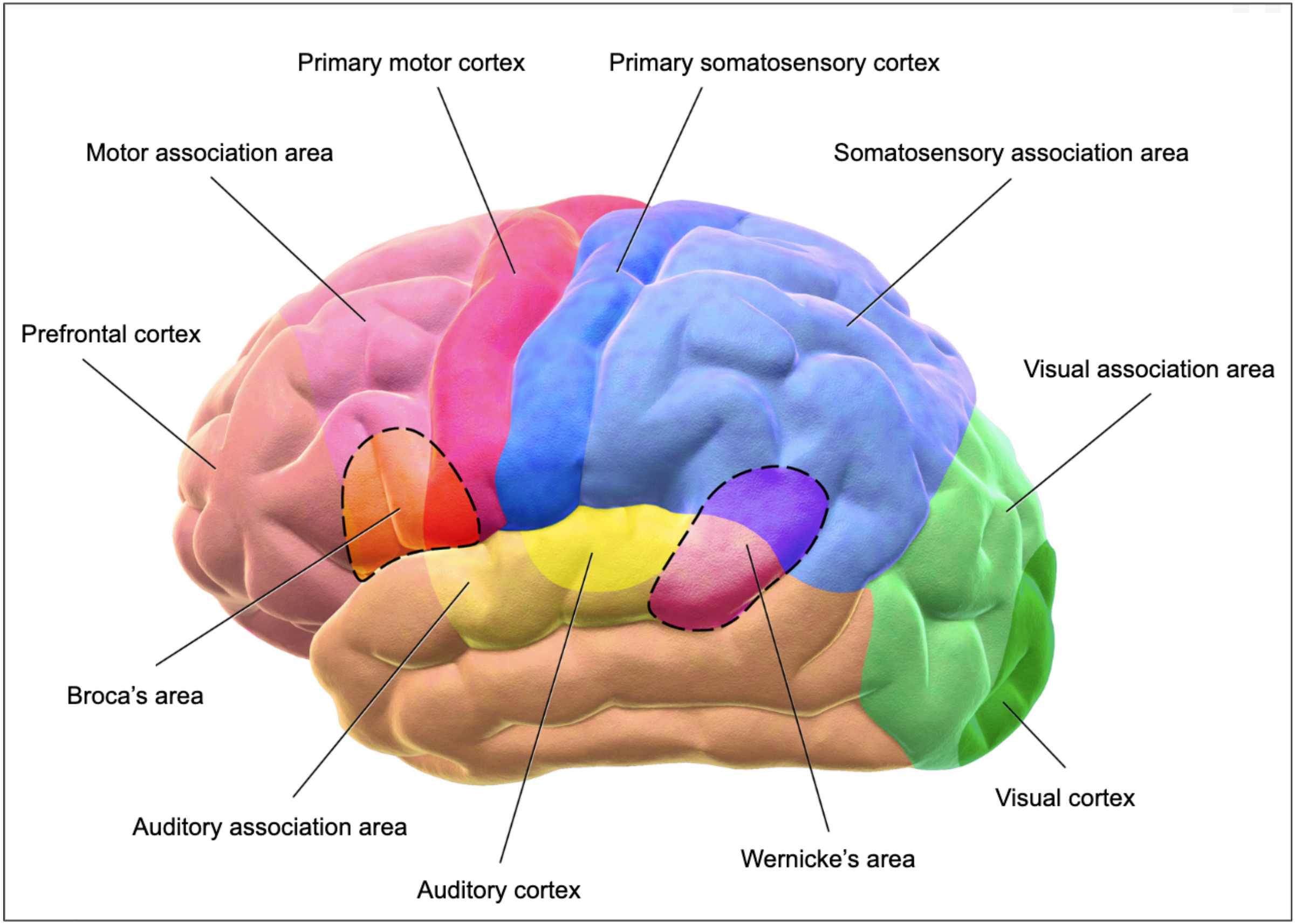
What would serve as a model of the human human brain?[†]

[†]Another of those naive rhetorical questions we promised you.

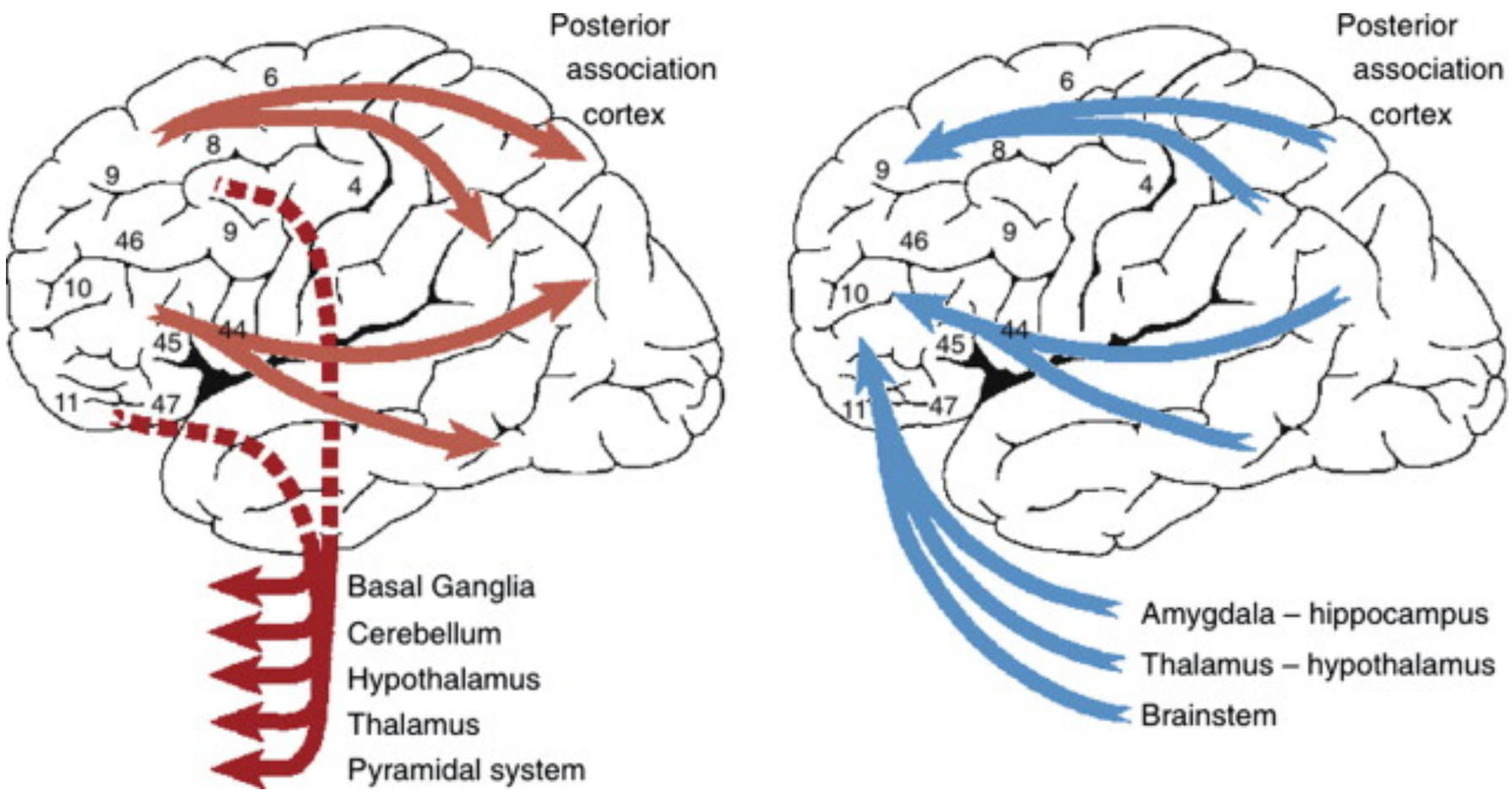
[‡]Part II: Stanford CS379C © 2020 Thomas Dean


Dorsal (superior)
Rostral (anterior)
Caudal (posterior)
Ventral (inferior)
Rostral
Caudal


Lateral
Medial


Primary motor cortex
Primary somatosensory cortex
Motor association area
Somatosensory association area
Prefrontal cortex
Visual association area
Broca's area
Auditory association area
Auditory cortex
Wernicke's area
Visual cortex

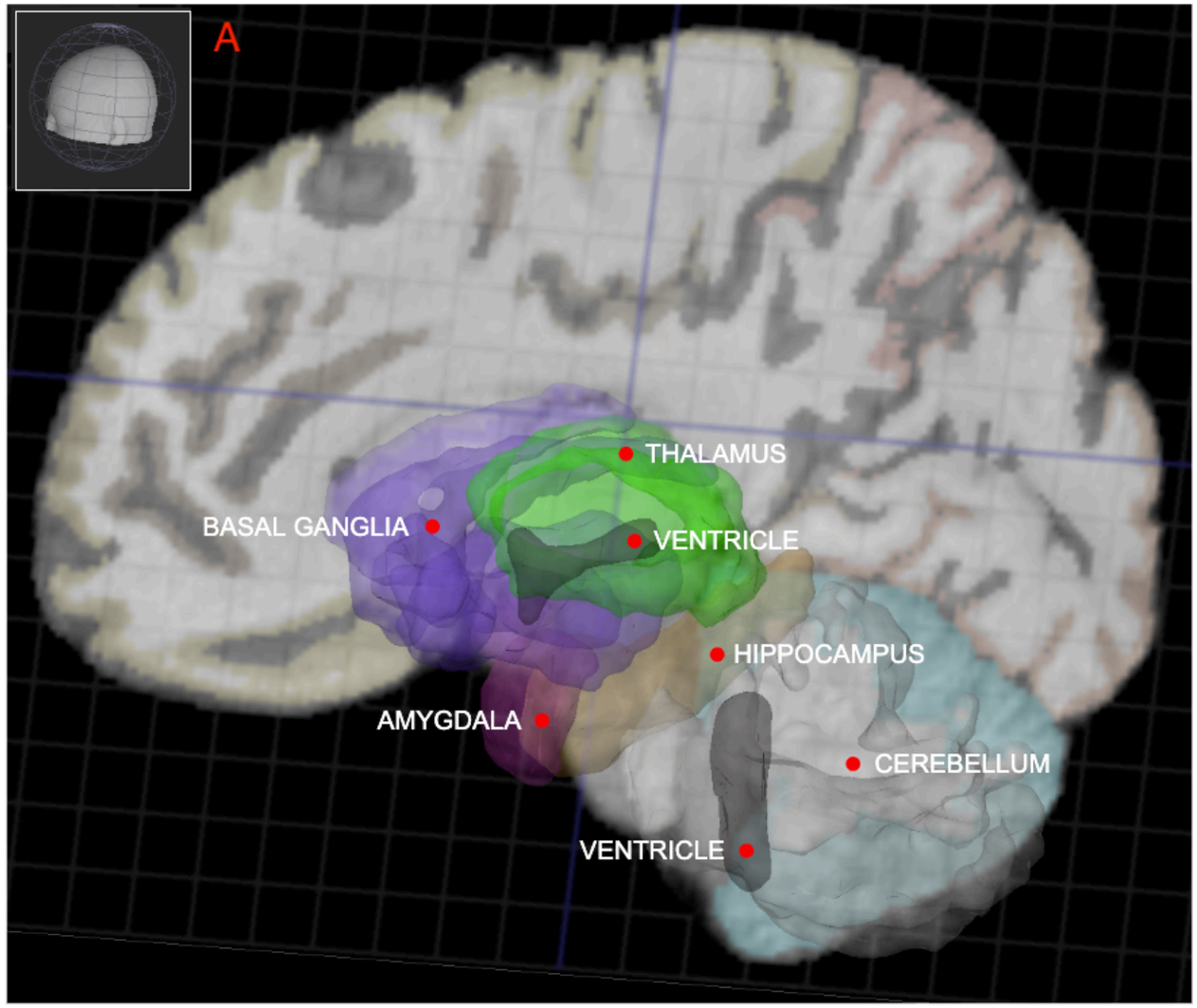
Vladimir Betz's Anterior / Posterior Functional Dichotomy

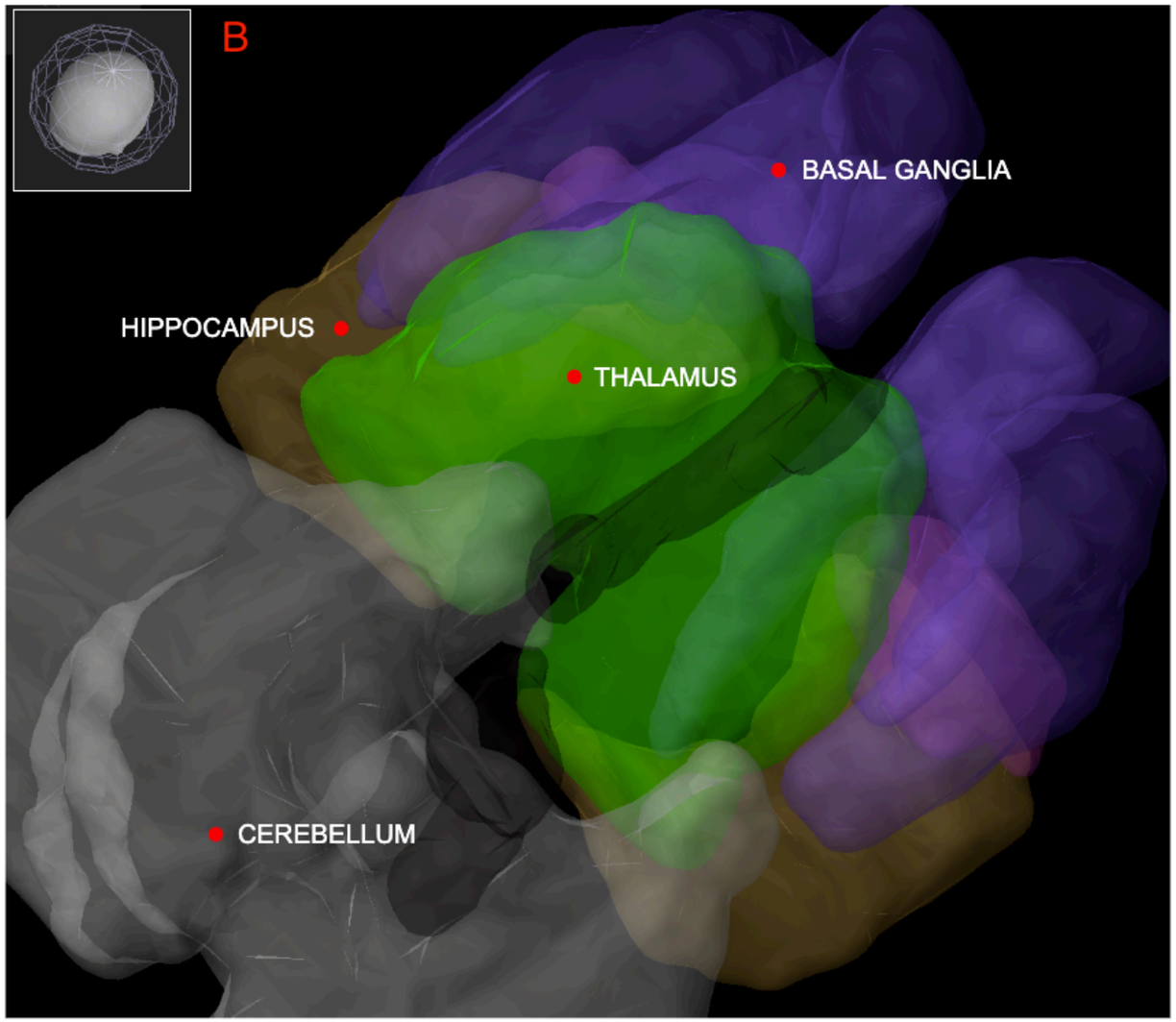


Betz, W. (1874) Anatomischer Nachweis Zweier Gehirn Centra. *Central-blatt für die medizinische Wissenschaften*. 12, 578–580.

Sergiy V. Kushchayev, Vitaliy F. Moskalenko, Philip C. Wiener, Vitaliy I. Tsymbaliuk, Viktor G. Cherkasov, Irina V. Dzyavulska, Oleksander I. Kovalchuk, Volker K. H. Sonntag, Robert F. Spetzler, and Mark C. Preul. The discovery of the pyramidal neurons: Vladimir Betz and a new era of neuroscience. *Brain*, 135(1):285-300, 2011.

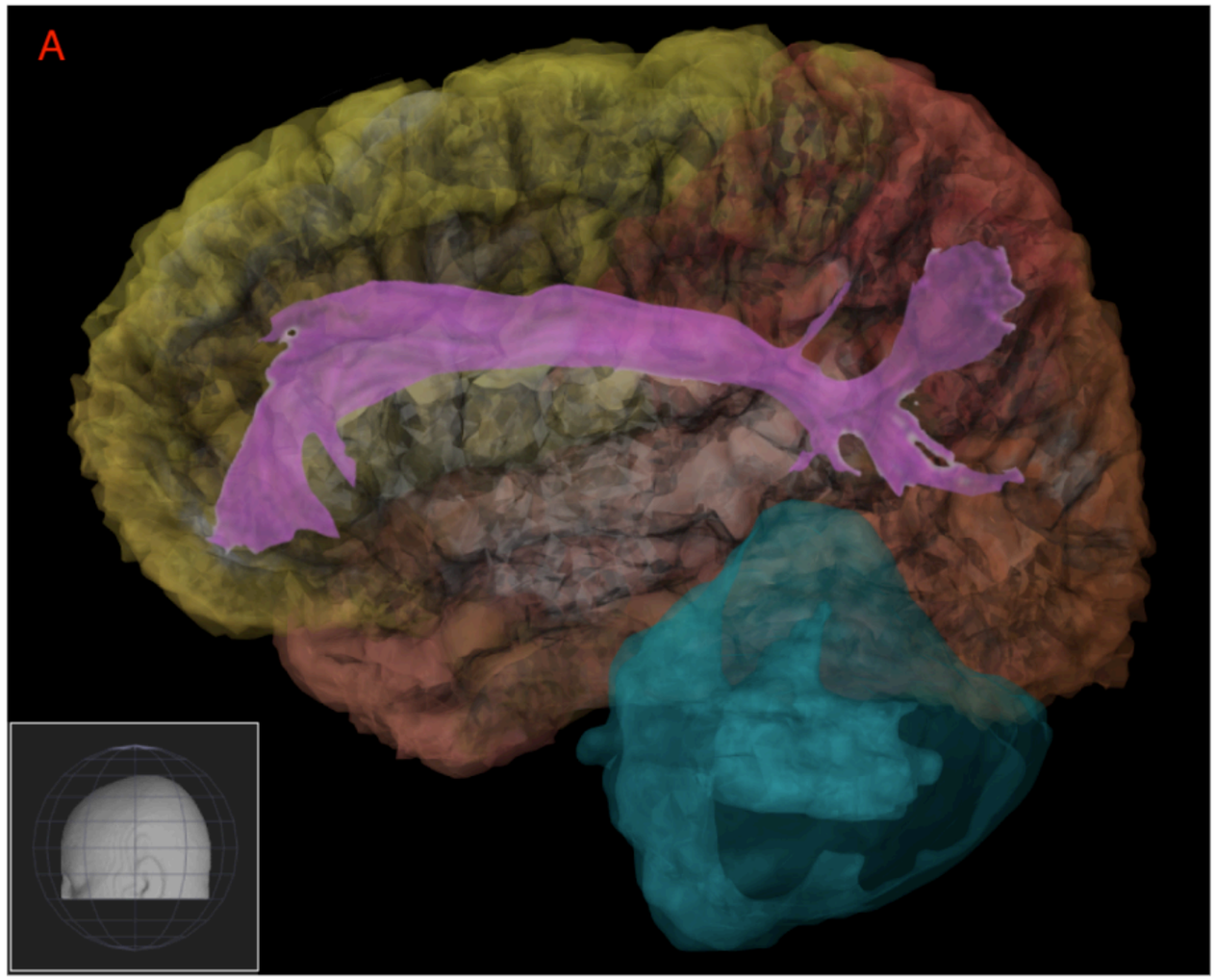
Subcortical Functional Circuits

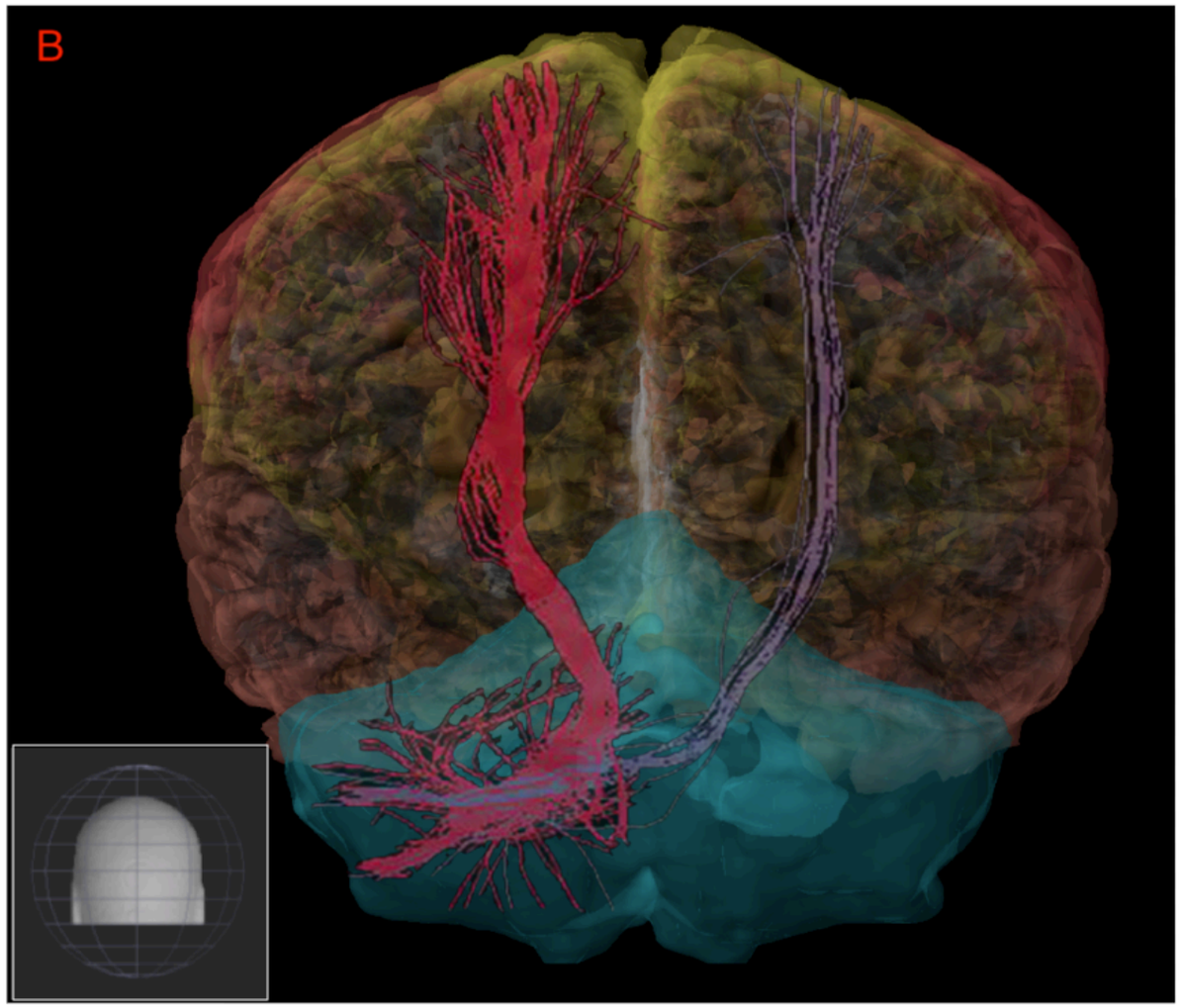




Michael J. Hawrylycz, Ed S. Lein, Angela L. Guillozet-Bongaarts, Elaine H. Shen, Lydia Ng, Jeremy A. Miller, Louie N. van de Lagemaat, Kimberly A. Smith, Amanda Ebbert, Zackery L. Riley, Chris Abajian, Christian F. Beckmann, Amy Bernard, Darren Bertagnolli, Andrew F. Boe, Preston M. Cartagena, M. Mallar Chakravarty, Mike Chapin, Jimmy Chong, Rachel A. Dalley, Barry David Daly, Chinh Dang, Suvro Datta, Nick Dee, Tim A. Dolbeare, Vance Faber, David Feng, David R. Fowler, Jeff Goldy, Benjamin W. Gregor, Zeb Haradon, David R. Haynor, John G. Hohmann, Steve Horvath, Robert E. Howard, Andreas Jeromin, Jayson M. Jochim, Marty Kinnunen, Christopher Lau, Evan T. Lazarz, Changkyu Lee, Tracy A. Lemon, Ling Li, Yang Li, John A. Morris, Caroline C. Overly, Patrick D. Parker, Sheana E. Parry, Melissa Reding, Joshua J. Royall, Jay Schulkin, Pedro Adolfo Sequeira, Clifford R. Slaughterbeck, Simon C. Smith, Andy J. Sodt, Susan M. Sunkin, Beryl E. Swanson, Marquis P. Vawter, Derric Williams, Paul Wohnoutka, H. Ronald Zielke, Daniel H. Geschwind, Patrick R. Hof, Stephen M. Smith, Christof Koch, Seth G. N. Grant, and Allan R. Jones. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489:391--399, 2012.

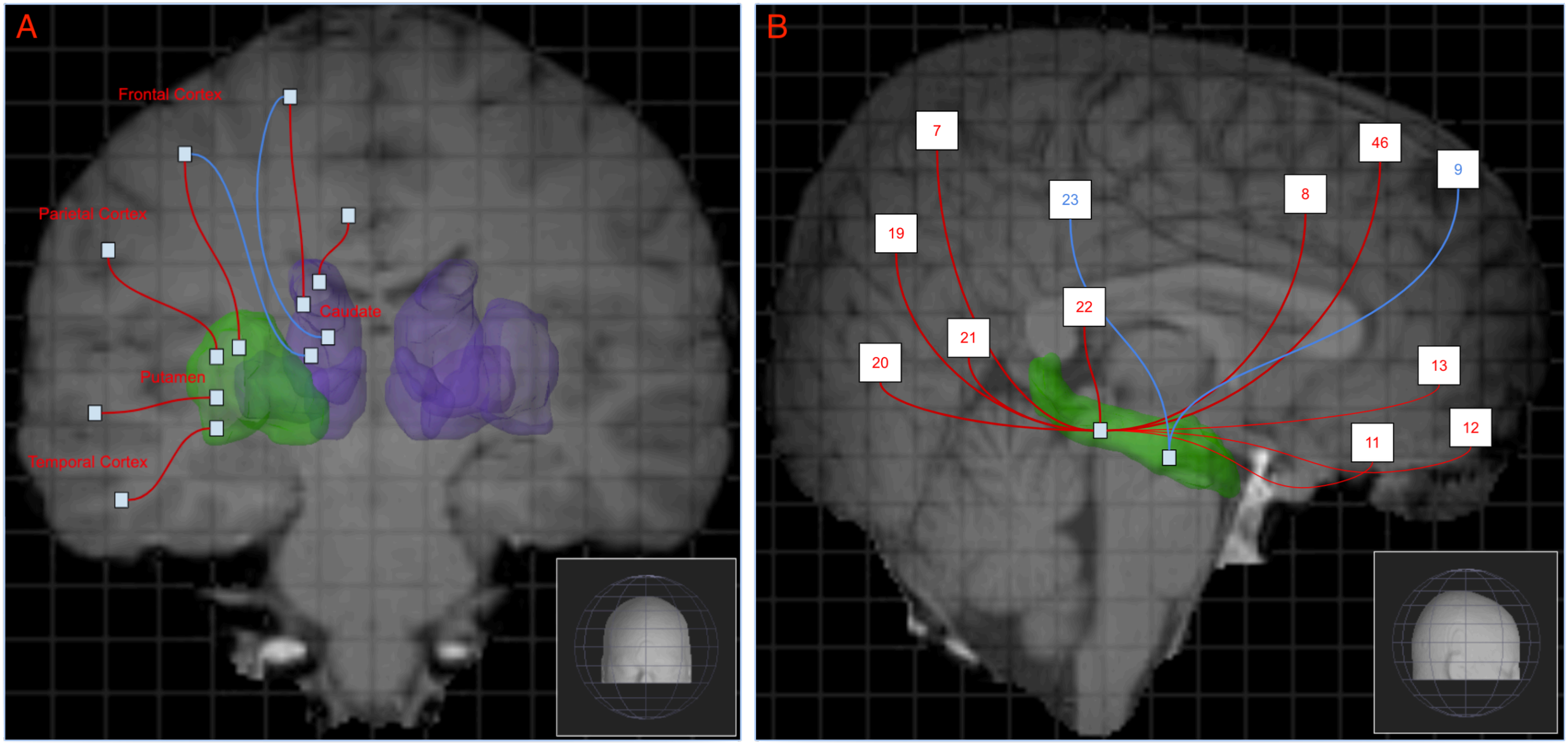
Major White Matter (Myelinated) Fiber Tracts





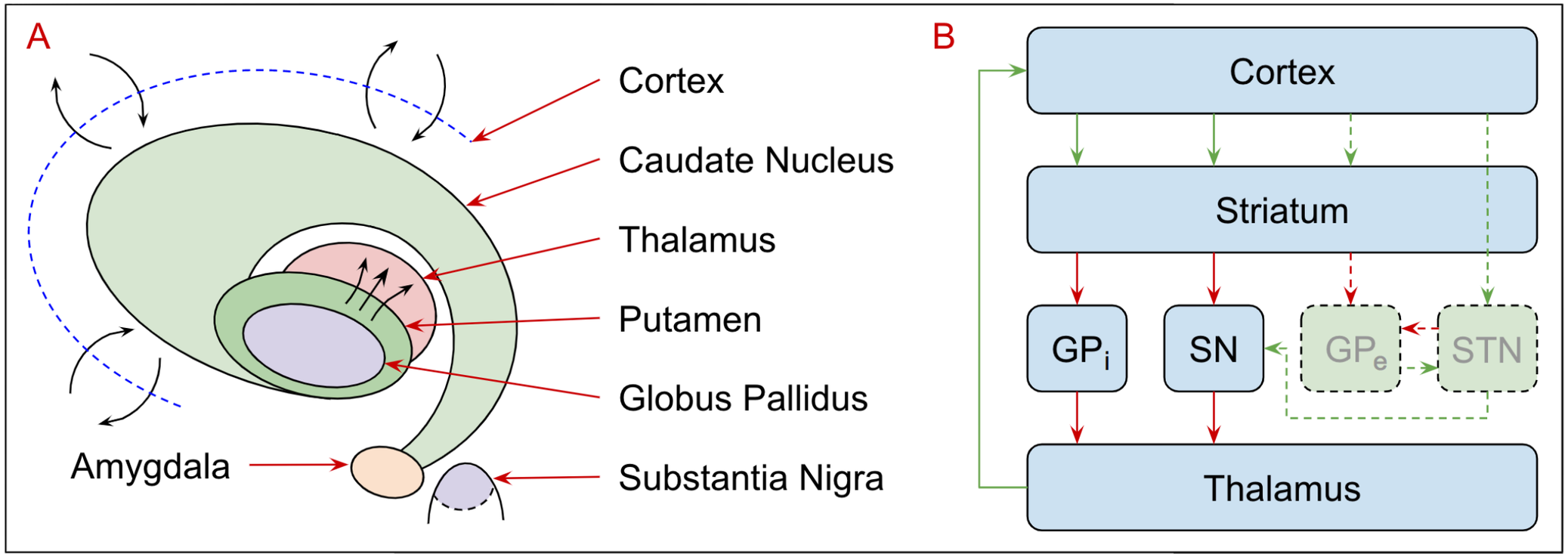
Michael J. Hawrylycz, Ed S. Lein, Angela L. Guillozet-Bongaarts, Elaine H. Shen, Lydia Ng, Jeremy A. Miller, Louie N. van de Lagemaat, Kimberly A. Smith, Amanda Ebbert, Zackery L. Riley, Chris Abajian, Christian F. Beckmann, Amy Bernard, Darren Bertagnolli, Andrew F. Boe, Preston M. Cartagena, M. Mallar Chakravarty, Mike Chapin, Jimmy Chong, Rachel A. Dalley, Barry David Daly, Chinh Dang, Suvro Datta, Nick Dee, Tim A. Dolbeare, Vance Faber, David Feng, David R. Fowler, Jeff Goldy, Benjamin W. Gregor, Zeb Haradon, David R. Haynor, John G. Hohmann, Steve Horvath, Robert E. Howard, Andreas Jeromin, Jayson M. Jochim, Marty Kinnunen, Christopher Lau, Evan T. Lazarz, Changkyu Lee, Tracy A. Lemon, Ling Li, Yang Li, John A. Morris, Caroline C. Overly, Patrick D. Parker, Sheana E. Parry, Melissa Reding, Joshua J. Royall, Jay Schulkin, Pedro Adolfo Sequeira, Clifford R. Slaughterbeck, Simon C. Smith, Andy J. Sodt, Susan M. Sunkin, Beryl E. Swanson, Marquis P. Vawter, Derric Williams, Paul Wohnoutka, H. Ronald Zielke, Daniel H. Geschwind, Patrick R. Hof, Stephen M. Smith, Christof Koch, Seth G. N. Grant, and Allan R. Jones. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489:391--399, 2012.

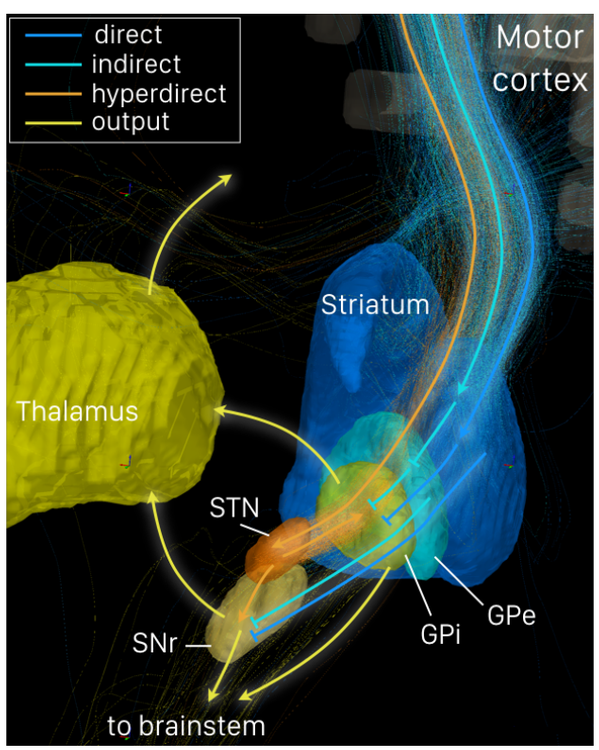
Localizing Function in the Cerebral Cortex

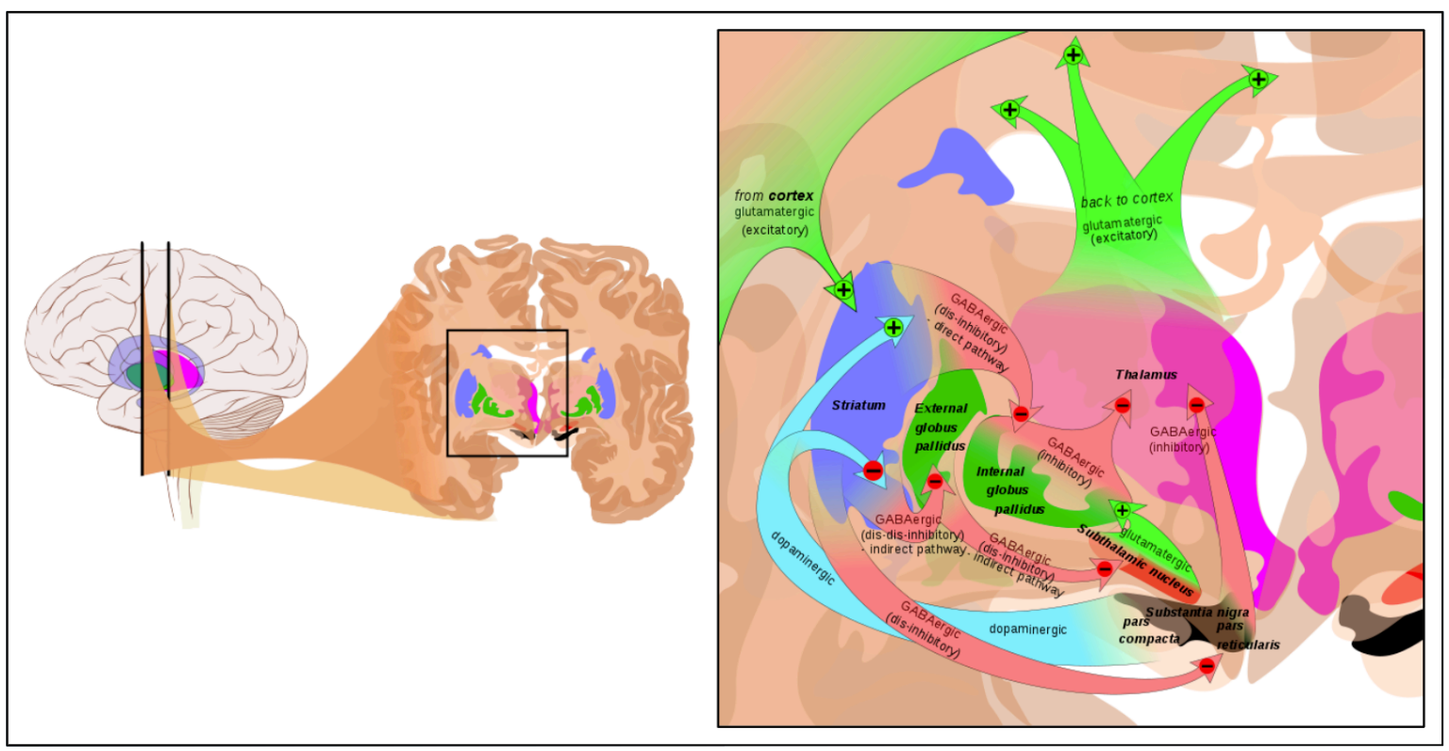


Michael J. Hawrylycz, Ed S. Lein, Angela L. Guillozet-Bongaarts, Elaine H. Shen, Lydia Ng, Jeremy A. Miller, Louie N. van de Lagemaat, Kimberly A. Smith, Amanda Ebbert, Zackery L. Riley, Chris Abajian, Christian F. Beckmann, Amy Bernard, Darren Bertagnolli, Andrew F. Boe, Preston M. Cartagena, M. Mallar Chakravarty, Mike Chapin, Jimmy Chong, Rachel A. Dalley, Barry David Daly, Chinh Dang, Suvro Datta, Nick Dee, Tim A. Dolbeare, Vance Faber, David Feng, David R. Fowler, Jeff Goldy, Benjamin W. Gregor, Zeb Haradon, David R. Haynor, John G. Hohmann, Steve Horvath, Robert E. Howard, Andreas Jeromin, Jayson M. Jochim, Marty Kinnunen, Christopher Lau, Evan T. Lazarz, Changkyu Lee, Tracy A. Lemon, Ling Li, Yang Li, John A. Morris, Caroline C. Overly, Patrick D. Parker, Sheana E. Parry, Melissa Reding, Joshua J. Royall, Jay Schulkin, Pedro Adolfo Sequeira, Clifford R. Slaughterbeck, Simon C. Smith, Andy J. Sodt, Susan M. Sunkin, Beryl E. Swanson, Marquis P. Vawter, Derric Williams, Paul Wohnoutka, H. Ronald Zielke, Daniel H. Geschwind, Patrick R. Hof, Stephen M. Smith, Christof Koch, Seth G. N. Grant, and Allan R. Jones. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489:391--399, 2012.

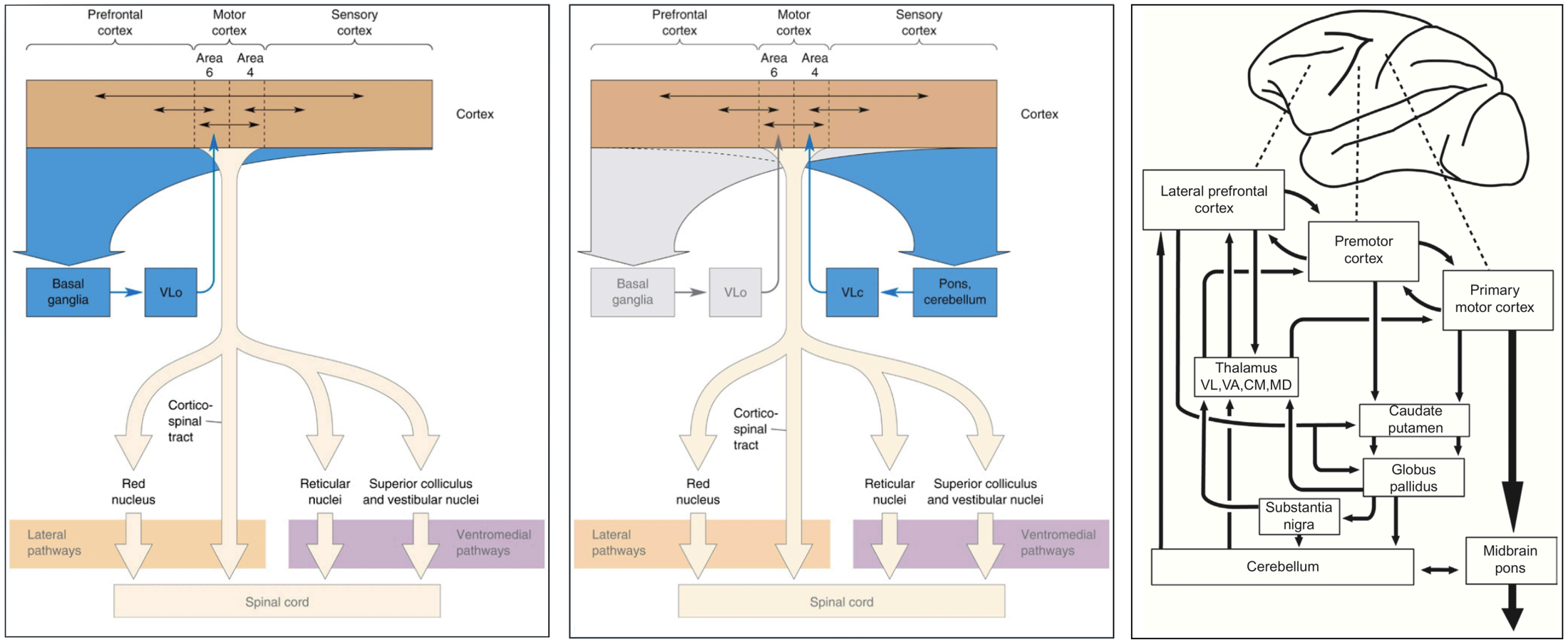
Localizing Function in the Cerebral Cortex: Basal Ganglia







Prefrontal Cortex Basal Ganglia Working Memory (PBWM)

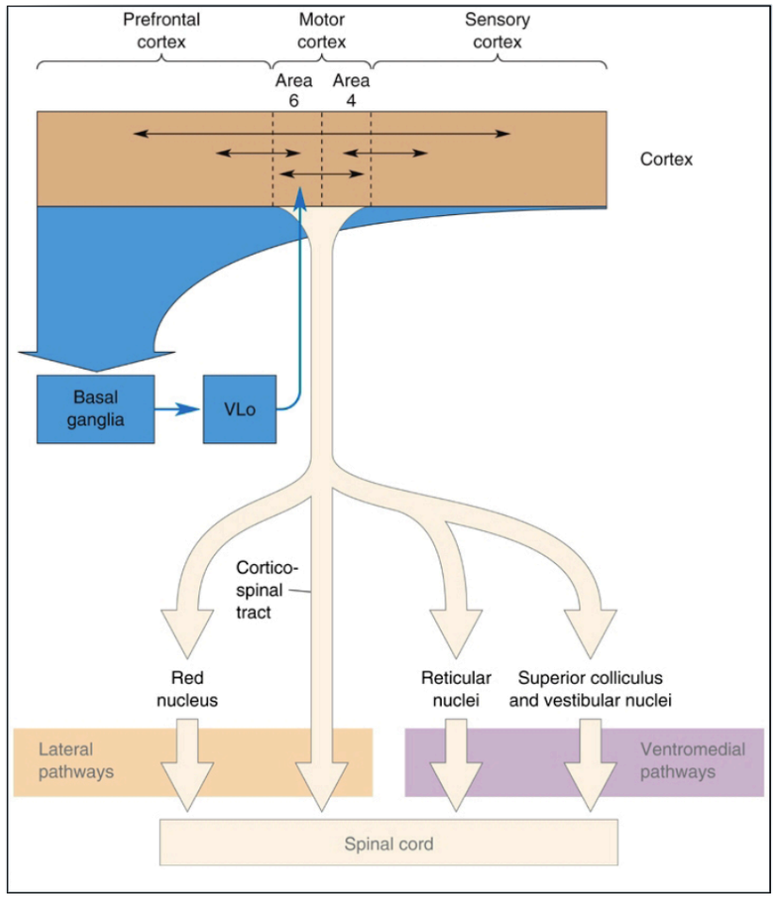


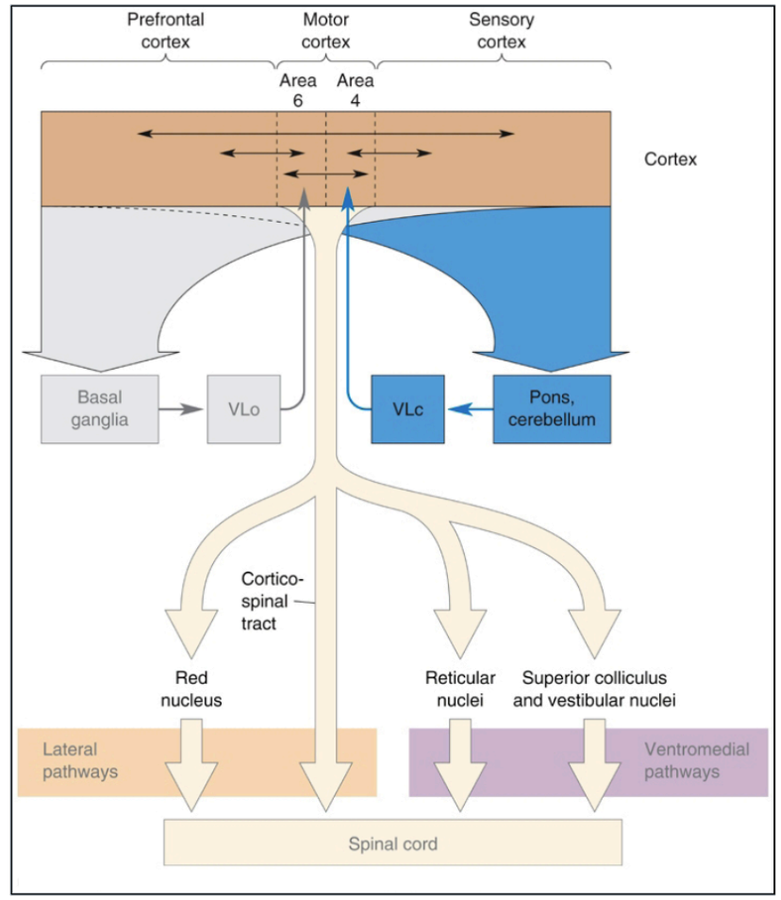
Brodmann Area 4 refers to the primary motor cortex of the human brain. It is located in the posterior portion of the frontal lobe. Brodmann Area 6 is located anterior to primary motor cortex, composed of the premotor cortex and supplementary motor area.

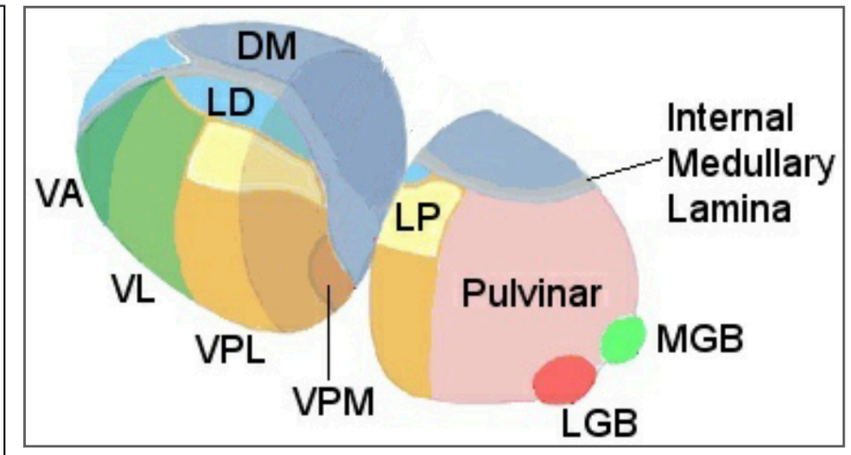
Mark F. Bear, Barry Connors, and Michael Paradiso. *Neuroscience: Exploring the Brain (Fifth Edition)*. Lippincott Williams & Wilkins, 2015.

Prefrontal Cortex Basal Ganglia Working Memory (PBWM)

For students looking for an introductory textbook, Bear, Connors, and Paradiso. *Neuroscience: Exploring the Brain,* is highly recommended, see Figure 14.10 below describing the motor loop from the cortex to basal ganglia to the thalamus and back to Brodmann Area 6 in the motor cortex. *Neuroscience Online* provides an excellent open-access alternative.

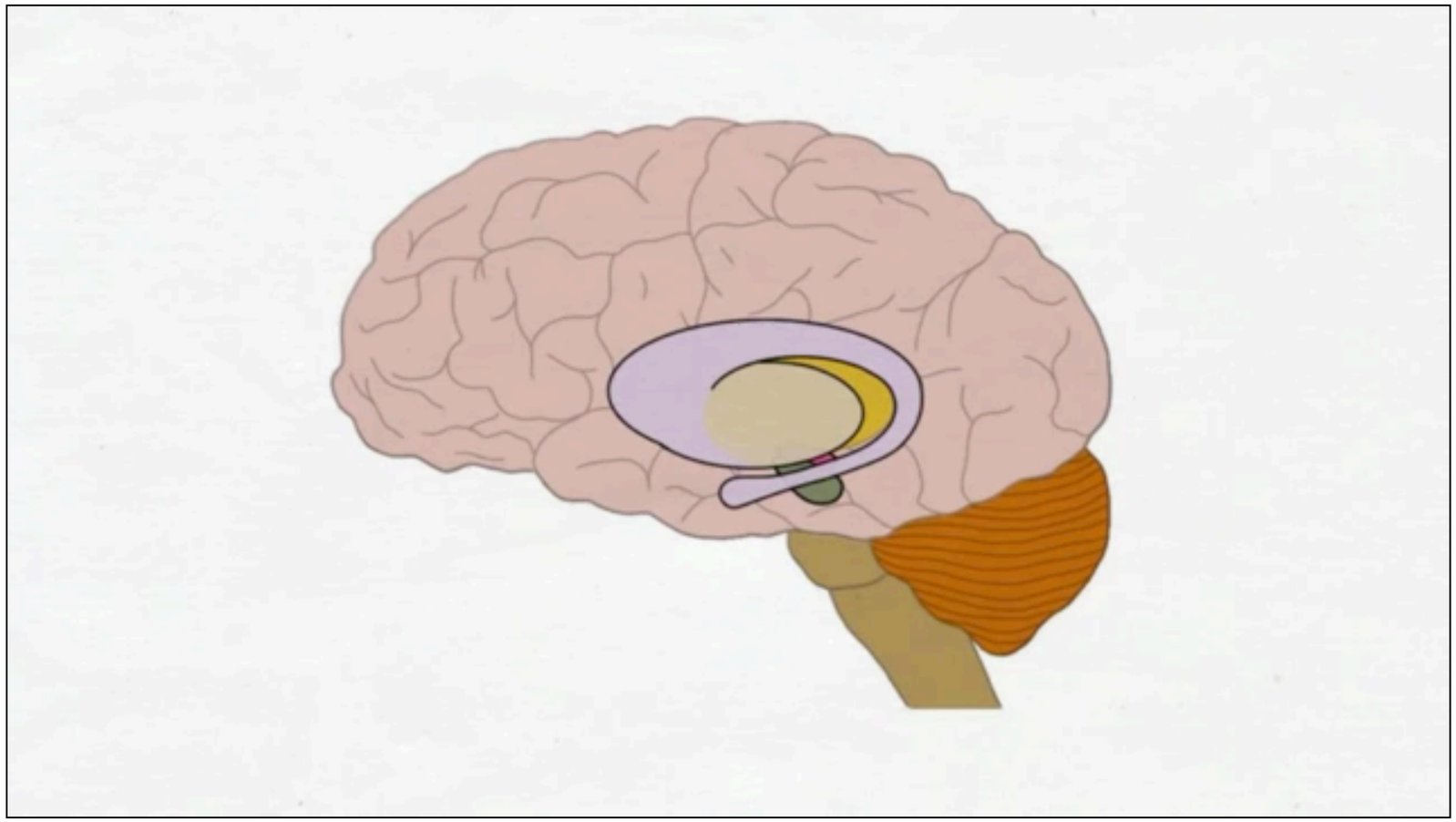


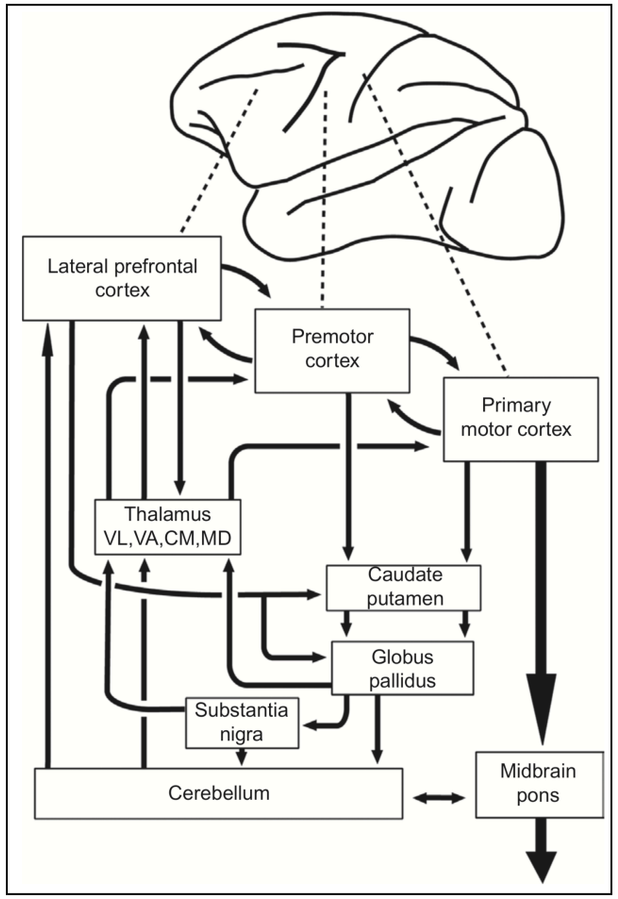




Mark F. Bear, Barry Connors, and Michael Paradiso. *Neuroscience: Exploring the Brain (Fifth Edition)*. Lippincott Williams & Wilkins, 2015.

Prefrontal Cortex Basal Ganglia Working Memory (PBWM)

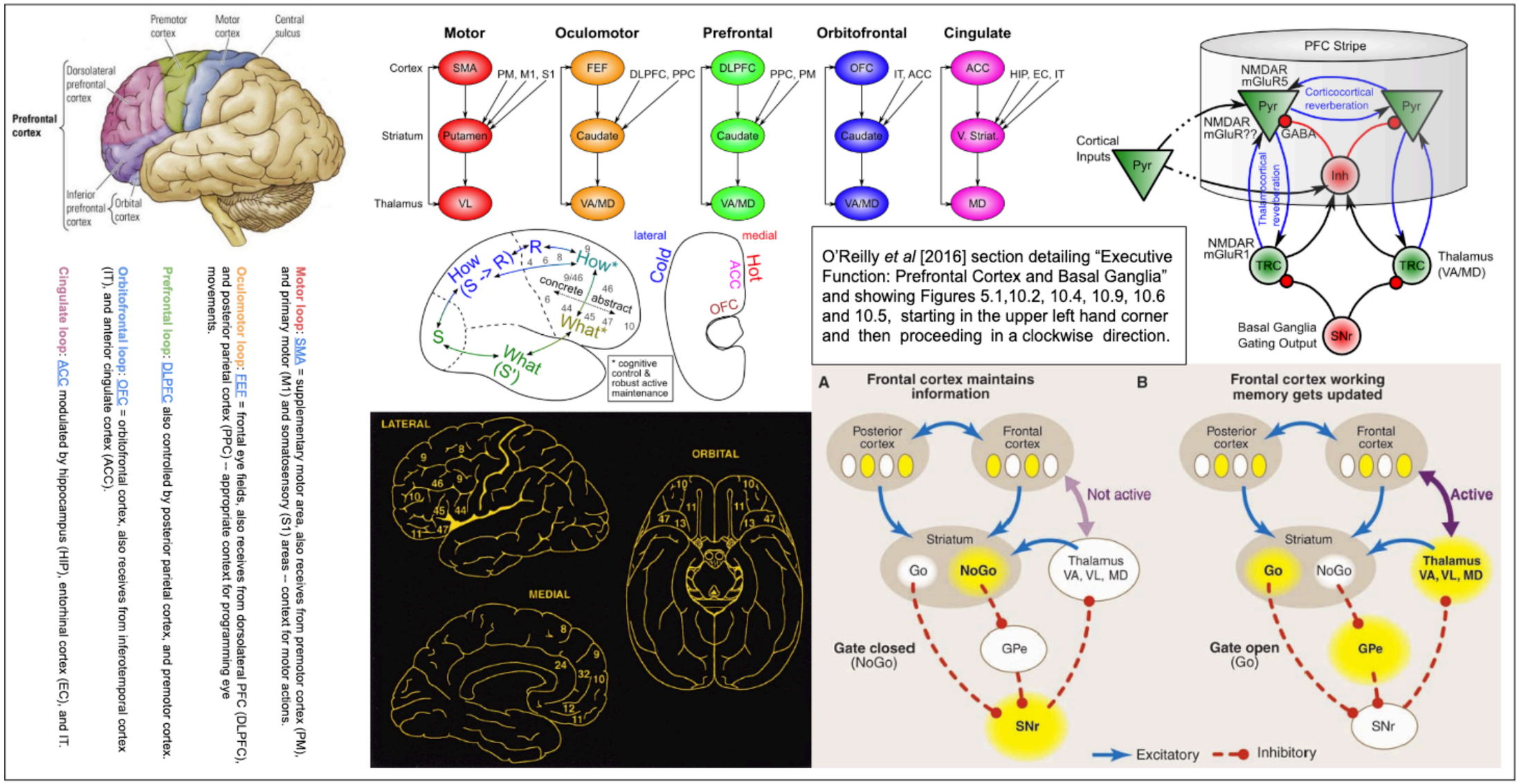




Mark F. Bear, Barry Connors, and Michael Paradiso. *Neuroscience: Exploring the Brain (Fifth Edition)*. Lippincott Williams & Wilkins, 2015.

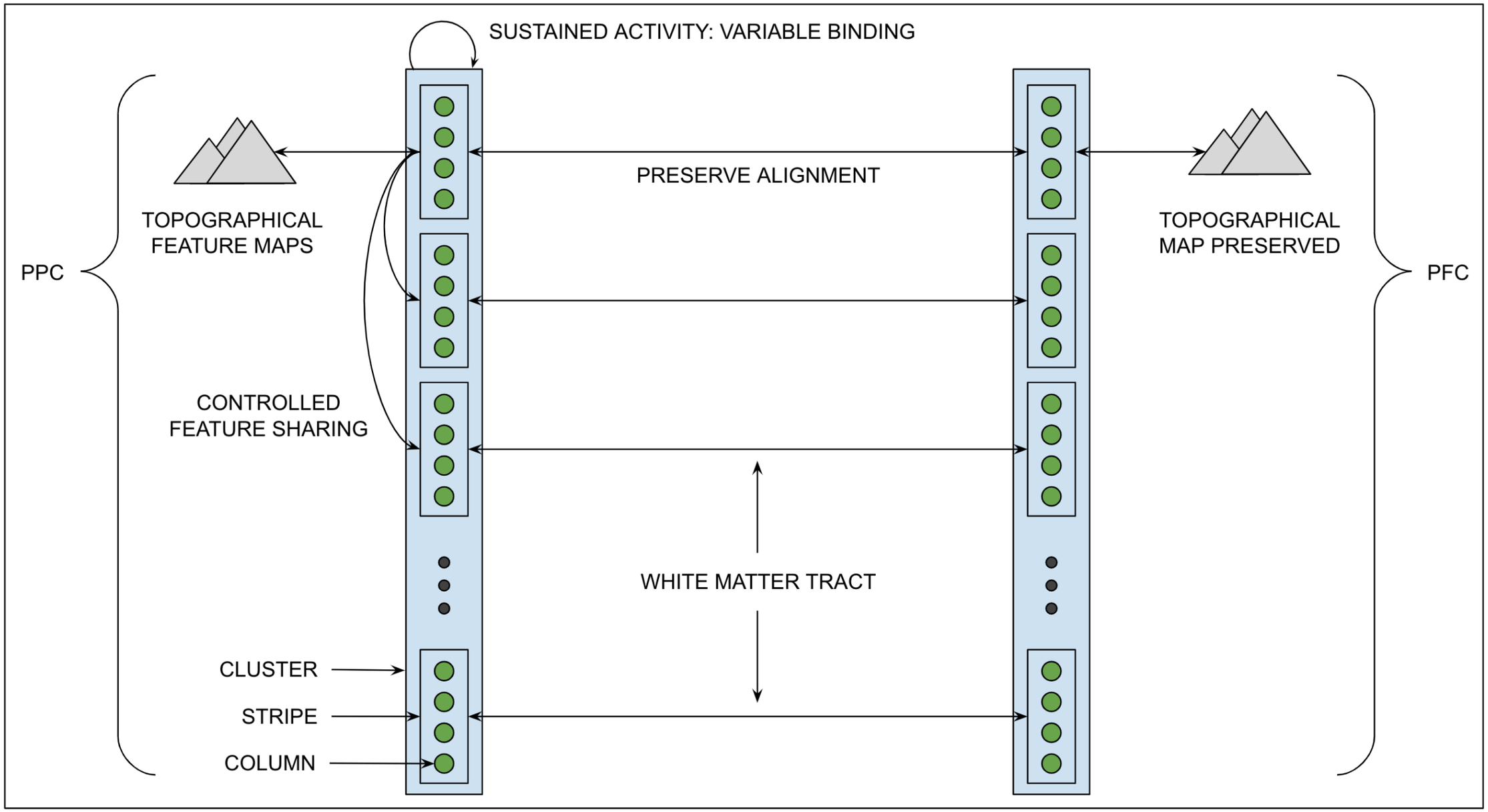
Prefrontal Cortex Basal Ganglia Working Memory (PBWM)

Students taking the course for credit will lead discussion sections focussed on core topics providing background material in neuroscience and machine learning. The text *Computational Cognitive Neuroscience*[1] covers the selected topics and serves as a primary reference to simplify explanations in the discussion sections. Here we see a sequence of graphics from the text for the discussion section on the Prefrontal Cortex Basal Ganglia Working Memory (PBWM) model[1].



[1] Randall C. O'Reilly, Yuko Munakata, Michael Frank and Thomas E. Hazy. *Computational Cognitive Neuroscience* (Third Edition) 2016.

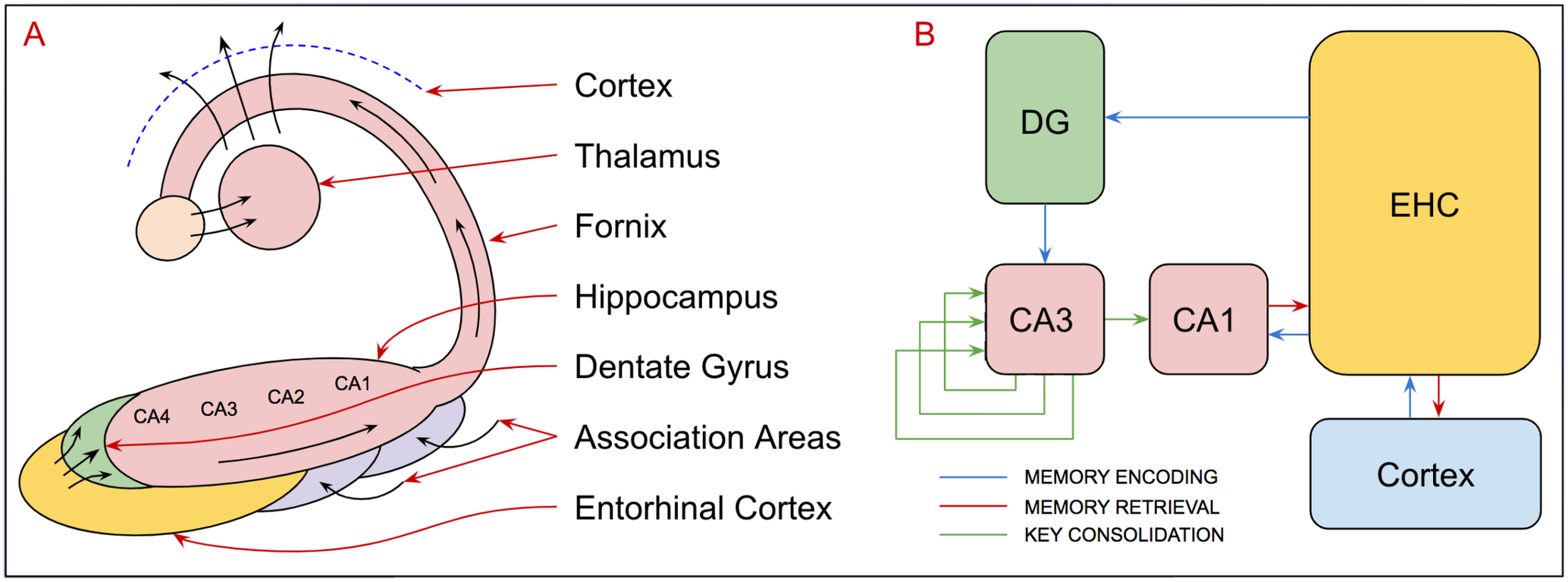
Localizing Function in the Cerebral Cortex: Stripe Clusters



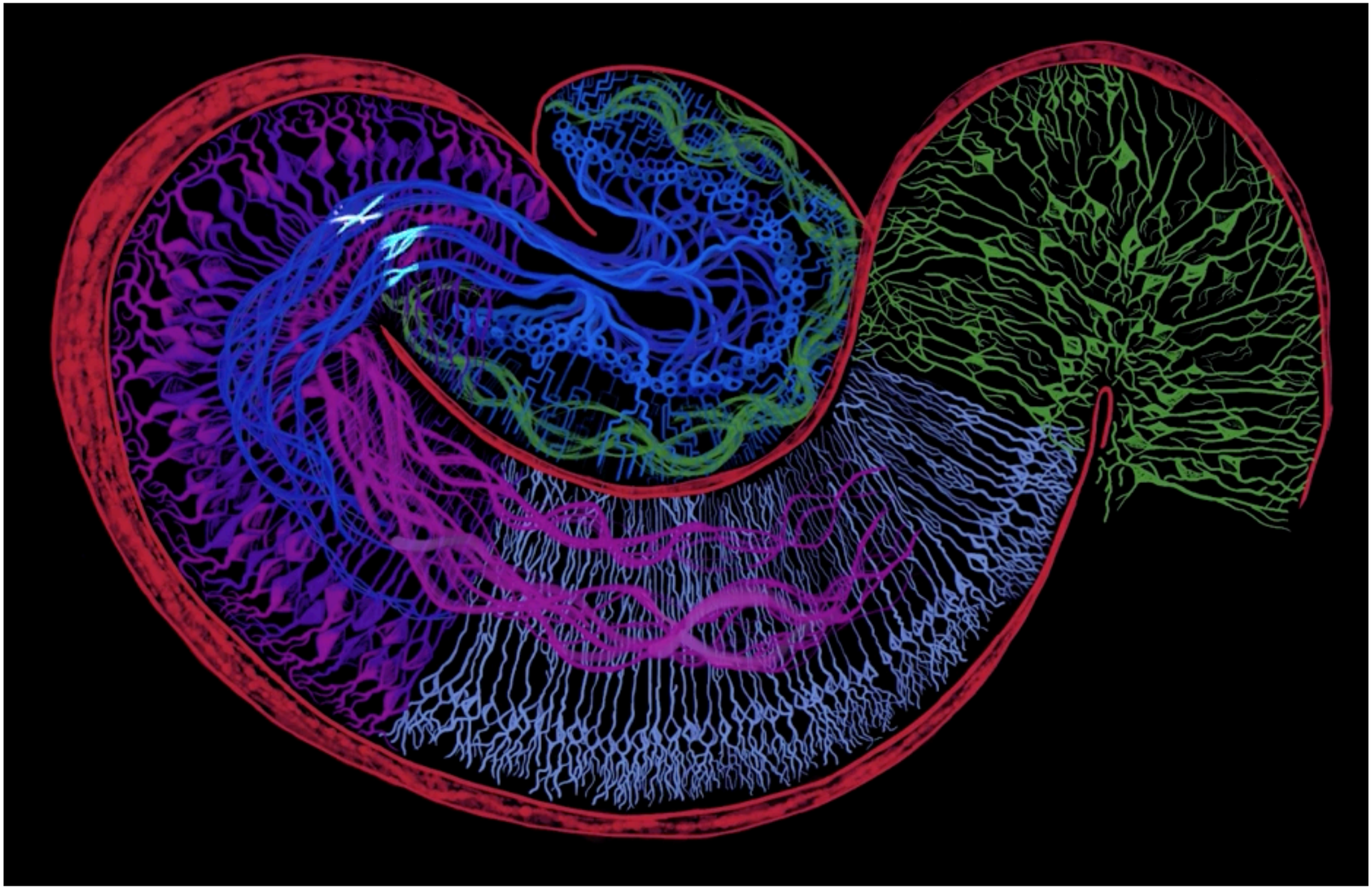
H. Barbas and M. A. Garcia-Cabezas. How the prefrontal executive got its stripes. *Current Opinion in Neurobiology*, 40:125-134, 2016.

David A. Lewis, Darlene S. Melchitzky, and Guillermo-Gonzalez Burgos. Specificity in the functional architecture of primate prefrontal cortex. *Journal of Neurocytology*, 31(3):265-276, 2002.

Localizing Function in the Cerebral Cortex: Hippocampus

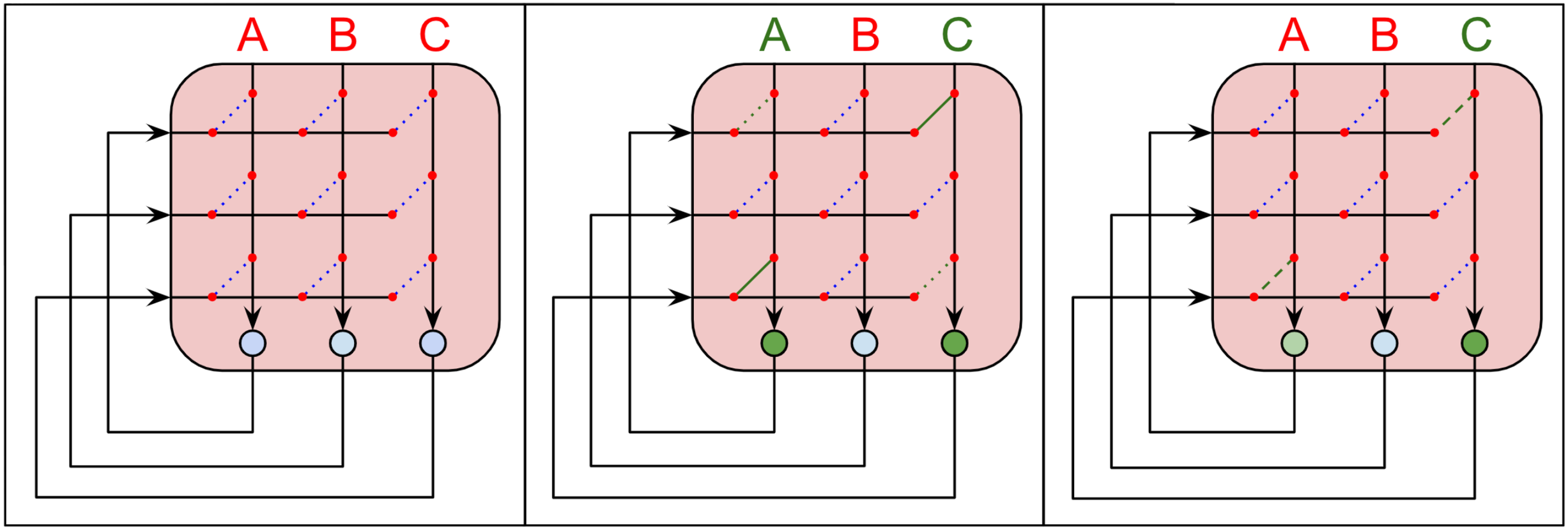


Localizing Function in the Cerebral Cortex: Hippocampus



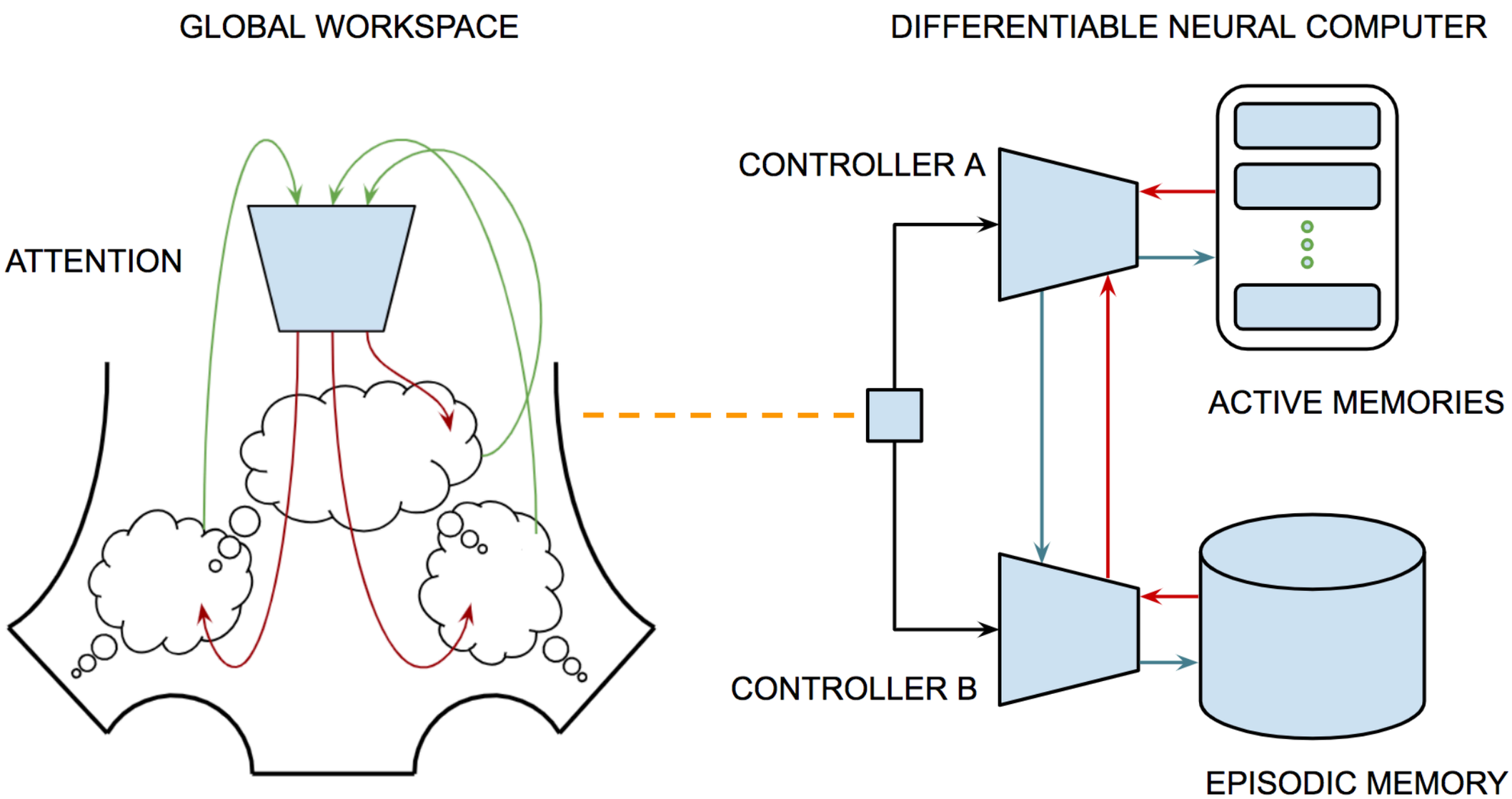
Updated final version of "Hebb's Three Postulates: from Brain to Soma": Barsen, S. (Director, Designer, Animator), Barnes, S. J. (Producer), Ritland, L. (Editor). (2015, Oct 18). Animating Hebb's Three Postulates: from Brain to Soma [hebb.metaplasticity.com].

Hippocampal Pattern Completion as Auto Associative Network



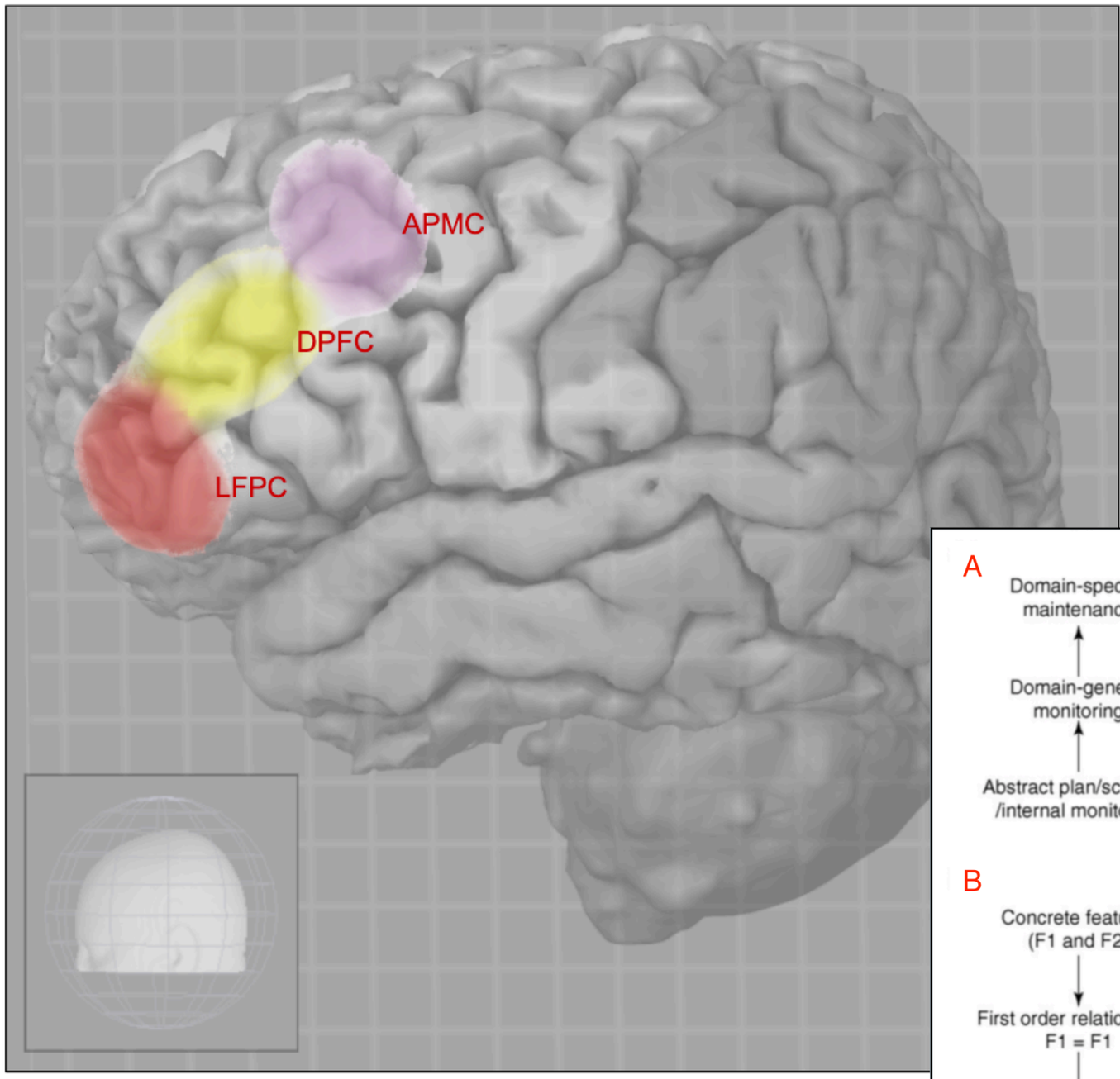
Edmund Rolls. The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, 7:74, 2013.

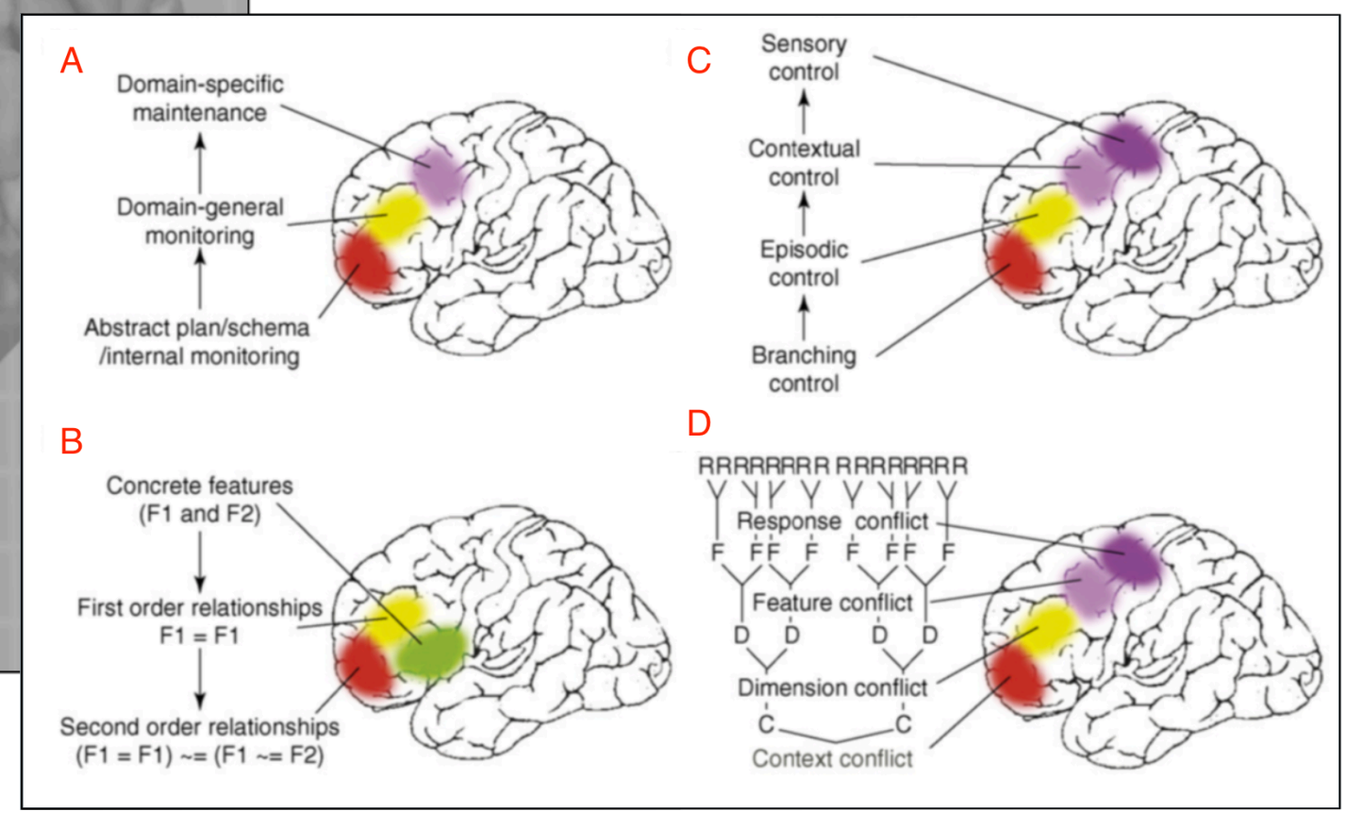
Hippocampal Complex / Episodic Memory Modeled as a DNC



Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, Adrià Puigdoménech Badia, Karl Moritz Hermann, Yori Zwols, Georg Ostrovski, Adam Cain, Helen King, Christopher Summerfield, Phil Blunsom, Koray Kavukcuoglu, and Demis Hassabis. Hybrid computing using a neural network with dynamic external memory. *Nature*, 538:471-476, 2016.

Rostro-Caudal Gradient: Hierarchical & Relational Complexity

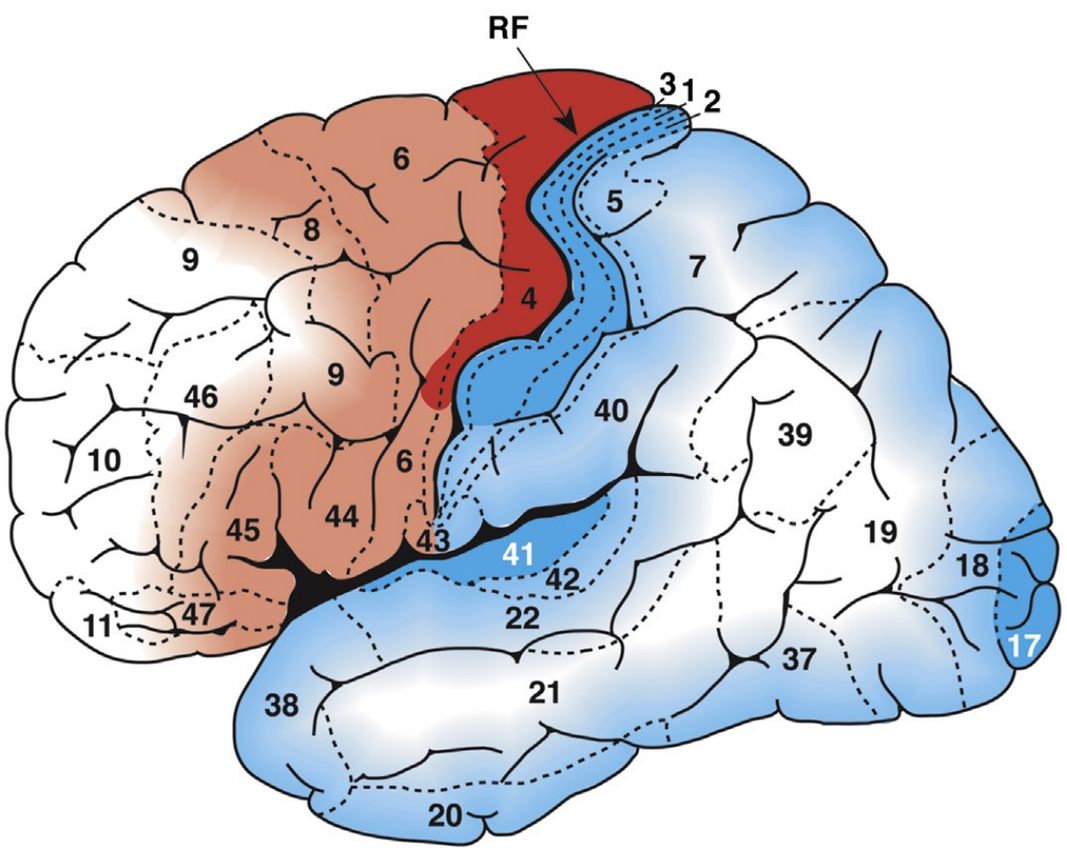




Koechlin, Ody, and Fredérique Kouneiher. The architecture of cognitive control in the human prefrontal cortex. *Science*, 302:1181-1185, 2003.

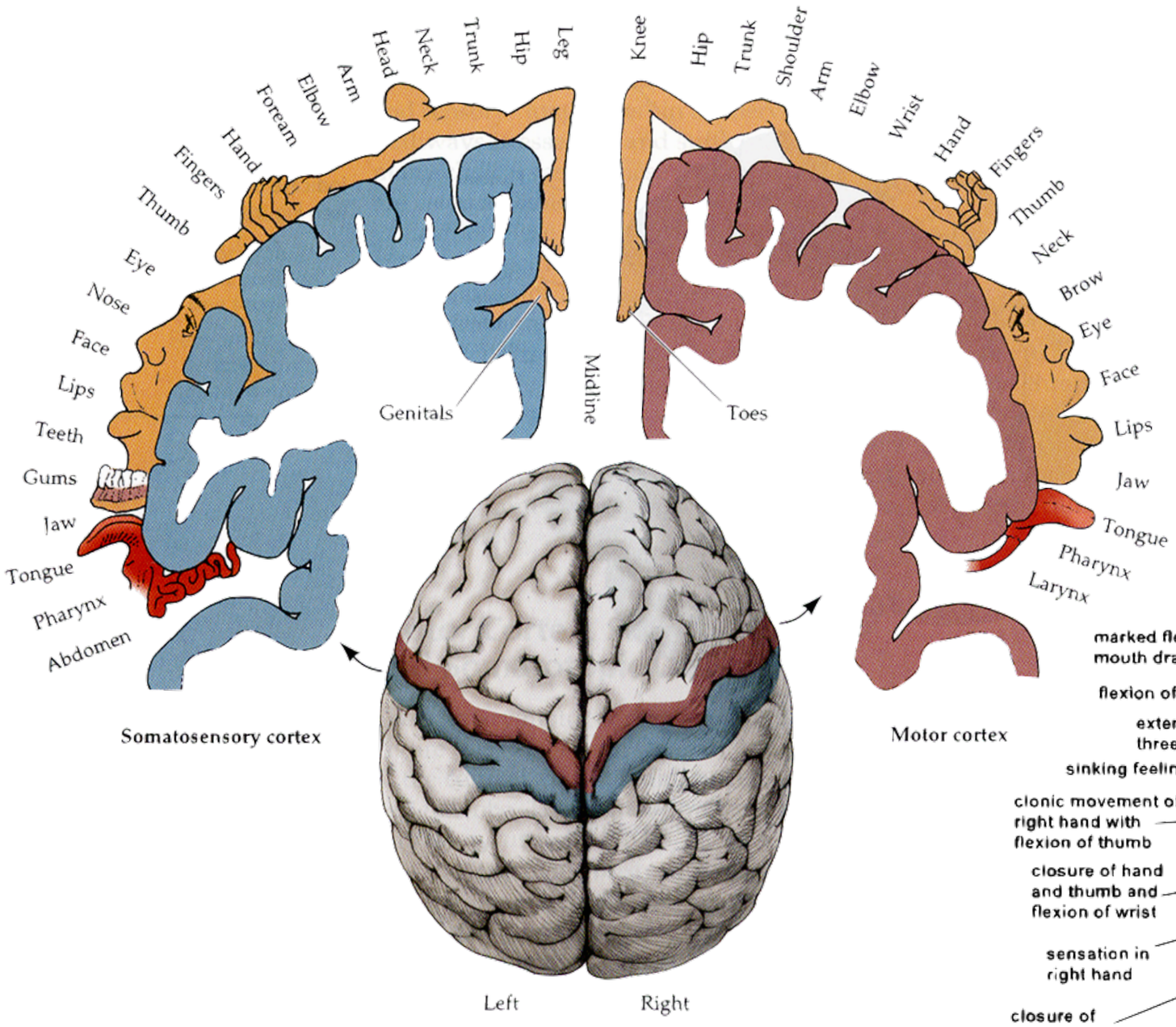
David Badre. Cognitive control, hierarchy, and rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, 12(5):193-200, 2008.

Hierarchical Organization of Memory Networks in Human Cortex



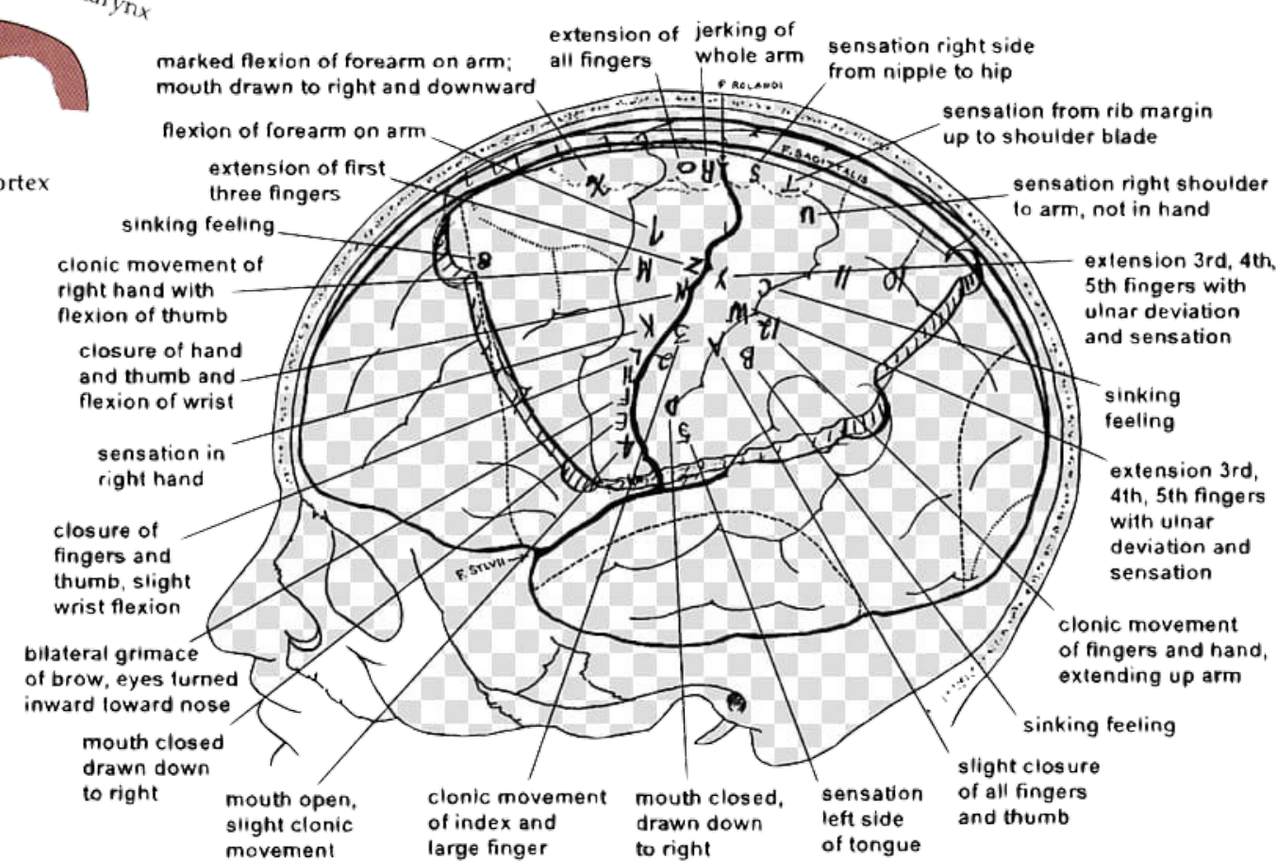
Joaquín M. Fuster. Chapter 8: An Overview of Prefrontal Functions, In *Prefrontal Cortex, 5th Edition*. pages 375-425. Elsevier, London, 2015.

Wilder Penfield's Motor & Somatosensory Homunculi

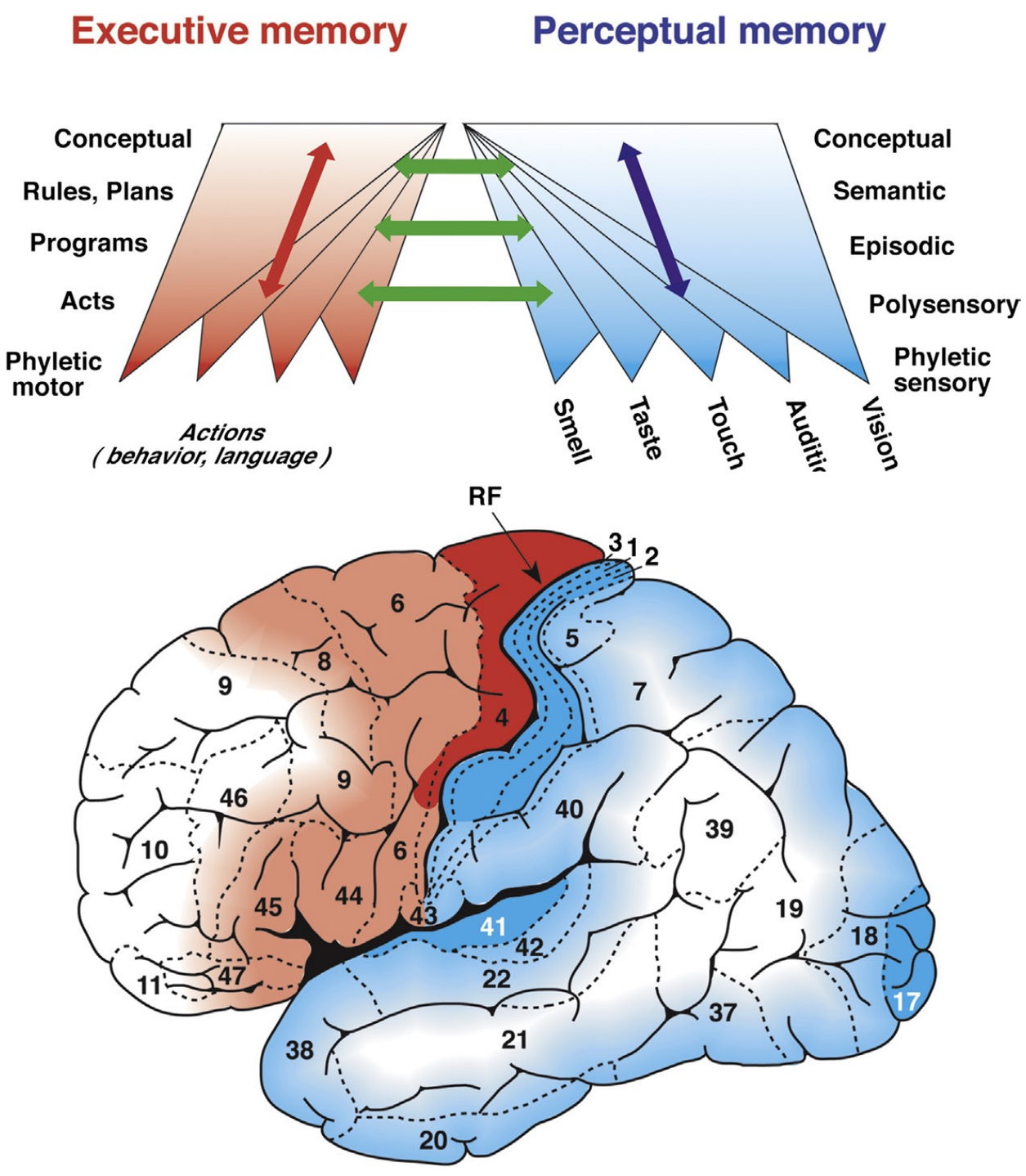


Anatomical / Topographical Registration

Distribution of Sensory & Motor Activity

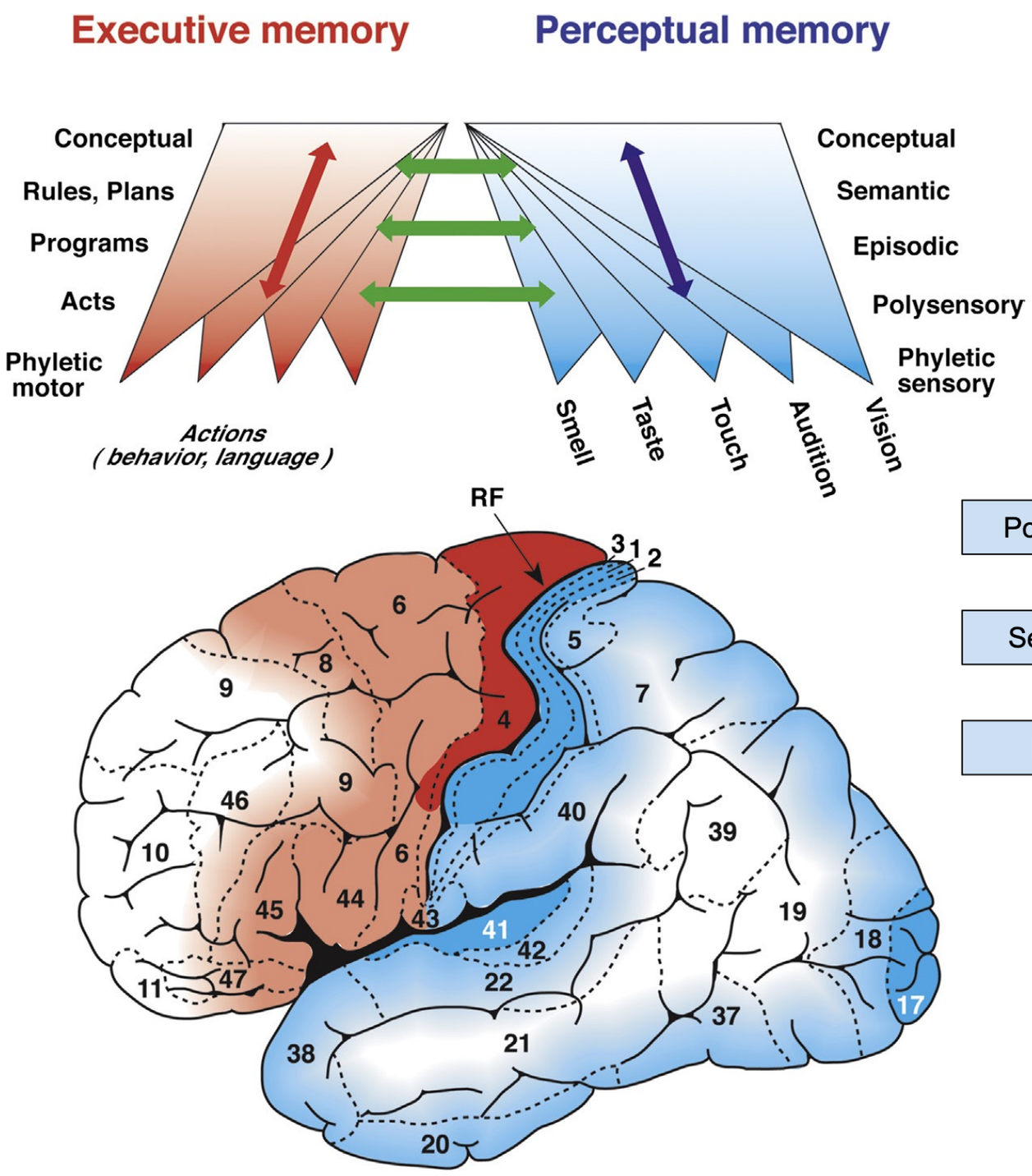


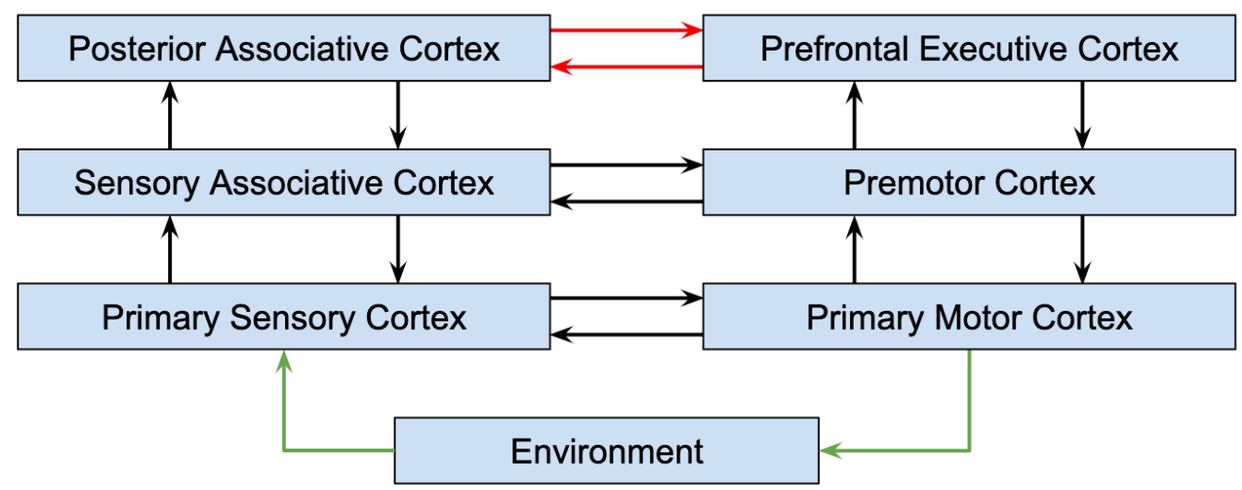
Hierarchical Organization of Memory Networks in Human Cortex



Joaquín M. Fuster. Chapter 8: An Overview of Prefrontal Functions, In *Prefrontal Cortex, 5th Edition*. pages 375-425. Elsevier, London, 2015.

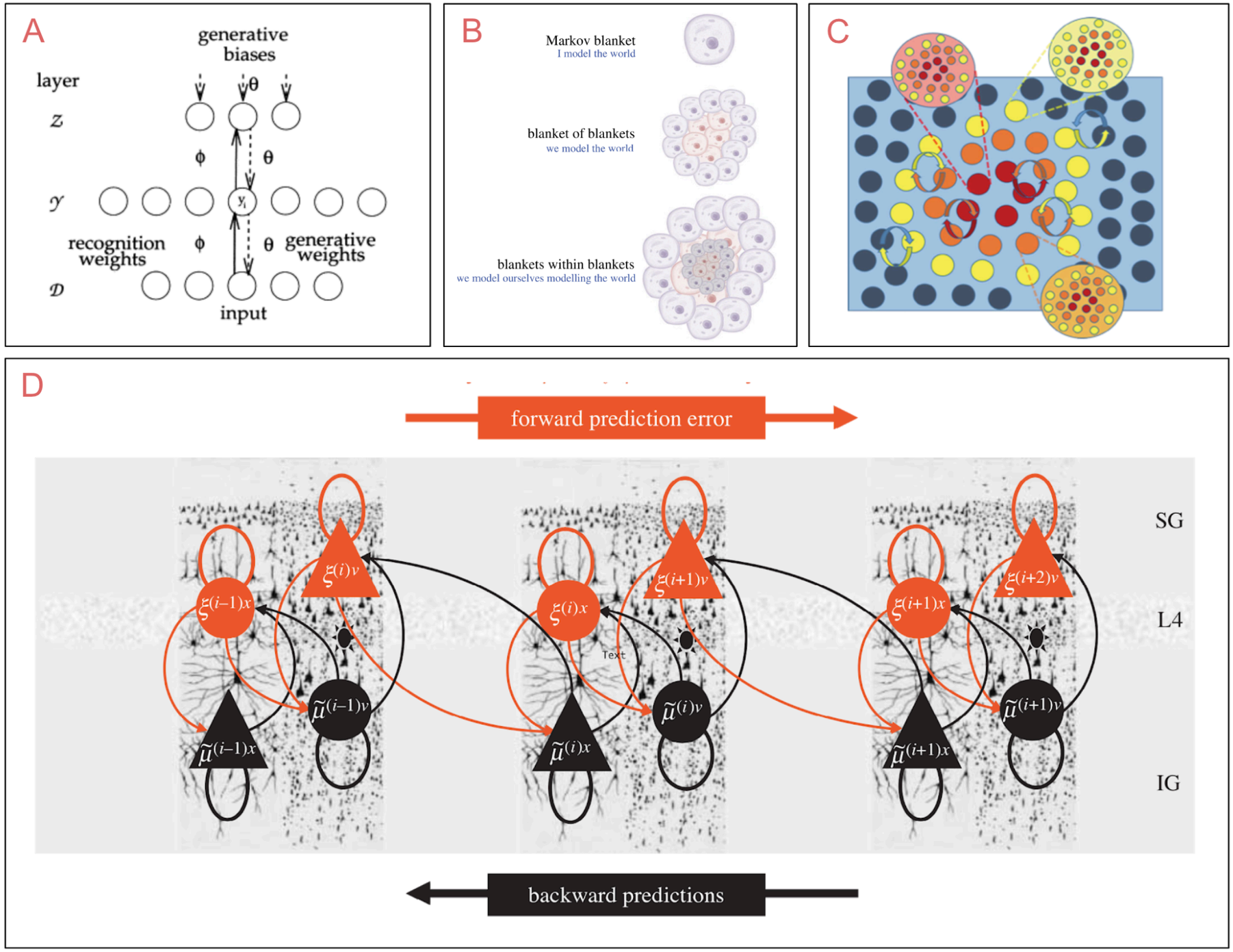
Hierarchical Organization of Memory Networks in Human Cortex





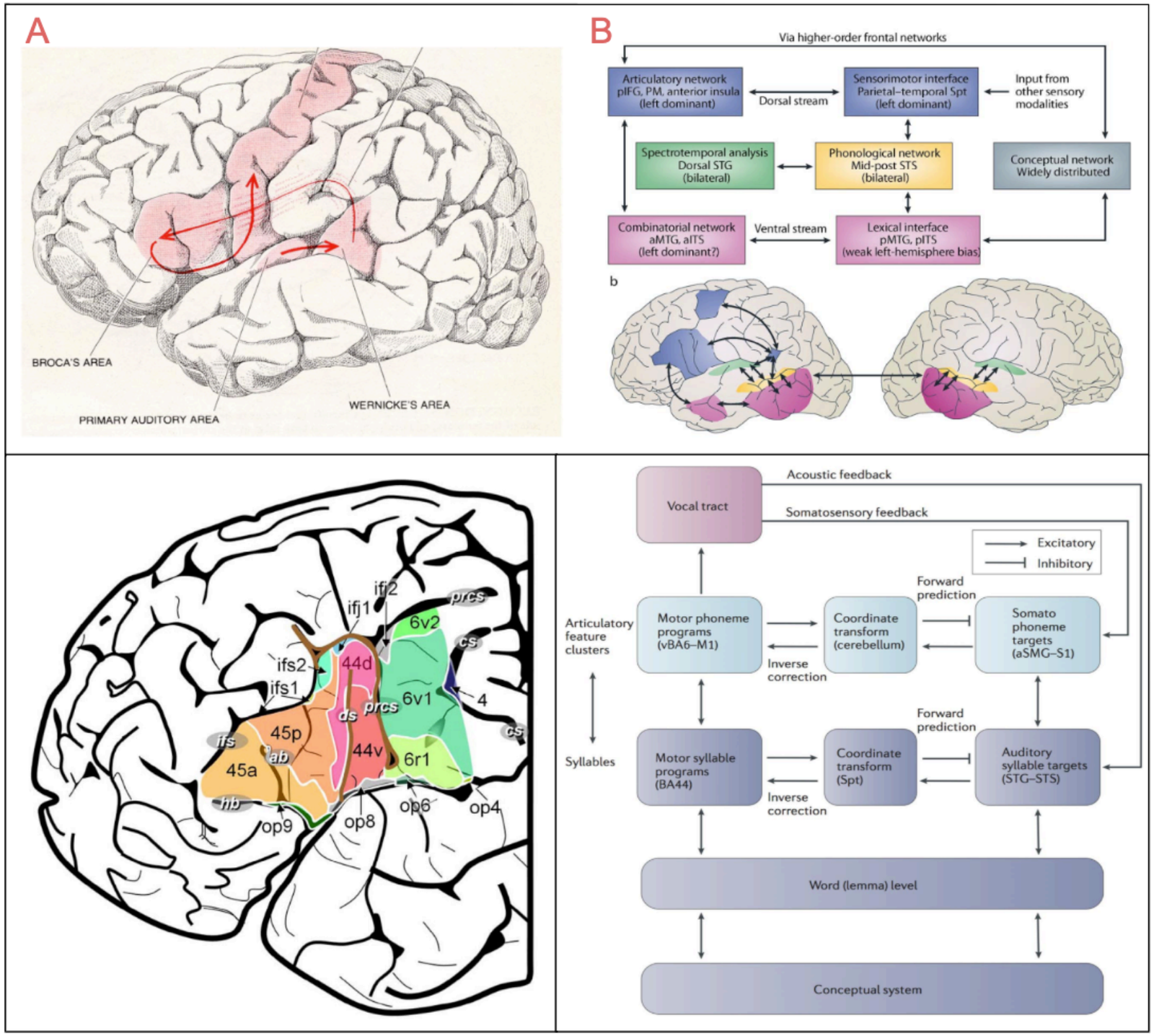
Joaquín M. Fuster. Chapter 8: An Overview of Prefrontal Functions, In *Prefrontal Cortex, 5th Edition*. pages 375-425. Elsevier, London, 2015.

Helmholtz Machines, Predictive Coding & Minimizing Free Energy



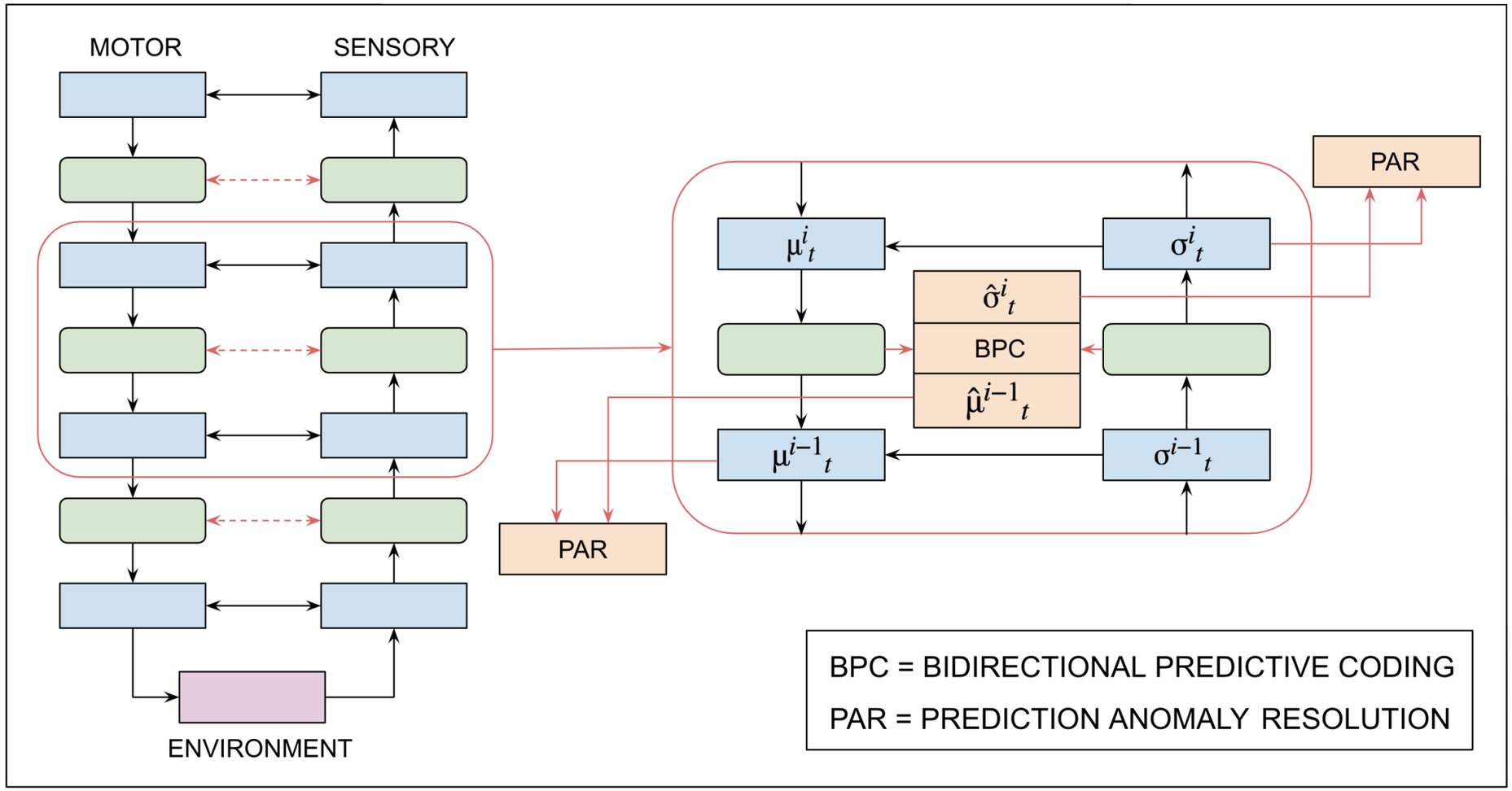
P. Dayan, G.E. Hinton, R.M. Neal, and R.S. Zemel. The Helmholtz machine. *Neural Computation*, 7:889-904, 1995.
Karl Friston and Stefan Kiebel. Predictive coding under the free-energy principle. *PT-RSL-B*, 364:1211-1221, 2009.
Karl Friston, Jean Daunizeau, and Stefan Kiebel. Reinforcement learning or active inference? *PloS ONE*, 4:6421, 2009.
Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex. *Nature Neuroscience*, 2:79-87, 1999.

The Classical & Dual Streams Model of Language Processing



Gregory Hickok and David Poeppel. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8:393, 2007.
David Poeppel, Karen Emmorey, Gregory Hickok, and Liina Pylkkanen. Towards a new neurobiology of language. *The Journal of Neuroscience: The official journal of the Society for Neuroscience*, 32:14125-14131, 2012.

Predictive Coding & The Perception-Action Cycle



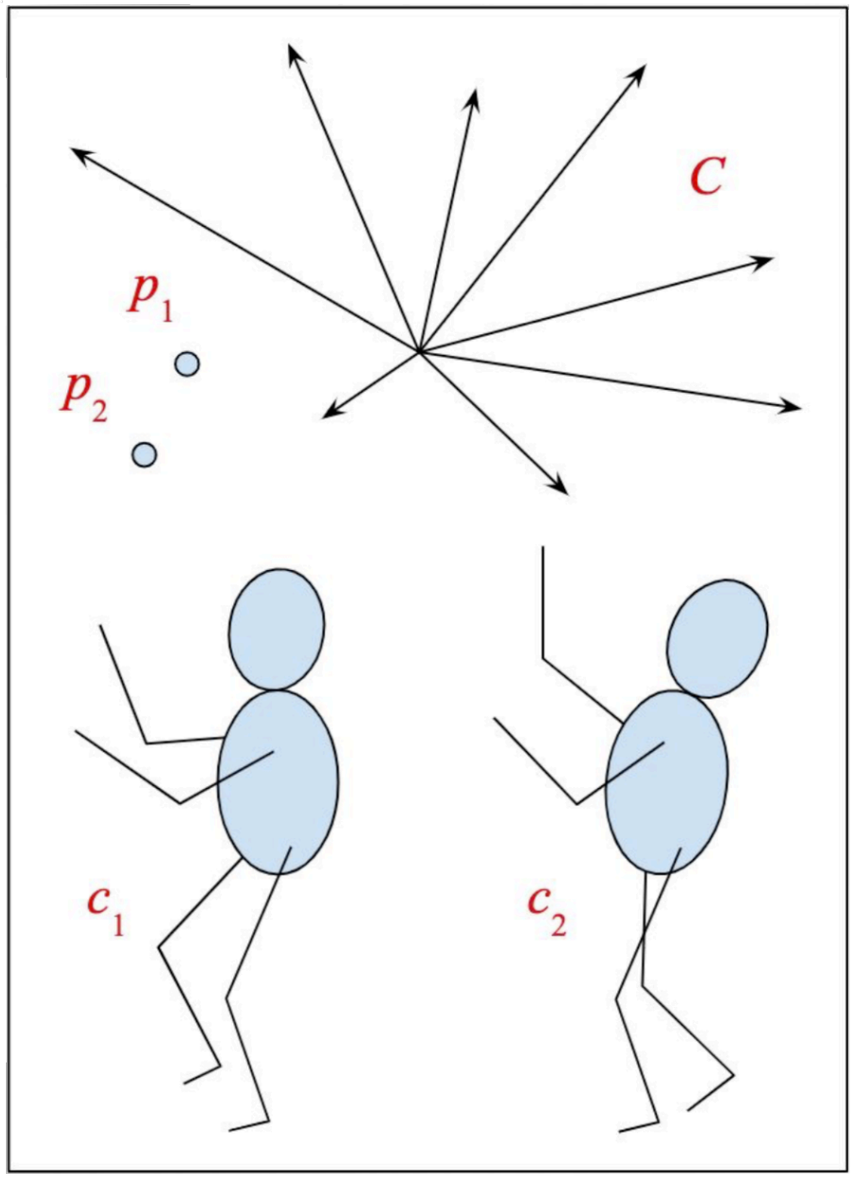
William H. Alexander and Joshua W. Brown. Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, 8:3843, 2018.

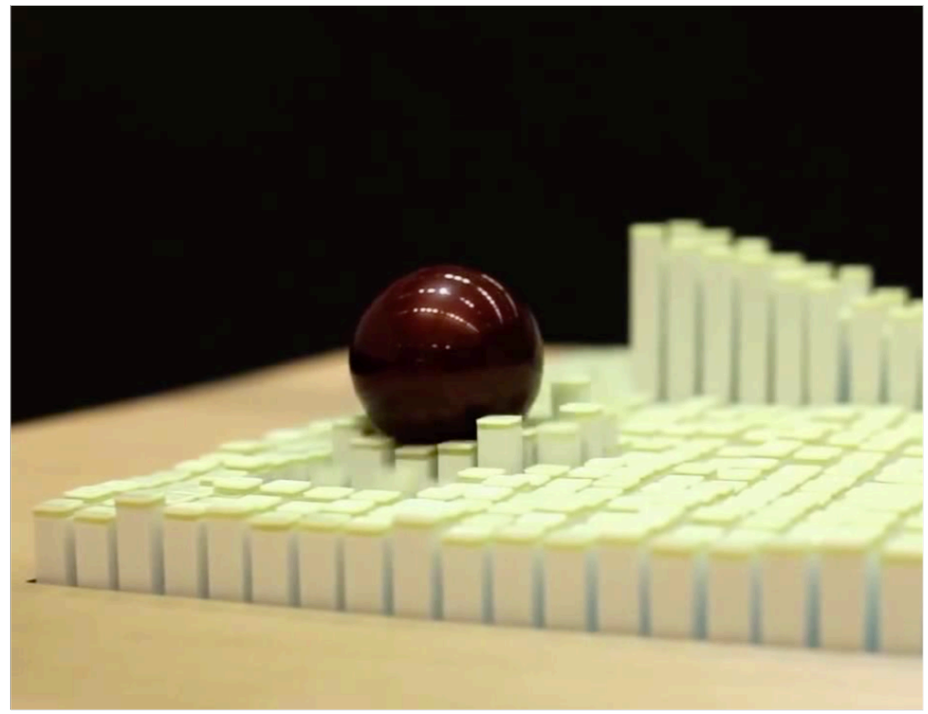
Stefan J. Kiebel, Jean Daunizeau, and Karl J. Friston. A hierarchy of time-scales and the brain. *PLOS Computational Biology*, 4(11):1-12, 2008.

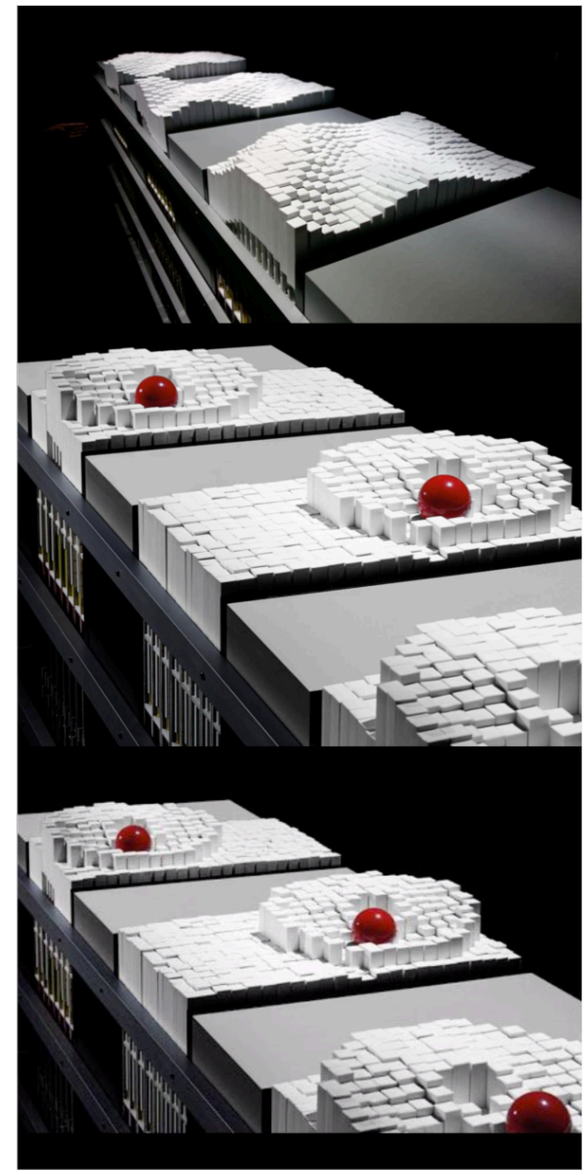
Rajesh P. N. Rao and Dana H. Ballard. Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, 9:721-763, 1996.

Kimberly L. Stachenfeld, Matthew M. Botvinick, and Samuel J. Gershman. The hippocampus as a predictive map. *Nature Neuroscience*, 20:1643, 2017.

Environmentally Grounded Embodied Cognition



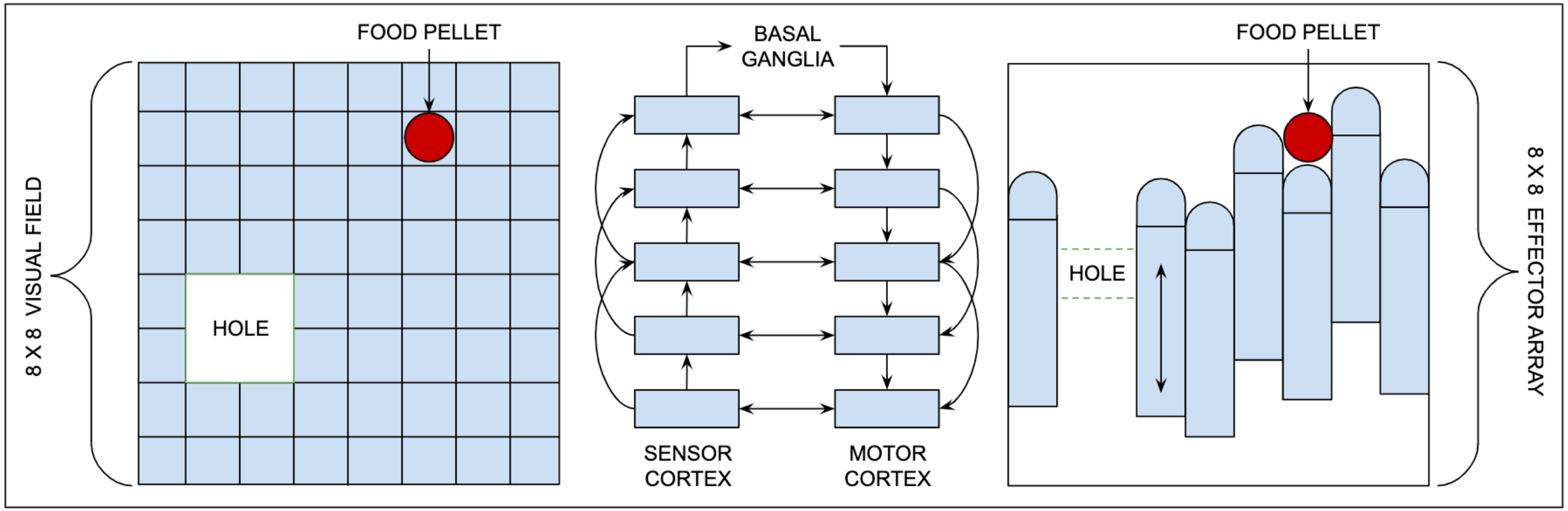




Daniel Leithinger, Sean Follmer, Alex Olwal, Hiroshi Ishii. Physical telepresence: Shape capture and display for embodied, computer-mediated remote collaboration. *Symposium on User Interface Software and Technology*, 2014.

Miguel Lázaro-Gredilla, Dianhuan Lin, J. Swaroop Guntupalli, and Dileep George. Beyond imitation: Zero-shot task transfer on robots by learning concepts as cognitive programs. *CoRR*, arXiv:1812.02788, 2018.

Environmentally Grounded Embodied Cognition



Supplementary Material

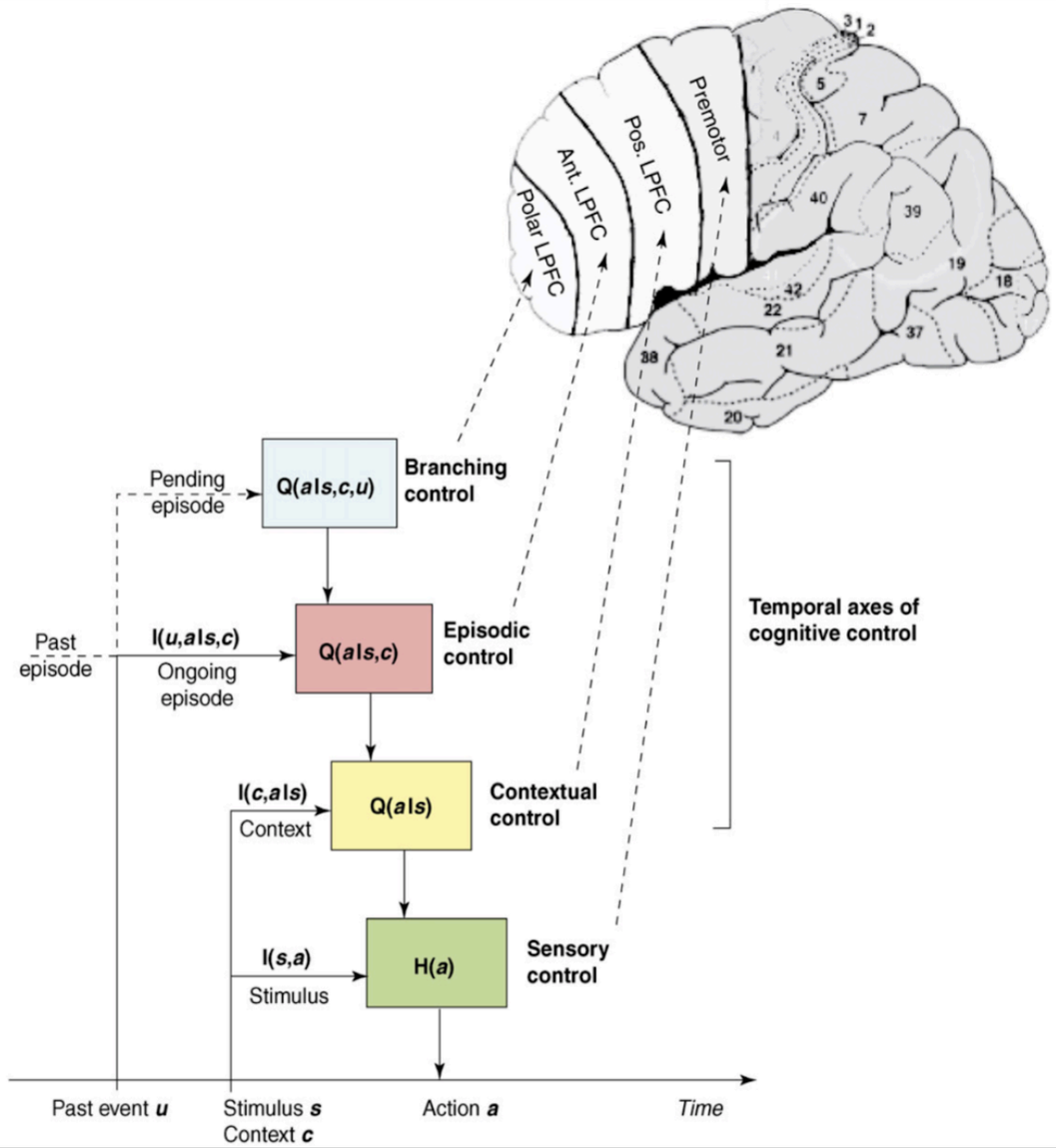
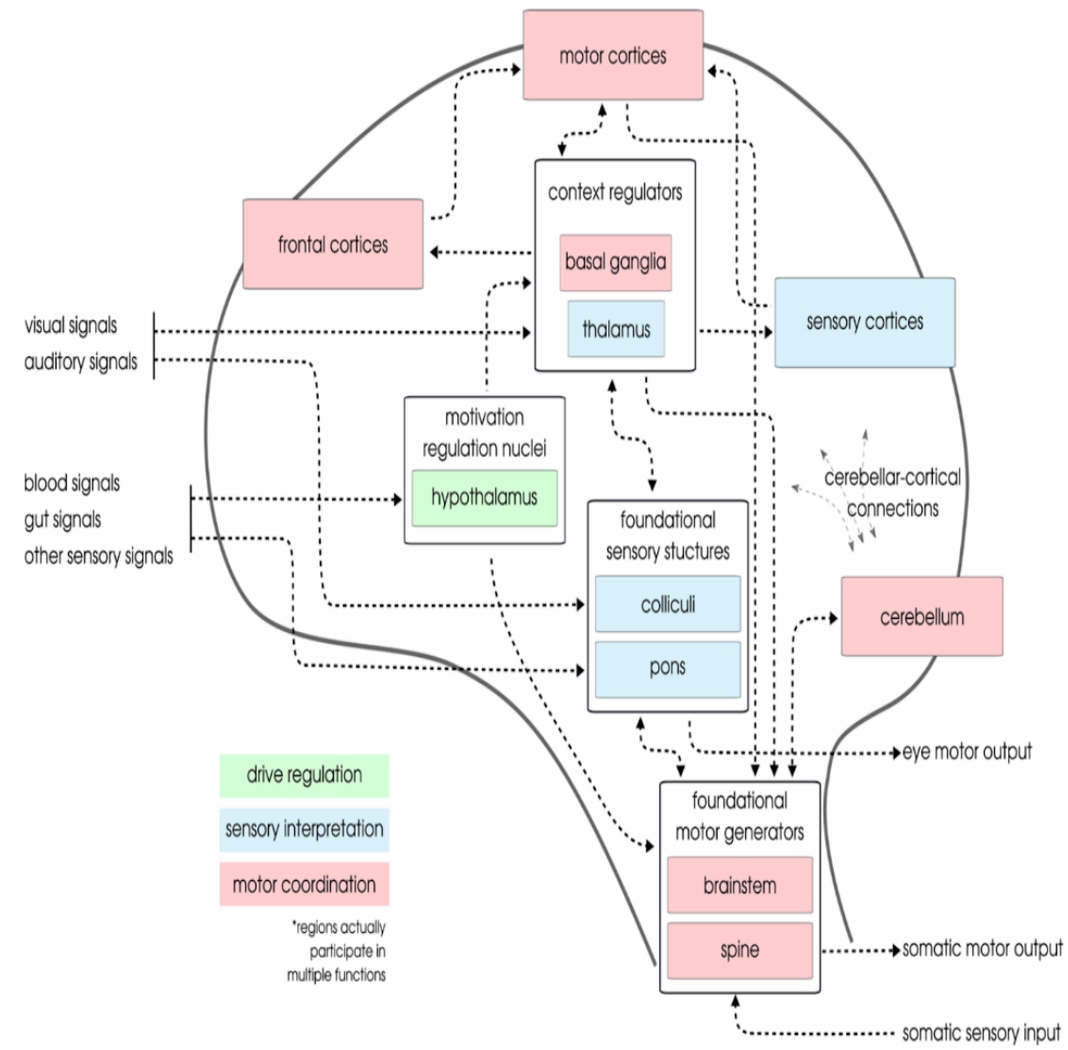


Figure 1. Model overview. Cognitive control operates according to three nested levels of control processes (contextual, episodic and branching) implemented from posterior to polar prefrontal regions. *H(a)* measures the total amount of control information required for selecting action A and processed in the premotor cortex. *H(a)* is the sum of two control terms: bottom-up information conveyed by a stimulus S [*I(s,a)*, sensorimotor control) and the remaining top-down information *Q(a|s)* processed in the posterior lateral PFC (Pos. LPFC), and measuring cognitive control. Cognitive control, in turn, is the sum of two control terms: bottom-up information conveyed by the context *c* in which stimulus *s* occurs (I(c,*a*|s), contextual control); and the top-down remaining information Q(*a*|s,c) processed in the anterior LPFC (Ant. LPFC). Finally, this latter control term is the sum of bottom-up information conveyed by a past event u (I(*u,a*)|s,c), episodic control) and the remaining top-down information processed in the polar LPFC (branching control). Branching control is related to the information conveyed by events preceding *u* and maintained in a pending state until completion of the ongoing episode.

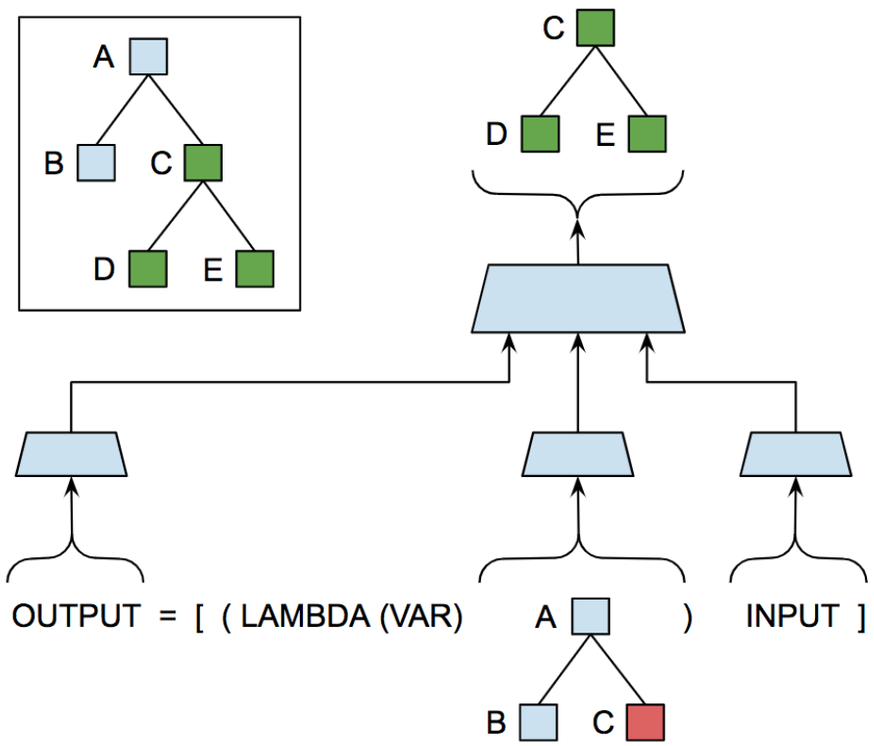


The diagram depicts an abstraction of the hierarchical anatomy of the mammalian nervous system. The scheme is, insofar as possible, a consensus view of previous hierarchical interpretations[3,4,6,69], with the intent of serving as an uncontroversial foundation. A natural entry point is the motivation regulation nuclei. The central nervous system receives information about the body via signals from the gut, level of hydration, hormones, blood sugar levels, and other measures. Much of this information arrives via structures such as the hypothalamus, which then communicates information related to motivational state to other parts of the brain. These signals related to basic drives (hunger, arousal, etc.) directly or indirectly will guide behavior. Subcortical structures, such as the basal ganglia, are responsible for regulating behavioral context and modulate the activity of more foundational motor generators in the brainstem and spine, which also receive limited sensory information via subcortical sensory structures. In parallel, motivational ("drive") information and sensory information are processed in cortical areas which in turn modulate behavioral context and ultimately allow for the use of more processed information to inform motor coordination via motor cortical areas.

A common motif across specific hierarchical models that have been proposed is the presence of multiple routes of information transmission and motor coordination. In terms of sensory input, dual sensory input pathways transmit information along a subcortical pathway as well as a cortical pathway[4]. Similarly, there are direct subcortical pathways from motivational centers (or what has been referred to as the limbic system) to brainstem nuclei that activate motor patterns, as well as indirect routes, either via the basal ganglia or through frontal cortices[3]. This multi-pathway motif structurally reflects some of hierarchical control principles, with multiple layers to the system being partially autonomous, each having access to partial and differently processed information.

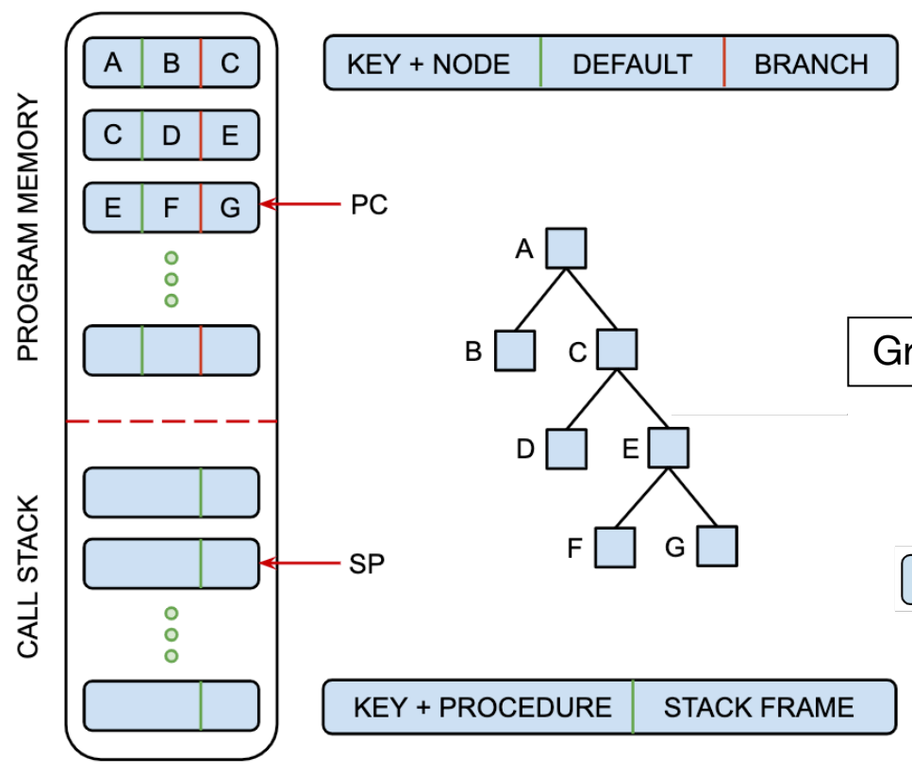
Etienne Koechlin and Christopher Summerfield. An information theoretical approach to prefrontal executive function. Trends in Cognitive Science, 11(6):229-235, 2007.
Josh Merel, Matthew Botvinick, and Gregory Wayne. Hierarchical motor control in mammals and machines. Nature Communications, 10(1):5489, December, 2019.

Differentiable Programming Tools from Standard Components



Differentiable Program Emulation

Differentiable Structured Programs



Graph Nets Imagination Coding