Trends in
Neurosciences

CellPress

Review

The 'neat' and 'messy' in task-dependent neural geometry and computation

Cheng Xue (薛澄)[1,*] and Gouki Okazawa[2,*]

To solve diverse real-world tasks, the brain must flexibly switch between task rules and adjust computations. Recent advances in analyzing neural data and modeling neural networks have revealed their 'neat' features: neuronal population activity encodes distinct task states and forms structured, interpretable representations of task variables, enabling efficient task switching. However, 'messy' features are also observed: task-irrelevant variables shape neuronal responses, representations become entangled, and behaviors exhibit apparent suboptimalities such as switch costs. We review these dual facets of experimental observations of task-dependent computation in the brain, particularly in primates. Recognizing this duality points to two future directions: refining theories to better capture brain-specific constraints and leveraging network models fitted directly to behavioral and neural data.

Highlights

Advanced analyses of neuronal population activity, along with comparisons to artificial neural networks, have transformed our understanding of how the brain performs and switches between multiple cognitive tasks.

Many 'neat' features of the brain's computations have been identified: neural populations encode task-relevant information in a structured way, supporting efficient task switching.

However, 'messy' features have also been observed: neural activity is modulated by task-irrelevant variables, representations often defy simple normative interpretation, and behavior shows suboptimalities such as switch costs.

We propose two paths forward: (i) pursue experiments, analyses, and models that better capture brain-specific constraints and (ii) generate and test hypotheses from artificial network models directly constrained by behavioral and neural data.

Advances in understanding of task-dependent neural computations

Animals have evolved remarkable cognitive abilities to solve and switch between diverse tasks. In different contexts, the brain engages in distinct computations that support specific stimulus-response associations, decision policies, and solutions [1]. At the same time, it leverages shared computational elements across tasks to enable efficient switching and generalization [2–4]. A growing body of research is shedding light on these sophisticated processes. In particular, advanced analysis techniques to visualize neuronal population activity provide new frameworks for interpreting complex neuronal signals that multiplex many task variables [2,5–9]. Moreover, artificial neural networks (ANNs) offer implementable hypotheses of how neural circuits perform tasks and now serve as benchmarks for behavioral and neural data [10–12]. These advances have occurred in tandem with richer task designs capable of teasing apart competing hypotheses [1].

This review aims to provide an overview of both 'neat' theoretical frameworks and 'messy' biological realities of task-dependent cognition learned from recent advances. We begin by summarizing recent insights into how the brain 'neatly' achieves flexible task switching—by forming distinct task states and representing task variables in a disentangled manner that supports efficient switching. We then turn to more complex (or 'messy') facets of neural computations; the brain also encodes task-irrelevant variables, forms entangled representations, and exhibits apparent suboptimalities such as switch costs. Throughout, we use the term 'neat' to denote simplified, interpretable computational frameworks and 'messy' to denote intricate experimental observations that resist immediate accommodation within existing normative models. This distinction is somewhat arbitrary and depends on perspective, but we find it useful for organizing current understandings and identifying unresolved problems. Of note, the term 'messy' is not intended to carry a negative connotation in this context; on the contrary, we view these observations as especially valuable for envisioning how future research might unfold through both hypothesis-driven and data-driven approaches, as discussed toward the end of

[1]Department of Neurobiology, University of Chicago, Chicago, IL, USA
[2]Institute of Neuroscience, Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Center for Excellence in Brain Science and Intelligence Technology, International Center for Primate Brain Research, Chinese Academy of Sciences, Shanghai, China

*Correspondence:
cxue@uchicago.edu (C. Xue) and
okazawa@ion.ac.cn (G. Okazawa).

https://doi.org/10.1016/j.tins.2026.04.012

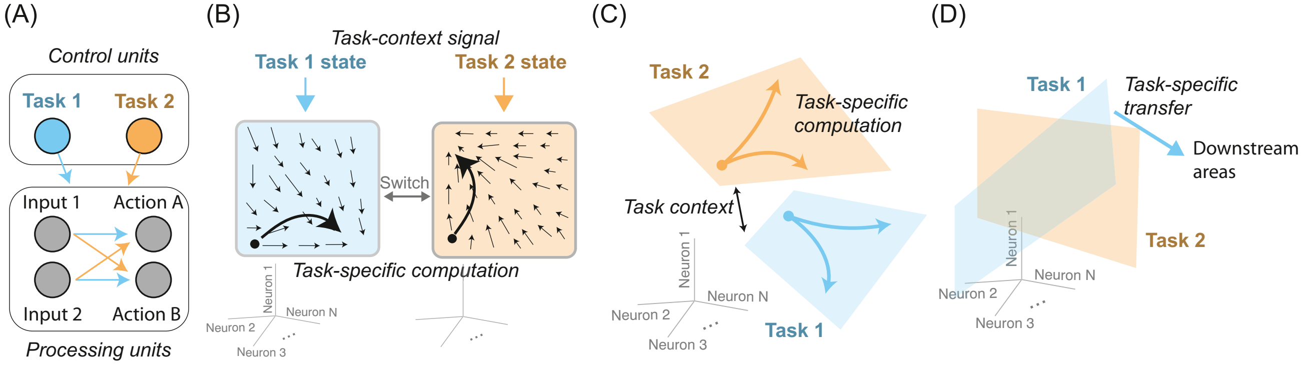
the article. We focus primarily on non-human primate research and related computational studies, while also referencing human and rodent work.

‘Neat’ elements in task-dependent neural computations

Neural geometry for task state and task computation

Task switching requires the brain to flexibly route information flow and adjust computations based on task demands [13]. For example, in a typical task-switching paradigm, subjects alternate between two distinct rules that require mapping the same sensory inputs to different behavioral outputs (e.g., categorizing either the motion or the color of colorful moving dots [5,14]). To account for this flexibility, classical models introduce dedicated control units that encode the task state (e.g., two neurons, each representing one of the two tasks) and modulate information routing in processing units that transform the same inputs into different outputs [13,15] (Figure 1A). This idea appeared in earlier connectionist models that used simplified network architectures [15], but it can also be instantiated in more biologically realistic models, as exemplified in a circuit model of the prefrontal cortex [21].

Another emerging approach to explain the behavior is computation through neural population dynamics [6,8,9,16] (Figure 1B,C). Neuronal population activity recorded from an area can be visualized in a neural state space, whose axes correspond to each neuron’s activity, or in a subspace obtained by dimensionality-reduction techniques (Figure 1B). In this space, the population activity (a point in the space) evolves over time to form trajectories, the shape of which is referred to as the neural geometry of the population response. The trajectories are under the influence of a dynamical landscape (small arrows in Figure 1B) determined by network connectivity and states. Task-context signals modulate the dynamical landscape, such that the same sensory input can lead to different neural trajectories and thus different outputs [5,22]. If task-context signals are themselves embedded in state space, the population dynamics may appear to occupy separable neural subspaces or ‘manifolds’ across tasks (Figure 1C; see [6], for a discussion on the distinction between subspaces and manifolds), with different dynamical landscapes governing neural trajectories and, consequently, the computations [22,23]. This perspective has



Trends in Neurosciences

Figure 1. An overview of ‘neat’ computational frameworks to explain the neural mechanisms of flexible sensorimotor associations. (A) Classical network models of flexible decision making often posit that control units maintain task states (colored circles) and modulate processing units (gray circles) that associate different sensory inputs to their respective action (colored arrows activated by the unit of the same color) [13,15]. (B) More recent network models explain this modulation as changes in the dynamical landscape in the neural state space, whose axes correspond to each neuron’s activity [5,8,16]. Neuronal population activity (a point in this space) evolves over time to form trajectories (large arrow) shaped by underlying dynamical landscape (small arrows), which are modulated by external control or task-context signals. (C) Alternatively, distinct computations may occur in separate neural manifolds, switching between tasks. Neuronal population activity follows different dynamics in different subspaces, enabling task-specific computations. (D) Task-dependent computations can also be implemented through selective communications between brain areas [17–20]; different subspaces transmit signals to downstream targets in different tasks.

emerged through the interplay between population-level neural analyses and recurrent neural network (RNN) models, which form internal dynamics akin to those observed in actual neural data when trained on the same tasks [5].

The description of two-task scenarios can be further generalized to a broader 'task space' that spans diverse computations. For example, if RNNs are trained on many tasks (e.g., parametric working memory, multi-sensory integration, and delayed match-to-sample [24,25]), their preparatory activity encodes the momentary task rule in a graded manner, creating a continuous task space. This space can recruit different computational elements for different tasks, such as point attractors for discrete classification, line attractors for continuous memory, and limit cycles for rhythmic sequence generation [24,26]. Experimentally, human neuroimaging studies reveal graded multi-task representations across large networks of brain regions [27,28], although characterizing the precise dynamical landscape of these representations in animals remains a challenge [29].

Such a 'neat' account of task-dependent computations also extends to communication across brain areas. During attentional control, associative brain areas maintain task states and modulate the processing of task-relevant information in the sensory cortex [30]. Task states can also enhance communication between specific subsets of population activity across areas (or 'communication subspaces') (Figure 1D) [17–20] or enhance relevant information through oscillation [31], thereby enabling selective information transmission. Similar forms of hierarchical control are observed beyond attentional control; for example, in a value-based decision-making task, the hippocampus encodes abstract task context and relays it to the orbitofrontal cortex that encodes task-specific item values [32].

These explanatory frameworks for flexible information routing (e.g., attractor dynamics, top-down signal modulation, and communication subspace) have often been introduced by separate lines of literature, and it remains to be seen whether and how they can be unified. Since similar representational geometry and computation can arise from different network implementations, investigating the circuit structures and connectivity behind neuronal population dynamics would help integrate these frameworks [33].

Normative task-switching behavior and structured task-variable encoding

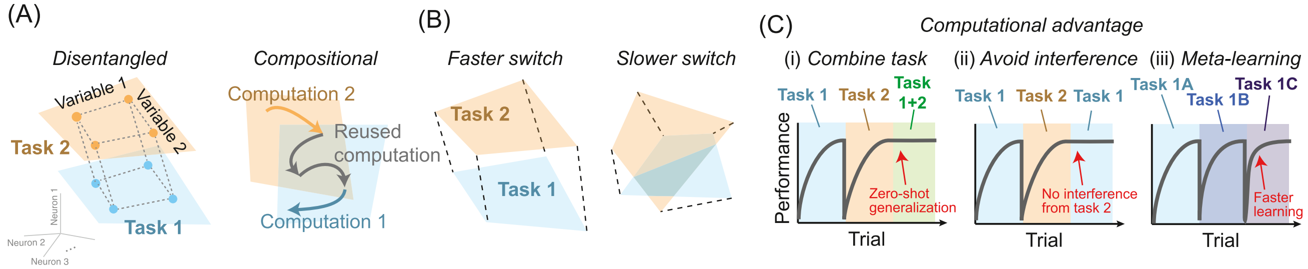
How does the brain successfully switch between task states? It must maintain the rules of multiple tasks separately without interference and rapidly activate the relevant task state upon recognizing the context. If there is no explicit environmental cue, the brain must infer the current task context from behavioral outcomes (specifically, reward or punishment).

Many studies have demonstrated that animals can perform such an inference and switch task rules efficiently [34–36]. This form of rapid, inference-based switching is often contrasted with slow rule learning or model-free reinforcement learning [37], which gradually refines stimulus-response associations or action values. Inference-based switching also enables more sophisticated updates of behavior due to internal models of hidden task structures. For example, after receiving error feedback for one stimulus, monkeys can infer a rule switch and apply the new rule to a different, unseen stimulus [38,39]. Another example is the ability of monkeys to evaluate their own performance according to stimulus difficulty in perceptual tasks and decide to switch rules based on whether the error occurred on an easy or difficult stimulus [40–42].

What neural representations support rapid and flexible rule switching? A notable emerging idea is the use of 'disentangled' (or 'factorized') [39,43] and 'compositional' [25,44] encodings of task-relevant variables in the brain. 'Disentangled' representations encode distinct task variables in dissociable formats (Figure 2A, left), enabling each variable to have a generalizable decoder independently of changes in other variables. This facilitates rapid switching of computations. 'Compositional' representations build complex functions from simpler, reusable components [49] (Figure 2A, right). Such modularity enables the flexible recombination of basic operations, supporting rapid adaptation to novel and complex problems. These encoding formats are also conceptually related to cognitive maps [50] and schemas [51], which are often invoked to explain computations in the medial temporal lobe and medial/orbital prefrontal cortex, although the present review does not discuss their relationships in detail (see [3,51]).

Evidence from neurophysiological studies supports both disentangled task variable representations and compositionality [29,39,43–45,52]. In a study involving monkeys, when the task required flexible switching of associations between stimuli, reward, and actions, the hippocampus and prefrontal cortex formed disentangled representations of these task variables that generalized across conditions [39]. Notably, later studies found that such disentangled representations are correlated with the subject's ability to rapidly switch task contexts [43] (Figure 2B) and can also facilitate compositional generalization, enabling zero-shot learning [44] (Figure 2C(i)). In another study, when monkeys were trained to switch between perceptual tasks with partially shared elements (e.g., relevant visual features and saccade targets), the prefrontal cortex reused neural subspaces that encoded these elements across tasks [29]. Yet another study found that monkeys trained to draw complex symbols employed strategies that decomposed symbols into simpler motor elements, a process reflected in premotor cortical activity [45].

These disentangled encodings of task variables can also be replicated with simple ANNs trained to perform cognitive tasks. For example, when multi-layer networks are trained to perform multiple tasks without interference, disentangled representation naturally emerges [46], underscoring the importance of training networks with multiple objectives (Figure 2C(ii)). When RNNs are trained to perform many tasks, they also exhibit common dynamic motifs that allow compositional generalization [24,25]. Even when they are repeatedly trained to perform tasks having the same trial structure but different sensorimotor mappings, they learn each task faster and faster (a phenomenon referred to as meta-learning; Figure 2C(iii)) due to the emergence of shared dynamics [48]. Furthermore, explicit training of RNNs with language instructions (through large-language model inputs)—analogous to human instruction—enhances compositionality and supports zero-



Trends in Neurosciences

Figure 2. Structured neural encoding of task variables enables efficient task switching. (A) Neuronal activity during flexible task switching often exhibits disentangled (or factorized) encoding, in which multiple task-dependent variables are represented in a separable form (left). It can also form compositional encoding, where shared computational elements are reused across tasks (right). (B) Disentangled encoding facilitates task switching. For example, human studies have shown that individuals with disentangled representations of task variables in the hippocampus can switch task rules immediately, whereas those with entangled representations switch more slowly [43]. (C) Disentangled and compositional encodings offer various computational advantages: (i) With compositional code, a novel task (task 1 + 2) can be solved immediately by combining solutions from previously learned tasks [29,44,45]. (ii) Learning a new task (task 2) does not interfere with performance on a previously learned task (task 1) [46,47]. (iii) As more tasks are learned, the learning rate increases [48].

shot transfer [53]. This is consistent with the finding that neuronal populations in the hippocampus recorded from human epileptic patients formed disentangled task-variable representations immediately after the patients received explicit task instructions [43].

In summary, advances in population-level neural analysis and artificial network modeling have provided a 'neat' view of how the brain represents and switches between tasks. In the neural state space, the formation of distinct subspaces and/or the modulation of attractor dynamics enables task-specific computations. In addition, the brain employs disentangled and compositional population codes to flexibly switch between task rules and generalize to novel contexts. These encoding motifs can be replicated by RNNs trained on similar tasks. These findings provide reason for optimism that advancing along this path will lead to a clearer understanding of task-dependent cognition.

Complexity on the computational level: suboptimal task switching

However, the 'neat' view does not explain many cases of suboptimal behavior in complex task control, which, in fact, offer important mechanistic insight [54,55]. A prominent example is 'switch costs': humans respond more slowly and less accurately immediately after a task switch [56,57] (but see [58–60] for mixed results in monkeys). Different psychological theories attribute these costs to an additional 'task-set reconfiguration' process [61] or the 'inertia' of previous task sets [62]. But how do such deviations from the 'neat' task control emerge in neuronal circuits?

One physiological observation related to suboptimal task switching is the graded and sluggish transition of neural signals reflecting task states (Figure 3A). In one study, for example, monkeys were trained to switch between spatial frequency and position discrimination tasks and use different sets of choice targets for different tasks [40]. Despite this requirement for an immediate switch in behavioral reports, task states encoded in the posterior parietal cortex shifted gradually over several trials before and after the behavioral switch. This graded transition impaired both behavioral accuracy and the orthogonality of V1 representations for orientation and position around the time of the switch [40,63]. A similar graded encoding of task states has been reported in other brain regions [29], supporting the notion that the brain does not necessarily represent task rules as discrete states.

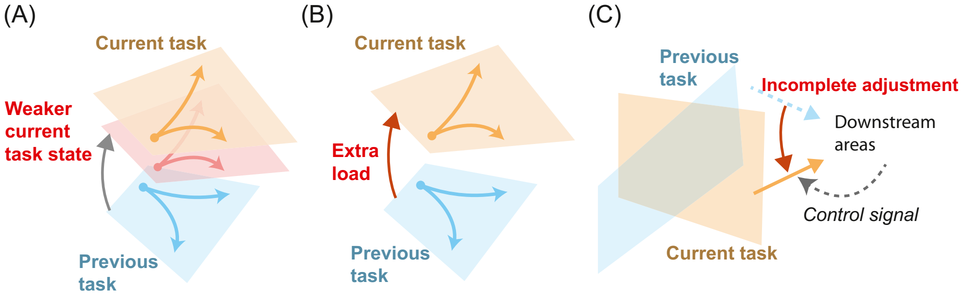


Figure 3. Complexity on the computational level: suboptimal task switching. In many task-switching paradigms, humans exhibit switch cost: slower reaction times and reduced accuracy immediately after a task switch. The 'neat' perspective does not immediately explain the sources of the cost. (A) One possible cause is that the brain's encoding of task state is not discrete even under two distinct task rules. Weaker task-state signals may lead to suboptimal behavior [40,63]. (B) Also, the transition between task states requires extra neural processes, leading to delayed responses after switching. Such a delay does not arise in task-optimized RNNs [64]. (C) Alternatively, even after switching a task state, control signals cannot fully adjust the information routing, leading to reduced performance. RNNs: recurrent neural networks.

Physiological studies in monkeys further support the existence of additional computations at the moment of a task switch (Figure 3B), observed as characteristic switch-selective neuronal responses in multiple brain regions [41,65–67]. These switch-selective responses predict whether monkeys successfully switch a rule, suggesting that they are required for changing task sets or strategies [65,67]. Interestingly, however, standard task-optimized RNNs do not require specific processes at task switches, as they can instantly adjust their internal states in response to context cues [63,64]. To produce RNNs that exhibit switching costs, one must instead fit the models directly to behavioral data [64]. These behaviorally fit RNNs display prolonged transitions between task states in response to external cues [64,68], resembling the reaction-time delays in humans. Alternatively, RNNs exhibit a switch cost when they are trained on task sequences where only one response is needed and then evaluated on sequences where responses in two consecutive trials are needed [69]. Networks retrained with these double-trial sequences learn an active reconfiguration strategy and exhibit switch-related signals similar to brain activity.

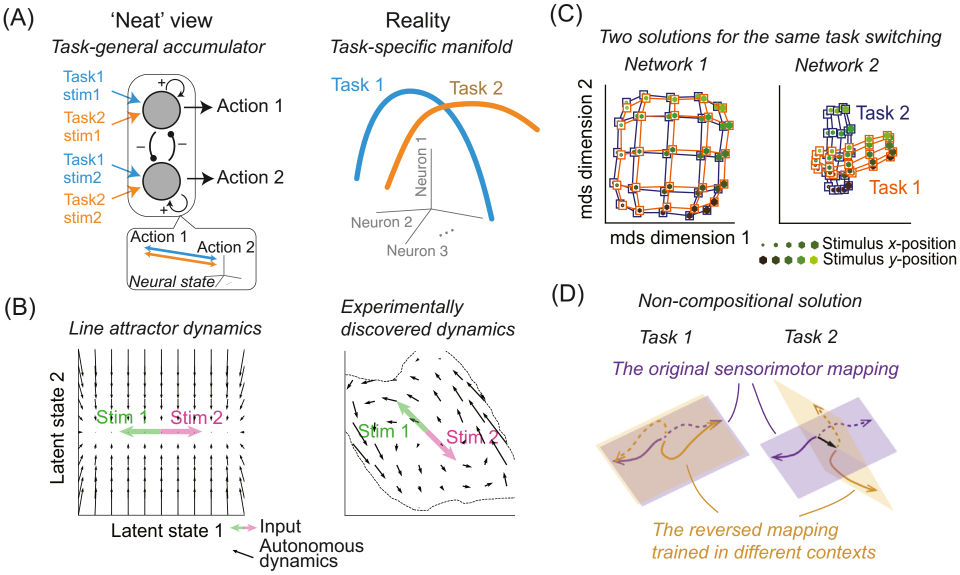
Yet another possible source of switch cost is the incomplete adjustment of information flow by top-down signals (Figure 3C). Switch cost is typically reduced when subjects are given more time to switch, but interestingly, it often persists even after ample preparation time—a phenomenon known as ‘residual switch cost’ [61]. Studies using perceptual tasks have suggested that this cost arises because sensory weighting cannot be fully adjusted based on a context cue alone [70,71]. Consistently, neuronal recordings have revealed lingering influences of task-irrelevant signals in sensory and associative areas [72–74], as well as neural signals that appear to actively suppress task-irrelevant information after a switch [75]. These findings suggest that task switching is not merely accomplished by an instantaneous transition of task states in a centralized control hub. Instead, it also requires adjustments across a distributed network of brain regions spanning the sensorimotor hierarchies [1,40].

Finally, while the brain struggles with switching rules, it can also fail to maintain a consistent rule even under stationary experimental conditions. In perceptual tasks, this manifests as erroneous responses to unambiguous stimuli, reflected in the lapse rates of psychometric functions [76,77]. One interpretation suggests that these lapses are an outcome of an exploration–exploitation trade-off that enables occasional deviations from the current rule to detect environmental changes [77,78]. Supporting this view, individuals with higher lapse rates tend to adapt more quickly when task demands shift [78]. Thus, apparent suboptimality in rule switching and maintenance might reflect the brain’s adaptive strategy to balance flexibility and stability [57,68]. If so, RNNs would also face the same trade-off, raising an intriguing question about what modifications to RNNs would lead to brain-like solutions.

Complexity on the implementation level: entangled representations

Non-canonical dynamics and the encoding of task-irrelevant variables

Analyses of neural geometry have shown that the seemingly ‘messy’ mixed selectivity of individual neurons can be visualized as ‘neat’ combinations of interpretable elements [5,6,8], but population analyses have also revealed deviations from the picture depicted by some of the ‘neat’ cognitive models. One example involves models of binary perceptual decision making: dominant models posit two neural modules competing for two choices via self-excitation and mutual inhibition, regardless of input or task context differences [79] (Figure 4A, left). These models give rise to line or bistable point attractor dynamics, in which neuronal population activity converges to one of two states corresponding to each choice [81,82] (Figure 4B, left). However, neural geometry observed in multiple studies does not align well with this ‘neat’ model of competing choice modules. In the monkey parietal cortex, neuronal activity was found to form a task-specific,



Trends in Neurosciences

Figure 4. Complexity on the implementation level: entangled representations. (A) 'Neat' binary decision-making models have proposed two neural modules competing for choices across different tasks [79] (left). The neural modules yield task-general manifolds between the encoding of two action plans (inset on the left). But in reality, parietal neurons in monkeys show task-specific, curved neural manifolds misaligned with action encoding [80] (right). Blue and orange curves represent task-dependent manifolds drawn by connecting neuronal responses to different stimulus strengths during decision formation. (B) 'Neat' models in A exhibit line or bistable point attractor dynamics for decision making [81] (left). But the actual dynamics estimated from neural populations often reveal more complex patterns [82] (right). (C) Disentangled representations (Figure 2A, left) are not the only solution for task switching. Two neural networks trained on the same context-switch task can reach distinct solutions [83]. (D) Compositionality (Figure 2A, right) does not necessarily arise either. When monkeys perform the same visuomotor response in two adaptation contexts, the brain uses two distinct neural subspaces, rather than reusing computational elements [84].

curved neural manifold to represent the strength of evidence for decisions, misaligned with the neurons' encoding of preferred saccade choices [80] (Figure 4A, right). This curvature did not appear to have a functional role and could result from biological constraints of spiking activity. Similar curvature was also found in the ventromedial prefrontal cortex during value decision making, explaining irrational choice behavior [85].

Another recent study directly estimated underlying dynamical landscapes from neural population data in rats performing an auditory discrimination task and found that the dynamics are complex and different from those predicted by previously proposed models [82] (Figure 4B, right). The observed dynamics instead reflect a transition from an initial phase dominated by sensory inputs to a phase dominated by autonomous dynamics, marking the moment of decision commitment. The complex geometries in these studies reveal brain-specific computations that could not be deduced from existing normative models.

More broadly, neural activity in decision-making circuits often reflects various task-irrelevant variables, which are typically overlooked by normative models. When population activity is decomposed into task-dependent and task-independent components, the majority of variance is attributed to the task-independent part [86]. For example, during cognitive tasks, a substantial portion of prefrontal activity in monkeys correlates with seemingly unrelated factors such as body posture [87], a pattern also observed in rodent V1 [88], though to a much lesser extent in primate

V1 [89]. Another source of task-irrelevant signals can arise from fluctuations in internal states such as engagement, which can profoundly influence neuronal responses [90–93]. Population-level neural analyses have indeed revealed interference between slow neural drifts and task-dependent computations [91,92]. It remains largely unclear how such task-'irrelevant' signals can be incorporated into models of cognition.

Entangled, non-compositional encoding

The 'neat' view posits that disentangled encoding of task variables underlies flexibility [43] (Figure 2A), yet network implementations do not strictly conform to such representations. For example, networks trained on identical context-switching tasks can develop distinct geometries [83]: one disentangled (Figure 4C, left) and another exhibiting complex, entangled patterns (Figure 4C, right), depending on their initial connection weights. Indeed, such variability in solutions is pervasive in biological networks [94–98] and artificial networks [95,99,100], challenging the necessity of a specific 'neat' geometry to solve a task.

Compositionality is also not guaranteed to be the best solution for all situations. For instance, even when monkeys perform reaching tasks with the same visuomotor mapping, posterior parietal neural activity occupies distinct subspaces between two adaptation contexts (Figure 4D) [84]. Thus, when learning a new task, the brain does not necessarily reuse the same resources but sometimes uses a separate neural subspace to minimize interference between tasks [23,98,101].

Indeed, in various task scenarios, disentangled or compositional encoding may not be advantageous. First, since disentangled representations should be recombined based on task demands in each trial, it can be more efficient to instead form a conjunctive code (task-specialized representations) when subjects are well trained on tasks [102,103]. Second, task-specific, rather than task-general, structures can help encode unique experiences. For example, in the monkey hippocampus, the population code for spatial position does not generalize across tasks [104]; instead, it nonlinearly mixes spatial and non-spatial information in a task-dependent manner, potentially supporting the maintenance of task-specific memory. Third, forming shared representations across tasks can be detrimental when the brain is engaged in multitasking, as overlapping information incurs interference [68]. In a working memory task requiring the memorization of a sequence of target locations, the prefrontal cortex uses both shared and distinct neural resources across targets [105], possibly as a solution to the trade-off between interference and representational cost. Hence, the brain must strike a delicate balance between shared and independent representations across tasks to support both flexible and robust behavior [68,106].

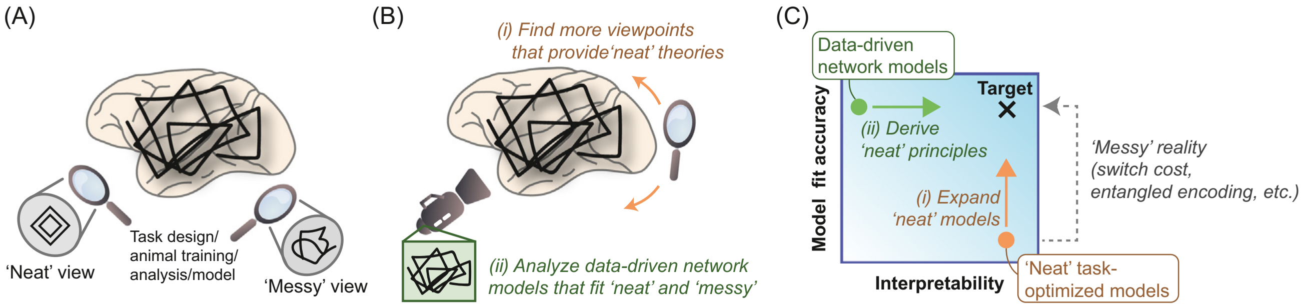
Two future research approaches to integrate the 'neat' and 'messy'

In the preceding sections, findings were deliberately organized into the 'neat' and 'messy' categories to highlight their multifaceted nature. Understanding the neurobiology of flexible cognition is arguably one of the most formidable challenges in neuroscience, yet advances in population-level neural analysis [2,5–9] and artificial network modeling [10–12] have made it possible to characterize its mechanisms in interpretable forms (Figures 1 and 2). However, studies have also found intricate complexity that defies simple normative accounts (Figures 3 and 4). These lines of evidence may seem contradictory, for instance, by revealing disentangled versus entangled representations, or task-dependent information routing versus suboptimal effects of task-irrelevant information. Yet they are unlikely to reflect mutually exclusive theories; rather, they collectively illuminate different facets of the brain's computations, each revealed through the particular lens of task design, analysis approach, and model choice. It is as if one is observing the projection of a complex sculpture (Figure 5A); it appears clearly structured when viewed from some angles (a magnifier in Figure 5A as an analogy for task designs, analyses, and models), whereas it looks messy from other angles. Most likely, this 'messy' side constitutes a substantial portion of the phenomenon, and thus cannot be ignored. How, then, can these divergent observations be integrated? In what follows, we explore two possible major directions for future research, drawing on this analogy (Figure 5B,C).

Expand the 'neat' interpretation (a hypothesis-driven approach)

One strategy is to expand the 'neat' interpretation through a hypothesis-driven approach. This approach focuses on finding better normative or task-optimized computational models that capture the biological constraints, turning apparent 'messiness' into an elegant interpretation (Figure 5B; 'moving the magnifier').

On the behavioral level, seemingly suboptimal strategies can actually be optimal under certain constraints or objective functions (i.e., what an organism seeks to optimize) [68,107,108]. Bayesian models of perception, for instance, have often been refined to account for data by modifying their prior distributions [109]. This strategy extends to ANNs through the neuroconnectionist program [11], which seeks to better align network models with experimental data by varying architectures, objective functions, and learning rules [10] together with training schedules (Figure 5C). For example, the human switch cost discussed earlier can be reproduced by RNNs in some cases if specific training sequences are used [69]. In another example, RNNs trained under ecologically relevant conditions—where errors are more frequent than correct outcomes—replicate rats' suboptimal reliance on correct feedback in learning tasks [107]. However, this modeling flexibility introduces a caveat: model assumptions can be iteratively adjusted to explain away the 'messy' results post hoc [55]. To guard against this, it will be increasingly important to identify unifying computational principles that explain diverse findings across multiple tasks and naturalistic conditions. Future work should also treat individual variability [94,95,97,98] and strategic fluctuations [93,110] not as nuisance noise but as meaningful features of behavior and models. Extending models in this direction may also capture dysfunction in neural computations, which opens avenues for clinical



Trends in Neurosciences

Figure 5. Two research directions for integrating the 'neat' and 'messy'. (A) Depending on the task designs, analysis methods, and models used to investigate brain computations, behavior and neural representations appear sometimes normative and structured ('neat') and sometimes not ('messy'). It is as if researchers are viewing a complex sculpture (represented by the tangled string on the brain schematic) from different angles (represented by magnifiers). (B) This analogy helps envision possible paths forward. One approach is to find more viewpoints that offer 'neat' theories ('moving the magnifier') by developing task designs, analysis tools, and models that better capture the brain's constraints and objective functions (i.e., what an organism seeks to optimize). The other approach is to employ large-scale network models that directly fit behavioral or neural data (represented by a video camera). Analyzing the fitted models may uncover interpretable principles of brain function. (C) One of the future goals is to construct models that are interpretable but also accurately capture the brain's task-dependent computations ('target' at the top right). 'Neat', task-optimized models (bottom right) can be interpretable but do not explain the 'messy' reality. The two directions represented in B take complementary paths toward achieving this goal.

applications (Box 1). At the same time, these approaches demand rigorous task designs to reduce interpretational uncertainty [119,120].

At the neuronal level, analysis of neural geometry has been an important tool for finding the 'neat' neuronal representation [6,8,9,16]. Its strength lies in the flexibility it offers for visualizing and interpreting neural activity. Conventional 'single-cell' analyses view a neural state space only along specific axes, such as population-averaged firing rate, whereas geometry analyses explore alternative axes (determined by dimensionality reduction techniques) and describe neural activity using richer language—such as rotation, orthogonality, and disentanglement—to relate neural activity to models. However, similar trial-average geometries do not necessarily reflect the same underlying computations [99]; thus, characterizing more fine-grained, single-trial dynamics may be required to interrogate the mechanisms [121]. Moreover, as we summarized earlier, the observed geometries often take forms that are uninterpretable by existing task-optimized models (Figure 4A,B). If brain-specific constraints (e.g., minimizing energy cost [105]) or broader objective functions (e.g., solving many tasks) are at play, better alignment with neural data may be achieved through the development of network models that are more biologically realistic [21] or perform many tasks [24,25].

Embrace the 'messy' reality (a data-driven approach)

An alternative direction is a data-driven approach that seeks to capture the full complexity of the system, both 'neat' and 'messy', using artificial network models (Figure 5B). At the behavioral level, RNNs can be trained to reproduce behavioral responses rather than being optimized for task performance. These data-driven networks serve as surrogate models to infer hidden internal dynamics in apparently 'messy' behavioral data [63,64,69,122]. For instance, this approach has been used to understand the neural mechanisms underlying switch cost [63,64]. At the neuronal level, a growing set of tools now supports fitting ANNs [123–126] or dynamical systems [99,127] to raw neural data as well as joint fitting of neural and behavioral data [128]. While such data-driven approaches are currently applied to individual datasets, it may eventually be possible to construct 'foundation' models that capture behavioral and neural data across a broad range of tasks. For example, a recent study [129] developed a large language model that fits human choice behavior across 160 psychological tasks, generalizes to held-out participants and unseen tasks, and aligns with neural activity.

Box 1. Integrating the 'neat' and 'messy' for future translational breakthroughs

Tests for cognitive flexibility—such as the Wisconsin Card Sorting Task—have long been considered candidate tools for diagnosing mental disorders such as Alzheimer's disease and obsessive-compulsive disorder [111], suggesting that the two future directions outlined in the main text (Figure 5) may also contribute to translational research.

The hypothesis-driven approaches (expand 'the neat') emphasize how neural computations align with or deviate from normative/task-optimized models, naturally extending to questions about how such computations are specifically altered in clinical populations. For example, models of reinforcement learning and decision making (e.g., the drift diffusion model) have been applied to clinical groups to pinpoint altered mechanistic components [112]. This approach can be extended to task-optimized ANNs; by adjusting their architectures, objective functions, and learning rules [10], one can investigate which changes in these components best approximate the differences between healthy and clinical populations [11].

The second, data-driven approaches propose using artificial network models to capture the full complexity of biological data. Most mental disorders resist explanation as simple breakdowns of isolated modules or causal chains [113], suggesting that normative/task-optimized models may fall short in accounting for all relevant details. Larger-scale networks, by contrast, may directly fit 'big data' from large patient cohorts and yield foundation models that generalize across individuals and tasks [114,115]. Currently, deep networks are used in translational research primarily to improve diagnostic accuracy based on biomarkers such as structural MRI, electroencephalography, and genetic data [116,117]. In these applications, deep networks are used as data analysis tools, without necessarily capturing the neural mechanisms of cognitive dysfunction. Future research should integrate task-related (cognitive) and task-irrelevant biological data [118].

There have also been growing efforts to derive the 'neat' computational principles from data-driven models (Figure 5C). At the behavioral level, tiny RNNs [122] with only one to four units can outperform classical cognitive models in fitting behavioral data and uncover interpretable computations that deviate from standard theories like Bayesian estimation. For neural data, employing low-rank architectures [130,131] or latent variable models [124,132] in RNNs facilitates mapping their complex dynamics to interpretable components. While these methods have proven powerful for simple tasks, extending them to complex settings remains a critical future direction. Researchers should also remain cautious: mere focus on data fit performance risks obtaining models that largely deviate from genuine biological mechanisms [133,134].

Concluding remarks

In this review, we provided an overview of recent advances in the study of task-dependent computations in systems neuroscience, with a focus on non-human primate research and relevant computational work. The popularization of techniques for visualizing neural geometry and modeling artificial networks has contributed to a 'neat' view of how mixed and heterogeneous neuronal responses can be interpreted as disentangled, structured representations of task variables in the neural state space, enabling efficient task switching. At the same time, brain computations also involve intricate complexity ('messy' aspects) that resists simple normative accounts, such as the encoding of task-irrelevant variables, entangled representations, and suboptimal behaviors. These diverse findings are perhaps unsurprising, given the diversity of task designs, analyses, and modeling used to study task-dependent computations (Figure 5A).

This perspective motivates two major directions for future research (Figure 5B). The first seeks to expand the 'neat' frameworks by refining task design, data analysis, and computational modeling. The second leverages ANNs to directly fit behavioral and neural datasets, capturing both the 'neat' and the 'messy'. In either of the two directions, future models should aim to incorporate brain-specific constraints and objective functions to explain the rich diversity of phenomenology (see Outstanding questions). These expanded models may also offer insights into altered neural mechanisms in clinical populations, strengthening links between basic and translational research.

Outstanding questions

Under what conditions do simplified models fail to explain animals' task-dependent behavior? What metrics should be used to determine whether a simplified model is insufficient?

Can the geometrical perspective be expanded to interpret population neural activity across a broad continuum of tasks and strategies beyond the simplified experimental settings used to date?

What is the best approach to establish the causal role of biological constraints (e.g., wiring limitations) and brain-specific objective functions (e.g., exploration–exploitation balance) in task-dependent computations?

Does training history—or even life experience—affect how individuals solve cognitive tasks, leading to cross-individual variability? How can models account for such variability?

How can researchers scale up binary task-switching tasks and models to more naturalistic, complex behavioral settings while maintaining experimental control?

How can simple, principled frameworks and models be reliably extracted from large network models trained on neural or behavioral data?

Can features of task-dependent neural geometry (e.g., entangled representations) serve as biomarkers for diagnosing cognitive deficits in clinical populations?

Acknowledgments

We would like to thank Sina Tafazoli, Thomas Luo, and Bin Min for their helpful feedback on an earlier version of the manuscript. G.O. was supported by the Brain Science and Brain-like Intelligence Technology—National Science and Technology Major Project (Grant No. 2021ZD0203702), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB1010202), the Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX02), the National Natural Science Foundation of China (Grant No. 32371077), and the National Natural Science Fund for Excellent Young Scientists Fund Program (Overseas). C.X. was supported by the Simons Foundation grant (542961SPI).

Declaration of interests

The authors declare no competing interests.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT to proofread the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Okazawa, G. and Kiani, R. (2023) Neural mechanisms that make perceptual decisions flexible. *Annu. Rev. Physiol.* 85, 191–215
2. Ostojic, S. and Fusi, S. (2024) Computational role of structure in neural activity and connectivity. *Trends Cogn. Sci.* 28, 677–690
3. Vaidya, A.R. and Badre, D. (2022) Abstract task representations for inference and control. *Trends Cogn. Sci.* 26, 484–498
4. Yang, G.R. *et al.* (2019) How to study the neural mechanisms of multiple tasks. *Curr. Opin. Behav. Sci.* 29, 134–143
5. Mante, V. *et al.* (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84
6. Ebitz, R.B. and Hayden, B.Y. (2021) The population doctrine in cognitive neuroscience. *Neuron* 109, 3055–3068

7. Barack, D.L. and Krakauer, J.W. (2021) Two views on the cognitive brain. *Nat. Rev. Neurosci.* 22, 359–371
8. Vyas, S. *et al.* (2020) Computation through neural population dynamics. *Annu. Rev. Neurosci.* 43, 249–275
9. Perich, M.G. *et al.* (2025) A neural manifold view of the brain. *Nat. Neurosci.* 28, 1582–1597
10. Richards, B.A. *et al.* (2019) A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770
11. Doerig, A. *et al.* (2023) The neuroconnectionist research programme. *Nat. Rev. Neurosci.* 24, 431–450
12. Yang, G.R. and Wang, X.-J. (2020) Artificial neural networks for neuroscientists: a primer. *Neuron* 107, 1048–1070
13. Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
14. Siegel, M. *et al.* (2015) Cortical information flow during flexible sensorimotor decisions. *Science* 348, 1352–1355
15. Cohen, J.D. *et al.* (1990) On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97, 332
16. Duncker, L. and Sahani, M. (2021) Dynamics on the manifold: identifying computational dynamical activity from neural population recordings. *Curr. Opin. Neurobiol.* 70, 163–170
17. Semedo, J.D. *et al.* (2019) Cortical areas interact through a communication subspace. *Neuron* 102, 249–259
18. Srinath, R. *et al.* (2021) Attention improves information flow between neuronal populations without changing the communication subspace. *Curr. Biol.* 31, 5299–5313.e4
19. Ruff, D.A. and Cohen, M.R. (2016) Attention increases spike count correlations between visual cortical areas. *J. Neurosci.* 36, 7523–7534
20. Barbosa, J. *et al.* (2023) Early selection of task-relevant features through population gating. *Nat. Commun.* 14, 6837
21. Liu, Y. and Wang, X.-J. (2024) Flexible gating between subspaces in a neural network model of internally guided task switching. *Nat. Commun.* 15, 6497
22. Xu, M. *et al.* (2024) Dynamic tuning of neural stability for cognitive control. *Proc. Natl. Acad. Sci. U. S. A.* 121, e2409487121
23. Sun, X. *et al.* (2022) Cortical preparatory activity indexes learned motor memories. *Nature* 602, 274–279
24. Driscoll, L.N. *et al.* (2024) Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nat. Neurosci.* 27, 1349–1363
25. Yang, G.R. *et al.* (2019) Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* 22, 297–306
26. Khona, M. and Fiete, I.R. (2022) Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* 23, 744–766
27. Nakai, T. and Nishimoto, S. (2020) Quantitative models reveal the organization of diverse cognitive functions in the brain. *Nat. Commun.* 11, 1142
28. Ito, T. and Murray, J.D. (2023) Multitask representations in the human cortex transform along a sensory-to-motor hierarchy. *Nat. Neurosci.* 26, 306–315
29. Tafazoli, S. *et al.* (2026) Building compositional tasks with shared neural subspaces. *Nature* 650, 164–172
30. Ruff, D.A. *et al.* (2018) Cognition as a window into neuronal population space. *Annu. Rev. Neurosci.* 41, 77–97
31. Fries, P. (2015) Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235
32. Elston, T.W. and Wallis, J.D. (2025) Context-dependent decision-making in the primate hippocampal–prefrontal circuit. *Nat. Neurosci.* 28, 374–382
33. Langdon, C. *et al.* (2023) A unifying perspective on neural manifolds and circuits for cognition. *Nat. Rev. Neurosci.* 24, 363–377
34. Goudar, V. *et al.* (2024) A comparison of rapid rule-learning strategies in humans and monkeys. *J. Neurosci.* 44, e0231232024
35. Bouchacourt, F. *et al.* (2022) Fast rule switching and slow rule updating in a perceptual categorization task. *Elife* 11, e82531
36. Mansouri, F.A. *et al.* (2020) Emergence of abstract rules in the primate brain. *Nat. Rev. Neurosci.* 21, 595–610
37. Bartolo, R. and Averbeck, B.B. (2020) Prefrontal cortex predicts state switches during reversal learning. *Neuron* 106, 1044–1054.e4
38. Saez, A. *et al.* (2015) Abstract context representations in primate amygdala and prefrontal cortex. *Neuron* 87, 869–881
39. Bernardi, S. *et al.* (2020) The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* 183, 954–967.e21
40. Xue, C. *et al.* (2022) Dynamic task-belief is an integral part of decision-making. *Neuron* 110, 2503–2511.e3
41. Sarafyazd, M. and Jazayeri, M. (2019) Hierarchical reasoning by neural circuits in the frontal cortex. *Science* 364, eaav8911
42. Purcell, B.A. and Kiani, R. (2016) Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4531–E4540
43. Courellis, H.S. *et al.* (2024) Abstract representations emerge in human hippocampal neurons during inference. *Nature* 632, 841–849
44. Ito, T. *et al.* (2022) Compositional generalization through abstract representations in human and artificial neural networks. *Adv. Neural Inf. Proces. Syst.* 35, 32225–32239
45. Tian, L.Y. *et al.* (2025) Neural representation of action symbols in primate frontal cortex. *bioRxiv* https://doi.org/10.1101/2025.03.03.641276
46. Johnston, W.J. and Fusi, S. (2023) Abstract representations emerge naturally in neural networks trained to perform multiple tasks. *Nat. Commun.* 14, 1040
47. Losey, D.M. *et al.* (2024) Learning leaves a memory trace in motor cortex. *Curr. Biol.* 34, 1519–1531
48. Goudar, V. *et al.* (2023) Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving. *Nat. Neurosci.* 26, 879–890
49. Fodor, J.A. (1975) *The Language of Thought*, Harvard University Press
50. Whittington, J.C. *et al.* (2022) How to build a cognitive map. *Nat. Neurosci.* 25, 1257–1272
51. Bein, O. and Niv, Y. (2025) Schemas, reinforcement learning and the medial prefrontal cortex. *Nat. Rev. Neurosci.* 26, 141–157
52. Park, J. *et al.* (2025) Compositional architecture: orthogonal neural codes for task context and spatial memory in prefrontal cortex. *bioRxiv* https://doi.org/10.1101/2025.02.25.640211
53. Riveland, R. and Pouget, A. (2024) Natural language instructions induce compositional generalization in networks of neurons. *Nat. Neurosci.* 27, 988–999
54. Acerbi, L. *et al.* (2014) On the origins of suboptimality in human probabilistic inference. *PLoS Comput. Biol.* 10, e1003661
55. Rahnev, D. and Denison, R.N. (2018) Suboptimality in perceptual decision making. *Behav. Brain Sci.* 41, e223
56. Monsell, S. (2003) Task switching. *Trends Cogn. Sci.* 7, 134–140
57. Egner, T. (2023) Principles of cognitive control over task focus and task switching. *Nat. Rev. Psychol.* 2, 702–714
58. Avdagic, E. *et al.* (2014) Rapid cognitive flexibility of rhesus macaques performing psychophysical task-switching. *Anim. Cogn.* 17, 619–631
59. Caselli, L. and Chelazzi, L. (2011) Does the macaque monkey provide a good model for studying human executive control? A comparative behavioral study of task switching. *PLoS One* 6, e21489
60. Stoet, G. and Snyder, L.H. (2003) Executive control and task-switching in monkeys. *Neuropsychologia* 41, 1357–1364
61. Rogers, R.D. and Monsell, S. (1995) Costs of a predictable switch between simple cognitive tasks. *J. Exp. Psychol. Gen.* 124, 207
62. Allport, A. *et al.* (1994) Shifting intentional set: exploring the dynamic control of tasks. In *Attention and performance XV: conscious and nonconscious information processing* (Umiltà, C. and Moscovitch, M., eds), MIT Press
63. C. Xue, *et al.*, Feature interference as a neuronal basis for the behavioural cost of task uncertainty. *Nat. Neurosci.*, in press
64. Jaffe, P.I. *et al.* (2023) Modelling human behaviour in cognitive tasks with latent dynamical systems. *Nat. Hum. Behav.* 7, 986–1000
65. Kamigaki, T. *et al.* (2009) Cognitive set reconfiguration signaled by macaque posterior parietal neurons. *Neuron* 61, 941–951
66. Premereur, E. *et al.* (2018) Functional MRI in macaque monkeys during task switching. *J. Neurosci.* 38, 10619–10630

67. Fascianelli, V. *et al.* (2020) Neural correlates of strategy switching in the macaque orbital prefrontal cortex. *J. Neurosci.* 40, 3025–3034
68. Musslick, S. and Cohen, J.D. (2021) Rationalizing constraints on the capacity for cognitive control. *Trends Cogn. Sci.* 25, 757–775
69. Ritz, H. *et al.* (2024) Humans actively reconfigure neural task states. *bioRxiv* https://doi.org/10.1101/2024.09.29.615736
70. Luo, T. *et al.* (2025) Limitation of switching sensory information flow in flexible perceptual decision making. *Nat. Commun.* 16, 172
71. Elchlepp, H. *et al.* (2017) Shifting attention between visual dimensions as a source of switch costs. *Psychol. Sci.* 28, 470–481
72. Yeung, N. *et al.* (2006) Between-task competition and cognitive control in task switching. *J. Neurosci.* 26, 1429–1438
73. Yao, T. and Vanduffel, W. (2022) Neuronal congruency effects in macaque prefrontal cortex. *Nat. Commun.* 13, 4702
74. Donahue, C.H. and Lee, D. (2015) Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nat. Neurosci.* 18, 295–301
75. Buschman, T.J. *et al.* (2012) Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* 76, 838–846
76. Gupta, D. *et al.* (2024) Trial-history biases in evidence accumulation can give rise to apparent lapses in decision-making. *Nat. Commun.* 15, 662
77. Pisupati, S. *et al.* (2021) Lapses in perceptual decisions reflect exploration. *Elife* 10, e55490
78. Ebitz, R.B. *et al.* (2019) Tonic exploration governs both flexibility and lapses. *PLoS Comput. Biol.* 15, e1007475
79. Wang, X.-J. (2002) Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36, 955–968
80. Okazawa, G. *et al.* (2021) Representational geometry of perceptual decisions in the monkey parietal cortex. *Cell* 184, 3748–3761
81. Wong, K.-F. and Wang, X.-J. (2006) A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* 26, 1314–1328
82. Luo, T.Z. *et al.* (2025) Transitions in dynamical regime and neural mode during perceptual decisions. *Nature* 646, 1156–1166
83. Flesch, T. *et al.* (2022) Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron* 110, 1258–1270
84. Guo, H. *et al.* (2025) Sensorimotor environment but not task rule reconfigures population dynamics in rhesus monkey posterior parietal cortex. *Nat. Commun.* 16, 1116
85. Jurewicz, K. *et al.* (2024) Irrational choices via a curvilinear representational geometry for value. *Nat. Commun.* 15, 6424
86. Kobak, D. *et al.* (2016) Demixed principal component analysis of neural population data. *Elife* 5, e10989
87. Tremblay, S. *et al.* (2023) Neural cognitive signals during spontaneous movements in the macaque. *Nat. Neurosci.* 26, 295–305
88. Stringer, C. *et al.* (2019) Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364, eaav7893
89. Talluri, B.C. *et al.* (2023) Activity in primate visual cortex is minimally driven by spontaneous movements. *Nat. Neurosci.* 26, 1953–1959
90. Allen, W.E. *et al.* (2019) Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* 364, eaav3932
91. Hennig, J.A. *et al.* (2021) Learning is shaped by abrupt changes in neural engagement. *Nat. Neurosci.* 24, 727–736
92. Cowley, B.R. *et al.* (2020) Slow drift of neural activity as a signature of impulsivity in macaque visual and prefrontal cortex. *Neuron* 108, 551–567
93. Mochol, G. *et al.* (2021) Prefrontal cortex represents heuristics that shape choice bias and its integration into future behavior. *Curr. Biol.* 31, 1234–1244
94. Fascianelli, V. *et al.* (2024) Neural representational geometries reflect behavioral differences in monkeys and recurrent neural networks. *Nat. Commun.* 15, 6479
95. Pagan, M. *et al.* (2025) Individual variability of neural computations underlying flexible decisions. *Nature* 639, 421–429
96. Nakai, T. *et al.* (2025) Cortical representational geometry of diverse tasks reveals subject-specific and subject-invariant cognitive structures. *Commun. Biol.* 8, 713
97. Latimer, K.W. and Freedman, D.J. (2023) Low-dimensional encoding of decisions in parietal cortex reflects long-term training history. *Nat. Commun.* 14, 1010
98. Holton, E. *et al.* (2025) Humans and neural networks show similar patterns of transfer and interference during continual learning. *Nat. Hum. Behav.* 10, 111–125
99. Soldado-Magraner, J. *et al.* (2024) Inferring context-dependent computations through linear approximations of prefrontal cortex dynamics. *Sci. Adv.* 10, eadl4743
100. Huang, A. *et al.* (2025) Measuring and controlling solution degeneracy across task-trained recurrent neural networks. *arXiv* https://doi.org/10.48550/arXiv.2410.03972
101. Menghi, N. *et al.* (2025) The effects of task similarity during representation learning in brains and neural networks. *Nat. Commun.* 16, 10812
102. Kikumoto, A. *et al.* (2025) Practice reshapes the geometry and dynamics of task-tailored representations. *Cereb. Cortex* 35, bhaf125
103. Mill, R.D. and Cole, M.W. (2025) Dynamically shifting from compositional to conjunctive brain representations supports cognitive task learning. *Nat. Commun.* 16, 10084
104. Gulli, R.A. *et al.* (2020) Context-dependent representations of objects and space in the primate hippocampus during virtual navigation. *Nat. Neurosci.* 23, 103–112
105. Li, S. *et al.* (2025) Flexible use of limited resources for sequence working memory in macaque prefrontal cortex. *Nat. Commun.* 16, 10386
106. Panichello, M.F. and Buschman, T.J. (2021) Shared mechanisms underlie the control of working memory and attention. *Nature* 592, 601–605
107. Molano-Mazón, M. *et al.* (2023) Recurrent networks endowed with structural priors explain suboptimal animal behavior. *Curr. Biol.* 33, 622–638
108. Ni, A.M. *et al.* (2022) A general decoding strategy explains the relationship between behavior and correlated variability. *Elife* 11, e67258
109. De Lange, F.P. *et al.* (2018) How do expectations shape perception? *Trends Cogn. Sci.* 22, 764–779
110. Yang, Q. *et al.* (2025) A language model of problem solving in humans and macaque monkeys. *Curr. Biol.* 35, 11–20
111. Snyder, H.R. *et al.* (2015) Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Front. Psychol.* 6, 328
112. Seriès, P. (2020) *Computational Psychiatry: A Primer*, MIT Press
113. Rust, N.C. (2025) *Elusive Cures: Why Neuroscience Hasn't Solved Brain Disorders—and How We Can Change That*, Princeton University Press
114. Wang, R. and Chen, Z.S. (2025) Large-scale foundation models and generative AI for BigData neuroscience. *Neurosci. Res.* 215, 3–14
115. Wang, E.Y. *et al.* (2025) Foundation model of neural activity predicts response to new stimulus types. *Nature* 640, 470–477
116. Cabrera-León, Y. *et al.* (2024) Neural computation-based methods for the early diagnosis and prognosis of Alzheimer's disease not using neuroimaging biomarkers: a systematic review. *J. Alzheimer's Dis* 98, 793–823
117. Aljuhani, M. *et al.* (2024) Use of artificial intelligence in imaging dementia. *Cells* 13, 1965
118. Monosov, I.E. *et al.* (2024) Ethological computational psychiatry: challenges and opportunities. *Curr. Opin. Neurobiol.* 86, 102881
119. Waskom, M.L. *et al.* (2019) Designing and interpreting psychophysical investigations of cognition. *Neuron* 104, 100–112
120. Fetsch, C.R. (2016) The importance of task design and behavioral control for understanding the neural basis of cognitive functions. *Curr. Opin. Neurobiol.* 37, 16–22
121. Galgali, A.R. *et al.* (2023) Residual dynamics resolves recurrent contributions to neural computation. *Nat. Neurosci.* 26, 326–338
122. Ji-An, L. *et al.* (2025) Discovering cognitive strategies with tiny recurrent neural networks. *Nature* 644, 993–1001
123. Pandarinath, C. *et al.* (2018) Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 15, 805–815

124. Langdon, C. and Engel, T.A. (2025) Latent circuit inference from heterogeneous neural responses during cognitive tasks. *Nat. Neurosci.* 28, 665–675
125. Kim, T.D. *et al.* (2025) Flow-field inference from neural data using deep recurrent networks. In *Proceedings of the 42nd International Conference on Machine Learning* (Vol. 267), pp. 30567–30590, PMLR
126. Schneider, S. *et al.* (2023) Learnable latent embeddings for joint behavioural and neural analysis. *Nature* 617, 360–368
127. Glaser, J. *et al.* (2020) Recurrent switching dynamical systems models for multiple interacting neural populations. *Adv. Neural Inf. Proces. Syst.* 33, 14867–14878
128. Mathis, M.W. and Mathis, A. (2026) Joint modelling of brain and behaviour dynamics with artificial intelligence. *Nat. Rev. Neurosci.* 27, 87–100
129. Binz, M. *et al.* (2025) A foundation model to predict and capture human cognition. *Nature* 644, 1002–1009
130. Mastrogiuseppe, F. and Ostojic, S. (2018) Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* 99, 609–623
131. Valente, A. *et al.* (2022) Extracting computational mechanisms from neural data using low-rank RNNs. *Adv. Neural Inf. Proces. Syst.* 35, 24072–24086
132. Zhang, Y. *et al.* (2025) Elucidating the selection mechanisms in context-dependent computation through low-rank neural network modeling. *Elife* 13, RP103636
133. Bowers, J.S. *et al.* (2023) Deep problems with neural network models of human vision. *Behav. Brain Sci.* 46, e385
134. Serre, T. and Pavlick, E. (2025) From prediction to understanding: will AI foundation models transform brain science? *Neuron* 113, 3504–3508