8

Functional Fractionation and Integration

Physiology, Networks, and Behaviors

Elisabeth A. Murray, Bruno B. Averbeck, David Badre, Christos Constantinidis, Roshan Cools, Clayton E. Curtis, Lesley K. Fellows, Anna S. Mitchell, John D. Murray, Erin L. Rich, and Mark D'Esposito

Abstract

From the early 1900s onward, anatomists have parcellated the cerebral cortex, including the frontal cortex. Initial approaches were based on both the features of stained cell bodies and the pattern of myelinated fibers, together called architectonics. The labels provided by these architectonic investigations are still widely used today. This chapter considers the extant evidence for functional fractionation of the frontal lobes, and whether the organization of the frontal lobes should be conceptualized in terms of functional and anatomical gradients, instead of discrete areas with well-delineated boundaries. Discussion includes how the frontal lobes interact with other parts of the brain to influence behavior as well as the identification of critical gaps in knowledge. The authors conclude that a greater understanding of frontal lobe function would emerge from advances in theory that connects different levels of explanation, that take into account evolutionary perspectives, and that lead to the development of a common cognitive-behavioral ontological framework.

General Introduction

The frontal lobes remain a formidable frontier in neuroscientific study, both literally and figuratively. Frontal cortex forms the furthest extent of the brain,

Group photos (top left to bottom right) Elisabeth Murray, Mark D'Esposito, Lesley Fellows, David Badre, John Murray, Clayton Curtis, Anna Mitchell, Christos Constantinidis, Roshan Cools, Bruno Averbeck, Elisabeth Murray, Mark D'Esposito, Erin Rich, Roshan Cools, Clayton Curtis, David Badre, Bruno Averbeck, Erin Rich, Lesley Fellows, Anna Mitchell, and John Murray

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

anteriorly, providing guidance in decisions ranging from the mundane—like what to eat for breakfast—to the profound—like the selection of a life partner. Frontal cortex is also very much an outer limit in the field of neuroscientific study, one in which the opportunities for research and development, and the promise of understanding and treating maladaptive behavior—whether arising from brain injury or dysfunctional neural circuits—have not been fully realized. Until we have identified and modeled the functions of frontal areas and their circuit interactions, we cannot fulfill one of the key objectives of translational neuroscience: effective treatments of neurological and psychiatric disorders.

Over the last few decades, the field has made substantial progress in defining the functional neuroanatomy of the frontal lobes. The underlying premise of this work is that localization of function arises in part because each frontal cortex region has a unique pattern of afferent and efferent connections. Here we discuss progress toward understanding frontal lobe function not only from identifying functions of single areas, but also in identifying the functions and computations of the networks in which those areas are embedded. We first address the evidence for functional specializations within the frontal lobe and whether the identified functions align with identified anatomical subdivisions. We then explore organizational principles of frontal cortex and how the frontal lobes influence behavior. Finally, we discuss what approaches might unravel the nature of circuit interactions involving the frontal lobe and how we might address gaps in our knowledge.

Anatomical Subdivisions in the Frontal Cortex

From the early 1900s onward, anatomists have parcellated the cerebral cortex, including the frontal cortex. Initial approaches were based on both the features of stained cell bodies (cytoarchitecture) and the pattern of myelinated fibers (myeloarchitecture), together called *architectonics*. Although the number of parcellations in frontal cortex has varied across investigators, as do the locations of boundaries, the labels provided by these architectonic investigations are still widely used today. This is in large part because the architectonic labels provide a common framework for presenting findings across experimental approaches. Recently, chemoarchitectonics has been added to the roster of methods, based on histochemical stains or patterns of receptors. Where cell types are similar across cortical areas, it is also possible that the relative distribution of those cell types could help delineate functionally distinct cortical fields. These new methods can refine classical cortical maps and offer an additional basis for generating hypotheses regarding the functions of these regions.

There is no consensus on whether anatomically identified regions in the frontal cortex correspond to meaningful functional zones. Neuropsychological evidence in humans and animals has generally pointed to a division of labor;

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

it seems likely that, in the frontal cortex, as in other parts of the brain, there is functional specialization, though the granularity of this evidence for regional specificity tends to be coarser than the fine cytoarchitectural parcellations suggested by anatomy. Recently, some have questioned whether architectonic fields have relevance to function at all (Hayden 2023), harkening back to similar arguments by Lashley and his colleagues in the 1940s. However, such challenges to orthodox frontal lobe maps have yet to offer precise and testable alternatives for defining the organization of function within the frontal lobes. A related consideration is whether the organization of frontal cortex should be conceptualized in terms of functional and anatomical gradients, instead of discrete areas with well-delineated boundaries. Whereas early sensory areas often have physiological properties that allow one to define clear areal boundaries, the extent to which this can be extrapolated to frontal areas (or indeed, other regions of association cortex) is unclear. Here, we revisit these questions and consider the extant evidence, as well as critical gaps in knowledge.

If There Is Functional Fractionation of the Frontal Lobes, What Would It Look Like?

The underlying premise of frontal cortex neurobiology is that localization of function arises from the unique pattern of afferent and efferent connections within each region, as well as local differences in connectivity and cellular properties, and that this spatial variation supports unique cognitive operations. If this premise holds, we should be able to gain insights into the functional organization of the frontal lobes from the convergence of anatomical and functional methods. Tract-tracing studies in macaques have provided anatomical data to support this idea. In identifying the major connections of individual areas, investigators have observed different patterns of connections across architectonic fields. Modern anatomical methods based on structural and functional magnetic resonance imaging (fMRI) have also contributed to our understanding of the anatomical organization of the frontal cortex. Diffusion-weighted imaging, which allows mapping of the patterns of diffusion of (mainly) water molecules in brain tissue, has been used to study white matter connectivity and integrity. Although this method was initially thought to hold promise for mapping the connections of the human brain, its accuracy is known to be limited by technical factors that are unlikely to be overcome by improved data acquisition or analysis methods (Thomas et al. 2014). Another approach has been to examine "connectional fingerprints" of different frontal lobe regions based on resting-state covariance of activations acquired during fMRI (Mars et al. 2016). This method is particularly useful because it can be applied in both macaques and humans and used to infer homology across frontal cortex regions. The downside of this approach is that covariation in physiological signals between areas does not necessarily reflect actual anatomical connectivity. However,

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

using this approach, the architectonic delineations of Price and colleagues (Öngür et al. 2003) and Petrides and Pandya (1999, 2002) have been largely supported by the resting-state functional connectivity studies in macaques and humans (Mars et al. 2016; Sallet et al. 2013). The one exception is the lack of a rostral lateral region in macaques with a connectional fingerprint matching that of the lateral frontal polar cortex of humans (Balsters et al. 2020; Neubert et al. 2015). Thus, macaques and presumably other simians most likely lack a homologue of human lateral frontal polar cortex. An alternative is that its homologue is relatively small in macaques and related simians. Figure 8.1 illustrates the frontal architectonic subdivisions in humans and macaques.

If frontal areas perform specialized functions, then the anatomical maps should align with functional data. Outside of the motor and premotor areas (see below), no single method has provided a reliable index of functional boundaries, so this information is often inferred from converging techniques. Arguably, within the frontal cortex, the strongest evidence for functional specialization comes from studies of people or animals with brain damage. Whether brain damage is accidental or experimental, it provides unique insight into whether a given bit of the brain is essential for a given behavior (Murray and Baxter 2006; Vaidya et al. 2019). Other loss-of-function experimental methods include reversible inactivations of cells with GABA agonists like muscimol, locally applied pharmacological agents that selectively increase or decrease cell activity, and recently developed chemogenetic methods that use virally delivered constructs in combination with systemically administered activators to shut down processing. Temporary disruption of function in humans is accomplished noninvasively with transcranial magnetic stimulation (TMS). TMS can be used to alter regional cortical function in frontal areas, exhibiting a higher degree of anatomical precision than in patients who have experienced accidental brain injuries. For example, Blumenfeld et al. (2014) observed distinguishable effects on memory function when applying TMS to neighboring sites within the middle versus inferior frontal gyrus. A relatively recently developed noninvasive method for producing regional inactivation of neural tissue is focused transcranial ultrasound stimulation (Folloni et al. 2019; Tufail et al. 2011; Yoo et al. 2011). Unlike TMS, focused transcranial ultrasound stimulation can be applied to deep structures in the brain (Folloni et al. 2019). Like experimentally induced cortical ablations, these methods have provided valuable insights into structure-function relationships within the frontal lobe. Another method for gaining causal insights into structure-function relationships involves applying electrical microstimulation to targeted regions of the brain, which can be performed in animals as well as in patients with depth electrodes placed for neurological disorders (e.g., epilepsy and Parkinson disease). Using logic analogous to that used for revealing topographic maps (such as primary visual cortex, V1), electrical stimulation can reveal systematic body maps in motor and premotor cortex areas, as well as somatic sensory areas (Halley et al. 2020), and might also provide evidence regarding the localization of function

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

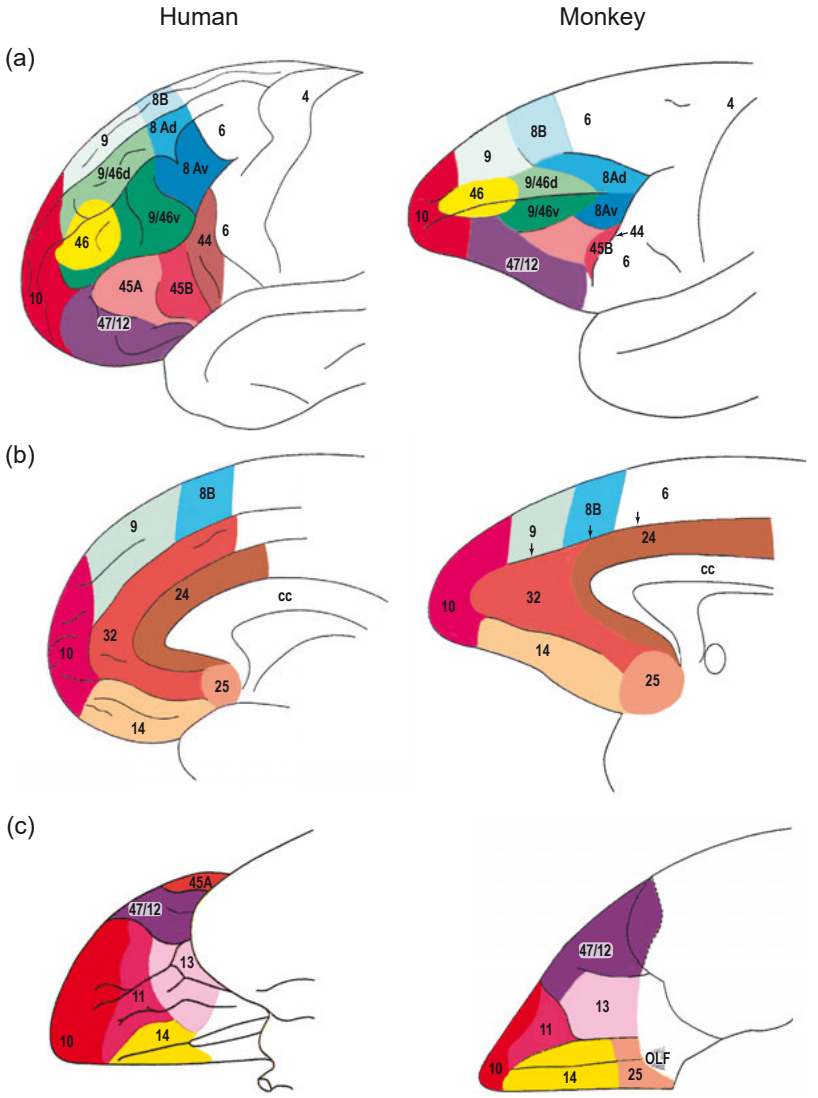

Figure 8.1 Schematic diagram of anatomically defined areas of human (left) and macaque monkey (right) frontal cortex, depicted on lateral, medial, and ventral surface views of the frontal lobe (top to bottom, respectively). Numerals refer to different architectonic fields; rostral is to the left. Adapted from Petrides and Pandya (1999, 2002).

in frontal areas outside the motor areas. Beyond manipulations, electrophysiological recording of the activity of neurons is commonly used to understand brain-behavior relationships. This is most common in awake-behaving monkeys, usually macaques, though there are also some opportunities to record neural signals directly from the brains of neurosurgical patients. Much more

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

widely used, although far less regionally precise methods that also provide evidence for functional specialization include scalp EEG and functional MRI. These methods reveal correlations of neural activity with behavior and can establish whether neural activity patterns differ across areas under particular behavioral conditions (e.g., evidence for encoding of distinct variables in different PFC areas).

Evidence for Functional Fractionation of the Frontal Lobe in Primates

In this section we discuss the evidence for functional fractionation within the frontal cortex of human and nonhuman primates. Detailed reviews are available elsewhere, as are book-length treatments of the topic (e.g., Passingham 2021). Here we evaluate the strength and consistency of the evidence within and across methods as well as the extent to which functional dissociations respect anatomical boundaries, illustrated with some examples.

Electrical Stimulation

Although no single approach is definitive, brain stimulation maps and neuropsychological studies have provided the most compelling data regarding the fractionation of function in the frontal lobes. Electrical stimulation of primary motor cortex, M1, and supplementary motor cortex (SMA), or M2, reveal body maps (Graziano et al. 2002; Halley et al. 2020; Mitz and Wise 1987; Penfield 1954; Woolsey 1963; Woolsey et al. 1952). Within these cortical areas in primates there are well-characterized and consistent stimulation-elicited movements arranged systematically according to body part. These data provide clear evidence of modularity of function within the anatomically defined areas M1 and SMA. In nonhuman primates, in addition to the SMA, five additional premotor areas have been identified (Dum and Strick 2002; Luppino et al. 1991): the dorsal and ventral premotor areas (PMd and PMv) and three cingulate motor areas. Like M1, each of these premotor areas has substantial direct projections to the spinal cord. It is possible to evoke movements of the distal and proximal forelimb using intracortical stimulation at relatively low currents in all six of the established premotor areas.

In addition, stimulation of the frontal eye fields (FEF), which reside in the arcuate sulcus of the macaque, yields a systematic map of the contralateral visual field in monkeys (Bruce and Goldberg 1985) as does stimulation of its presumed homologue in humans (Blanke et al. 1999). Electrical stimulation of FEF reliably induces saccades of a particular direction and amplitude (Bruce et al. 1985; Robinson and Fuchs 1969), providing evidence for modularity of function within FEF that is based on saccade direction relative to the current eye position.

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

More recently, investigators have used modified population receptive field modeling of fMRI measurements to define visual areas across individuals and species. Using this technique, two visual field maps of contralateral space have been identified along the superior and inferior portions of the precentral sulcus in humans (Mackey et al. 2017). The map in the superior precentral sulcus is thought to be the homologue of the macaque FEF (Vernet et al. 2014); alternatively, this region could be one of several premotor oculomotor representations (Passingham 2021; Schall et al. 2020). At least some evidence suggests the macaque FEF might also contain two topographic maps (Savaki et al. 2015). Critically, these visual maps in frontal cortex, together with the body movement and eye movement maps evoked by electrical stimulation, constitute well-defined anatomical units that researchers can reliably target for study; they serve not only as a basis for alignment of maps generated by fMRI and other methods, but also offer a view of the flow of information from prefrontal areas to output effectors (e.g., eyes, head, forelimbs, hindlimbs) in humans and macaques.

Neuropsychological Studies

Human neuropsychological studies based on patients with accidental brain injury, such as traumatic penetrating head injury, brain damage incurred by stroke, tumor removal, or ruptured aneurysms, have been pivotal in identifying functional zones within the frontal lobes. These cases have provided causal evidence for functional fractionation of the human frontal cortex, with a level of explanation that has immediate relevance to the clinic. This method also has constraints: lesions incurred in humans most often are moderate in size, with varying degrees of overlap at regional/subregional levels, as well as involvement of underlying white matter which may lead to dysfunction beyond the anatomically defined lesion boundaries. Voxel-based lesion-symptom mapping can, in principle, improve the spatial resolution of functional inferences, but in practice is limited by available sample size and non-independence of how patterns of damage relate to the etiology of the damage. Lesion etiology limits what can be tested: patterns of injury typically segregate in lateral frontal (LF), dorsomedial prefrontal (DMF), and ventromedial-orbitofrontal lobes (VMF/OF). Damage to VMF/OF and DMF is often bilateral (to varying degrees); even unilateral lesions likely disrupt callosal integrity, thereby introducing the possibility of disconnecting regions in the other hemisphere. LF damage, in contrast, is rarely bilateral. The level of anatomical resolution that can be tested in human lesion studies is typically no finer than 3–6 regions: motor/premotor, DMF, VMF/OF, LF, and frontal pole. Studies variably consider the effects of damage to the left versus right hemisphere versus both.

Given the possibility of nonspecific or "off target" effects of brain injury, or illness more generally, double dissociation provides the strongest evidence for regional specialization. In such studies, two cohorts with lesions affecting

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

different frontal cortex areas are compared on two or more functional assessments, administered as tasks. If one cohort demonstrates impairment on task A but not B, while the second cohort is impaired on task B but not A, this double dissociation is strong evidence that the functions assessed by tasks A and B are independent and depend critically on different neural substrates. Single dissociations (e.g., when a lesion of a particular area causes impairment in task A but not B) are also relevant. Compared to double dissociations, however, single dissociations are more open to alternative, nonspecific explanations such as general differences in task difficulty or reliability (Vaidya et al. 2019).

There are multiple examples of double dissociation in humans with regionally specific focal damage within the frontal lobes. For example, patients with VMF/OF damage are impaired in probabilistic stimulus-reward reversal learning, i.e., choices between two "objects" (decks of cards) yielding different monetary outcomes. They perform similarly to healthy controls in a task with the same dynamic reward structure, but where a reward is associated with one of two actions. The opposite pattern of results was obtained in humans with damage centered in the dorsal anterior cingulate cortex (ACC), part of the DMF (Camille et al. 2011b). Adding further assurance (and allowing more anatomical specificity), this finding replicates a similar observation in macaques with experimentally induced lesions to either orbitofrontal cortex (OFC) or ACC (Rudebeck et al. 2008b). There is a larger literature showing regionally specific lesion effects (single dissociations) across prefrontal cortex (PFC), tested with either multiple regions of interest or voxel-based lesion-symptom mapping (e.g., Gläscher et al. 2009; Tsuchida and Fellows 2013).

Lesion studies can also help to dissect distinct structure-function mappings that separately contribute to overall task performance. For example, damage to DMF and left LF, but not VMF/OF, disrupts different aspects of working memory performance in a two-back task (Tsuchida and Fellows 2009). Lesion studies can also fail to find task dissociations, which could be considered evidence that two tasks are drawing on the same component process. For example, performance on both Stroop and task-switching tasks is impaired after left LF lesions (Tsuchida and Fellows 2013), suggesting that these tasks tap into a common underlying function carried out by LF. Given the inherent heterogeneity of lesions, however, together with individual differences in structure-function relationships, studies of this type provide relatively weak evidence. Stronger conclusions can only be drawn by considering multiple lines of evidence.

As indicated above, brain damage in humans rarely respects anatomical boundaries and may affect underlying white matter pathways. As a result, based on human neuropsychological studies alone, it has been difficult to refine frontal cortex function beyond the broad anatomical regions outlined above. TMS in humans can produce more localized effects, but its influence is limited to frontal areas at or near the surface of the cranium. More fine-grained causal tests for structure-function relationships, respecting areal boundaries, require experimentally controlled lesions or other causal regional manipulations in

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

nonhuman animals. Because new PFC areas emerged in early primates (Preuss and Wise 2022), and because extant nonhuman primates like macaques (Old World monkeys), marmosets (New World monkeys), and humans inherited these new PFC regions from a common ancestor, nonhuman primates have been indispensable for unraveling the function of the PFC.

Neuropsychological studies in macaques and marmosets have identified specialized functions for several parts of granular PFC. Here we focus on a few studies and paradigms that yield key findings. For example, within the ventral PFC, multiple studies have found doubly dissociable effects of lesions of OFC (areas 11/13/14) versus ventrolateral PFC (area 12/47) (Baxter et al. 2009; Dias et al. 1996a; Rudebeck et al. 2017b). In one study, monkeys performed two different tasks requiring them to take into account stimulus-reward-value associations while performing object choices (Rudebeck et al. 2017a). Both tasks manipulated reward value and in each, task performance indexed the ability to update rapidly object-reward-value associations. However, one task required updating of the desirability of food based on internal state, whereas the other required updating of the availability of food based on external contingencies. Selective lesions of OFC (areas 11/13/14) led to severe impairments on the task requiring value updating based on internal state but no impairment on the task requiring updating based on external contingencies, whereas selective lesions of ventrolateral PFC (area 12/47) led to the opposite pattern of results. This result supports the idea of fractionation of function in the ventral frontal cortex and is consistent with the idea that different types of specialized representations reside in OFC versus ventrolateral PFC (for review, see Murray and Rudebeck 2018; Rudebeck et al. 2017a).

An even finer fractionation of function within these areas was achieved using reversible inactivation. In one study, temporary inactivation of caudal OFC area 13 but not rostral OFC area 11 led to impairments in updating values based on changes in internal state (i.e., satiety). By contrast, temporary inactivation of rostral OFC but not caudal OFC led to a selective impairment in choosing between visually presented objects based on that updated value (Murray et al. 2015). In this case, the increased temporal specificity of pharmacological infusions over permanent lesions was critical to revealing a finer dissociation of processes involved in reward updating; inactivations delivered to different regions at different points in the task (before versus after satiation) produced distinct effects. Consistent with this finding, a human fMRI study found dissociable activations within OFC in a stimulus-reward task employing multiple foods. Participants in the experiment first learned a variety of arbitrary image-food associations. An important aspect of the design was that multiple images mapped onto individual foods. Then, using a repetition-suppression design, the investigators showed that rewards (i.e., specific foods) activated caudal OFC area 13 whereas stimulus-reward associations led to activation of rostral OFC area 11 (Klein-Flugge et al. 2013). This finding supports the idea that these OFC subregions have analogous functions in macaques and humans.

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

Additional fMRI studies in humans yield findings consistent with those in macaques; OFC activations reflect changes in food value that accompany object choices (Howard and Kahnt 2017). Finally, damage to the VMF/OF in humans, like damage to OFC in macaques, results in impairments in choice behavior in humans that resemble what is observed in macaques (Reber et al. 2017); they are impaired at switching their choices from the objects leading to a sated food to those leading to a nondevalued food. There is, however, an obvious difference between the human and macaque studies: We can ask humans why they made their choice. Humans with damage to VMF/OF explicitly indicate that after selective satiety, they no longer want the sated food, despite the fact that they usually make the choice that leads to getting that food. This points to a disconnection of knowledge and action that is evident in the choice behavior of humans with damage to VMF/OF sectors of the frontal lobe, and which resembles the choices of macaques with selective inactivations within OFC (Murray et al. 2015; Reber et al. 2017).

Notably, there are dissociable functions *within* areas consequent to selective neurotransmitter depletions. Using tasks known to be dependent on OFC—in this example, stimulus discrimination extinction—it has been shown that depletions of either serotonin or dopamine produce different patterns of behavioral impairment. For example, marmosets with OFC serotonin depletion showed an inability to overcome their bias toward responding to the previously rewarded stimulus, whereas those with OFC dopamine depletion were not biased toward the previously rewarded stimulus but nevertheless persisted in responding in the absence of reward (Walker et al. 2009). These and related results point to ways in which monoamine neurotransmitters can influence PFC-dependent behavior in regionally specific ways (Clarke et al. 2004, 2007; Walker et al. 2009).

There is also evidence for fractionation of function at the resolution of architectonically defined regions of medial frontal cortex and OFC with respect to threat reactivity. Activation of marmoset subgenual cingulate area 25 appears to induce an overall negative state, biasing basal cardiovascular activity toward sympathetic control, increasing reactivity to predictable as well as unpredictable threats, and enhancing avoidance of threats in an approach-avoidance task (Alexander et al. 2019, 2020; Wallis et al. 2019). In contrast, activation or inactivation of areas 14, 11, or 13 has no impact on basal cardiovascular activity and has more selective effects on threat responsivity. Specifically, whereas activation of area 14 produces little reactivity to predictable, certain threat, it increases reactivity to uncertain threat (Stawicka et al. 2020). Enhanced reactivity to uncertain threats is also seen in relation to areas 13 and 11, but in contrast to area 14 and 25, it is inactivation of these regions rather than their activation that heightens reactivity to uncertain threats (Stawicka et al. 2022). There are also clear distinctions between the effects of inactivation of area 11 and area 12/47 with respect to negative biasing as a consequence of threats in approach avoidance. Whereas

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

inactivation of area 12/47 biases responding away from threats at the time of threat exposure, inactivation of area 11 had no such effect; instead its effects, which involve the enhancement of negative bias in responding, are only observed the next day, likely the result of an altered threat memory (Clarke et al. 2015). This overall pattern of results whereby granular PFC regions, as distinct from agranular cingulate cortex, have a greater role in contexts of uncertain threat is consistent with the findings from Mobbs and colleagues in humans. According to these investigators, prefrontal regions are only engaged when a threat is distal (in time, space or probability) and there is time to engage PFC mechanisms (e.g., in OFC and ventrolateral PFC) compared to when the threat is proximal and rapid response selection is required (Mobbs et al. 2020).

Within dorsal PFC regions there is also evidence for functional specializations that map onto anatomical subdivisions. Here, the parcellation of Petrides and Pandya shows three distinct subdivisions: area 9, area 46, and area 9/46. As shown in Figure 8.1, area 46 occupies the banks of the rostral half of the principal sulcus in macaques, area 9/46 occupies the caudal half, and area 9 lies above (dorsal and medial to) the principal sulcus. It has long been known that areas 46 and 9/46, typically referred to collectively as dorsolateral PFC, are essential for performance of delayed response tasks, including delayed response and delayed alternation. Although the original reports were based on aspiration lesions (Goldman et al. 1971), more recently the result has been confirmed using a more selective method: chemogenetic inactivation (Upright et al. 2018). There may be even further fractionation of function within this region. It has been suggested that the two regions have specialized functions. Passingham points out that based on anatomical projections, the more rostral of these dorsal PFC regions, area 46, is likely involved in identifying goals for saccadic eye movements (Passingham 2021:226). These regions are active when monkeys need to learn to perform sequences of movements (Averbeck et al. 2006) and are essential when monkeys make judgments about temporal order (Petrides 1991).

A key point related to the foregoing discussion is that the circuitry dedicated to reaching movements and eye movements is not only dissociable, but has different functions in choice behavior. The breakthrough concept is that saccades are not a movement so much as a mechanism for orienting attention (overt attention in this case), whereas reaching is not a mechanism for orienting attention. Selective pressures would have operated differently on circuits for eye movements and those for arm movements, for the simple reason that no primate ever "grasped" anything with an eye movement.

The findings reviewed above—based on effects of lesions, temporary inactivations, and pharmacological manipulations—provide strong evidence for functional fractionation within the frontal lobes. Next we consider the evidence for task-based regionally specific patterns of activity, first in humans (fMRI), then in nonhuman primates (neurophysiology).

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

Functional Magnetic Resonance Imaging

In general, fMRI in humans takes the approach of investigating differences in the blood-oxygen-level-dependent (BOLD) response across experimental conditions as a means of testing functional fractionation. Such differences can be tested as a univariate difference in overall voxel activity, differences in decoding or similarity matrices (i.e., representational similarity analyses), or patterns of functional connectivity. Fractionation is considered evident in region by effect interactions, such that the differences between conditions can be shown to change as a function of region. Although these interactions are sometimes interpreted as dissociations, it is rare to find full cross-over double dissociations (Chatham and Badre 2012, 2020; Fletcher and Henson 2001). Evidence of single dissociations distinguishing regions of the frontal lobe is much more common from fMRI. Although the latter suggest functional differentiation, they come with limitations to inference.

An advantage of fMRI is that it allows for measurement of activity while humans are performing a wide range of tasks. This is particularly important for studying the human frontal lobe, as it permits the study of the kinds of complex and higher-order tasks for which the frontal lobes are thought to be crucial. It follows that fMRI is one of the primary sources of evidence for functional differences across regions of the frontal lobe. Studies using fMRI have located evidence for the coarser frontal lobe distinctions that are supported by multiple sources of evidence, such as between VMF/OF, LF, and DMF. However, it is also a source of evidence for finer grained differences in function. For example, differences in fMRI activity within the LF, specifically between dorsal premotor cortex and dorsolateral PFC, have been observed based on demands for sensorimotor versus cognitive control (Badre and D'Esposito 2009; Nee and D'Esposito 2016; Badre, this volume). In rarer cases, these finer regional differences observed in fMRI have been supported by convergent evidence, such as from lesions in human patients. For example, patients with lesions in regions overlapping the zones activated in the aforementioned fMRI studies exhibited a pattern of behavioral deficits consistent with a hierarchical relationship between sensorimotor and cognitive control (Azuar et al. 2014; Badre et al. 2009). Likewise, TMS of these subregions, guided by fMRI, yielded a similar pattern of deficits (Nee and D'Esposito 2017).

More routinely, however, observations of regional differentiation with fMRI at a finer scale have not seen convergent evidence from other methods. For example, several fMRI studies have consistently reported activation in the LF polar cortex (i.e., the most rostral portion of the LF cortex), when processing abstract functions such as exploration over exploitation (Badre et al. 2012; Daw et al. 2006) or counterfactual predictive task-set processing (Koechlin and Hyafil 2007). However, corresponding lesion evidence in humans or lesion or physiological evidence in animals has not been reported.

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

Crucially, in this case, this failure may be due to methodological limitations, such as the ability to train animals in tasks hypothesized to involve frontal pole or the lack of homologous areas across species. As indicated earlier, it appears that macaques lack a homologue of human LF polar cortex (Neubert et al. 2015). Nonetheless, at finer granularity, fractionation of function in PFC is often supported only by fMRI evidence.

Of course, evidence from fMRI is limited in several ways. These include the indirect and correlative nature of the signal, its lack of temporal resolution, and in some cases, smaller effect sizes with limited samples. With regard to understanding the fractionation of frontal lobe, outside of the premotor and motor regions discussed above (see section on Electrical Stimulation) it is challenging to localize activity acquired from fMRI with reference to a map that aligns with anatomical features and allows for comparison across individuals and species. Thus, findings of functional differences in patterns of activation across regions observed using fMRI, even ones found repeatedly and reliably, are only a starting point. Investigations using causal methods and detailed physiological analysis are crucial. Such work would benefit from a more refined anatomical framework to align findings across individuals.

One recent advance has come from anatomically aligning fMRI data with tertiary sulci, which are small, shallow sulci that show a good deal of variation in presence and location across individuals (Weiner 2023). In one example, investigators aligned data from individual subjects to their (variably located) paraintermediate frontal sulcus and found that the sulcus marked a transition in function within the LF cortex (Willbrand et al. 2023a). Thus, anchoring data to tertiary sulci may be a way to overcome at least some individual differences in brain shape and structure. Rather than averaging data across brains, which tends to blur the pattern of activations, anchoring activations to a tertiary sulcus before averaging allows finer structure-function mapping.

Neurophysiology

The suggestion from methodologies such as neuropsychology or fMRI that a particular region of PFC subserves a given function has, in many cases, led researchers to seek neurophysiological correlates of those functions in the same area to understand underlying mechanisms. These approaches have led to a wealth of data showing that the activity of neurons in PFC can encode or represent a wide range of information, from external stimuli or motor responses to reward expectations to abstract concepts and rules. Based on the findings of double and single dissociations, investigators have expected to observe neuronal activity that not only reflects the differences in function, but also serves as the origin of it. To date, distinctions in the neurophysiology of different frontal regions have, however, been much less clear cut than many would have expected. As reviewed by Rich and Averbeck (this volume), this is true even at a coarse level of anatomical parcellation, where evidence for

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

functional dissociations using other methods is quite strong. For instance, the lateral PFC is strongly implicated in cognitive control functions, including the use of rules and strategies; however encoding of rules and strategies is found not only in lateral PFC, but also in other areas such as OFC (Fascianelli et al. 2020; Wallis et al. 2001; Yamada et al. 2010). Conversely, OFC and neighboring PFC regions are involved in evaluation and value-based decision making, but decision-relevant information is strongly represented not only by OFC neurons but by those in lateral PFC (Leon and Shadlen 1999; Roesch and Olson 2003; Tsutsui et al. 2016b; Watanabe 1996) as well as medial PFC (Cai and Padoa-Schioppa 2012; Chien et al. 2023; Kennerley and Wallis 2009a; Matsumoto et al. 2003).

Despite the encoding of many variables across multiple frontal regions, there are some counter examples. Some reports show clear-cut differences in the activity of neurons across regions. For example, in a study by Tsujimoto and colleagues, who recorded neurons in three frontal cortex regions while monkeys performed a cued strategy task, only neurons in frontal polar cortex signaled responses that were correct according to the cued strategy (before feedback); only OFC neurons signaled the response that had been made (after feedback), whether correct or incorrect; and dorsolateral PFC encoded responses in a modality specific way. These signals support a role for dorsolateral PFC in generating responses, a role for OFC in assigning outcomes to choices, and a role for frontal polar cortex in assigning outcomes to cognitive processes (Tsujimoto et al. 2012). In addition, a few consistent trends across studies can be found. Perhaps most clearly, dorsolateral regions tend strongly to represent factors related to space, including action or attention that is directed in space, and these variables are typically poorly represented by ventral regions such as the OFC (reviewed by Rich and Averbeck, this volume). This is generally consistent with the idea that LF cortex plays a role in translating goals to actions (Averbeck and Murray 2020; Cai and Padoa-Schioppa 2014). In addition, there have been many reports of small but significant distinctions in the proportion of neurons encoding different types of information. For instance, when monkeys chose a rewarding cue or rewarding action, more neurons in the dorsal ACC, compared to OFC, tended to encode actions, whereas more OFC neurons encoded stimuli (Luk and Wallis 2013). This is consistent with the human and monkey neuropsychology data reviewed in the section above, where damage to OFC and ACC disrupt the assignment of value to stimuli or actions respectively. However, in the neurophysiology study, the magnitudes of the encoding biases were small and only found briefly, in one phase of the task. Most studies focus on the small differences because the differences are consistent with the hypothesis of functional localization. However, the strongest patterns in the data indicate widespread encoding in PFC of most variables at roughly comparable levels.

In another example, analysis of neural activity during a baseline "hold period" in a reinforcement learning task, rather than during the trial itself,

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

revealed that OFC neurons maintain a representation of values and target stimuli, whereas lateral PFC regions had only a weak representation of these variables (Tang et al. 2022a). Once the choice options were presented, directional analyses indicated that value and identity information flowed to dorsal circuits. This is consistent with other cases, where differences in the timing of responses can suggest a flow of information from one region to another. For instance, similar proportions of neurons in OFC and dorsolateral PFC encode rewards, but encoding begins about 80 ms earlier in OFC, again suggesting that this information is passed from OFC to dorsolateral PFC to influence behavior (Wallis and Miller 2003b). Taken together, there are small differences in neural encoding across prefrontal areas, and these support the idea of functional fractionation. However, these differences are embedded in a predominant pattern of similarity across regions that to date has made neurophysiology one of the less useful methodologies for distinguishing functional regions of PFC or establishing finer grained parcellations of functional areas. Viewed from another perspective, perhaps differences in encoding have more to do with differences in the areas to which each subregion of PFC projects, or to the distributed nature of representations, as opposed to functional parcellations. Nonetheless, neurophysiology is the most direct method of investigating mechanisms that produce complex cognition and behavior and is, therefore, a critical component of understanding the functional organization of PFC.

Taken together, convergent findings from multiple methods have led to a widespread consensus that there are distinct functional specializations within frontal cortex. What remains to be elucidated, however, are the degree and particulars of finer parcellations, the computation that each region contributes, how it participates in larger networks, and how behavior emerges from interactions of those distributed networks.

What Are the Organizational Principles of the Frontal Lobe?

Although much remains to be learned about the organization of function in the frontal lobes, several organizational principles are evident. Here we consider these general principles in the hope they will inform theory and thereby speed progress toward a more thorough understanding of frontal lobe function.

As mentioned above, the functional fractionation of PFC regions emerges in part from specialized, topographically defined inputs and outputs. In this regard, one can consider connectivity between frontal lobe regions, cortico-cortical connectivity more broadly, and cortico-subcortical circuits. After addressing potential hemispheric specialization of function, we discuss cortico-subcortical connectivity, in part because these subcortical inputs and outputs substantially influence frontal lobe neuronal activity and 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

Hemispheric Specialization

All mammals have a hemispheric specialization of premotor and motor areas that control the movement of the contralateral limbs and eye movements that direct gaze into the contralateral visual field. Beyond this specialization for motor control, there is abundant evidence for hemispheric specialization of function in humans but little or no evidence for hemispheric specialization in macaques. In humans, one of the strongest specializations involves speech and language processing in the left hemisphere. Even here, however, specialization is relative; with few exceptions, both hemispheres can process most types of information.

In both humans and macaques, evidence suggests that visual working memory in the PFC operates largely independently within each hemisphere, with each processing information in the contralateral visual hemifield. For instance, working memory capacity limitations depend on the number of memoranda per visual hemifield and are generally unaffected by stimuli presented in the opposite (unattended) field (Buschman et al. 2011; Delvenne 2005; Umemoto et al. 2010). Similarly, neurons tend to show stronger encoding of contralaterally presented cues (Brincat et al. 2021; Funahashi et al. 1990; Kastner et al. 2007; Kornblith et al. 2016; Luria et al. 2016; Rainer et al. 1998). However, beyond spatial specificity, the processes carried out in each hemisphere appear similar, and under natural viewing conditions, information is likely to be transferred rapidly from one hemisphere to the other (Brincat et al. 2021).

Although the organizing principles of PFC lateralization remain unclear, there are some examples of lateralized structure-function lesion effects in the PFC, beyond language and motor processes, where lateralization is very well established. In one human lesion study (Geddes et al. 2014), effective interference resolution was found to require either right or left lateral PFC, depending on the nature of the task. In another study (Stuss and Alexander 2007), the left lateral PFC was found to play a pivotal role in task-setting—a function that entails the establishment of a stimulus-response relationship—whereas the right lateral PFC was engaged in monitoring processes involving the continuous assessment of task performance for quality control and the implementation of required behavioral adjustments.

Another interesting domain of specialization involves affect. In humans, evidence suggests that posterior regions of the right hemisphere are specialized for the interpretation of emotional information, including information contained in tone of voice and facial expressions. In addition, anterior regions of the right hemisphere are specialized for the production of emotional cues (e.g., facial expressions) that serve a communicative function. Correlates of mood states, while represented bilaterally, show some asymmetry. For example, fMRI studies suggest that greater activation of left than right frontal regions is associated with positive mood and approach behaviors. In contrast, greater activation of the right than left frontal regions is associated with

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

negative mood and avoidance behavior (Davidson 1992). It has been proposed that this asymmetry is due to the asymmetric autonomic innervation of the heart (Craig 2009). In addition, there is an asymmetry in the effects on autonomic output following electrical stimulation of the insular cortex where, in humans, there appears to be right-sided dominance for sympathetic effects (Oppenheimer et al. 1992).

Cortico-Basal Ganglia-Thalamic Loops

All cortical areas, including the frontal lobes, participate in cortical-basal ganglia-thalamocortical "loops" (Alexander et al. 1986). This fundamental loop architecture involves a series of projections from cortex to striatum, striatum to pallidum, pallidum to thalamus, and, finally, thalamus back to cortex. Importantly, the loops project back to the same regions of cortex from which they originated. Because this feature of cortical organization is well known, and discussed extensively elsewhere, we will not repeat it here. That said, we note that cortico-striatal connections are more complex and less segregated than stated above, and that interactions between functional territories are extensive. Thus, within the striatum, there appears to be integration of information across what are classically considered reward, cognitive, and motor territories of the frontal cortex (Haber 2016). In addition, the existence of "focal" and "diffuse" cortical projections to the striatum opens the possibility that these two termination patterns serve different functions (Haber et al. 2006; Watakabe et al. 2023).

Cortico-Thalamo-Cortical Connectivity

In mammals, the frontal cortex and thalamus are anatomically interconnected and share a common developmental trajectory. Several thalamic nuclei connect directly with the frontal lobes including the mediodorsal (MD) thalamus, motor thalamus, anterior thalamus, pulvinar, intralaminar nuclei, and the nucleus reuniens. Different thalamic neurons provide either targeted, or more diffuse, frontal inputs, replicating patterns of thalamocortical connectivity across different thalamic nuclei, now referred to as thalamocortical motifs (Halassa and Sherman 2019). Each thalamic nucleus also has reciprocal modulation with Layer VI of the frontal lobes via the reticular thalamic nucleus (Halassa and Sherman 2019). The entire frontal cortical mantle is reciprocally interconnected to different MD subdivisions. These MD thalamocortical projections in primates target deep Layer III and Layer IV, while Layer V projects directly back to each of these MD subdivisions, or indirectly via cortico-striatal-thalamic loops (Barbas et al. 1991; Giguere and Goldman-Rakic 1988; Goldman-Rakic and Porrino 1985; Porrino et al. 1981; Ray and Price 1993; Saunders et al. 2005; Schwartz et al. 1991; Timbie and Barbas 2015; Xiao et al. 2009). In human neuroimaging, multi-domain thalamic network hubs have now been identified

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

(Hwang et al. 2021; Shine et al. 2023). These cortico-thalamo-cortical circuits are consistent with the idea that frontal cortico-thalamic interactions are essential to cognitive function (e.g., for review, see Mitchell 2015; Perry et al. 2021).

Gradients

Within PFC, at least two spatially organized gradients of anatomical circuitry can be discerned. First, two large-scale anatomical circuits are evident: dorsal and ventral. The ventral PFC regions, including OFC (areas 11, 13, 14) and ventrolateral PFC (areas 12/47), are part of a larger network that has prominent connections with the amygdala and inferior temporal visual cortex, the ventral striatum, the medial portion of MD, and the hypothalamus. The dorsal PFC is part of a network that has prominent connections with parietal cortex, the dorsal striatum, the lateral portion of MD, and few connections with the hypothalamus (Averbeck and Murray 2020). This pattern of connections suggests that ventral and dorsal PFC regions have distinct functions. Specifically, it has been proposed that, operating in the networks in which they are embedded, the ventral and dorsal PFC define behavioral goals and orchestrate behavior to achieve behavioral goals, respectively (Averbeck and Murray 2020; Giarrocco and Averbeck 2023; Marquand et al. 2017; O'Reilly 2010).

The dorsal-ventral dichotomy outlined above can be viewed as part of a larger medial versus lateral pattern that is evident in all mammals. Comparative neuroanatomical studies have revealed that cerebral cortical organization can be viewed as a set of concentric rings around a core of eulaminate cortex, with the core containing, among other things, primary sensory areas S1, A1 and V1. Medial to the core is cortex with one developmental origin, and lateral to the core is cortex with a different developmental origin. Thus, the mammalian neocortex can be described as two sheets (Cisek 2022). As indicated earlier, after the divergence of rodent and primate lineages—roughly 80 million years ago—additional frontal cortex regions emerged in primates. We note that calling this pattern a "gradient" is a convenient label; there is no evidence that the frontal neocortex evolved in an ordered sequence (Murray et al. 2017). The emergence of new frontal and parietal cortex areas eventually led to the long-range frontoparietal connectional networks described next (Figure 8.2).

A second spatially organized gradient of circuitry in frontal cortex involves rostro-caudal patterns of connections (Murray and Constantinidis, this volume). Setting aside the details of point-to-point projections makes it easier to see this organization, which essentially looks like a series of reciprocally related concentric bands. Specifically, frontoparietal circuits are topographically organized such that primary somatosensory cortex and primary motor cortex are reciprocally related, the posterior parietal and premotor areas are reciprocally related, the inferior parietal and ventral premotor areas are reciprocally related, and, finally, the medial parietal and adjacent areas in the posterior intraparietal sulcus and dorsolateral prefrontal regions are interconnected

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

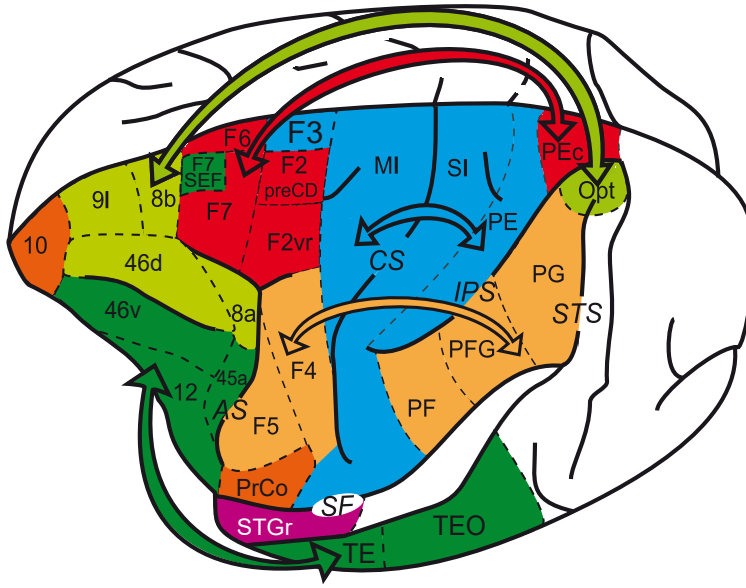

Figure 8.2 Patterns of frontoparietal and frontotemporal connections in macaques. Arrows indicate reciprocal anatomical projections between regions. Adapted from Giarrocco and Averbeck (2021). Sulcal abbreviations: AS, arcuate sulcus; CS, central sulcus; IPS, intraparietal sulcus; SF, Sylvian fissure; STS, superior temporal sulcus. F2-F7, areal abbreviations for premotor cortex regions identified by Matelli et al. (1985, 1991). PF, PG, PFG, Opt, PE, PEc, areal abbreviations for parietal cortex areas identified by Pandya and Seltzer (1982). Areas 8a, 8b, 9, 45a, 46v, 46d, 10, 12, areal abbreviations for areas identified by Petrides and Pandya (2007). PrCo, precentral opercular area; STGr, rostral superior temporal gyrus; TE, rostral inferior temporal cortex; TEO, caudal inferior temporal cortex.

(Cavada and Goldman-Rakic 1989; for review, see Giarrocco and Averbeck 2021).

Additional gradients may be evident in neurotransmitters and their respective receptor distributions within frontal cortex. For example, a recent report revealed frontal cortex regional differences in receptor densities, based on analysis of 14 distinct receptor types. In general, rostral frontal areas were characterized by higher receptor densities, whereas more caudal areas had lower receptor densities. This information was combined with information about MR-based connectional fingerprints and cytoarchitecture to suggest novel cortical subdivisions. Thus, the rich information about laminar and regional receptor distributions may provide additional insight into the molecular structure underlying the fractionation of function within the frontal cortex (Rapan et al. 2023). Similarly, a consideration of the combined morphological, electrophysiological and transcriptomic properties of neurons may yield insight into the functional organization of the frontal lobes (Gouwens 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

Networks

Functional imaging studies have led to the description of several large-scale systems or networks of functionally interconnected brain regions. As discussed earlier, these functionally connected networks emerge from observations of covariation in fMRI activations (i.e., the temporal association between the patterns of activations in two or more brain regions) and have been proposed to be important for particular aspects of brain function. As summarized by Gratton (this volume), the PFC possesses several networks, which has led to the idea that different networks—as opposed to different architectonic areas—might carry out different aspects of PFC function (e.g., specific types of executive function or cognitive control). The network approach has also identified "hubs," regions that have connections distributed across multiple networks. Although there is as yet no consensus regarding assignment of particular functions to specific networks, the network approach may offer insights into regional interactions both within the PFC and between the PFC and other brain regions at a systems level (Menon and D'Esposito 2022).

Hierarchy

Anatomically, the laminar origin and termination of projections have been used to classify a connection as "feedforward," "feedback," or "lateral." This classification was initially used in the visual system, although it applies to other sensory systems as well. For example, projections identified as feedforward have their origin in deep layer 3, whereas projections identified as feedback originate in (typically) layers 5 or 6.

The feedforward and feedback architecture has been used to infer cortical hierarchy: feedforward connections are efferents from regions lower in the hierarchy toward regions higher in the hierarchy, and feedback are the inverse. Another idea is that hierarchy can be based on the asymmetry of connections: regions higher in the hierarchy exhibit more efferent connections to regions lower in the hierarchy than to those higher in the hierarchy. These ideas have important implications for computational models of cortical and network interactions.

In general, the frontal cortex exhibits a feedforward pattern of projections from rostral eulaminate to caudal dysgranular and agranular regions. For example, according to the laminar-based hierarchy, rostral OFC area 11 exhibits a feedforward projection to caudal OFC area 13 (i.e., mainly layer 3 neurons in area 11 give rise to projections to the deep layers in area 13), which in turn feeds forward to the caudal orbital agranular insular areas (Barbas 2000; Carmichael and Price 1996). On the lateral surface, this model suggests that area 10 is located at a higher level than more posterior regions, namely areas 45, 46 and 8A.

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

Unfortunately, the laminar- and asymmetry-based definitions of hierarchy do not always agree. For example, the asymmetry-based model reveals that area 10 does not sit atop the hierarchy as would be predicted (Goulas et al. 2014). In addition, both laminar- and asymmetry-based classifications may mask other fine-grained differences in connectivity that inform modes of communication (Rockland 2022).

As noted above, an anterior-posterior hierarchical specialization has been suggested within the lateral PFC based on anatomical and imaging studies, with more abstract operations localized anteriorly on the prefrontal surface (Badre, this volume; Badre et al. 2009; Koechlin et al. 2003). Neurophysiological evidence supports this idea: neurons with shorter response latencies, smaller receptive fields, and greater selectivity for stimulus properties are encountered in posterior regions of the PFC, but neurons responsive to more abstract qualities are more frequent in anterior areas (Riley et al. 2017). Plasticity of responses, dictated by task demands, is also more prominent in anterior areas (Riley et al. 2018).

Direct evidence of systematic variation of plasticity markers between eulaminate and agranular areas has been documented in the PFC. For example, the expression of calcium/calmodulin-dependent protein kinase II (CaMKII), which is essential for plasticity, is greater in medial frontal areas 25 and 32 relative to polar PFC area 10 and dorsolateral PFC area 46. By contrast, markers of cortical stability, including intracortical myelin, perineuronal nets, and parvalbumin show the reverse pattern (Garcia-Cabezas et al. 2017). Changes in neuronal morphology, molecular profiles of the synaptic apparatus, and the influence of neuromodulator systems have also been implicated in long-term prefrontal plasticity (Laroche et al. 2000; McEwen and Morrison 2013) and may differ between areas. Finally, short-term synaptic plasticity, depression, or facilitation has been documented in the PFC, and this too may be critical, particularly for task-related plasticity (Hempel et al. 2000).

How Do the Frontal Lobes Influence Behavior?

There is an extensive literature on the role of the PFC in executive function. This summary term provides a succinct way to discuss the planning and control of behavior (sometimes called cognitive control), the withholding of behaviors, and the pursuit of both immediate and long-term goals over hours, days, months, or years into the future, including embedded, intermediate, and nested goals and strategies for achieving such goals (for review, see Friedman and Robbins 2022). We have discussed frontal lobe function in other terms, but for readers interested in a consideration of executive function, we recommend discussions offered by Shenhav et al. (this volume) and Duncan and Friedman (this volume). Here we focus on just a few of the many ways in which the frontal lobes interact with other regions to influence 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

Cortico-Striatal-Thalamocortical Interactions

As indicated in the prior section, the PFC is embedded in a larger network of areas including cortical-cortical and cortical-subcortical projections. Efforts to understand the functional organization of PFC, therefore, need to take these larger networks into account. As discussed by Rich and Averbeck (this volume), there is a topographic organization of the cortical-cortical and cortical-subcortical circuits. At the highest level, ventral-medial PFC (e.g., area 25) and caudal OFC (area 13) are connected to the ventral striatum, ventral pallidum, and medial, magnocellular mediodorsal thalamic nucleus. The lateral PFC (e.g., area 46) is connected to the dorsal striatum, dorsal GPi, and lateral, parvocellular MD.

Consistent with this network organization, lesions of the caudate produce deficits in patients that closely resemble those that follow damage to dorsolateral PFC (Sandson et al. 1991). This finding mirrors early lesion work in monkeys which suggested similar behavioral impairments following lesions to either structure. Experiments in monkeys that have simultaneously recorded in area 46 and the caudate, to which area 46 projects, have shown similar responses in tasks that require spatial learning, although the caudate did have stronger correlations with the values of specific actions (Seo et al. 2012). Similar activity has also been seen across OFC and the ventral striatum in tasks in which monkeys have to learn the values of images (Costa et al. 2019; Tang et al. 2022a). In related work, lesions to medial MD thalamus, the part of MD most prominently connected to OFC, have shown deficits similar to those seen following lesions to ventrolateral PFC areas 12/47 (Chakraborty et al. 2016; Rudebeck et al. 2017b). Thus, a consistent set of findings, across human lesion, animal lesion, and neurophysiology in monkeys, have demonstrated that connected prefrontal cortical and subcortical areas show similar neurophysiological responses, as well as similar effects of lesions.

Cortico-Cortical Interactions

The extensive anatomical connections of the PFC place it in a privileged position to send feedback signals to the rest of the brain. Empirical support for the existence of such signals was obtained by Joaquin Fuster. In one study, a cooling probe was used to disrupt PFC function while neural activity was simultaneously recorded in the visual association cortex of monkeys performing a delayed match-to-sample task (Fuster 1985). When PFC was cooled, there was a reduction in delay-related neural activity in the temporal cortex. This finding indicated that PFC modulated the activity of the temporal cortex. In addition, PFC cooling affected the selectivity of neural responses in the temporal cortex. For example, neurons in the temporal cortex that originally coded for distinct color attributes displayed reduced selectivity for color following PFC cooling, consistent with similar recent studies. These findings have been

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

replicated in human studies, both through fMRI investigations in healthy individuals utilizing TMS to perturb PFC function and through scanning patients with focal PFC lesions (Buschman et al. 2011; Lee and D'Esposito 2012).

Additionally, frontal-temporal interaction in macaques is essential for rapid acquisition of visual stimulus-reward, stimulus-stimulus, and stimulus-action associations (Bussey et al. 2002; Clark et al. 2013; Eacott and Gaffan 1992) and for the retrieval of stimulus-stimulus associations (Tomita et al. 1999). Lesions that surgically disconnect frontal and temporal cortex disrupt associative learning while leaving intact basic visual sensory, motor, and reward processing. Thus, cortico-cortical interactions involving the frontal lobe appear to be involved in a variety of functions, including top-down modulation and/or attention, learning, and retrieval.

Cortico-Amygdala Interactions

The ventral (areas 12, 11, 13 and 14) and medial (areas 25, 32 and 24) frontal cortex areas have extensive reciprocal connections with the basolateral portion of the amygdala (for review, see Murray and Fellows 2022; Aggleton et al. 2015). In macaques, studies using crossed disconnection surgeries have examined the consequences of functional disconnection of OFC and ACC from the amygdala. Crossed lesions of the amygdala and frontal cortex regions (i.e., involving removal of the amygdala in one hemisphere and a frontal cortex region in the other hemisphere) produce a functional disconnection because the amygdala projections to frontal cortex are ipsilateral. Using this approach, Murray and colleagues found that a network composed of the OFC, amygdala, and medial MD thalamus is critical for performing the devaluation task, a task in which changes in food value need to be taken into account before making object choices; this circuit is essential for linking objects in the environment with food value and adjusting those valuations in real time based on current biological needs (Murray and Rudebeck 2013).

Crossed surgical disconnection of the amygdala from ACC yields a different impairment, one involving loss of social interest/and/or social signaling (Pujara et al. 2022). Consistent with this finding, simultaneous recording in amygdala and ACC during a social reward allocation task reveals neural signatures (e.g., based on coherence between ACCg spikes and BLA local field potentials) of prosocial behavior and of vicarious versus experienced rewards (Dal Monte et al. 2020; Putnam et al. 2023). These data point to a role for ACC in social evaluation and, together with the information about the effects of OFC disconnection from amygdala mentioned above, provide a neural framework for distinct value assignment processes in the PFC.

Little information is available about the mechanisms underlying amygdala interactions with frontal cortex. Studies combining electrophysiology with causal manipulations indicate that amygdala inputs are important for acquiring as well as maintaining representations of the value of liquid rewards

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

(both anticipated and received) in OFC but not ACC (Rudebeck et al. 2013a, 2017a). It is known that neurons in both OFC and amygdala of macaques signal the value of anticipated and received foods and fluids, as well as types of fluid, during choice tasks and appetitive Pavlovian conditioning (Morrison and Salzman 2009; Padoa-Schioppa 2011; Paton et al. 2006). Thus, these data suggest the possibility that the basolateral amygdala plays a general role in maintaining representations in frontal cortex, according to the types of representations stored in that area. Consistent with this idea, neurons in rat gustatory cortex lose representations of taste palatability but not identity after temporary inactivation of amygdala (Piette et al. 2012).

Although the amygdala is often considered responsible for processing "emotion" and neocortex for processing "cognition," this division of labor is almost certainly incorrect. Instead, it seems likely that emotional and cognitive parameters are inextricably linked and represented in dynamic neural circuits within amygdalo-frontal circuits, among other regions (Salzman and Fusi 2010).

What Are the Key Knowledge Gaps?

A lack of synthesis and coordination among disparate fields of research hampers progress in understanding the PFC. Research from the cellular and molecular level up to the systems level would benefit from closer integration, as would translational work in animal models and human subjects. Evolutionary, cross-species comparative research could provide the broader perspective needed for progress along these lines. There is also a large gap in our understanding of how subregions of frontal cortex interact with each other and within larger networks. Theoretical work, both conceptual and computational, could play an important role in bridging these gaps.

Theory That Aims to Connect Different Levels of Explanation

While there are exceptions, an ongoing gap in the field's approach to understanding functional fractionation of the frontal lobe has been a failure to explicitly bridge across levels of analysis and to integrate work done in cells and circuits with that at the systems and functional level. Computational neuroscience is an indispensable part of any strategy to overcome these obstacles and draw these links in a formal way (Badre et al. 2015). In neuroscience, computational modeling is often pursued at individual levels of analysis, from biophysically realistic models at the cellular level to abstract mathematical descriptions of behavior. Over the last decade, however, there has been fruitful progress in developing computational models that bridge levels of analysis (for detailed discussion, see Frank 2015 as well as Koechlin and Wang, this volume). These approaches allow modeling frontal function at one level to inform questions and models of other adjacent levels (e.g., Frank and Badre 2012;

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

Moolchand et al. 2022; Neymotin et al. 2020; Shin et al. 2017). Thus, theory provides a principled way of linking levels of analysis, going from molecules and cell types to systems-level function to complex behavior.

Theory and Experiments on Multiregional Interactions

Another major gap to be bridged concerns how regions of the frontal lobe interact, and how they interact with the cortical-cortical and subcortical networks in which they are embedded. For example, given that lesions to connected areas of cortex, thalamus, and striatum can lead to similar deficits, what is specific about the contribution of frontal cortex? Progress on these questions will require development and synthesis of both theoretical and empirical research. Some work has looked at computational implementations of PFC cognitive processes (Hart and Huk 2020; Wimmer et al. 2014), but these models lack biophysical detail and are confined to a single prefrontal subregion.

Ideally, the development of models that represent computations distributed across multiple areas would be a coordinated effort between theorists and experimentalists, and there would be consensus about how to assess the usefulness of such models. Such models should be able to perform a set of core PFC-dependent tasks and serve as a platform to integrate a wide range of experimental findings to achieve a cross-level mechanistic understanding of frontal lobe function. In practice, it would be beneficial to see studies of multiregional systems (e.g., cortico-cortical, cortico-striatal, or cortico-thalamocortical) in which high-level cognitive processes can be mapped onto specific regions or circuits, together with a computational model of how behavior is implemented in neurons, networks, and systems.

There has been progress on several theoretical questions in PFC related to representation and processing, mostly within specific regions. For example, there has been a shift in focus toward understanding representation and computation at the level of neuron populations, with the perspective that the fundamental unit of the brain's computation is a collection of interacting neurons that create dynamical activity. As discussed by Rich and Averbeck (this volume), the activity of large neural populations can often be captured by a low-dimensional manifold exhibiting a particular geometry (Chung and Abbott 2021). Examining these geometries can reveal dimensions that emphasize certain types of information over others, or that maintain information in orthogonal subspaces. For example, a recent study in mice found that within a high-dimensional space of neural activity, different subspaces were functionally connected with different networks of brain regions (MacDowell et al. 2023). This allowed a single area to interact simultaneously with multiple circuits. In addition, changing the alignment of the subspace (i.e., changing the geometry of neural responses) switched communication among networks, suggesting a mechanism that could support cognitive flexibility. These types of analyses are based on the notion that properties of neural populations are

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

fundamental to brain computation, and that those computations cannot be studied by examining single neurons in isolation. The conceptual step from single neurons to populations is part of a larger step from one area to a network of interconnected areas. If our level of analysis is mismatched to the level of the PFC computation, we will fail to understand the link between neural activity and behavior. An alternative possibility is that the homogeneity of neural responses across the frontal lobes could reflect computation that is truly distributed. In this case, an open challenge will be to determine how modular functions arise from distributed systems. Intersections with theory and modeling will be critical to further these ideas.

Another series of theoretical and experimental studies has shown that a unique property of PFC is that it houses complex, high-dimensional representations (Fusi et al. 2016; Rigotti et al. 2013). These high-dimensional representations encode combinations of actions, stimuli, contexts, and outcomes in a way that allows downstream areas to decode any combination of the information. One view suggests that, from these high-dimensional representations, the striatum learns to select actions or action sequences specific to those stimuli in a given context that led to advantageous behavior (Parker et al. 2022). To further connect this idea to the discussion about functional organization, different prefrontal-basal ganglia circuits may be specialized for selection of different types of information. For instance, OFC-ventral striatal circuits may specialize in identifying or selecting goals based on combinations of stimuli and outcomes. By contrast, circuits involving medial and lateral PFC may combine actions or hierarchically organized action sequences and/or cognitive mechanisms (e.g., working memory or response inhibition) that allow goals to be achieved. If so, the striatum would play an integral part in learning to link the stimulus and context representations to the actions and cognitive mechanisms required to reach a given goal. The thalamus would be relevant to relaying the associations formed by the basal ganglia back to the cortex, perhaps for action execution through descending motor systems or for updating frontal representations. Consistent with an updating process, recent evidence has shown that the high-dimensional representations in PFC become lower-dimensional with learning, presumably because they become more sculpted to task demands (Mack et al. 2017; Wojcik et al. 2023). Testing the role of thalamocortical-basal ganglia circuits in this model is not straightforward, however, as disrupting any single element (e.g., frontal representations, striatal learning mechanisms, or thalamic relay mechanisms) would be expected to lead to deficits in behavior. Additional assumptions would have to be made about how each of these processes is implemented. Notably, some studies that have examined cortical and striatal representations of actions, selected to achieve specific goals, have shown that the cortex represents the chosen actions before the striatum, at least under some conditions (Seo et al. 2012; cf. Pasupathy and Miller 2005).

Although computational models have been developed that account for these specialized coding properties, these models have not taken into account the

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

dynamical properties of PFC. Thus, the previous models have only formally described these processes, without embedding them in dynamical systems, or more ambitiously, multi-area dynamical systems models. More work is needed to continue to develop these models. Increasingly sophisticated computational models that respect the functional organization of the networks within which specific prefrontal areas are embedded will allow the development of quantitative predictions that can be tested empirically, using, for example, simultaneous recordings across multiple areas. Currently, few such predictions exist.

Theory That Aims to Connect PFC Function with Evolutionary Perspectives

In general terms, the frontal cortex is thought to store knowledge about behavioral goals and actions that could achieve them, along with outcomes that should result from such actions (Miller and Cohen 2001; Passingham 2021; Passingham and Wise 2012). As mentioned in the prior section, high-dimensional representations in PFC combine this information and confer several adaptive advantages, such as empowering individuals to learn from specific (less averaged) events (Massi et al. 2018). For example, when a new type of representation brings together previously unassociated stimuli, contexts, goals, and action sequences, selective forces can favor such representations to generate the cortical maps characteristic of each species (Murray et al. 2017). A consideration of representations and their adaptive advantages could provide the theoretical perspective to bridge several current gaps in knowledge, such as why cortical areas in the PFC are so much more difficult to define than in sensory areas of cortex.

It is well established that new PFC areas appeared in early primates and more emerged later in anthropoid primates. Accordingly, as discussed by Weiner et al. (this volume), there are many more frontal areas in primates than in rodents. However, a collection of sensory-cortex-like areas might not be the best way to think about the PFC. Although it is appealing to think that PFC areas are organized like early sensory areas, which have a well-defined function and discrete boundaries, that is not the only way representations can be distributed in the cortex. Biologically significant representations may be more widely distributed within the PFC. In these instances, it will be difficult or impossible to discover area-function relationships that look like maps of visual areas. The variation among published architectonic maps of the PFC, the lack of agreement about the precise number of areas or their boundaries, and the distributed encoding of variables observed in neurophysiological studies strongly suggest that there is some other organizing principle underlying PFC function. Accordingly, future work may benefit from a renewed focus on neural representations that smaller units of cortex, such as individual columns, generate and store, as well as the advantages such representations confer on animals in their natural habitats.

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

Consideration of the adaptive advantages provided by specific representations could lead to the development of new tasks that are more closely related to ethologically relevant behaviors. For example, recent work in freely moving macaques has set the stage for bringing together behavior (e.g., pose estimation), wireless electrophysiology, and autonomic measures in social contexts (Hayden et al. 2022; Maisson et al. 2023; Milton et al. 2020). Further work along such lines promises to bridge another key gap in knowledge: that between laboratory or clinical settings and the natural habitats of species that serve as animal models.

Conceptual Theory

A general challenge for studying structure-function relationships within the frontal lobes has been to develop a common cognitive-behavioral ontological framework, especially one that crosses human and animal models. A first problem is that typical experimental paradigms systematically remove elements that likely require the frontal lobes. That is, tasks meant to assess frontal cortex function often provide simple, salient stimuli, strongly constrain possible responses, use instruction to set explicit expectations about the task, provide practice to refine performance before the "real" task starts, and provide trial-by-trial feedback which, with enough repetitions, may eliminate the need for the frontal lobes entirely by converting the behavior to a well-learned "habit." Despite that limitation, there are many tasks that rely on the frontal lobes, but little standardization of what may be the crucial details of instruction, practice, timing, and trial number. The standardized human tasks (i.e., from clinical neuropsychology) tend to be grounded in a more classic theoretical framework, largely aiming to tap lateral PFC-mediated attentional and set-shifting abilities, which may not be readily mapped to current conceptual or computational models that include decision making, social behaviors, and flexible learning from reward. It would be helpful to develop a set of more standardized tasks grounded in current theories and useful for modeling. Tasks that address how we navigate novel or changing environments, select and pursue ecologically relevant motivational goals, and learn rapidly from real or vicarious experience may be especially useful, particularly for linking across humans and other primates. Further validation could come from considering the correspondence of such tasks and the clinical phenomenology that we think may relate to fractionated frontal lobe function, setting a direction that could connect with the clinic.

One measure of our understanding of frontal lobe fractionation is our ability to predict how prefrontal regions would be activated by cognitive tasks. This can be quantitatively formalized as the challenge of predicting cortical maps of activation across conditions of a given arbitrary task design (e.g., via an encoding model). Encoding models in fMRI develop voxel-level tuning functions that identify the features of a task that drive activation of a given voxel. (These are similar to tuning functions, for example, in visual cortex, that describe

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

the properties of a visual stimulus that activate a neuron.) Meta-analytic approaches that combine task fMRI data across many different tasks provides the current state of the art. For example, NeuroQuery (Dockès et al. 2020) predicts brain-wide maps associated with various neuroscience terms (e.g., "working memory" or "reward"), from compiling activation coordinates and extracting text terms, across many fMRI studies. The resulting maps tend to be much coarser than the group-level activation contrast maps of any particular fMRI study, revealing finer regional differentiation than can currently be predicted *a priori* for a novel task. These meta-analytic approaches use neuroscientific terms from publications rather than a standardized description of the task itself. Thus, these approaches will be limited in their ability to capture neural effects of task manipulations.

Prediction of task fMRI maps from task description is limited by our ability to represent computationally a novel task in relation to other tasks. An embedding of tasks in some latent space could potentially capture how tasks differentially engage cognitive processes in a way that enables a mapping into the space of neural activations. This encoding model approach has been fruitful in the study of naturalistic perception. For instance, encoding models can predict voxel-wise cortical map activations by natural visual images (via receptive field models) and by spoken text (via semantic category labels) (Huth et al. 2016; Kay et al. 2008). A challenge for the study of frontal cortex function is to apply similar approaches to cognitive tasks.

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