



Published in final edited form as:

Curr Opin Neurobiol. 2010 April ; 20(2): 257–261. doi:10.1016/j.conb.2010.01.008.

Computational Models of Cognitive Control

Randall C. O'Reilly, Seth A. Herd, and Wolfgang M. Pauli

Department of Psychology and Neuroscience, University of Colorado Boulder, 345 UCB, Boulder, CO 80309

Randall C. O'Reilly: randy.oreilly@colorado.edu

Abstract

Cognitive control refers to the ability to perform task-relevant processing in the face of other distractions or other forms of interference, in the absence of strong environmental support. It depends on the integrity of the prefrontal cortex and associated biological structures (e.g., the basal ganglia). Computational models have played an influential role in developing our understanding of this system, and we review current developments in three major areas: dynamic gating of prefrontal representations, hierarchies in the prefrontal cortex, and reward, motivation, and goal-related processing in prefrontal cortex. Models in these and other areas are advancing the field further forward.

Introduction

Computational models are important for making explicit links between biological mechanisms and the cognitive and behavioral phenomena that they produce. In the domain of research on cognitive control (e.g., the ability to perform task-relevant processing in the face of other distractions or absence of strong environmental support), there is a rich history of computational modeling that has served to focus empirical and other theoretical work on specific biological mechanisms and their functional roles. For example, early models showed how active maintenance of information in a working memory system could be accounted for in terms of reverberatory excitation among a set of interconnected neurons [1,2], and how this actively maintained information can provide a strong “top down bias” to influence processing throughout the brain [3,4,5,6]. These models resonated with a growing body of data from many methodologies, and associated theoretical ideas, to provide a coherent account of the involvement of the prefrontal cortex (PFC) in cognitive control [7,8,9]. This work has been highly influential in the field, and represents one of the most significant success stories for the contributions of the computational modeling approach.

Building upon this first wave of synthesis, a new generation of computational models have been extending the theoretical and computational framework to include multiple other systems that the PFC interacts with to achieve cognitive control, and elaborating the contributions that the PFC itself makes. Some particularly active areas of research include:

- Monitoring and feedback mechanisms that can regulate the application of PFC-mediated cognitive control, associated with the anterior cingulate cortex [10,11,12].

© 2010 Elsevier Ltd. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

- Dynamic gating mechanisms that can influence the updating and active maintenance of information in PFC, associated with the basal ganglia [13,14,15] and the neuromodulator dopamine [16,7].
- The hierarchical organization of PFC representations as a means of efficiently performing complex cognitive tasks [17,18,19].
- The interactions between PFC and subcortical areas involved in reward and motivation, to organize cognitive control around desired goals, and provide higher-level control and integration of reward and punishment signals [20,21,22,23].

We review some key ideas from the last three of these areas below.

Dynamic Gating of PFC Representations

A remarkable consensus across multiple different levels of analysis has converged on the general idea that an important element of cognitive control is a dynamic gating mechanism that can determine when PFC representations are updated to reflect new information, versus when they continue maintaining older information. From a purely computational perspective, the LSTM (Long Short-Term-Memory) model advanced this gating idea to provide a more robust active memory system [24]. From a biological perspective, the notion that the basal ganglia act as a gating system in the context of motor control has long been accepted [25,26,27], and this can be directly extended to account for a dynamic gating role in working memory [13,14]. Interestingly, these neurally-based ideas have converged with more top-down driven cognitive modeling work in the ACT-R framework, which postulates a gating-like role for the basal ganglia as well [15]. The ACT-R modeling work shows how basal ganglia gating can initiate the firing of cognitive productions, as in the classical production system models from symbolic AI. Thus, it seems that several major threads of computational modeling work are on the verge of a significant new integration [28,9].

Several recent models have explored the dynamically-gated working memory system in the context of rich mathematical frameworks such as bilinear mappings and partially-observable markov decision processes (POMDP), which provide important insights into the computational properties of these models [21,29,30]. The recent work of Dayan [21,29] is particularly interesting in attempting to bridge between basic gating-like working memory updating mechanisms and the ability of the system to perform arbitrary tasks rapidly through verbal instruction. Dayan draws nice connections between the ACT-R work and the habit versus goal-directed action framework in animal learning [31]. In this context, most existing neural models can be considered to be performing according to a complex habit that is inflexible and cognitively impenetrable, whereas much of human performance is more goal-directed or rule-governed (flexible and cognitively driven). Dayan shows how this rule-governed behavior can emerge from more basic gating-like mechanisms, though many difficult problems remain to be solved in this area.

At a more cognitive level, the work of Ashby and colleagues has shown how a PFC/BG system can explain important findings in the categorization literature, adding further empirical avenues for testing such models [32]. More generally, the empirical data on biological and cognitive neuroscience studies of this PFC/BG system is growing rapidly, and providing considerable support for the central ideas behind these computational models [33,34,35,36]. Thus, this area of research and modeling represents another important emerging success story for the computational modeling approach.

One outstanding question in this area is the relative role of dopamine as a gating signal in the basal ganglia and PFC. Some models focus on the role of dopamine in training a selective gating signal in PFC that is capable of gating new information into some regions of PFC, while

leaving others to actively maintain older information [9,37]. Other models focus on the role of a global dopamine signal in PFC for gating [16,38]. Future work should examine the interplay between a global dopamine signal, and more focal gating signals from BG.

Hierarchies in PFC

Models of behavioral control have long involved hierarchies [39]. Only recently, however, have hierarchical models made explicit contact with relevant neuroscientific data, which is accumulating quickly [40,18,41]. Convergent findings suggest a hierarchy organized along the posterior-anterior anatomical axis of the PFC, with more anterior areas providing higher-level control representations, consistent with the original ideas of Fuster [42]. However, the best way to characterize this hierarchy is still a matter of considerable debate, and computational models are providing some important insights.

Botvinick [17] showed how learning hierarchically structured tasks in a network with connectivity similar to PFC can give rise to a semi-hierarchical organizational structure. The shortest route from stimulus to response in the model was directly from stimuli to premotor areas, with increasingly more indirect routes available through more anterior PFC areas (consistent with anatomical data [40,42]). When this network learned several complex behavioral tasks, such as making tea or coffee, representations of more temporally protracted task elements developed in higher-level layers. Critically, however, the model was not strictly hierarchical, and thus can capture important subtleties in human performance, such as the context sensitivity of subordinate actions.

Biologically detailed models of PFC active maintenance with basal ganglia gating mechanisms have also been shown to develop hierarchical task representations [19], in the context of the hierarchically-structured 12-AX task [14]. Two architectural variations of the model were used. In the first, a posterior region of PFC received input directly from sensory areas, and projected to motor areas, while an anterior PFC region received only from posterior PFC, similar to the architecture used by [17]. The second architectural manipulation made the BG connectivity hierarchical, with anterior PFC driving posterior BG gating, but not vice-versa. This manipulation is consistent with findings that PFC/BG loops exhibit a spiral-like structure [43,44]. Both manipulations caused specialization for more temporally extended (“outer loop”) representations in anterior PFC, with additive effects of each.

These studies show how initial connectivity, combined with learning, can explain some aspects of a hierarchical organization in PFC, within an overall framework of common underlying mechanisms across the PFC. However, the empirical literature is unclear about the precise nature of the PFC hierarchical structure: e.g., is it about rule complexity and relational integration [45,46,18], or is it more about abstract representations [47,48,49], or both? One intriguing prospect for future research is that basic differences in duration of active maintenance across PFC areas (e.g., due to more anterior areas being further removed from sensory perturbations) may interact with learning mechanisms to naturally produce a gradient of abstraction [9,50].

Reward, Motivation, and Goals

The PFC is considered the “executive” of the brain for multiple reasons, among them that it is strongly concerned with the “bottom line” – satisfying the basic needs and goals of the organism. This occurs through bidirectional interactions with subcortical and other “limbic” brain areas that are known to be involved in processing basic affective signals such as reward, punishment, fear, etc. The ventral and medial areas of the PFC are most directly involved in these interactions, and considerable attention has been focused on the roles of the orbital prefrontal cortex (OFC), and the anterior cingulate cortex (ACC).

The OFC is strongly interconnected with the basolateral amygdala (BLA), which is known to be important for processing a range of basic emotions. From a computational perspective, the OFC may play a similar role in this system as PFC does more generally: active maintenance and top down biasing, but in this case maintenance of reward-related information, and biasing of function in the BLA. Also, because subcortical areas may have relatively slow synaptic learning, rapid working memory updates in the OFC may be required to keep track of recent changes in outcome probabilities and magnitudes [51,22,52]. Models implementing this idea have been able to account for a range of different data [22,52]. For example, patients with OFC damage have been shown to exhibit poor decision making in gambling tasks, which seems to be due to an underlying difficulty in integrating reward magnitude with its probability, and/or reversing initial positive reward associations [53]. The OFC model of Frank & Claus [22] was able to account for this data in terms of a reduced active maintenance system that otherwise facilitates reversal learning of emotional associations, in a manner consistent with other such reversal learning models [49,54], and OFC lesions in rats [55,52].

Although the anterior cingulate cortex (ACC) is often thought of as a conflict monitoring area [12], a growing body of data suggests that this may be a subset of a more general computational function as an OFC-like area that is concerned with associations between motor actions and reward/punishment outcomes [56] (whereas OFC is concerned with associations between stimuli and reward/punishment outcomes). One recent model of ACC suggests that it arbitrates between model-based and model-free responses encoded in the dorsomedial vs. dorsolateral striatum, respectively [20,21]. Model-based responding is goal directed, and is sensitive to manipulations such as devaluation of the unconditioned stimulus, whereas model-free responding is more habitual and insensitive to outcome-related manipulations [57]. A neural network model of instrumental and Pavlovian conditioning shows how widely-accepted biological learning mechanisms can result in a shift from dorsomedial goal-directed behaviors early in skill acquisition to dorsolateral stimulus-response solutions during consolidation of a skill [23].

More work is needed in this area to explore how different types of affective and goal-related signals are encoded and processed in different PFC areas — the ventral and medial areas of PFC are extensive, and current research has focused somewhat narrowly on reward-related processing within a restricted range of behavioral paradigms.

Conclusions

In effect, computational models act like very precise theories, working along with other theoretical ideas at various levels of description. A crucial benefit of computational models as theoretical tools is that they enable simulation of complex cognitive processes, and principled generation of novel hypotheses for testing. The models reviewed above, covering a wide range of different levels of abstraction, contribute important insights and predictions that are being actively tested. This new body of data is then informing new versions of the models, perpetuating the classic cycle of theory development, testing, and revision that underlies all good science.

In addition to the domains mentioned above that have received significant coverage in the literature, we suggest that there are several important areas that are ripe for future exploration, including models exploring the interactions between hippocampal episodic memory and PFC working memory [58,59,60], and more work examining the nature of learning and rapid vs. slow adaptation of PFC representations over time — the prevalent focus on active maintenance as the main form of memory in the PFC may be causing us to miss out on a range of other important memory mechanisms that may have important explanatory roles in a wide range of

cognitive control phenomena, especially as one considers the developmental spectrum [61, 62].

Acknowledgments

Supported by ONR grant N00014-07-1-0651 and NIH grant MH079485.

References

1. Zipser D. Recurrent network model of the neural mechanism of short-term active memory. *Neural Computation* 1991;3:179–193.
2. Dehaene S, Changeux JP. A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience* 1989;1:244–261.
3. Cohen JD, Dunbar K, McClelland JL. On the control of automatic processes: A parallel distributed processing model of the stroop effect. *Psychological Review* 1990;97(3):332–361. [PubMed: 2200075]
4. Munakata Y. Perseverative reaching in infancy: The roles of hidden toys and motor history in the aoverlineb task. *Infant Behavior and Development* 1997;20(3):405–416.
5. Stedron JM, Sahni SD, Munakata Y. Common mechanisms for working memory and attention: the case of perseveration with visible solutions. *Journal of cognitive neuroscience* 2005;17:623–631. [PubMed: 15829082]
6. Herd SA, Banich MT, O'Reilly RC. Neural mechanisms of cognitive control: an integrative model of stroop task performance and fMRI data. *Journal of cognitive neuroscience* 2006;18:22–32. [PubMed: 16417680]
7. O'Reilly, RC.; Braver, TS.; Cohen, JD. A biologically based computational model of working memory. In: Miyake, A.; Shah, P., editors. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. New York: Cambridge University Press; 1999. p. 375-411.
8. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annual review of neuroscience*. 24
9. O'Reilly RC. Biologically based computational models of high-level cognition. *Science (New York, N.Y.)* 2006;314:91–94.
10. Cohen JD, Braver TS, O'Reilly RC. A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 1997;351:1515–1527.
11. Brown JW, Braver TS. Learned predictions of error likelihood in the anterior cingulate cortex. *Science (New York, N.Y.)* 2005;307(5712):1118–1121.
12. Botvinick MM, Cohen JD, Carter CS. Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences* 2004;8(12):539–546. [PubMed: 15556023]
13. Frank MJ, Loughry B, O'Reilly RC. Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience* 2001;1:137–160.
14. O'Reilly RC, Frank MJ. Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation* 2006;18:283–328. [PubMed: 16378516]
15. Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y. An integrated theory of the mind. *Psychological Review* 2004;111(4):1036–1060. [PubMed: 15482072]
16. Cohen JD, Braver TS, Brown JW. Computational perspectives on dopamine function in prefrontal cortex. *Current opinion in neurobiology* 2002;12:223–229. [PubMed: 12015241]
17. Botvinick MM. Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, affective & behavioral neuroscience* 2008;7:356–366.
18. Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences* 2007;11(6):229–235. [PubMed: 17475536] Proposes a simple mathematical framework classifying decision complexity, and proposes that increasingly anterior areas of PFC are brought into play for more complex decisions.

19. Reynolds JR, O'Reilly RC. Developing PFC representations using reinforcement learning. *Cognition*. 2009
20. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience* 2005;8(12):1704–1711.
21. Dayan P. Simple substrates for complex cognition. *Frontiers in neuroscience* 2008;2(2):255. [PubMed: 19225599] Synthesizes numerous different threads of computational modeling to understand the difference between rule-driven vs. habitual task performance, which is a critical issue that computational models are currently tackling.
22. Frank MJ, Claus ED. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological review* 2006;113(2):300–326. [PubMed: 16637763]
23. Pauli, WM.; Atallah, HE.; O'Reilly, RC. Integrating what & how/where with instrumental and pavlovian learning: A biologically-based computational model. *Proceedings of the XXIX, International Congress of Psychology*; (in press)
24. Hochreiter S, Schmidhuber J. Long short term memory. *Neural Computation* 1997;9:1735–1780. [PubMed: 9377276]
25. Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience* 1986;9:357–381.
26. Middleton FA, Strick PL. Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain and cognition* 2000;42:183–200. [PubMed: 10744919]
27. Redgrave P, Prescott TJ, Gurney K. The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience* 1999;89:1009. [PubMed: 10362291]
28. Jilk D, Lebiere C, O'Reilly R, Anderson J. SAL: an explicitly pluralistic cognitive architecture. *Journal of Experimental & Theoretical Artificial Intelligence* 2008:197–218. Philosophically oriented treatment of levels of analysis issues in computational modeling, arguing for a pluralistic approach, and description of an attempt to develop models that synthesize across these different levels.
29. Dayan P. Bilinearity, rules, and prefrontal cortex. *Frontiers in computational neuroscience* 1(1)
30. Todd, MT.; Niv, Y.; Cohen, JD. Learning to use working memory in partially observable environments through dopaminergic reinforcement. In: Koller, D., editor. *Advances in Neural Information Processing Systems (NIPS)* 21. Curran Associates; 2008.
31. Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 1999;37:407–419. [PubMed: 9704982]
32. Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychological review* 1998;105:442. [PubMed: 9697427]
33. Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* 2004;306:1940–1943. [PubMed: 15528409]
34. Frank MJ, Samanta J, Moustafa AA, Sherman SJ. Hold your horses: impulsivity, deep brain stimulation, and medication in Parkinsonism. *Science (New York, N.Y.)* 2007;318:1309–1312. Tests important predictions from computational models of the basal ganglia, with practical implications for Parkinson's patients.
35. Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M. Genetically determined differences in learning from errors. *Science (New York, N.Y.)* 2007;318(5856):1642–1645.
36. Moustafa AA, Cohen MX, Sherman SJ, Frank MJ. A role for dopamine in temporal decision making and reward maximization in parkinsonism. *Journal of Neuroscience* 2008;28(47):12294–12304. [PubMed: 19020023]
37. Frank MJ, O'Reilly RC. A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience* 2006;120:497–517. [PubMed: 16768602]
38. Durstewitz D, Seamans JK. The computational role of dopamine d1 receptors in working memory. *Neural networks : the official journal of the International Neural Network Society* 2002;15:561–572. [PubMed: 12371512]
39. Rumelhart DE, Norman DA. Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science* 1982;6(1):1–36.

40. David B, Mark D. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature reviews* 2009;10(9):659–669. 10.1038/nrn2667.
41. Botvinick M, Niv Y, Barto AC. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*. 2009 Nice review of computational modeling work on hierarchical representations and organization of behavior.
42. Fuster, JM. *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe*. 3rd Edition. New York: Lippincott-Raven; 1997.
43. Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *The Journal of Neuroscience* 2000;20:2369–2382. [PubMed: 10704511]
44. Haber SN, Calzavara R. The cortico-basal ganglia integrative network: The role of the thalamus. *Brain research bulletin* 78
45. Bunge SA, Helskog EH, Wendelken C. Left, but not right, rostromedial prefrontal cortex meets a stringent test of the relational integration hypothesis. *NeuroImage* 2009;46(1):338–342. [PubMed: 19457362]
46. Christoff K, Gabrieli JDE. The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology* 2000;28:168–186.
47. Badre D, D'Esposito M. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of cognitive neuroscience* 19
48. Christoff K, Keramatian K, Gordon AM, Smith R, Madler B. Prefrontal organization of cognitive control according to levels of abstraction. *Brain research* 2009;1286:94–105. [PubMed: 19505444] Intriguing data showing hierarchical organization in PFC according to abstraction, even at the level of word semantics.
49. O'Reilly RC, Noelle DC, Braver TS, Cohen JD. Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cerebral cortex (New York, N.Y. : 1991)* 2002;12:246–257.
50. Rougier NP, Noelle D, Braver TS, Cohen JD, O'Reilly RC. Prefrontal cortex and the flexibility of cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences* 2005;102(20):7338–7343.
51. Holland PC, Gallagher M. Amygdala-frontal interactions and reward expectancy. *Current opinion in neurobiology* 2004;14(2):148–155. [PubMed: 15082318]
52. Pauli, WM.; Hazy, TE.; O'Reilly, RC. Division of labor among multiple parallel cortico - basal ganglia - thalamic loops in pavlovian and instrumental tasks: A biologically-based computational model. Poster presented at the Multidisciplinary Symposium on Reinforcement Learning;
53. Fellows LK, Farah MJ. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral cortex (New York, N.Y. : 1991)* 2004;15(1):58–63.
54. Dehaene S, Changeux JP. The Wisconsin card sorting test: theoretical analysis and modeling in a neuronal network. *Cerebral cortex (New York, N.Y. : 1991)* 1992;1:62–79.
55. Stalnaker TA, Franz TM, Singh T, Schoenbaum G. Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron* 2007;54
56. Rushworth MFS, Behrens TEJ, Rudebeck PH, Walton ME. Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in cognitive sciences* 2007;11(4):168–176. [PubMed: 17337237] Suggests a clear distinction between anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC) in terms of action vs. stimulus associations (respectively) with outcome information.
57. Yin HH, Mulcare SP, Hilario MRF, Clouse E, Holloway T, Davis MI, Hansson AC, Lovinger DM, Costa RM. Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature neuroscience* 2009;12
58. Reynolds, JR.; Braver, TS. Computational and neural mechanisms in task-switching. Poster presented at the annual Cognitive Neuroscience Society Conference; San Francisco, California. 2002.
59. Ranganath C, Cohen MX, Dam C, D'Esposito M. Inferior temporal, prefrontal, an hippocampal contributions to visual working memory maintenance and associative memory retrieval. *Journal of Neuroscience* 2004;24(16):3917–3943. [PubMed: 15102907]

60. Hasselmo ME, Stern CE. Mechanisms underlying working memory for novel information. *Trends in cognitive sciences* 10
61. Chatham CH, Frank MJ, Munakata Y. Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106(14):5529–5533. [PubMed: 19321427] Shows that young children use a qualitatively different (reactive) strategy for cognitive control compared to adults.
62. Paxton JL, Barch DM, Racine CA, Braver TS. Cognitive control, goal maintenance, and prefrontal function in healthy aging. *Cerebral cortex* (New York, N.Y. : 1991) 18