10

Computational Models of Prefrontal Cortex

Two Complementary Approaches

Etienne Koechlin and Xiao-Jing Wang[1]

Abstract

The frontal lobe cortex is among the brain regions that evolve the most across mammals. In rodents, the prefrontal cortex (PFC) comprises the orbitofrontal cortex, the anterior cingulate complex (ACC), as well as the prelimbic and infralimbic areas in the medial wall. In primates, the PFC has evolved with the addition of the lateral PFC. In humans, the PFC features the further development of its most anterior part, especially in the lateral sector, and is often named the frontopolar cortex. Human patients with PFC lesions exhibit little impairments in basic sensorimotor, memory, learning, and language functions. Thus, the PFC function fulfills additional, more abstract functional demands. Its characterization has long remained elusive through the use of poorly defined notions such as executive/cognitive control, working memory, or cognitive flexibility. Here, computational models are shown to overcome these theoretical shortcomings by providing more precise accounts, predictions, and simulations of PFC function at the neuronal and behavioral levels. Two approaches have been developed in neurobiology and cognitive neuroscience, respectively. Time is ripe to integrate the two for a cross-level understanding of PFC function.

Introduction

Computational approaches of prefrontal cortex (PFC) function may start from a simple postulate: PFC function has evolved to enhance animal adaptive behavior. From that respect, computational models of PFC function should address two key overarching issues: (a) how PFC basic cognitive operations emerge from the neural networks that have evolved in the PFC and (b) which

[1] Alphabetical listing: both authors contributed equally to this work.

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

are the key functional limitations of basic adaptive processes external to the PFC that PFC processes overcome in the service of enhanced adaptive behavior. The first issue is addressed through neural network models of PFC operations, primarily based on neurophysiology of single cells from animals performing a task and neural circuit dissection. The second issue is addressed through computational cognitive models of PFC function, primarily informed by behavior and brain imaging like fMRI, in the tradition of cognitive psychology. The interactions between these perspectives are necessary to achieve an understanding of the PFC (Miller and Cohen 2001; Wang 2013). Because PFC is implicated in many psychiatric disorders, progress in this area has spurred translational research that gave rise to the nascent field of computational psychiatry (Wang and Krystal 2014).

Neural Network Models of PFC Operations

Fundamental Cognitive Processes

Biologically based neural circuit modeling strives to build mathematical models across levels, from molecules and cell types to collective neural circuit dynamics to functions. In frontal cortex research, this approach was initially developed for working memory, the brain's ability to internally hold and manipulate information that is essential to enable mental processes separate from direct sensory stimulation. Working memory is commonly studied in the laboratory using delay-dependent tasks, where information about a sensory stimulus must be held internally across a delay period to guide a behavioral response later. Since the discovery of stimulus-selective persistent neural activity during a mnemonic delay period (Fuster and Alexander 1971), its circuit mechanism was investigated experimentally by Patricia Goldman-Rakic (1995) and others, as well as through computational models (Amit and Brunel 1997; Brunel and Wang 2001; Compte et al. 2000a). The main idea is that working memory in the absence of any external input can be actively sustained by recurrent synaptic excitation. Modeling work found that recurrent excitation must be slow and depend on NMDA receptors (Wang 1999), a theoretical prediction that was supported by monkey experiments (Figure 10.1a) (Wang et al. 2013). Thus, slow reverberation is now considered as a characteristic of PFC. This finding is of clinical interest, because NMDA receptor hypofunction is implicated in PFC deficits associated with schizophrenia (Coyle et al. 2003).

In the cortex, excitation is balanced by inhibition, which is mediated by multiple subtypes of GABAergic cells. Motivated by the need for a working memory system to "gate out" behaviorally irrelevant stimuli, Wang et al. (2004c) proposed a disinhibitory motif (Figure 10.1b) composed of three interneuron subclasses. While parvalbumin-positive interneurons control spiking output of pyramidal neurons, interneurons that express somatostatin or

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

calbindin target dendrites are well positioned to gate inputs to pyramidal cells. When pyramidal cells are inhibited by interneurons that express calretinin or vasoactive intestinal peptide, the "gate" would be open, allowing for inputs to enter the circuit. This theoretically predicted disinhibitory motif has now been well-established experimentally (Tremblay et al. 2016). It is noteworthy that, compared to primary sensory areas, the ratio of input-controlling and output-controlling interneurons is much higher in the PFC, presumably tailored to its functional requirements (Wang 2020).

Furthermore, the recurrent neural circuit model initially proposed for working memory turned out to be suitable to account for key computational processes in decision making, which depends on the PFC, posterior parietal cortex and other associated brain regions. Experiments revealed that quasi-linear ramping neural activity over time underlies accumulated information in perceptual decision making (Roitman and Shadlen 2002), which in the model

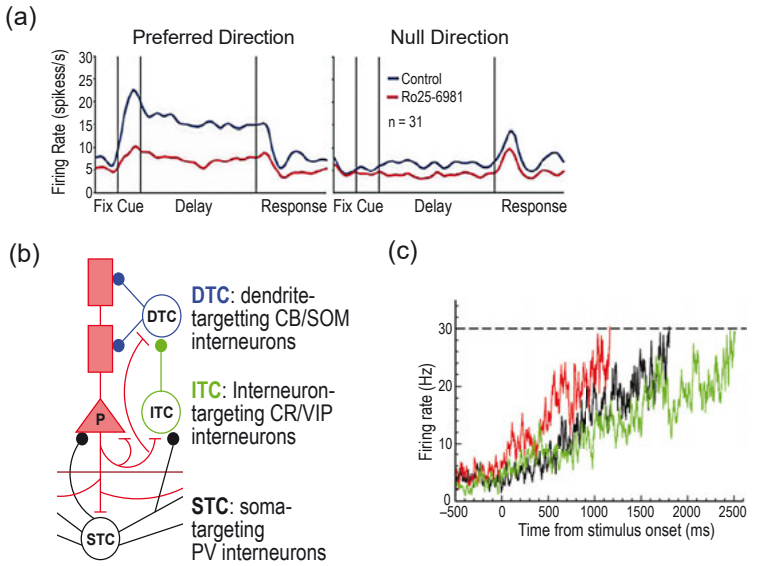


Figure 10.1 Intrinsic circuit properties and dynamics in the prefrontal cortex. (a) Dependence of delay period persistent activity on the NMDA receptors in a monkey experiment using the ODR task. Average response showing the mean firing patterns of 31 dlPFC delay cells for their preferred (left panel) versus nonpreferred directions (right panel) under control conditions (blue) and after iontophoresis of Ro25-6981, a selective antagonist of NR2B-containing NMDA receptors (red). Ro25-6981 markedly decreased task-related firing, especially for the neurons' preferred direction. Reproduced from Wang (2013). (b). The model scheme from Wang et al. (2004) with three inhibitory cell subclasses in addition to pyramidal (Pyr) cells: perisoma-targeting (parvalbumin-containing, PV), interneurons express somatostatin (SST) or calbindin (CB), VIP or calretinin (CR)-containing interneurons. (c) Ramping activity of a recurrent neural circuit model for working memory and decision making (Wang 2002).

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

is realized by slow reverberation (Figure 10.1c). Attractor dynamics underlying selective persistent activity during working memory produces a categorical choice in a decision process (Wang 2002). These results led to the proposal of "cognitive-type" cortical microcircuit (Wang 2013). Mathematically, the strength of recurrent excitation must exceed a threshold level, when a sudden transition called bifurcation takes place, leading to the functional capability to subserve working memory and decision making.

In summary, neural circuit modeling across levels has yielded several surprises: the idea of slow reverberation mediated by the NMDA receptors, the disinhibitory motif, and a common circuit mechanism for working memory and decision making.

Behavioral Flexibility

The PFC plays a central role in behavioral flexibility, illustrated by the Wisconsin Card Sorting Test as a clinical assessment of frontal lobe function. Can the attractor network model be generalized to rule-guided flexible behavior? Consider a simplified version of the Wisconsin Card Sorting Task. Given a sensory cue (a colored shape, e.g., red circle), a subject selects one of two test stimuli that matches the cue either in color or shape, depending on the task rule (color or shape) (Mansouri et al. 2006). Presumably, the rule that is currently valid, say color, is represented internally by persistent activity of "color rule cells," which must be maintained across trials, but switched off when the rule has changed (e.g., from color to shape), signaled by a negative feedback. To illustrate the problem (Figure 10.2a), assume that the neural activity (high or low, H or L) is determined by two types of inputs: recurrent drive which is high or low depending on whether the neuron is active or not (i.e., the internally maintained rule is color or shape), and feedback signal which can be positive (in which case the activity should stay) or negative (in which case the activity should switch). The required input-output mapping amounts to the exclusive OR operation.

The key to solving this problem is to introduce neurons that show conditional responses; for instance, having firing that is selectively high for a particular stimulus only when rule 1 but not rule 2 is currently valid. This reasoning led Rigotti et al. (2010) to propose the concept of mixed selectivity, by adding to a decision-making circuit a large "reservoir" of randomly connected neurons (RCNs) (Figure 10.2b). The basic idea is that by virtue of random connections, RCNs are naturally activated by a combination of synaptic inputs from external stimuli as well as rule-coding neurons (e.g., the color rule is currently in play *and* the network receives a negative feedback signal), and such mixed selectivity is exactly what is needed to solve the task. This model provides a general framework for describing context- or rule-dependent tasks (Rigotti et al. 2010). Figure 10.2c–d shows such a network model for the simplified Wisconsin Card Sorting Test. Notable is the high degree of variability

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

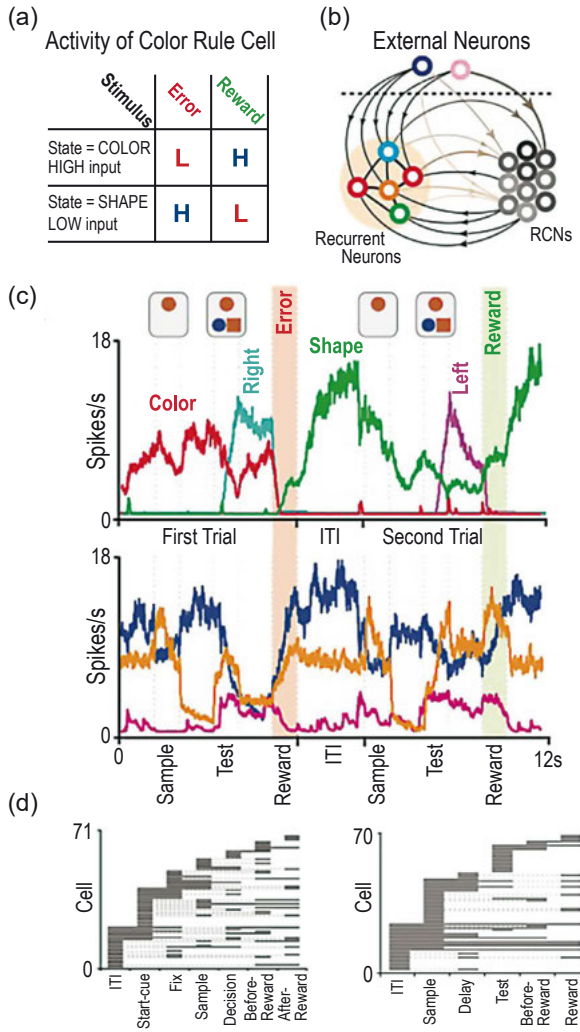


Figure 10.2 A network for rule-based behavior. (a) Exclusive or (XOR) computation by a cell that encodes the rule "color" in a simple variant of the Wisconsin Card Sorting Task; see text for further details. (b) Neural network architecture: randomly connected neurons (RCNs) naturally display mixed selectivity. (c) Firing activity time course for five sample neurons. Light pink vertical line: rule switch; light green line: rule stay. Top: two rule selective neurons; bottom: three RCNs. (d) Rule selectivity pattern is heterogeneous over time and across neurons. Left: rule selectivity for 70 simulated cells in the model. For every trial epoch (x-axis) a black bar is shown when the neuron had a significantly different activity in shape and in color rule blocks. Neurons are sorted according to the first trial epoch in which they show rule selectivity. Right: rule selectivity for spiking activity of single units recorded in prefrontal cortex of monkeys performing an analog of the Wisconsin Card Sort Task (Mansouri et al. 2006). Adopted from Rigotti et al. (2010).

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

of firing activity, across cells as well as for a single neuron across task epochs. Heterogeneity and mixed selectivity are salient yet puzzling characteristics of frontal cortical neurons recorded from behaving animals. Our model suggests that mixed selectivity is computationally desirable as it allows the network to encode a large number of facts, memories, events, and importantly their combinations, the latter being critically important for enabling the PFC to subserve context- and rule-dependent flexible behavior. The theoretical proposed concept of mixed selectivity has been supported by analysis of PFC single-neuron activity in behaving monkeys, establishing another principle for understanding how the PFC works.

More recent work investigated how a single brain area like the PFC may subserve many cognitive tasks. With the help of machine learning (Figure 10.3a), Yang et al. (2019) built a recurrent neural network capable of performing 20 cognitive tasks that are commonly used in monkey physiological experiments and which engage various core cognitive functions, including working memory, rule-based decision making, categorization, and inhibitory control of responses. This model made it possible to examine subportions of the model that represent neural clusters engaged in different types of cognitive building blocks. Concretely, the extent of engagement in a task by each model neuron is measured by a quantity called normalized task variance (Figure 10.3b). The task variances of each unit form a vector in the 20-dimensional space of tasks, and relationships between units can be assessed using clustering algorithms. Units were self-organized into distinct clusters through learning; those belonging to the same cluster are mainly selective in the same subset of tasks. For instance, inhibitory control is often studied using an anti-response task paradigm where a salient stimulus is shown, orienting toward it is prepotent but must be suppressed; instead, the correct action is a more deliberate response diametrically opposite to the stimulus. Three anti-response tasks (Anti-, reaction time-Anti, and delayed-Anti) primarily engage a distinct cluster #3 (purple), and computationally inactivating units in that cluster impairs only anti-response tasks but not the others.

This model needs to be biologically elaborated to provide insights into the brain mechanism of rule-guided behavior. First, it should obey Dale's law, which states that a given neuron contains and releases only one type of neurotransmitter, so that circuit wiring diagram can be identified with separate excitatory and inhibitory neurons. Second, as discussed above, gating may involve different inhibitory cell types which can be incorporated into the model. Third, the model can be extended to multiple modules that differentially encode task representation and task rule, guided by sensorimotor mapping.

Distributed Process with Functional Specificity

While neural correlates of a cognitive function, such as working memory, are commonly observed in PFC, they are also present in other parts of the cortex,

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

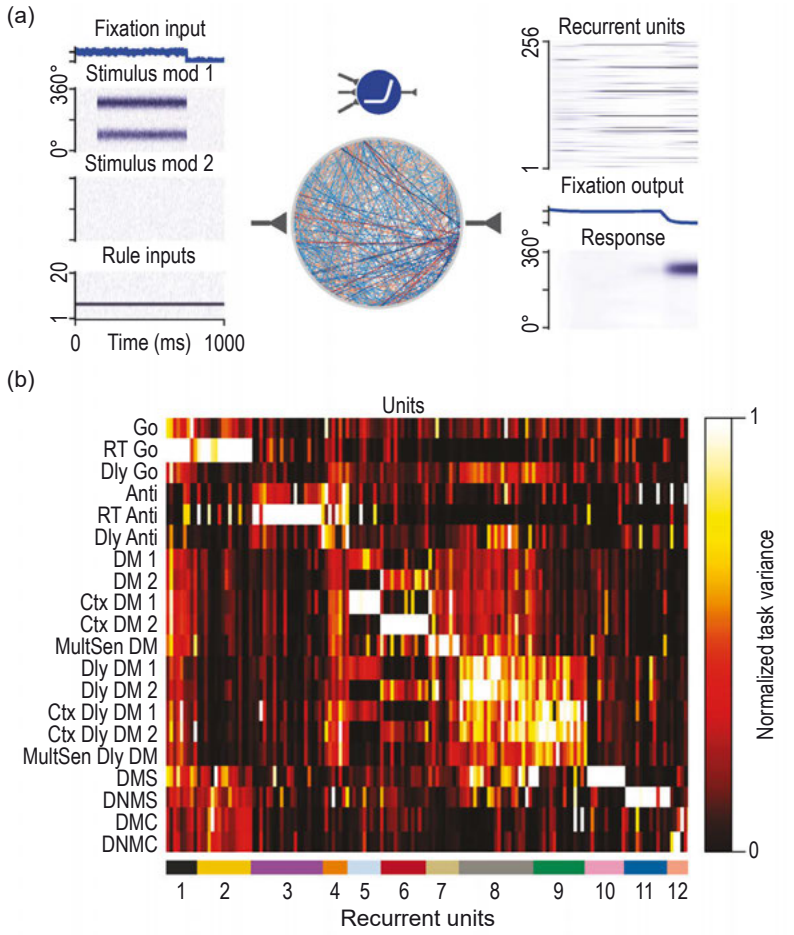


Figure 10.3 A recurrent neural network (RNN) trained to perform 20 cognitive tasks. (a) Schematic of model setting. Left: in a trial, the RNN receives a rule cue, sensory stimuli, and a fixation signal when the network should not produce a motor response. In this example of motion direction discrimination, the stimulus is shown in mod 1 pathway. Right: network dynamics with RNN units (top), fixation unit (middle), motor response unit (bottom). (b) Task variances across all tasks and active units, normalized by the peak value across tasks for each unit. Units form distinct clusters identified based on normalized task variances. Each cluster is specialized for a subset of tasks, such as those that involve a mnemonic delay (Dly). A task can involve units from several clusters; for example, delayed match-to-sample (DMS) engages clusters #1, 2, 8, and 10. Units are sorted by their cluster membership, indicated by colored lines at the bottom. Adapted from (Rigotti et al. 2010).

including the posterior parietal cortex (Leavitt et al. 2017). Because neurons are recorded from the intact brain where areas are interconnected, it is not a given that neural firing in an area related to working memory, even PFC, is

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

generated locally or depends on interactions between multiple areas. As a matter of fact, studies using modern tools for neurophysiological recording and calcium imaging appear to show widespread neural correlates of behaviorally relevant attributes, thus raising the question of how distributed representation can be reconciled with functional localization.

Mejias and Wang (2022) developed a large-scale model of distributed working memory using a directed and weighted connectivity for macaque monkey cortex of Markov et al. (2014). Figure 10.4a–b show model simulation of a visual delayed response task. Notably, responses to an input during stimulus presentation occur in the portion of the model that simulates posterior parts of the cortex, whereas persistent activity during the delay period displays a spatial pattern involving frontal, parietal, and temporal areas of the model. Persistent activity of each area plotted as a function of its hierarchical position exhibits a gap in the firing rate that separates the areas that exhibit mnemonic activity

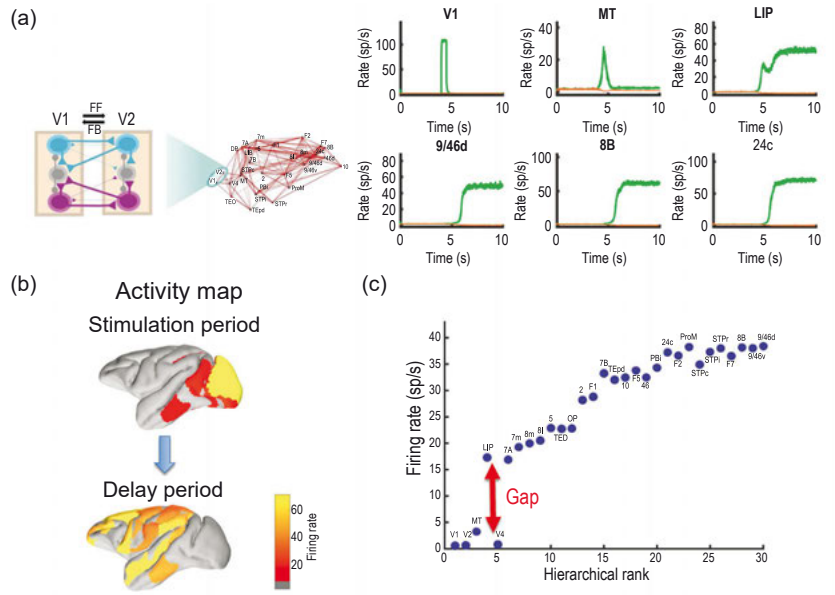


Figure 10.4 Distributed working memory representation in a large-scale monkey cortex model when none of isolated areas is capable of generating persistent activity. (a) Model schema is shown on the left; a model simulation of a visual delayed response task is shown on the right, where activities of the two excitatory neural populations are given for six sample areas. Green: preferred to the shown stimulus; red: nonpreferred to it. (b) Activity map is confined to the posterior part of the cortex during stimulus presentation. By contrast, it is distributed in the frontal, parietal, and temporal areas during the delay period after stimulus withdrawal. Firing rate is shown in color. (c) Mnemonic firing rate of the selective neural pool in each area during the delay period is plotted as a function of its hierarchical position. Those areas displaying persistent activity are separated from those that do not, by a gap in the firing rate. Reproduced from Wang (2020) with original data from Mejias and Wang (2022).

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

from those that do not (Figure 10.4c). This is reminiscent of a bifurcation, but it occurs in space rather than as a function of a parameter. The transition is robust: changes of network parameters would alter the location of cortical tissue where the transition occurs and show precisely which areas exhibit mnemonic persistent activity but would not abolish the transition itself.

This "bifurcation in space" phenomenon represents a mechanism for the emergence of functional modularity. In the model, parcellated areas follow an identical canonical local circuit organization, but certain properties, like the strength of synaptic excitation, vary systematically in the form of macroscopic gradients calibrated experimentally (Wang 2020). Interareal cortical interactions quantified by the connectomic analysis involve long-range connections, which makes it all the more remarkable that the sudden transition can occur locally in a multiregional cortex. Thus, working memory is distributed, yet depends on a specific subset of areas, in contrast to the absence of modularity manifest by merely graded variations of engagement across the entire cortical mantle. Furthermore, some areas show mnemonic activity as a result of sustained inputs from other core areas, including the PFC. These findings suggest a general principle for understanding functional specificity compatible with distributed cognitive processes.

Summary

In close interplay with experiments, theory has produced new concepts like slow reverberation, disinhibitory motif, cognitive-type microcircuits capable of working memory and decision making, mixed selectivity, and bifurcation in space as a mechanism for the emergence of functional modularity in a large-scale cortex endowed with a canonical circuit architecture. These concepts, derived from biologically based cross-level neural circuit modeling, have furthered our understanding of PFC function. Looking ahead, with the prospect of new availability of big data (ranging from genomic analysis to connectome to large-scale recordings), theory and mathematical modeling are poised to play an indispensable role in elucidating the complex inner working of the frontal lobe at the core of cognition and intelligence.

Computational Cognitive Models of PFC Function

Reinforcement learning (RL) is commonly viewed as describing animal (including human) basic adaptive behavior. Empirical evidence indicates that the basal ganglia interacts with the premotor cortex and lateral orbitofrontal cortex (OFC) to contribute to RL (and likely along with the insular cortex for punishments). RL is a simple, robust, and efficient adaptive process. RL, notably its temporal-difference algorithmic implementation (Sutton and Barto 1998), assumes the brain encodes stimulus-action and stimulus-reward associations

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

that reflect experienced rewards (or punishments). These associations adjust online according to the discrepancy between actual and expected rewards/punishments encoded in these associations and gradually guide action selection toward the most valuable course of actions. We refer to such courses of action as *selective models*. Computational models using RL can learn complex selective models to adapt to complex situations. In particular, when action outcomes depend only on current external states and actions, RL potentially converges toward the behavioral strategy maximizing rewards, regardless of the situation complexity (Sutton and Barto 1998). RL is robust to uncertainty and contingency changes. While more efficient adaptive processes exist and adjust faster to changing situations, their gains relative to RL performances are often weak compared to their increased computational complexity and are obtained at the cost of decreased versatility. For instance, adaptive processes based on Bayesian inferences regarding the external contingency volatility and its variations across time (e.g., Behrens et al. 2007), adjust relatively faster than RL in varying volatile environments but perform much less efficiently than RL in stable environments with sparse environment feedbacks (Findling et al. 2021).

Still, RL exhibits key adaptive limitations. First, RL algorithms learn from reward subjective values, which vary according to animals' internal states or needs. For instance, a thirsty animal may learn through RL an efficient course of actions to obtain water. When the animal becomes hungry, however, this course of action becomes ineffective to get food, and the animal is forced to relearn from scratch a new course of action to acquire food. More generally, the problem arises because RL algorithms learn from the value rather than identity of action outcomes. Overcoming this adaptive weakness requires learning world models, which we refer to as *predictive models*, that link stimuli, actions, and outcomes, irrespective of rewarding values. Such predictive models enable RL algorithms to operate covertly (a process named model-based RL), according to current animal internal states/needs, to build effective selective models on demand and subsequently act in an efficient manner (Gershman et al. 2014; Liu et al. 2021). Learning predictive models remains, in principle, a basic process that corresponds to register the environment statistics. Future research is needed to understand how the animal is driven to learn such predictive models, which appear critical for responding to the ever-changing internal states and needs of an animal.

Second, learning and adjusting selective and predictive models in RL is achieved by erasing previously learned information. This naturally allows these models to adapt to new situations but it also requires the animal to relearn entirely these models when situations encountered in the past reoccur at a later time. Our natural environment actually features a constant mixture of new and recurrent situations: for instance, access to water sources may periodically change according to seasons but also suddenly when unique events occur like forest fires. New and recurrent situations are potentially unlimited; that is, external contingencies form a potentially infinite-dimension space, which

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

prevents animals and actually any physical device from learning and parametrically adjusting only one comprehensive predictive model of the world. Thus, efficient adaptations require animals to gradually learn multiple predictive models as discrete entities, ideally as much as the number of encountered distinct situations: learned predictive models form a repertoire in long-term memory, thus defining a finite but expanding behavioral space, whose dimensions correspond to the number of situations encountered and perceived as distinct. This adaptive process, however, raises complex computational issues in terms of how animals identify situational changes or recurrent versus new situations as well as how they retrieve previously learned selective/predictive models or learn new models when facing recurrent versus new situations (Koechlin 2014).

These two RL adaptive limitations are tightly linked as learning predictive models unfold over time and rely on the assumption that the ongoing situation is identified as remaining unchanged. These limitations appear to be so fundamental for efficient adaptive behavior that we can reasonably assume that the PFC has evolved primarily to overcome them. Another RL functional limitation identified is the lack of learning rate adjustments according to the change frequencies of external contingencies, often referred to as volatility (Behrens et al. 2007). Indeed, efficient adaptive behavior requires learning rates to increase when volatility increases so as to discount previously learned information. Complex probabilistic inference models involving the PFC have been proposed to estimate volatility to make such learning rate adjustments (Behrens et al. 2007; Payzan-LeNestour and Bossaerts 2011). However, a recent computational study (Findling et al. 2021) shows that counterintuitively, such adjustments are likely to derive merely from neural computational imprecisions conforming to Weber's law (the more internal representations change, the more imprecise are representation updates) rather than from additional volatility estimate processes.

Medial OFC and ACC Overcome RL Adaptive Limitations

Empirical evidence suggests that through RL mechanisms, lateral OFC encodes/stores the *experienced* reward value of stimuli, irrespective of associated actions (i.e., in a Pavlovian fashion) (O'Doherty 2007; Rouault et al. 2019), while the premotor cortex encodes/stores stimulus-action associations. Accordingly, lateral OFC provides subjective values of actual action outcomes which enables the learning of stimulus-action associations in the premotor cortex likely via the basal ganglia. Lateral OFC, premotor cortex, and the basal ganglia thus form a basic functional network (possibly along with the insular cortex for punishments) that subserves the RL of *selective models* guiding behavior (Soltani and Koechlin 2022).

In contrast, empirical evidence suggests that medial OFC encodes/stores the identity of action outcomes (i.e., their probability of occurrences following

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

action selection), along with the associated *prospective* reward values (Jones et al. 2012; Rouault et al. 2019). Based on experienced reward values encoded in the lateral OFC, prospective reward values correspond to the current valuation of prospective action outcomes likely based on animals' internal states or needs. This indicates that the medial OFC is likely to encode *predictive models* subserving model-based RL (Chan et al. 2016) to enable RL to operate covertly, and through the basal ganglia to readjust selective models encoded in the premotor cortex to guide behavior according to animals' current internal states/needs.

Computational models introduced the notion of *state beliefs*; namely, probability distributions over predictive models measuring to which extend they apply to the current situation or equivalently, their posterior ability to predict actual action outcomes through standard probabilistic inference processes (Chan et al. 2016). Computational models further introduced the notion of actor reliability—the belief that the predictive model guiding ongoing behavior applies to the current situation relative to any *known* or *unknown* alternative predictive models based on the maximum entropy principle (predictions in unknown situations are at chance level) (Collins and Koechlin 2012). Critically, actor reliability assesses whether the current situation is likely to remain the same or has changed; that is, whether the current predictive/selective model guiding ongoing behavior (referred to as the actor task set) remains reliable or not. In the former case, the corresponding predictive model continues to guide behavior and to improve through online learning. In the latter case, this task set is inhibited and replaced by another one (see below). Empirical studies provide evidence that the medial OFC indeed monitors online actor reliability based on actual action outcomes (Domenech et al. 2020; Donoso et al. 2014a).

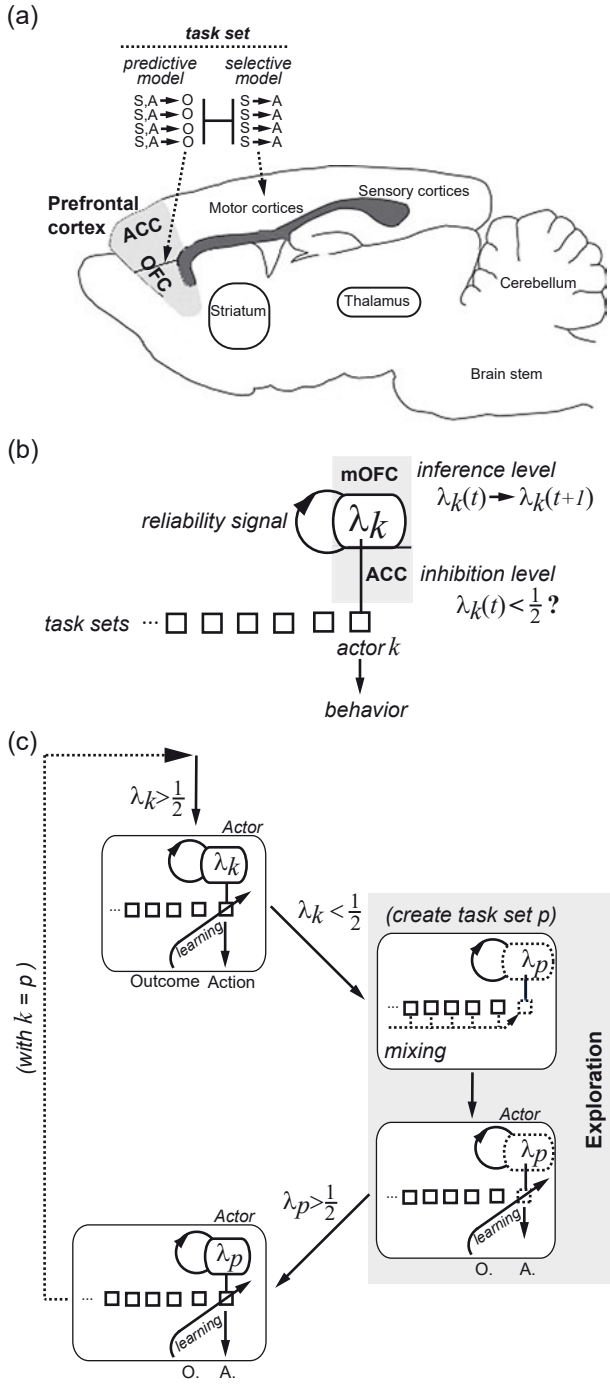
Thus, the OFC encodes several signals, including experienced stimulus values in the lateral OFC, prospective outcome values, outcome probabilities, and actor reliability in the medial OFC. All these signals and possibly others may potentially guide behavior. A classical view originating in the rational decision theory states that to achieve action selection, the signals encoded in medial OFC are combined together to compute the subjective expected utility of each related behavioral option—a common currency used as a decision variable to arbitrate between the options. Recent computational studies, however, suggest that instead, these different signals independently compete and concurrently contribute to action selection within the ACC, after each signal type is normalized across available actions (Cao and Tsetsos 2022; Farashahi et al. 2019; Rouault et al. 2019). These studies show that these contributions are not weighted equally at choice time with the predominance of medial OFC signals related to predictive models. The weighting also varies depending upon the environment characteristics. For instance, the more volatile the environment, the less outcome probability signals were shown to predominate, in agreement with the fact that volatile environments prevent an organism from forming precise predictive models (Farashahi et al. 2019). Exactly how the weighting

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

is determined remains an open issue. A possible hypothesis is that the weighting naturally arises from neural reciprocal interactions between the ACC and OFC regions rather than deriving from additional higher-order computational processes. For instance, less precise predictive models are likely encoded with increased neural variability in medial OFC, which in turn should weaken their remote influence on the ACC.

It is also likely that following action outcomes, actor reliability signals play a predominant role. Indeed, the medial OFC was observed to signal proactively that the situation might have changed right before experiencing action outcomes (i.e., the actor reliability is deemed as uncertain), leading the ACC to process actual action outcomes as confirming or denying this medial OFC prediction. ACC was observed to process actual action outcomes as a trigger to inhibit and switch away from the ongoing predictive model or to stay with the same predictive model to guide subsequent behavior (Domenech et al. 2020). Thus, the ACC is modeled as inducing behavior to switch to *undirected exploration* corresponding to the formation and learning of a new predictive and selective model (i.e., a new actor task set for guiding subsequent behavior). Computational models further propose that this new actor is first built from mixing previously learned predictive and selective models stored in long-term memory according to actual action outcomes and then adjusts subsequently to actual external contingencies (Collins and Koechlin 2012; Koechlin 2014) (see Figure 10.5). As the medial OFC monitors actor reliability, this new actor may eventually be deemed as reliable, in which case its selective and predictive model are consolidated in long-term memory to contribute to creating new actors in the future. Neural correlates of such covert confirmation events based on actor reliability were observed within the basal ganglia in the ventral striatum, which receives direct projections from medial OFC (Donoso et al. 2014a). In addition, once the new actor is deemed reliable, it will likely become unreliable at some point, in which case a new actor creation process will be triggered again. Although the neural mechanisms involved in the actor creation and consolidation processes remain poorly specified, we presume that these processes which rely on long-term memory involve a large network of brain regions, notably outside the PFC, which along basal ganglia certainly comprises the hippocampus, known for its central role in memory retrieval and world model constructs (e.g., Whittington et al. 2020). In this view, the medial OFC and ACC control only *when* to create and consolidate new actors whereas the creation and consolidation processes per se appear to unfold outside PFC control. Importantly, the computational model combining actor reliability monitoring, actor creation, and consolidation shows that the repertoire of task sets that comprise joint selective and predictive models, stored in long-term memory, extends in a way that associates more recurrent situations with task set replicas in long-term memory. As a result, actor creation relies more extensively on task sets associated with more recurrent situations. This computational model forms the optimal adaptive process with the constraint

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530



From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

that only actor reliability is inferred online from action outcomes (Collins and Koechlin 2012).

Lateral and Frontopolar PFC Overcome OFC and ACC Adaptive Control Limitations

As described above, the medial OFC and ACC, in association with basal ganglia, form a consistent and efficient system that controls adaptive behavior in uncertain, changing, and open-ended environments beyond basic RL processes. Nonetheless, this medial PFC system exhibits three key functional limitations:

1. Actor reliability is inferred only from actual action outcomes, so that switching away from the current actor occurs only after experiencing actual action outcomes. This might be especially detrimental in case of adverse action outcomes.
2. Actor creation ignores the context in which selective/predictive models were learned, which may lead actor creation to start guiding behavior using maladaptive task sets (i.e., selective/predictive models) stored in long-term memory. For instance, the selective/predictive models I learned when interacting with people at work might not be well adapted when I interact with my roommates and vice versa.
3. By monitoring only actor reliability, the system is constrained to make irreversible decisions, when switching away from the current actor and creating new actors (actor creation cannot be reversed to re-instantiate a new actor creation).

In other words, the medial OFC-ACC system lacks flexibility, which is especially detrimental when dealing with discrete entities such as task sets (i.e., in non-parametric inferences). We have proposed that the evolution of lateral

Figure 10.5 Model of rodent PFC function. (a) Schematic representation of the rodent brain; PFC includes the anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC) to manage the creation of actor task sets to guide behavior. Task sets comprise selective and predictive models (i.e., stimulus-action associations and stimulus-action-outcome associations, respectively). Selective models are encoded in motor/premotor cortices; the OFC encodes predictive models. (b) Diagram showing inferential and creative processes composing the rodent PFC function. OFC infers actor reliability λ (i.e., to predict action outcomes and monitor when the situation changes). While the actor remains reliable ($\lambda > 1-\lambda$), the actor drives behavior and adjusts its internal models (learning, exploitation periods). ACC detects when the actor becomes unreliable ($\lambda < 1-\lambda$) and the situation has presumably changed. ACC inhibits the unreliable actor and triggers the creation of a new actor. Actor creation results from mixing task sets stored in long-term memory (square) yielding to forming an unreliable actor. While this newly created actor remains unreliable, it drives behavior and learns external contingencies (exploration period). Once it becomes reliable, it is consolidated in long-term memory, and a new exploitation period starts to create new actors from long-term memory. Reproduced from Koechlin (2020).

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

PFC in primates overcomes the first two limitations, while further evolution of the frontopolar PFC in humans overcomes the third (Koechlin 2014).

There is ample empirical evidence that lateral PFC is involved in switching between sensorimotor mappings according to contextual cues. Computational cognitive studies in humans further show that stimulus-action associations are spontaneously learned and aggregated into clusters/chunks indexed by contextual cues (Collins and Frank 2013). Such clustering processes, which lead to hierarchical selective models, were shown to occur in posterior lateral PFC (Badre et al. 2010). These results also indicate that the actor task set guiding behavior comprises an additional internal model—the *contextual model*—that links the actor to contextual cues. Accordingly, actor contextual models can be modeled as learning to which extent external cues are predictors of actor reliability (Collins and Koechlin 2012; Koechlin 2014). Thus, lateral PFC enables actor reliability to be inferred from contextual cues and to switch away from the current actor proactively when such cues occur before acting and experiencing action outcomes (see Figure 10.6). Moreover, contextual models enable the contribution of task sets stored in long-term memory to actor creation to be weighted according to the current context of action. As a result, actor creation relies mostly on task sets which were possibly learned previously in similar contexts. In particular, when current contextual cues were previously associated with specific task sets, actor creation operates as if directly retrieving such task sets from long-term memory. Indirect empirical evidence from cognitive control and memory retrieval studies suggests that such cue-based inferences about the reliability of actors and actor creation involve mid-lateral PFC (Koechlin et al. 2003; Nee and D'Esposito 2016, 2017). The resulting executive system that spans the medial PFC (comprising medial OFC and ACC) and lateral PFC form an optimal adaptive system with the constraint that only actor reliability is inferred online (Koechlin 2014).

As noted above, this constraint implies irreversible decisions and yields a system that lacks flexibility to switch back and forth between multiple potential actors guiding behavior. Computational models indicate that overcoming this limitation requires inferring online the reliability of potential alternative task sets in addition to the current actor task sets guiding ongoing behavior (Collins and Koechlin 2012). For clarity, we refer to such potential alternative actors as counterfactual task sets, which as the current actor, consist of selective/predictive/contextual models forming consistent, discrete executive entities. Inferring in parallel the online reliability of multiple task sets is beneficial in many respects:

1. Reliability inference is improved as each task set now measures to which extent its predictive model applies to the current situation relative to the other task set predictive models along with any additional, unknown/unmonitored predictive models.

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

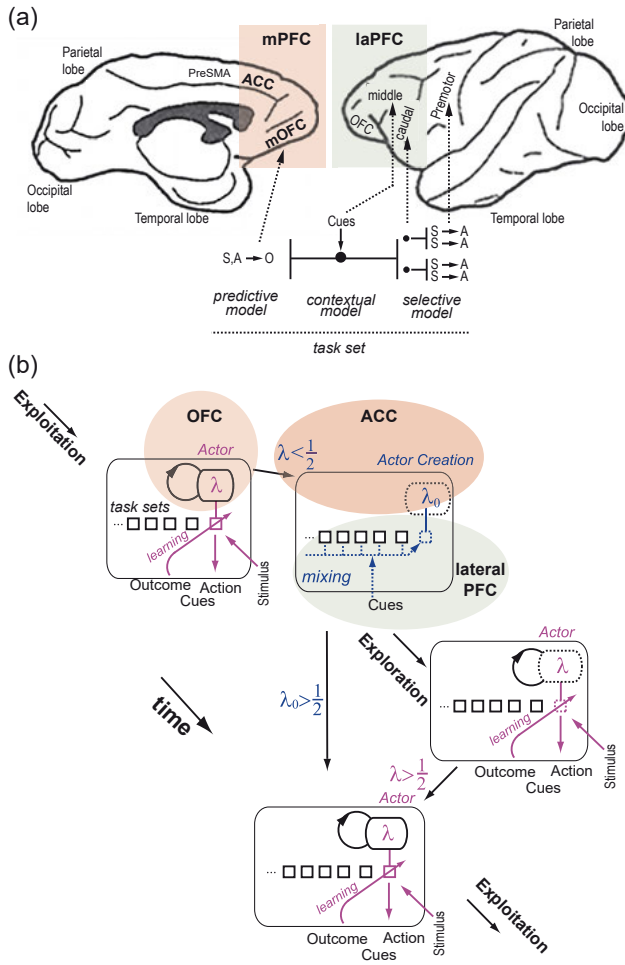


Figure 10.6 Model of monkey PFC function. (a) Schematic representations of the monkey medial and lateral cerebral cortex. Compared to rodents, monkey PFC has an additional lateral prefrontal cortex (laPFC) comprising a middle and caudal sector. In monkeys, task sets are assumed to comprise contextual models (associating task set to external cues) encoded in the laPFC. Contextual models indexing task sets are represented in the middle laPFC and allow chunking processes in caudal laPFC to operate within task sets (see text). (b) Illustration of the inferential and creative processes in the monkey PFC function. Inferential processes are similar to those in rodents (see Figure 10.5), except that contextual models enable the updating of actor reliability according to the occurrences of external cues (in addition to action outcomes). Actor creation may thus occur proactively. Contextual models also have a major role in refining actor creation: the mixture of task sets in long-term memory is now weighted by current external cues according to contextual models. As a result, new actors may be created as immediately reliable ($\lambda_p > 1 - (\lambda_p)$. In that event, the exploration period is skipped, leading to the ability to recreate new actors much more rapidly. Reproduced from Koechlin (2020).

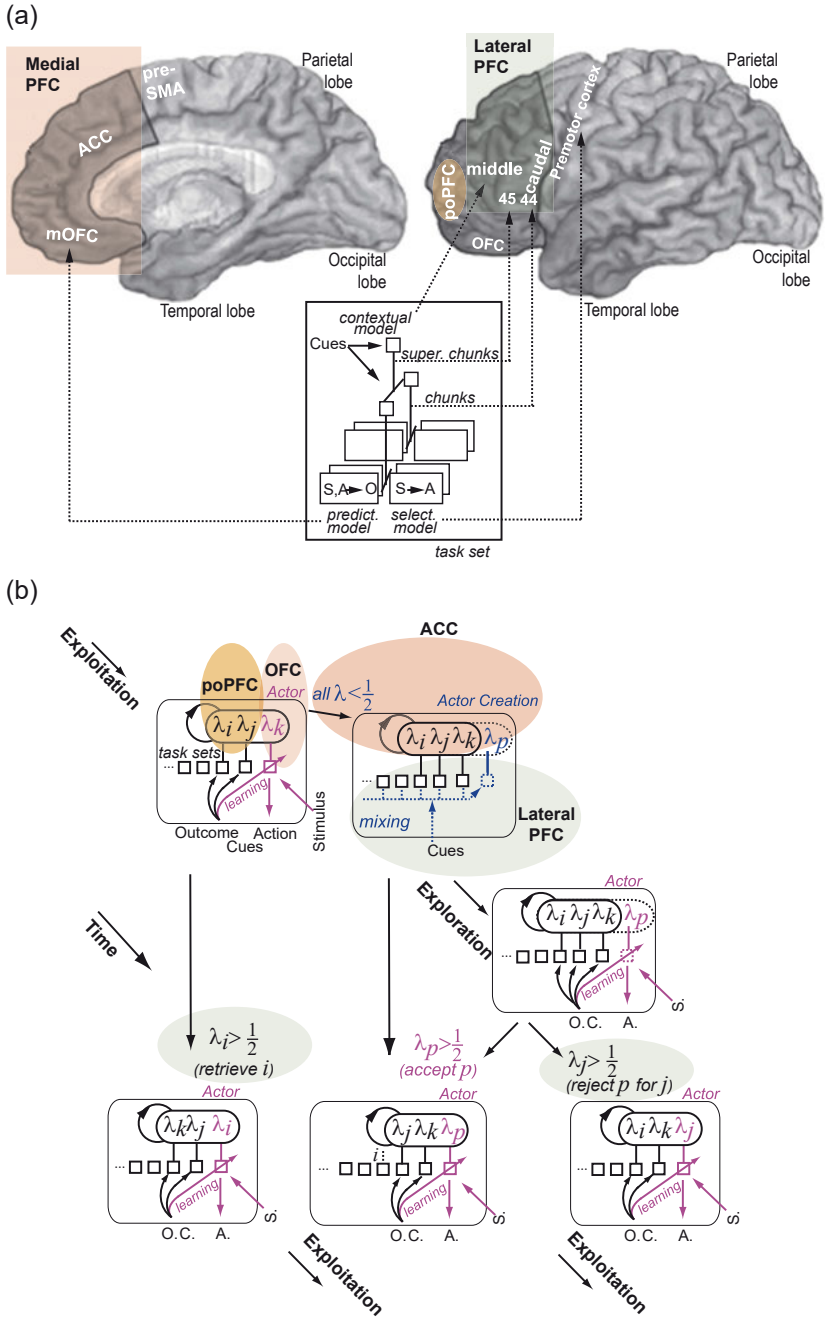
From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

2. When a counterfactual task set becomes reliable (implying that the current actor task set is deemed unreliable), the system simply switches to this counterfactual task set to replace the current actor and guide subsequent behavior, while the current actor guiding ongoing behavior becomes a counterfactual task set.
3. When no task sets are deemed as reliable, actor creation from long-term memory occurs for guiding subsequent behavior, while the current actor again become a counterfactual task set.
4. Actor creation may be rejected later on, whenever a counterfactual task set becomes reliable, while the newly created actor is still not deemed as reliable, preventing the newly created actor task set from being consolidated in long-term memory.

The resulting actor creation process thus resembles hypothesis testing: a new hypothesis (the newly created actor task set) is tested against alternative hypotheses (the counterfactual task sets) based on the acquisition of additional information from action outcomes or contextual cues. This computational model forms an optimal adaptive algorithm in uncertain, changing, and open-ended environments with the following constraint: only a limited number of counterfactual actors can be monitored online in parallel (see Figure 10.7). This computational model was shown to account well for human performances in such environments and performed better than several alternative models. Model fitting to human performances further suggests that humans monitor online no more than three counterfactual task sets in parallel (Collins and Koechlin 2012). When this capacity limit is reached, the least recently used counterfactual task set is simply discarded from online monitoring, while remaining stored in long-term memory. Furthermore, empirical evidence shows that the human frontopolar cortex monitors the reliability of counterfactual task sets (Donoso et al. 2014a; Mansouri et al. 2017), and as mentioned before, the current actor reliability is monitored in medial OFC. Rejecting actor creation to select a counterfactor task set monitored in the frontopolar cortex and deemed

Figure 10.7 Model of human PFC function. (a) Schematic representations of the human cerebral cortex. Compared to monkeys, human PFC comprises a frontopolar region (poPFC) in the lateral forefront of the PFC with no known homologues in monkeys. In humans compared to monkeys, task sets are likely to comprise two nested, abstract levels of chunking, involving BA 44 and 45, and may play a major role in language (see text). (b) Inferential, selective, and creative processes forming the human PFC function. Compared to monkeys (see Figure 10.6), the human poPFC forms an inferential buffer to infer and monitor the reliability of additional task sets (counterfactual task sets) in addition to the actor task set monitored in the medial OFC. This additional inferential capability endows humans with the ability to retrieve a counterfactual task set directly to drive behavior when it becomes reliable, in both exploitation and exploration periods. During exploration, this ability yields newly created actors to be rejected and disbanded and corresponds to hypothesis testing bearing upon task set creation. Reproduced from Koechlin (2020).

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530



From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

as reliable to guide subsequent behavior was further found to involve mid-lateral PFC (Donoso et al. 2014a). Thus, this computational model suggests that the human frontopolar cortex forms a capacity-limited online monitoring buffer that allows switching back and forth across several potential task sets to guide behavior and regulate the online creation and storage of task sets in long-term memory through hypothesis-testing processes.

It is worth noting that *the* true optimal adaptive model requires additional features:

- A monitoring buffer with unlimited capacity, so that the reliability of all created task sets is inferred in parallel.
- Reliability inferences are not limited to online forward inferences but also involve critically offline backward inferences to enable constant online revision of actor creation.
- Actor guiding behavior is the continual parametric mixture of all created task sets weighted by their reliability.

More precisely, the optimal adaptive model involves mixtures of Dirichlet processes that generalize Bayesian inferences to open-ended environments (Doshi-Velez 2009; Gershman et al. 2010; Teh et al. 2006), but whose computational costs are exorbitant and even intractable, thereby hindering its optimality in practice. Accordingly, we have reviewed evidence that the human PFC has evolved as capturing tractable algorithmic approximations of key computational components underlying optimal adaptive behavior:

- Monitoring (a limited number of) multiple potential task sets,
- A minimal form of backward inferences through hypothesis testing involved in actor creation (creating new actors may be revised later on), and
- Mixing all created task sets weighted by contextual models when actor creation occurs to guide behavior.

Newly created actors are thus parametric mixtures of previously learned selective, predictive, and contextual models. Note that in contrast, mixing the task sets monitored in a *capacity-limited* buffer according to their relative reliability to guide behavior is detrimental, because the proper task set might actually be stored in long-term memory without being monitored.

Higher cognition comprising planning, reasoning, and language production might simply reflect the functioning of this whole computational PFC architecture (Koechlin 2020). As noted above, planning amounts to covertly navigating within the current actor predictive model through model-free RL using the actor task set. Reasoning can amount to combining reliability inferences about several potential task sets viewed as multiple behavioral hypotheses with hypothesis-testing regarding actor creation viewed as hypothesis generation. Language production may amount to actor creation viewed as generating

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

linguistic sentences in reciprocal interactions with the superior temporal sulcus through the arcuate fasciculus (Rouault and Koechlin 2018).

What Drives Learning in Predictive Models

The preceding sections outline the key role of predictive models for efficient adaptive behavior. Predictive models as "world models" predict potential action outcomes, enable adjustment of selective models to internal states/needs, and have the ability to carry out planning covertly, and to detect situational changes that may result in actor changes through actor reliability inferences. We indicated above that learning predictive models is based simply on registering the experienced environmental contingencies. This could happen on the fly while other incentives, such as rewards, are driving animals' behavior. Given the critical role of predictive models, we reason that learning predictive models might also be an intrinsic motivation driving animals' behavior.

The classical theory is that animals/humans' behavior is primarily driven through the maximization of subjective rewards (e.g., Schultz 2015). To be efficient, reward maximization requires deviating episodically from what was learned as the most rewarding course of action and to explore alternative courses of action so as to avoid being trapped in local reward maxima. In this view, predictive models are learned on the fly; there are no specific incentives to learn them.

Another theory proposes that animals/humans' behavior is primarily driven by minimizing expected free-energy or "expected surprise" (Friston 2010): behavior aims at producing outcomes expected from predictive models, and predictive models are adjusted according to actual action outcomes. Under this view, potential subjective rewards are absorbed as highly expected outcomes in predictive models. The theory offers a general, principled view of adaptive behavior, revealing that behavior is centered on learning adequate predictive models and acting accordingly. The theory has, however, two key limitations. First, it assumes that agents have an exhaustive representation of all potential situations (latent states) they may encounter, corresponding to as many task sets that they monitor in parallel to form beliefs about their occurrences. This assumption is unrealistic in real-life environments that feature unlimited potential situations. As noted above, biological systems and physical devices are limited inasmuch as they only monitor a small fraction of potential situations/task sets. Discussion in the preceding sections actually outlines the optimal adaptive system, when the monitoring/inferential capacity is assumed to be limited and suggests that the evolution of PFC implements this capacity-limited adaptive system. Second, and more problematically, the theory relies on an *arbitrary* parametrization of potential subjective rewards aimed at transforming them into outcome expectations to absorb them into predictive models. This is problematic because parametrization actually determines the critical balance between reward- and information-seeking behavior;

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

that is, between exploitation and exploration. Accordingly, the theory appears to define this balance arbitrarily with no accounts of how it is determined and possibly controlled.

To address this issue, we proposed an alternative theory at FENS 2022, based on the fundamental principle of statistical physics. In contrast to Friston's free-energy theory, it distinguishes between the notion of energy and entropy which translate here to the notion of reward as energetic resource and information as negative entropy (Vaillant-Tenzer and Koechlin 2023). The general idea is that behavior aims at primarily maximizing expected information gain *with the homeostatic constraint* to maintain enough energetic resources (i.e., to get enough rewards compensating resource consumption) to pursue this information quest. (Note that unlike biological systems, physical systems "behave" in the converse way as maximizing their entropy with the constraint of maintaining their energy constant.) Within the computational framework outlined in the preceding sections, maximizing expected information gain means selecting actions where the outcomes are expected from the current actor predictive model to best improve predictive power or equivalently to best reduce its predictive entropy/uncertainty. The "statistical physics" formalization of this principle leads to the hypothesis that behavior aims at maximizing the weighted sum of expected subjective rewards and expected information gain within the current actor predictive model. Critically, the weighting of expected subjective rewards relative to information gain is fully determined by the Lagrangian multiplier relative to the homeostatic constraint. This Lagrangian multiplier is not computable in closed form but varies approximately as the inverse of the total amount of agent's energetic resources and consequently as the inverse of accumulated rewards over time. Accordingly, the more an agent is deprived, the more it will exhibit reward-seeking behavior. The more an agent accumulates rewards, the more it will exhibit information-seeking behavior. The more an agent acquires predictive knowledge of the current situation (i.e., expected information gain will vanish), the more it will exhibit reward-seeking behavior. Thus, the hypothesis predicts a complex dynamic balance between reward- and information-seeking behavior. For instance, when an agent faces a new, unknown situation, information-seeking behavior will first dominate as expected information gains within the current actor predictive model are initially at a maximum: thereafter, reward-seeking behavior will begin to dominate as expected information gains start to decline. Next, when received rewards start accumulating, information-seeking behavior will emerge again. And so on. The hypothesis thus predicts that the balance between reward- and information-seeking depends on the agent's homeostatic states and is likely mediated by brain regions monitoring such homeostatic states. A possible candidate where this occurs is the anterior insular cortex, which has been recently associated with homeostasis monitoring and which widely projects to medial PFC regions (Livneh et al. 2020).

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

Concluding Remarks

In this chapter, we have described the modeling of neural networks and cognitive computations subserving PFC function in two distinct sections. This division is unrelated to any distinctions between the classical Marr's levels of brain analysis—namely the physical, representational, and functional level—whereby the functional level describes the function of one system, the representational level how this function is achieved, and the physical level the material device realizing this function (Marr 1982). Both sections independently entail descriptions at all of the three levels. For instance, the first section describes the working memory function, whereas the second addresses the reliability monitoring function. The first section describes different classes of inhibitory neurons, whereas the second section describes different cortical areas in the PFC and so on. Instead, the first and second section address functional, representational, and physical issues at two distinct *scales* of brain organization: at the neuronal and cortical scale, respectively. The two sections reflect the idea that the functional, representational, and physical concepts differ between these two scales of analysis.

We view these conceptual differences across scales as similar to those present in physics. For instance, pressure makes sense at the scale of gas volumes but not at the level of gas molecular constituents. This does not imply that there are no connections between the elements describing the different levels of brain organization. On the contrary, quantitative models are especially useful, if not necessary, to understand how the different organization levels are connected and interact with each other. To date, however, there is little modeling work that aims to link the different neuronal and cortical levels in the PFC, in the way as in the visual system, models of cortical maps, and hierarchical visual processing have been developed. Filling this gap will certainly be an important future avenue in developing models and understanding frontal lobe function.

Acknowledgments

This work was partly supported by an NIH grant R01# MH062349.

From "The Frontal Cortex: Organization, Networks, and Function," edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530

From “The Frontal Cortex: Organization, Networks, and Function,” edited by Marie T. Banich, Suzanne N. Haber, and Trevor W. Robbins. 2024. Strüngmann Forum Reports, vol. 34, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262549530