Cell PRESS

A neurocomputational system for relational reasoning

Barbara J. Knowlton[1], Robert G. Morrison[2], John E. Hummel[3] and Keith J. Holyoak[1]

[1] Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2] Department of Psychology, Neuroscience Institute, Loyola University Chicago, Chicago, IL 60626, USA
[3] Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

The representation and manipulation of structured relations is central to human reasoning. Recent work in computational modeling and neuroscience has set the stage for developing more detailed neurocomputational models of these abilities. Several key neural findings appear to dovetail with computational constraints derived from a model of analogical processing, 'Learning and Inference with Schemas and Analogies' (LISA). These include evidence that (i) coherent oscillatory activity in the gamma and theta bands enables long-distance communication between the prefrontal cortex and posterior brain regions where information is stored; (ii) neurons in prefrontal cortex can rapidly learn to represent abstract concepts; (iii) a rostral-caudal abstraction gradient exists in the PFC; and (iv) the inferior frontal gyrus exerts inhibitory control over task-irrelevant information.

How is thinking realized in the human brain?

One of the deepest puzzles for cognitive neuroscience is to explain how the most distinctively human types of thinking and reasoning are realized in the brain. Humans can grasp analogies between disparate situations, infer hidden causes of observed events, apply general rules to novel situations, and learn new abstractions from experience [1–3]. Such intellectual abilities, which exceed those of any other extant primate species (perhaps in a qualitative manner [4]) are difficult to capture in any computational model, but pose particular challenges for those that aim for neural fidelity. How does the brain organize neurons, which are basically simple computing devices, so as to achieve the kinds of complexity manifested in human thinking and reasoning?

Research over the past decade and a half has begun to address this challenge. Cognitive neuropsychological and neuroimaging studies have implicated various subregions of the prefrontal cortex (PFC; Figure 1) as critical parts of a larger network supporting higher cognition (for reviews, see [5–9]). The most anterior lateral portion of the PFC, generally termed frontopolar or rostrolateral (RLPFC), is activated by tasks that require integration of multiple relations, processing relatively abstract concepts, or negotiating hierarchical goal structures [10–22]. More dorsal and inferior areas of the PFC have also been implicated in the systematic control of representations necessary for these processes [10,13,20,23–25].

Over roughly the same time period, a number of computational models [26–30] have attempted to explain aspects of human thinking and reasoning within neural systems, differing in their architectures and domains of application (Table 1). A substantial gap remains, however, between current theories of PFC function and computational models capable of actually performing tasks involving thinking

Glossary

Active memory: information in a state of current readiness for use in processing (including the active portion of LTM), typically over a time span of around 20 seconds.
Analogical mapping: the process of identifying systematic correspondences between elements of two situations (analogs) based on relational structure.
Cross-frequency coupling: interactions between different frequency bands, such as theta and gamma, which aid in integrating neural activity across different spatial and temporal scales.
Driver: in LISA, an analog that is currently in active memory and serves as a generator of spreading activation.
Phase set: in LISA, the set of mutually desynchronized role bindings represented by neuronal oscillations. The phase set corresponds to the current focus of attention and is the most significant bottleneck for reasoning with relations. The phase set is equivalent to working memory (WM) for relations.
Proposition: a predicate instantiated by binding its role(s) to particular arguments (objects or other propositions). A proposition is the smallest unit of representation that can have a truth value: intuitively, a 'complete thought'. In LISA, a proposition is represented by a hierarchy of structure units.
Proxy unit: a transient representation of a structure unit, formed in prefrontal cortex in order to support structured reasoning, such as an analogical comparison.
Recipient: in LISA, an analog that is currently receiving activation from the driver. There may be multiple recipients in long-term memory (during retrieval) or a single recipient in active memory (during mapping, inference, and schema induction).
Role-based relational reasoning: reasoning that depends on the active representation and manipulation of concepts involving roles and role binding (see 'proposition').
Role binding: the binding of a single argument (object or proposition) to a single role associated with a predicate.
Schema: a relatively abstract relational structure representing a category or class of situations (e.g., a schema for a type of problem). In LISA, schemas can be formed as a consequence of comparing two or more relatively specific analogs.
Semantic unit: a unit that represents a simple element of meaning, associated with neurons in posterior cortex. In LISA, semantic units are the sole conduits for the transmission of activation between distinct analogs.
Spike-timing-dependent plasticity: a phenomenon based on evidence that if a neuron is being driven at a high rate, as occurs in the high gamma band, the inputs driving it will be strengthened. It provides a neural mechanism by which the kind of synchronous activity that in the LISA model supports dynamic representations will also lead to synaptic strengthening.
Structure unit: in LISA, a unit representing a component of a proposition within an analog: P (proposition), RB (role binding), O (object), and R (role); or a correspondence between elements of two analogs (M). Such units may be associated with neurons in posterior cortex (when stored in LTM), but must also be associated with dynamically recruited neurons in prefrontal cortex (see 'proxy unit').

Corresponding author: Knowlton, B.J. (knowlton@psych.ucla.edu).

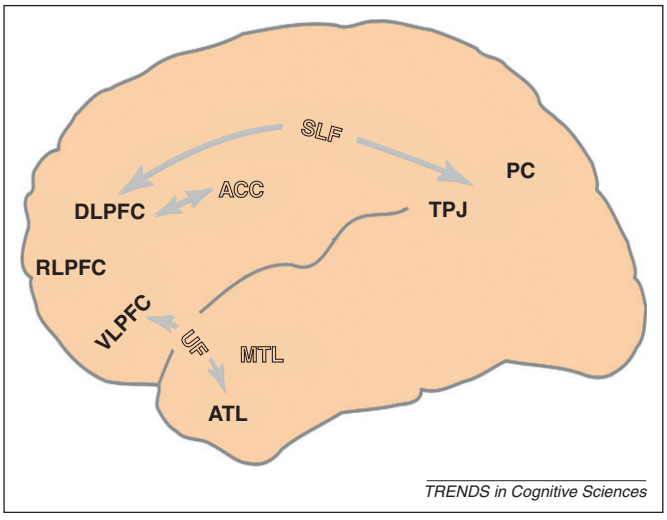

Figure 1. Anatomy and connections related to relational reasoning. Areas of the prefrontal cortex (PFC) frequently identified in reasoning studies include the rostrolateral prefrontal cortex (RLPFC; anterior region of the inferior frontal gyrus, approximately Brodmann area 10, sometimes referred to as frontopolar prefrontal cortex), the dorsolateral prefrontal cortex (DLPFC; anterior region of the middle frontal gyrus, approximately Brodmann areas 9/46), and the ventrolateral prefrontal cortex (VLPFC; posterior region of the inferior frontal gyrus, approximately Brodmann areas 47/45/44). The anterior temporal lobe (ATL; located on the anterior lateral surface of the temporal lobe, approximately Brodmann areas 20, 31, 38) is frequently associated with semantic memory (see [72]) and is important for reasoning about semantic relations [24]. The medial temporal lobe (MTL; located on the medial surface of the temporal lobe including the hippocampus and entorhinal cortex, approximately Brodmann areas 27, 28, 34, 35, 36) is critical for episodic memory [73], and thus is important for relational reasoning about specific events. The ATL and MTL are connected to areas in the VLPFC via the uncinate fasiculus (UF). Regions in the parietal lobe, such as areas around and including the precuneus (PC; approximately Brodmann area 7) and the temporal parietal junction (TPJ; approximately Brodmann area 39) have heavy reciprocal connections to the PFC via the superior longitudinal fasciculus (SLF). These areas are frequently associated with tasks requiring relational reasoning about visuospatial entitites. The anterior cingulate cortex (ACC; located on the medial surface of prefrontal cortex approximately, Brodmann areas 24, 32, 33) is frequently active during relational reasoning and has reciprocal connections to the DLPFC.

and reasoning. Functional theories often highlight very general concepts such as 'abstraction' and 'relational integration', which though potentially helpful remain ill-defined in the absence of computational instantiations.

Here we attempt to build an initial bridge across this gap. Focusing on computational mechanisms instantiated in a leading model of relational reasoning, 'Learning and Inference with Schemas and Analogies' (LISA; [27,31]), we review the neural literature to construct more specific hypotheses about how these mechanisms may be realized in the human brain. Even though our focus is on the LISA model, we will also note connections with other neurocomputational models. Our opinion article reveals a remarkable convergence between constraints on models of human reasoning derived from computational analyses, behavioral experiments, and neurophysiological investigations. Although necessarily preliminary, we hope that this effort will help the development of more detailed neural models of high-level cognition.

Role-based relational reasoning in LISA

The LISA model provides a computational account of role-based relational reasoning: inferences that depend on the roles that entities play, not just on perceptual similarity. For example, knowing that Sam is an enemy of Brian, and Dylan is a friend of Sam, a person might conjecture that Dylan may also be an enemy of Brian (Figure 2). This 'mutual support' schema may have itself been acquired through analogical reasoning, by comparing specific cases that share a common relational structure.

LISA codes an analog by binding distributed representations of roles to distributed representations of their fillers (coded on separate pools of semantic units; Figure 2). Semantic units are assumed to be coded by neurons in posterior regions. For each individual analog, a hierarchy of localist structure units represents objects (O), relational roles (R), individual role bindings (RB), and complete propositions (P). Structure units may be coded in long-term memory (LTM), but in order to be made available for active comparisons, they require a transient form (proxy units) in active memory [32]. During mapping, the emerging correspondences are also coded by proxy units, called M *(mapping)* units, that connect structure units of a given type across the two analogs (e.g., P units to P units). These explicit learned correspondences allow LISA to assess the overall similarity between two analogs [33] and to generate sensible

Table 1. Major neurocomputational models of human thinking

Model	Architecture	Domain(s) of Application	Unique Characteristics
SMRITI [28]	Localist connectionist network using binding by synchrony	Episodic memory encoding, storage, binding and retrieval	Corresponds to known architecture of hippocampus
LISA [27,31]	Distributed connectionist network using binding by synchrony	Relational reasoning	Integrates dynamic binding in WM with static binding in LTM, enabling symbolic processing to arise from a neural architecture
STAR-2 [26]	Connectionist network using tensor products to code bindings	Relational reasoning	Tensor rank (number of basis vectors that together define the tensor) corresponds to relational complexity, predicting difficulty of reasoning tasks
ACT-R with neural modules [30]	Production system integrated with modules for perception, motor responses, spatial representation, memory retrieval, and goal maintenance	Solving algebraic equations and related laboratory tasks	Provides a macro-level mapping between information-processing modules and brain areas, coupled with an analysis of the time course of activation for each module
SAL (Synthesis of ACT-R and Leabra) [29]	Production system with declarative memory (ACT-R); subsymbolic processes realized by a connectionist network based on a point-neuron activation function (Leabra)	Spatial navigation and search	Integration of symbolic control processes with subsymbolic representations and learning mechanisms, using a neurally-constrained architecture

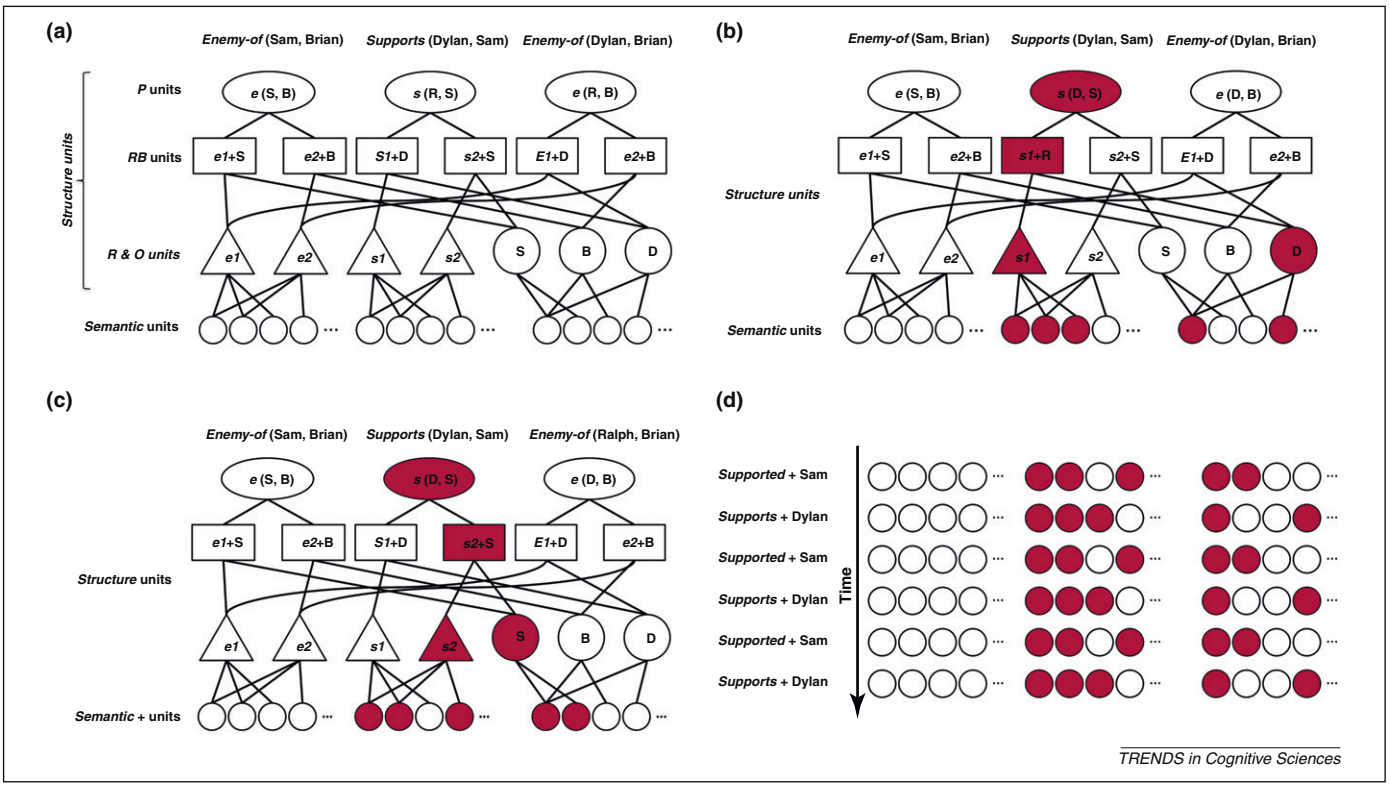

Figure 2. Representation of propositions in the LISA model. **(a)** A pool of semantic units (bottom), which are connected to localist structure units that capture bindings at successive levels of generality: individual roles and objects, bindings of objects to roles, and bindings of role/filler combinations into propositions. **(b)** In a single phase of the dynamic form of binding, one role binding of one proposition becomes active, along with its constituent O and R units and associated semantic units. **(c)** In a subsequent phase, a different role binding and its constituents are activated in synchrony with each other (and out of synchrony with the first role binding). **(d)** The overall pattern in which structure units for the proposition fire across a series of temporal phases.

structured inferences based on correspondences between elements of the two analogs.

LISA exploits dynamic binding coded by neural synchrony [34] to impose a hierarchical temporal structure on knowledge representations within working memory (WM). A small number of role bindings in one analog (the driver) can enter the phase set – the set of mutually desynchronized role bindings. The phase set corresponds to the current focus of attention, and is the most significant bottleneck in the system. Each individual phase (the smallest unit of WM) corresponds to one role binding (i.e., an RB unit and its constituents). The size of the phase set is determined by the number of role-filler bindings (phases) that can be simultaneously active but mutually out of synchrony. The maximum number is proportional to the length of time between successive peaks in a phase (the period of the oscillation) divided by the duration of each phase (at the level of small populations of neurons) and/or temporal precision (at the level of individual spikes) [35]. Binding may be accomplished by synchrony in the >30 Hz (gamma) range, with a neuron or population of neurons generating one spike (or burst) approximately every 25 ms, implying WM capacity of approximately 4-6 role bindings (roughly 2-3 propositions). This value is consistent with estimates of WM capacity based on behavioral evidence (e.g., [36]) and may have roots in the mechanisms by which information is processed throughout the brain, including lower-level posterior cortex [37].

Because of the strong capacity limit on its phase set, LISA's processing is necessarily highly sequential,

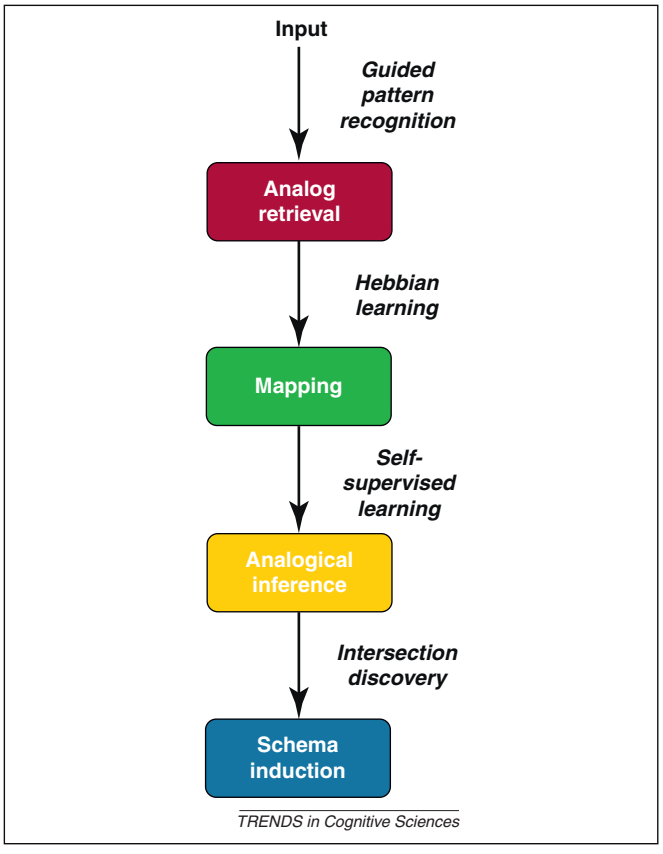

Figure 3. Operations on LISA's knowledge representations at major stages of relational reasoning and learning.

constituting a form of guided pattern recognition (Figure 3). At any given moment, one analog (the driver) is the focus of attention. As one or more driver propositions enter the phase set, synchronized patterns of activation are generated on the semantic units (one pattern per RB). In turn, these patterns activate propositions in one or more recipient analogs in LTM (during retrieval), or a single recipient in active memory (during mapping, inference and schema induction).

LISA provides a natural account of the loss of relational reasoning in populations with forms of brain damage, such as patients suffering from either frontal-lobe or temporal-lobe variants of Frontotemporal Lobar Degeneration [24]. The model has also been used to simulate changes in relational reasoning during cognitive development [38–40] and normal aging [41]. In the remainder of this article we review several key neural findings that appear to correspond to computational constraints that arise in LISA.

Role of oscillatory activity in reasoning

LISA fundamentally depends on the representation of information in a temporal structure. RB units must be activated in synchrony with O and P units (and their associated semantic units) to form dynamic representations, while these different representational complexes must be kept out of synchrony with each other to maintain distinct, non-overlapping role-filler bindings (Figure 4a, b). Temporal structure in the form of oscillatory activity is in fact prominent in the brain [42], although no direct evidence yet connects such activity to the coding of propositions. Rhythmic neural activity, as reflected in the firing of groups of neurons, can be detected throughout the central nervous system, both in local interactions within a neural ensemble, and across brain regions through long-distance connections between populations of neurons [43].

Oscillations can arise from the intrinsic circuit properties of the central nervous system, as neurons tend to be interconnected with numerous excitatory and inhibitory neurons, resulting in entrainment of firing of an ensemble of neurons at a specific frequency. These oscillations may reflect the firing rates of individual neurons, such as those that show bursts of firing in the gamma band. Slower oscillations, such as firing in the theta band (4–8 Hz), generally do not arise from a group of individual neurons firing at that frequency; rather, summed over a large group of neurons, peaks of firing will be apparent at this lower frequency due to slower feedback modulation.

In LISA, the smallest unit of WM is essentially defined as the synchronous firing of representational units. Importantly, LISA uses phase to maintain the separation of multiple role-filler bindings. Behavioral experiments using a priming paradigm suggest that synchrony underlies the representations of perceptual relations for humans [44]. Likewise, a recent EEG study suggests that phase synchrony within the fronto-parietal network can bind object properties together in WM [45]. Electrophysiological studies in nonhuman primates have revealed a link between WM and the synchronous firing of neural ensembles. For example, pairs of neurons have been shown to fire in synchrony in a task-dependent manner, consistent with the hypothesis that synchronous firing dynamically represents the representations needed in WM to perform the current task [46]. The fact that neural assemblies can

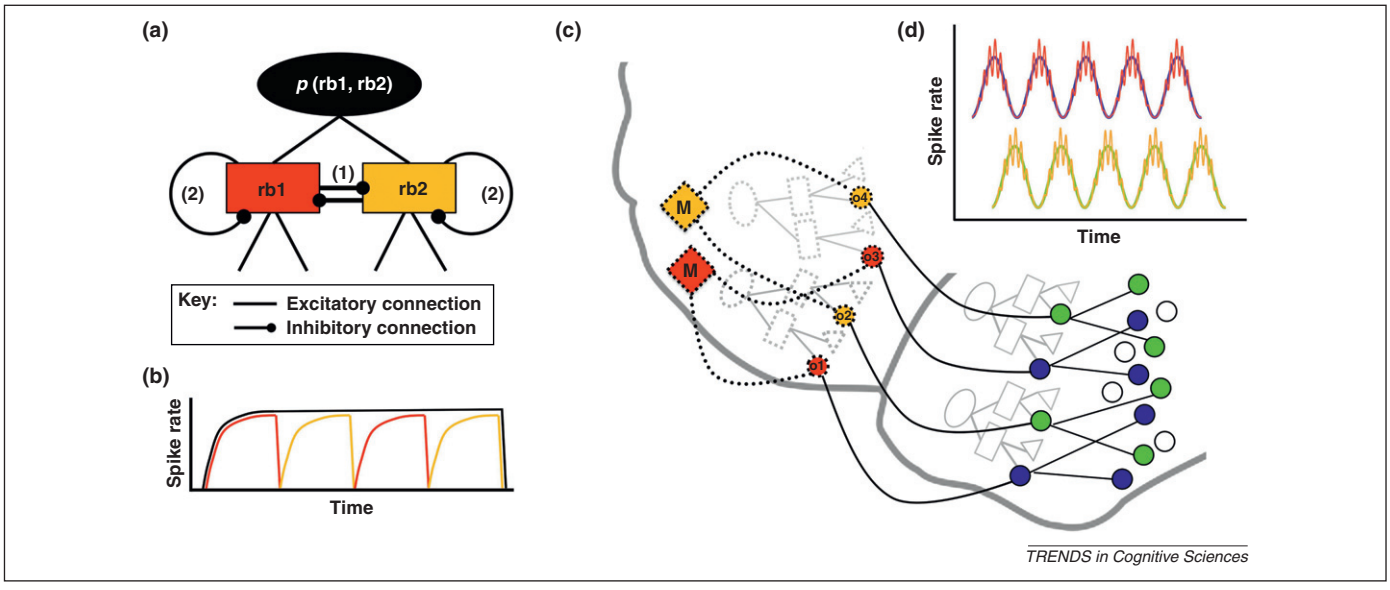

Figure 4. Oscillatory inhibition and cross-frequency coupling in LISA. **(a)** Oscillatory inhibition is central to LISA's ability to exploit temporal synchrony to discover relational mappings in WM. Role-binding (RB) units for a single proposition [e.g, *p* (rb1, rb2)] are kept out of phase by pools of inhibitory neurons that inhibit competing RBs (e.g., rb1 and rb2) **(1)**, and also apply inhibition to an RB after it fires **(2)**. **(b)** Thus, 'tonic' excitation from a single proposition unit in the driver [e.g, *p* (rb1, rb2)] causes all attached RB units (e.g., rb1 and rb2 here) to fire out of phase with each other, while allowing each RB an opportunity to fire as determined by accumulating mapping evidence (via the mapping connections). Oscillatory inhibition is critical for determining LISA's intrinsically limited WM capacity. **(c-d)** By using temporal synchrony as a binding mechanism, LISA is able to utilize cross-frequency coupling in conjunction with spike-timing-dependent plasticity to rapidly learn relational mappings (i.e., M units representing correspondences between objects in the driver and the recipient) via Hebbian learning. In this example, semantic units represented by populations of neurons in posterior cortex (blue units in panel **(c)** fire relatively slowly [slower blue wave in panel **(d)**], sending activation to populations of neurons representing object1 [red o1 unit in panel **(c)**]. These neurons fire at a much higher collective frequency [faster red waves in panel **(d)**]. In LISA, similar types of units in the driver and recipient are connected via mapping units [represented here by the dotted lines in panel **(c)**]. Synchronous rapid firing of the populations of neurons representing o1 and o3 result in rapid changes to the mapping units via spike-timing-dependent plasticity. Thus, LISA quickly learns that o1 and o3 are firing at the same time, and hence correspond relationally. A similar correspondence exists between o2 and o4 [orange units in panel **(c)**], which fire out of synchrony with o1 and o3 given their different roles in their respective propositions.

rapidly shift between different patterns of synchrony depending on the information being held in WM indicates that the method of representing propositions in LISA is neurally credible.

LISA also depends on long-distance communication between prefrontal regions and posterior regions of the cortex, where semantic information is stored. In order for prefrontal RB units to be activated by semantic information, there must be a means by which oscillatory activity in the temporal lobes engages circuits in PFC that represent this information. In the brain, synchrony in lower frequency bands, including theta, is detectable between sites separated by several millimeters, suggesting that entrainment of neural activity across brain regions occurs at lower frequency oscillations [47–49]. In contrast, synchronous activity within local neural circuits tends to be higher frequency, in the gamma range. Within the PFC, local circuits will require inhibition to maintain the phase relationships of different role-filler bindings.

Studies of learning and memory have shown that brain oscillatory activity is relevant to behavior. For example, successful memory formation is associated with the tighter coupling of the firing of individual neurons to the theta frequency [50]. In addition, stimulation during theta peaks is particularly effective in inducing long-term potentiation in the hippocampus, whereas blocking theta prevents the induction of long-term potentiation [51,52]. It thus appears that neural oscillations provide support for neural plasticity. Reasoning similarly requires the rapid formation of new representations. In LISA, M units are formed between structure units in the driver and recipient to capture correspondences between them. This type of rapid learning of connections has been observed in PFC, based on single-unit recordings with non-human primates [53,54].

It appears that, in addition to plasticity being related to the phase of the theta cycle, high gamma frequency itself can directly enhance neural plasticity through a Hebbian learning mechanism [55]. When neuronal circuits fire at this high rate, inputs to a neuron arrive shortly before the neuron fires, which thus becomes depolarized. In Hebbian plasticity, synaptic inputs that are active when the neuron is depolarized are strengthened. Thus, if the neuron is driven at a high rate, as occurs in the high gamma band, the inputs driving it will be strengthened. This phenomenon, termed 'spike-timing-dependent plasticity' [56], implies that there is a neural mechanism by which the kind of synchronous activity postulated by LISA will also lead to synaptic strengthening. Such increases in synaptic strength may underlie the strengthening of M units during the mapping process. The facilitatory influences of theta- and gamma-band activity on neural plasticity appear to be related [47,48]. Through cross-frequency coupling (see [57]), the phase of the low frequency theta wave modulates the power of the gamma band, such that the amplitude of EEG measured in the gamma band is greatest at a specific point in the theta wave (Figure 4c, d). The theta wave may serve to entrain bursts at gamma frequency by shifting the probability of spike timing. By this mechanism, long-distance communication across regions in the form of theta activity could modulate the timing of gamma bursts. It follows that information in regions of posterior cortex responsible for representing semantic information may influence local gamma activity in the PFC via theta frequency firing.

Additional evidence supports the hypothesis that the phase of neuronal oscillations in the PFC codes the representation of specific items in WM. Simultaneous recording of single units and the local-field potential in monkeys have demonstrated that spikes in response to a specific stimulus occur at a characteristic point in a 32-Hz cycle, corresponding to the gamma band. When the monkey's task was to keep more than one item in mind at a time, spikes corresponding to the second item occurred at a different point in the wave [58]. Interestingly, spike synchronization was also observed at approximately 3 Hz, at the lower end of the theta band. The fact that both theta and gamma band oscillations are synchronized in the PFC is consistent with the possibility that cross-frequency coupling is a means to coordinate activity in distant regions with the rapid oscillations that support local processing. The finding that phase-specific spiking in the gamma phase is related to segregation of information in WM is consistent with LISA's mechanism for representing propositions via temporal asynchrony of role bindings.

Proxy units in prefrontal cortex

In LISA, when propositions enter active memory, proxy units (the transient form of structure units) are rapidly formed in PFC. These proxy units that code individual analogs in turn connect to M units that represent correspondences between the elements of two analogs. The rapid learning required by these units could be supported by the spike-timing-dependent plasticity that can occur during high gamma-band activity synchronized to the theta rhythm [56]. The rapid changing of weights on these units makes them suitable to dynamically represent different stimuli depending on the information being processed at the moment. Neurons with the properties ascribed to proxy units have been identified in the primate lateral PFC [53,54,59]. About a third of neurons in the lateral PFC respond to categories of stimuli, which means that the neuron will increase firing rate when presented with members of a particular category (e.g., dogs). Importantly, these neurons fire based on the conceptual properties of the stimuli, and not simply on the basis of their visual features. Unlike neurons in inferotemporal cortex, neurons in lateral PFC respect the sharp boundaries between categories. These neurons respond similarly to typical and atypical members of a category, which suggests that they are sensitive to the rules defining the concept itself. This property is necessary for the proxy units in LISA, in that they must be able to represent high-level propositions that are abstract.

Furthermore, neurons in the PFC appear to also be able to code abstract relational rules. Cromer *et al*. [54] had monkeys perform two different tasks: one in which they had to respond to matching stimuli and another in which they had to respond if the stimuli did not match [13]. The most common type of neuron recorded in the prefrontal cortex (41% of all those recorded) responded selectively to the current rule, regardless of the stimuli that were present. These neurons thus appear to represent the relation

between stimuli and a rule governing the current task – not the stimuli themselves – a major requirement for the proxy units posited by LISA. Although neurons responding to abstract rules could be found in all regions of the PFC, the majority were located in the lateral subregion.

The apparent flexibility of these neurons is another property that makes them suitable as instantiations of the proxy units postulated by LISA. Individual neurons do not simply respond to one type of stimulus; rather, in the context of different tasks, they respond to different categories [54]. A neuron's response to the same stimulus can vary on a trial-by-trial basis depending on the task performed [60]. This flexibility stands in stark contrast to the firing properties of inferotemporal neurons, in which firing to a complex visual stimulus is relatively static [61]. Although caution is warranted in extrapolating from studies of monkeys to more complex human reasoning, such findings suggest that PFC neurons may act as representational elements for very different propositions depending on the task context.

Such dynamic repurposing of neurons is consistent with the flexible role played by proxy units in the LISA model. Proxy units are formed rapidly to represent propositions during reasoning. Spike-timing-dependent plasticity resulting from fast gamma-band activity in prefrontal neurons can support the kind of rapid changes in response properties that are necessary for proxy units. Importantly, synaptic strength can be rapidly decreased, as well as increased, based on the timing of presynaptic firing and post-synaptic depolarization. Long-term depression of synaptic strength allows neurons to be returned to a pool from which they can be recruited for the representation of new propositions. Moreover, the shift from long-term potentiation to long-term depression occurs as the result of a shift in spike timing on the order of milliseconds [62]. These findings suggest that spike-timing-dependent plasticity can support the rapid binding and unbinding that is fundamental to the LISA model.

Rostral-caudal abstraction gradient in PFC

Recently, an effort has been made to understand the subregions in PFC in terms of a hierarchy of action. Badre and D'Esposito [6] have argued that more caudal regions of the PFC are involved in generating specific stimulus-response motor actions, whereas progressively more rostral regions become more involved when actions must be based on more abstract information (e.g., a plan based on the integration of multiple subgoals) [6]. For example, although caudal regions of the PFC are sufficient to subserve the act of picking up a cup from which to drink, more rostral regions would become engaged in the act of deciding what to drink in order to satisfy more abstract goals (e.g., trying to be healthy). Similarly, Christoff *et al.* [16] have shown that a set to process more abstract concepts selectively activates RLPFC.

In LISA, the highest level of hierarchical organization is reflected in the M units, which form rapid associations between elements of propositions in the analogs being compared. These mapping units represent very abstract information, specifically, shared relational roles that can make otherwise dissimilar propositions analogous. Moreover, the very process of identifying relational commonalities can trigger the acquisition of more abstract schemas for classes of situations. The LISA architecture is thus based on representations at successive levels of abstraction, an overall structure that appears to be reflected in the organization of the PFC.

The role of inhibition in reasoning

The nervous system is characterized by the interplay of excitation and inhibition. At the circuit level, the tight coupling of inhibitory interneurons and excitatory neurons results in oscillatory activity that allows for temporal coding of information in LISA's WM. As discussed above, circuit-level inhibition is required to maintain role-filler bindings mutually out of synchrony, and thus distinct in WM (Figure 4a, b). In addition, inhibition plays a role in reducing interference from competing semantic concepts during analogical reasoning (e.g., [24,25,34,40]). Activation of propositions in the driver analog will trigger activation in related semantic units, which in turn will activate candidate recipient propositions. The most active recipient propositions will eventually enter WM and be available for analogical mapping.

However, if task-irrelevant propositions are activated in the driver, these may bias the system to find suboptimal correspondences. LISA postulates top-down inhibition of propositions tagged as low in goal-relevance, which helps prevent these propositions from entering the phase set. This selectivity increases the efficiency of the mapping process by enhancing the signal-to-noise ratio favoring goal-relevant matches. Regions in the PFC exhibit similar properties by selectively inhibiting semantic representations in posterior regions of cortex. The PFC has many reciprocal connections with posterior cortical regions, including in particular temporal and posterior parietal lobes (see [63]), and thus is able to influence processing in these regions. The primary evidence for the role of the PFC in inhibition is that a major consequence of prefrontal lesions is behavioral disinhibition. Patients with damage to the prefrontal cortex often fail to inhibit inappropriate behaviors and have difficulty maintaining cognitive control. By one view, the main role of the prefrontal cortex is the dynamic filtering of activations in posterior cortical regions to facilitate behavior directed towards a goal [64]. Different subregions appear to support inhibition in different domains. In particular, damage to the right inferior prefrontal cortex impairs ability to inhibit a prepotent response in cognitive tasks whereas damage to orbitofrontal regions results in social and emotional disinhibition [65].

Behavioral studies have shown that the presence of irrelevant relations in analogs can impact relational reasoning, suggesting the inhibition is a necessary component of analogical reasoning [24,25,38,40,41,66]. Neuroimaging studies have produced even more direct evidence of the engagement of prefrontal inhibitory control during analogical reasoning. In a number of studies, activity was observed in the inferior frontal gyrus while subjects were solving analogy problems [23,67]. This same region has been shown to be active in a number of tasks of inhibitory control or semantic competition (see [65]). Cho et al. (2010) found direct evidence for the involvement of this region in

inhibitory control during analogy, showing that activity in a region of right inferior frontal gyrus increased when the amount of interfering information increased [13]. Similarly, using recordings of scalp EEG, Sweis et al. (2012) found that when participants needed to ignore a distracting relation while solving a visual analogy, right PFC was modulated at late stages of processing, and the degree of modulation interacted with the reasoner's WM capacity [20]. The inferior frontal gyrus thus may be the anatomical source of the active inhibition of competing units postulated by the LISA model.

Further directions

Many open questions remain (Box 1). Although we have focused on implications of neural evidence for the LISA model, these findings have implications for other neurocomputational models. Moreover, aspects of several of these models might be integrated to broaden coverage of high-level human cognition. LISA and SMRITI [28] both use patterns of neural timing to encode binding information, and can be viewed as complementary (LISA focusing on prefrontal functions and reflective reasoning, SMRITI focusing on hippocampal functions). The macro-level neural modules postulated by ACT-R [30] are compatible with LISA; ACT-R can be viewed as a model of the control structure within which human relational reasoning may operate.

There is reason to hope that advances in neuroimaging techniques [68], combined with refined methods for analyzing temporal patterns of neural activity, will make it possible to directly test some of the hypotheses we have laid out concerning the neural basis of relational reasoning. In addition, the general LISA architecture may be extended to incorporate additional factors related to PFC function. For example, in order to capture information processing in the PFC more fully, the LISA model would need to incorporate the influences of neuromodulators. Monoamine neurotransmitters, including dopamine, norepinephrine and serotonin, each have complex neuromodulatory roles. Depending on the conditions, these compounds have very large inhibitory or excitatory effects on neural transmission within the PFC; moreover, their effects often seem to interact with one another [69]. A large number of psychiatric conditions that affect cognition involve monoamine dysregulation, and a suitably expanded LISA model could aid in understanding these disorders at the circuit level. It is likely that some individual differences in reasoning ability may be explained by polymorphisms in genes that code for monoamine receptors (e.g., [70]). In addition, the cognitive monitoring and evaluative functions of the anterior cingulate cortex may impact PFC processing via connections with neurons in the locus coeruleus that are the source of cortical noradrenergic modulation [71]. Thus, elaborating the LISA model to incorporate neuromodulatory effects could lead to further advances in understanding the conditions that lead to optimal (and suboptimal) reasoning in the human brain.

Box 1. Questions for future research

Recent work in the cognitive neuroscience of thinking and in computational modeling has raised new hypotheses about how thinking is realized in the human brain. Some current questions are:

- Can direct evidence be obtained to support the possible role of oscillatory neural activity in coding propositions in human PFC?
- Can computational models employing oscillatory algorithms such as cross-cortical coupling be used to predict the brain's complex network dynamics during relational reasoning as measured via electrophysiology?
- How are the various types of inhibition necessary for a model such as LISA implemented in the brain throughout the time course of relational reasoning?
- What functions does the RLPFC support in relational reasoning at a computational level and how do these relate to its functions in other tasks?
- What is the relationship between dynamic role binding in the PFC and the binding operations subserved by the hippocampus and medial temporal cortex?
- What roles do neurotransmitters play in relational reasoning and can their effects be mapped onto components of a computational model?
- What learning processes are involved in creating the pools of semantic units that code the meanings of objects and relations?
- What role do subcortical structures play in relational reasoning? Are fronto-striatal circuits particularly important for mapping units, which must be maintained without continuous attention over short periods of time during reasoning?

Acknowledgements

Preparation of this paper was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract number D10PC20022. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained hereon are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the U.S. Government. Additional support was provided by the American Federation of Aging Research and Arthur Gilbert Foundation, the Illinois Department of Public Health, the Loyola University Chicago Dean of Arts and Sciences and the Graduate School (to R.G.M.). We thank Krishna Bharani for help in preparing the manuscript, and Paul Kogut and the rest of the Lockheed Martin FRAMES team for many helpful discussions. Two anonymous reviewers provided valuable comments on an earlier draft.

References

1 Holyoak, K.J. (2012) Analogy and relational reasoning. In *The Oxford Handbook of Thinking and Reasoning* (Holyoak, K.J. and Morrison, R.G., eds), pp. 234–259, Oxford University Press
2 Holyoak, K.J. and Cheng, P.W. (2011) Causal learning and inference as a rational process: the new synthesis. *Annu. Rev. Psychol.* 62, 135–163
3 Penn, D.C. and Povinelli, D.J. (2007) Causal cognition in human and nonhuman animals: a comparative, critical review. *Annu. Rev. Psychol.* 58, 97–118
4 Penn, D.C. *et al.* (2008) Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130 discussion 130–178
5 Badre, D. (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn. Sci.* 12, 193–200
6 Badre, D. and D'Esposito, M. (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat. Rev. Neurosci.* 10, 659–669
7 Koechlin, E. and Hyafil, A. (2007) Anterior prefrontal function and the limits of human decision-making. *Science* 318, 594–598
8 Knowlton, B.J. and Holyoak, K.J. (2009) Prefrontal substrate of human relational reasoning. In *The Cognitive Neurosciences* (Gazzaniga, M.S., ed.), pp. 1005–1017, MIT Press
9 Morrison, R.G. and Knowlton, B.J. (2012) Cognitive neuroscience of higher cognition. In *The Oxford Handbook of Thinking and Reasoning* (Holyoak, K.J. and Morrison, R.G., eds), pp. 67–89, Oxford University Press
10 Bunge, S.A. *et al.* (2005) Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. *Cereb. Cortex* 15, 239–249

11 Bunge, S.A. *et al.* (2009) Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46, 338–342
12 Charron, S. and Koechlin, E. (2010) Divided representation of concurrent goals in the human frontal lobes. *Science* 328, 360–363
13 Cho, S. *et al.* (2010) Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cereb. Cortex* 20, 524–533
14 Christoff, K. *et al.* (2001) Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149
15 Christoff, K. *et al.* (2003) Evaluating self-generated information: anterior prefrontal contributions to human cognition. *Behav. Neurosci.* 117, 1161–1168
16 Christoff, K. *et al.* (2009) Prefrontal organization of cognitive control according to levels of abstraction. *Brain Res.* 1286, 94–105
17 Dumontheil, I. *et al.* (2010) Development of the selection and manipulation of self-generated thoughts in adolescence. *J. Neurosci.* 30, 7664–7671
18 Green, A.E. *et al.* (2010) Connecting long distance: semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cereb. Cortex* 20, 70–76
19 Kroger, J.K. *et al.* (2002) Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cereb. Cortex* 12, 477–485
20 Sweis, B.M. *et al.* (2012) The time course of inhibition in analogical reasoning: an event-related potential approach. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (Miyake, N. *et al.*, eds), Cognitive Science Society
21 Wendelken, C. and Bunge, S.A. (2010) Transitive inference: distinct contributions of rostrolateral prefrontal cortex and the hippocampus. *J. Cogn. Neurosci.* 22, 837–847
22 Wendelken, C. *et al.* (2008) 'Brain is to thought as stomach is to ??': investigating the role of rostrolateral prefrontal cortex in relational reasoning. *J. Cogn. Neurosci.* 20, 682–693
23 Watson, C.E. and Chatterjee, A. (2012) A bilateral frontoparietal network underlies visuospatial analogical reasoning. *Neuroimage* 59, 2831–2838
24 Morrison, R.G. *et al.* (2004) A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *J. Cogn. Neurosci.* 16, 260–271
25 Krawczyk, D.C. *et al.* (2008) Distraction during relational reasoning: the role of prefrontal cortex in interference control. *Neuropsychologia* 46, 2020–2032
26 Wilson, W.H. *et al.* (2001) The STAR-2 model for mapping hierarchically structured analogs. In *The Analogical Mind* (Gentner, D. *et al.*, eds), pp. 125–159, MIT Press
27 Hummel, J.E. and Holyoak, K.J. (2003) A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* 110, 220–264
28 Shastri, L. (2002) Episodic memory and cortico-hippocampal interactions. *Trends Cogn. Sci.* 6, 162–168
29 Jilk, D.J. *et al.* (2008) SAL: An explicitly pluralistic cognitive architecture. *J. Exp. Theor. Artif. Intell.* 20, 197–218
30 Anderson, J.R. *et al.* (2007) Information-processing modules and their relative modality specificity. *Cogn. Psychol.* 54, 185–217
31 Hummel, J.E. and Holyoak, K.J. (1997) Distributed representations of structure: a theory of analogical access and mapping. *Psychol. Rev.* 104, 427–466
32 Hummel, J.E. and Landy, D.H. (2009) From analogy to explanation: relaxing the 1:1 mapping constraint *very carefully*. In *New Frontiers in Analogy Research: Proceedings of the Second International Conference on Analogy* (Kokinov, B. *et al.*, eds), pp. 211–221, New Bulgarian University
33 Taylor, E.G. and Hummel, J.E. (2009) Finding similarity in a model of relational reasoning. *Cogn. Syst. Res.* 10, 229–239
34 Palva, J.M. *et al.* (2010) Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7580–7585
35 Singer, W. and Gray, C.M. (1995) Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586
36 Cowan, N. (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114 discussion 114–85
37 Buschman, T.J. *et al.* (2011) Neural substrates of cognitive capacity limitations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11252–11255
38 Morrison, R.G. *et al.* (2011) A computational account of children's analogical reasoning: balancing inhibitory control in working memory and relational representation. *Dev. Sci.* 14, 516–529
39 Doumas, L.A.A. *et al.* (2008) A theory of the discovery and predication of relational concepts. *Psychol. Rev.* 115, 1–43
40 Doumas, L.A.A. *et al.* (2009) The development of analogy: task learning and individual differences. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (Taatgen, N. *et al.*, eds), pp. 3133–3138, Erlbaum
41 Viskontas, I.V. *et al.* (2004) Relational integration, inhibition, and analogical reasoning in older adults. *Psychol. Aging* 19, 581–591
42 Uhlhaas, P.J. *et al.* (2009) Neural synchrony in cortical networks: history, concept and current status. *Front. Integr. Neurosci.* 3, 17
43 Buzsaki, G. (2006) *Rhythms of the Brain*, Oxford University Press
44 Lu, H. *et al.* (2006) Role of gamma-band synchronization in priming of form discrimination for multiobject displays. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 610–617
45 Phillips, S. *et al.* (2012) Visual feature integration indicated by phase-locked frontal-parietal EEG signals. *PLoS ONE* 7, e32502
46 Sakurai, Y. and Takahashi, S. (2006) Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks. *J. Neurosci.* 26, 10141–10153
47 Canolty, R.T. *et al.* (2006) High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313, 1626–1628
48 Canolty, R.T. and Knight, R.T. (2010) The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515
49 Siegel, M. *et al.* (2012) Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* 13, 121–134
50 Rutishauser, U. *et al.* (2010) Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature* 464, 903–907
51 Berry, S.D. and Seager, M.A. (2001) Hippocampal theta oscillations and classical conditioning. *Neurobiol. Learn. Mem.* 76, 298–313
52 Hyman, J.M. *et al.* (2003) Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to the peak of theta and long-term depression when delivered to the trough. *J. Neurosci.* 23, 11725–11731
53 Asaad, W.F. *et al.* (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21, 1399–1407
54 Cromer, J.A. *et al.* (2010) Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66, 796–807
55 Song, S. *et al.* (2000) Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* 3, 919–926
56 Markram, H. *et al.* (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215
57 Jensen, O. and Colgin, L.L. (2007) Cross-frequency coupling between neuronal oscillations. *Trends Cogn. Sci.* 11, 267–269
58 Siegel, M. *et al.* (2009) Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21341–21346
59 Freedman, D.J. *et al.* (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316
60 Warden, M.R. and Miller, E.K. (2010) Task-dependent changes in short-term memory in the prefrontal cortex. *J. Neurosci.* 30, 15801–15810
61 Freedman, D.J. *et al.* (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246
62 Sjostrom, P.J. and Hausser, M. (2006) A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron* 51, 227–238
63 Fuster, J.M. (2008) *The Prefrontal Cortex*, Academic Press
64 Shimamura, A.P. (2000) Toward a cognitive neuroscience of metacognition. *Conscious Cogn.* 9, 313–323 discussion 324–326
65 Aron, A.R. *et al.* (2004) Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177
66 Richland, L.E. *et al.* (2006) Children's development of analogical reasoning: insights from scene analogy problems. *J. Exp. Child Psychol.* 94, 249–273

67 Green, A.E. *et al.* (2006) Frontopolar cortex mediates abstract integration in analogy. *Brain Res.* 1096, 125–137
68 Palva, S. and Palva, J.M. (2012) Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs. *Trends Cogn. Sci.* 16, 219–230
69 Di Pietro, N.C. and Seamans, J.K. (2011) Dopamine and serotonin interactively modulate prefrontal cortex neurons in vitro. *Biol. Psychiatry* 69, 1204–1211
70 Tan, H.Y. *et al.* (2007) Epistasis between catechol-O-methyltransferase and type II metabotropic glutamate receptor 3 genes on working memory brain function. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12536–12541
71 Aston-Jones, G. and Cohen, J.D. (2005) Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *J. Comp. Neurol.* 493, 99–110
72 Visser, M. *et al.* (2010) Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J. Cogn. Neurosci.* 22, 1083–1094
73 Squire, L.R. *et al.* (2004) The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306