

How to study the neural mechanisms of multiple tasks

Guangyu Robert Yang, Michael W Cole and Kanaka Rajan

Most biological and artificial neural systems are capable of completing multiple tasks. However, the neural mechanism by which multiple tasks are accomplished within the same system is largely unclear. We start by discussing how different tasks can be related, and methods to generate large sets of inter-related tasks to study how neural networks and animals perform multiple tasks. We then argue that there are mechanisms that emphasize either specialization or flexibility. We will review two such neural mechanisms underlying multiple tasks at the neuronal level (modularity and mixed selectivity), and discuss how different mechanisms can emerge depending on training methods in neural networks.

Address

Zuckerman Mind Brain Behavior Institute, Columbia University, Center for Molecular and Behavioral Neuroscience, Rutgers University-Newark, Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, United States

Corresponding author: Rajan, Kanaka (kanaka.rajan@mssm.edu)

Current Opinion in Behavioral Sciences 2019, 29:134–143

This review comes from a themed issue on **Artificial Intelligence**

Edited by **Matt Botvinick** and **Sam Gershman**

<https://doi.org/10.1016/j.cobeha.2019.07.001>

2352-1546/© 2019 Published by Elsevier Ltd.

Why should we study multiple tasks?

The study of systems and cognitive neuroscience rely heavily on investigating neural systems as they perform various tasks. A task generally refers to the set of computations that a system needs to perform to optimize an objective, such as reward or classification accuracy. A classical cognitive task is the random-dot motion task [8], where an agent/subject needs to decide the moving direction of a group of coherently moving dots, amid a group of randomly moving dots. In neuroscience and cognitive science, each task is typically designed to shed light on the neural mechanism of a particular function. For example, the random-dot motion task is carefully designed so the moving direction cannot be inferred by the stimulus at any individual time point so this task can be used to study how agents integrate information over time. A substantial body of experimental and

computational work has been devoted to understand the neural mechanisms behind individual tasks.

Although neural systems are usually studied with one task at a time, these systems are usually capable of performing many different tasks, and there are many reasons to study how a neural system can accomplish this. Studying multiple tasks can serve as a powerful constraint to both biological and artificial neural systems (Figure 1a). For one given task, there are often several alternative models that describe existing experimental results similarly well. The space of potential solutions can be reduced by the requirement of solving multiple tasks.

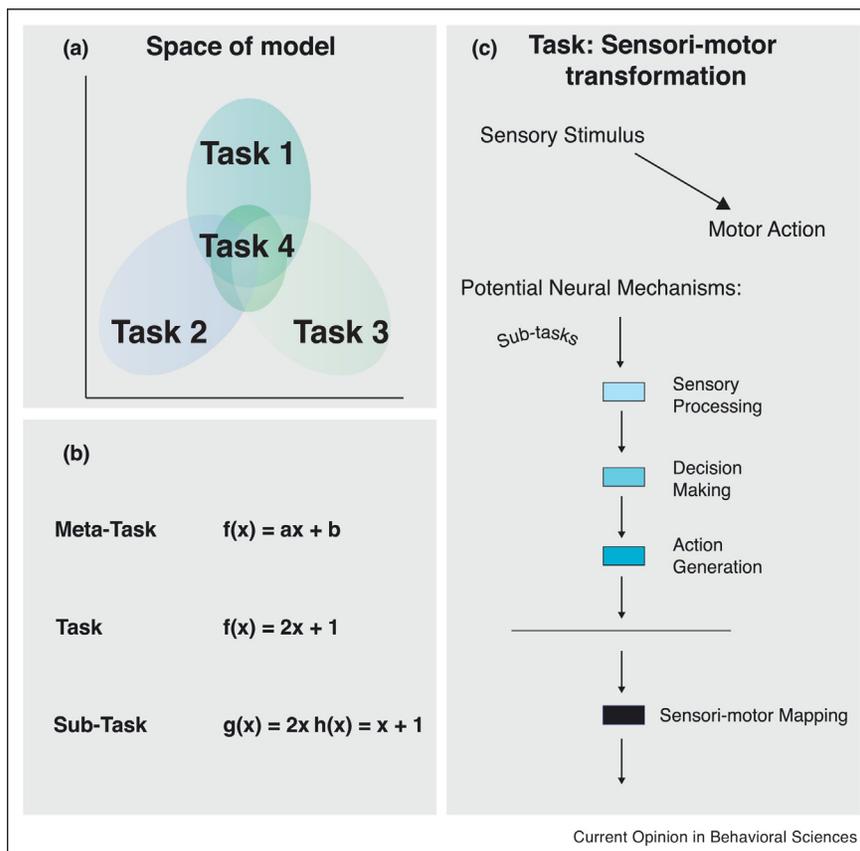
Experiments can uncover neural representation or mechanisms that appear sub-optimal for a single task. For instance, neural activity in prefrontal cortex during working memory tasks is often highly dynamical [43], even though such time-varying representations are not necessary for these tasks. In another example, selectivity of parietal cortex neurons can shift across days even when mice continue to perform the same task equally well [19]. These seemingly unnecessary features could potentially be better understood in the context of having a single system that needs to solve many varied tasks [44,19].

Studying multiple tasks also raises important questions that are not as salient when studying a single task. One such question is the issue of continual learning. Humans and animals can learn new tasks without rapidly forgetting all previous tasks learned. In contrast, traditional neural network models experience “catastrophic forgetting”, where learning of a new task can strongly interfere with performance of previously learned tasks. It remains to be understood how biological brains combat this issue of catastrophic forgetting.

When multiple tasks are learned sequentially, the learning of previous tasks can potentially lead to emergence of network architecture, neural representation, and learning rules that greatly facilitate learning of future tasks. Mechanisms and strategies for making this happen is the subject of transfer learning [45] and meta-learning (learning-to-learn). The topic of curriculum learning is concerned with finding a good set of tasks to pre-train on before training a difficult task, which can aid learning and transfer of task features.

Finally, studying a network capable of performing multiple tasks raises questions about how the neural representation of different tasks are related to one another. Similar to how a substantial amount of neuroscience work is

Figure 1



(a) In the space of models, every task can be solved by many different models, indicated by the colored areas. Solving multiple tasks provides a stronger constraint on the space of allowed models. (b) Any specific task usually comprises several subtasks, and can also be described as an instance of a meta-task. (c) The organization of tasks is agent-dependent. Which subtasks to break a task into depend on the presumed neural mechanism. For a sensori-motor transformation task, if the computation is carried out by multiple stages of network processing, it is more sensible to break the task into multiple subtasks. However, if the task is performed by a neural network with a single hidden layer, the previous subtasks may no longer be meaningful.

devoted to understanding how different stimuli and actions are represented in the same circuit, we can ask how tasks are represented in the same circuit. Are different tasks supported by overlapping populations of neurons? If so, how strong is the overlap, and what is the overlapping part responsible for? Understanding this question can potentially help us better understand continual learning and learning-to-learn, given that catastrophic forgetting presumably happens because the representation of tasks interfere, while transfer learning happens when representation of tasks can be reused.

Organization of tasks

Imagine we are studying how different tasks are represented in the brain. A visual task (for example object recognition) and a motor task (for example, an arm reaching task) will utilize largely non-overlapping populations of neurons. On the other hand, two visual tasks, for

example, object recognition and reading, will utilize largely overlapping populations of neurons. Why are the neural resources separated in one case, and shared in another case? Intuition tells us that the two visual tasks are closer to each other, therefore can reuse similar circuits, while the visual and motor tasks are farther apart. How do we make these intuitive concepts of task similarity more formal?

To answer this question, we argue that it is critical to develop a vocabulary so we can more rigorously discuss how tasks are related to each other. Understanding the relationship between tasks will then help us understand why some tasks interfere with other tasks, and why learning of one task can improve learning of another task. Here we describe two fundamental relationships that tasks can have with one another. Later we review how large sets of tasks can be constructed based on these

relationships. Tasks can be directly related to one another through at least two types of relationships: a part-whole relationship, and a specific-general relationship.

Part-whole relationship Each individual task (the whole) can comprise multiple subtasks (parts). To perform the whole task, it is necessary to perform all the subtasks. A task is a supertask of its subtasks. For example, inferring momentary moving direction is a subtask of the random-dot motion task, which requires integrating momentary evidence across time. The task of computing $f(x) = 2x + 1$ (Figure 1b) can be written as a combination of two subtasks $g(x) = 2x$ and $h(x) = x + 1$ such that $f(x) = h(g(x))$. A subtask is itself a task, and can typically be further divided into subtasks, forming a hierarchical tree of tasks [3].

Specific-general relationship Meanwhile, a more general task can be instantiated as a more specific task. We call the more general task a “meta-task”, and the more specific task, a task instance. Here we use “meta” as meaning *beyond* and *higher-order*, instead of *self-referential*. The task $f(x) = 2x + 1$ can be treated as a special case of the more general task $F(x) = ax + b$, with $a = 2$ and $b = 1$.

Relationships between tasks are agent-dependent

The above definitions of subtask and meta-task imply, for example, that processing a pixel is a subtask of processing an image, while recognizing a single image is a task instance of the meta-task of recognizing images. Following our previous example, $f(x) = 2x + 1$ can be divided into subtasks $f(x) = h(g(x))$, where $g(x) = \sqrt{x}$ and $h(x) = 2x^2 + 1$. It can also be viewed as an instance of the meta-task $F(x) = ax + a^2 + b$, where $a = 2$ and $b = -3$.

This conceptualization of the relationship between tasks is useful for describing different ways tasks can be represented by agents. In practice, it is useful to describe two computational processes as separate tasks if the neural mechanisms are different between those processes. In contrast, it is useful to describe multiple computational processes as part of the same task if those processes are carried out using an overlapping set of neural representations. If two computational processes are presumably supported by the same mechanism (for example, classifying two images from the same dataset), then conceptually there is no need to separate them into two tasks, instead they can be considered two conditions of the same task.

Agents (animals/networks) can have different underlying neural mechanisms for the same task. So there must be an agent-dependent view on how to decompose a task into subtasks, and whether one task is a meta-task of another task (Figure 1c). The task “driving” can be intuitively decomposed into subtasks such as object recognition, planning, and motor control. Yet if a computationally-poor agent

drives by mapping pixel inputs directly into motor outputs through a feedforward network with a single hidden layer, the recognition/planning/control decomposition would no longer be meaningful. Whether a task should be viewed as coming from a particular meta-task is similarly influenced by the neural mechanism.

Constructing large sets of tasks

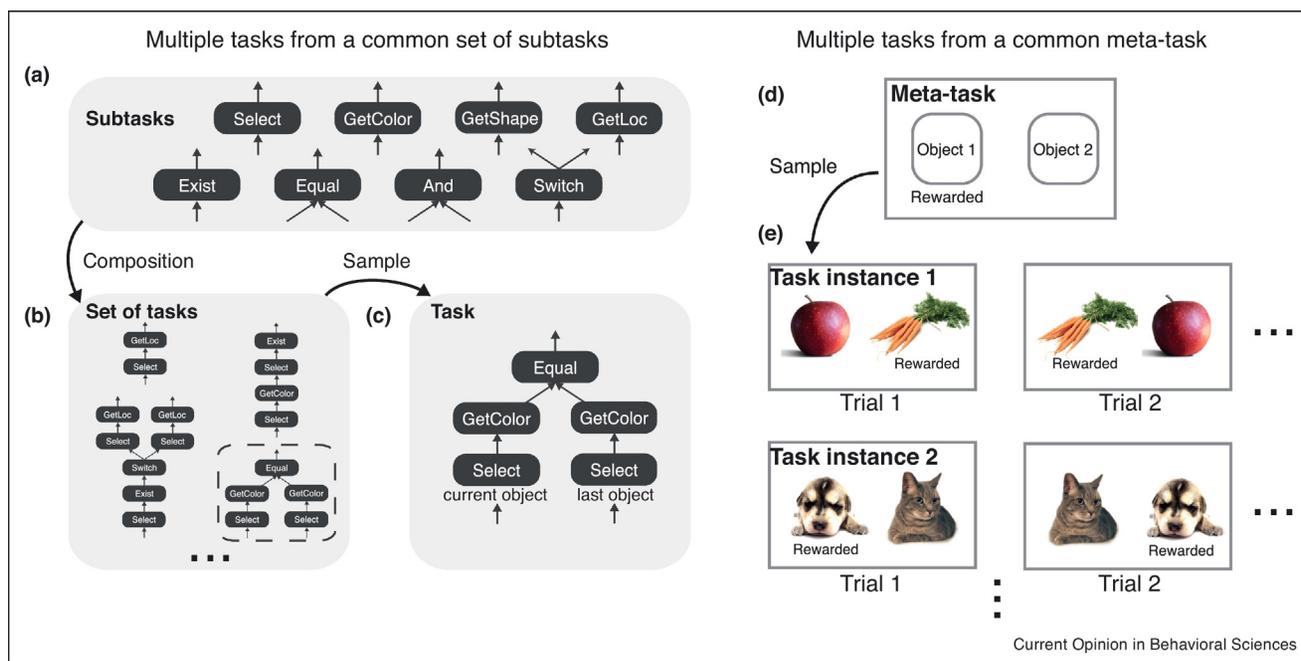
Neuroscience and cognitive science have benefited tremendously from carefully designed single tasks. To study the neural mechanism of one particular function, a neuroscience task is typically constructed such that other functions cannot be used to solve it. How can we extend this design principle to the study of multiple tasks? We review two methods to build large, controlled sets of tasks, one starts with a common set of subtasks, another starts with a common meta-task.

Tasks with common subtasks In visual perception, various tasks like object classification, localization, and segmentation [38] share many similar computations. These shared computations or subtasks are not necessarily named, but are typically embodied by common feature extraction layers in both animals (e.g., retina, V1) and neural network models. In motor control and robotics, many tasks involve the same lower-level subtasks like walking and grasping [58,9]. Many cognitive tasks involve working memory and decision making [67**].

Besides choosing tasks that already share common computations, we can construct many interrelated tasks starting from a small set of subtasks as building blocks [32**,66*,35,63] (Figure 2a). A task can be characterized as a graph of subtasks (Figure 2b). For example, combining the subtasks “select an object from an image”, “get the color of an object”, and “compare the value of two attributes”, we can construct a variant of the delayed-match-to-category task [24]: “Compare the color of the current object with that of the last object” (Figure 2c). The task graph describes the order in which the subtasks are composed together [32**,66*]. This approach allows for the compositional generation of a large number of tasks using a small set of subtasks. Cole and colleagues have studied how humans can perform many tasks using a dataset of 64 tasks that are generated compositionally from 4 sensory, 4 motor, and 4 logic subtasks [14,30].

Many questions we would like to ask with multiple tasks can benefit from having a collection of tasks with common subtasks. The total number of distinct tasks can grow exponentially with the number of subtasks, therefore providing strong constraint for training. It provides a way to test whether networks can transfer knowledge better when the new task has more common subtasks with previously learned tasks. It also allows us to ask whether a neural network can quickly identify the common subtasks from learning a group of tasks.

Figure 2



(a-c) Generating multiple tasks from a shared set of subtasks. A small subset of subtasks (a) can be used to generate a large number of tasks through composition (b), from which a single task can be sampled (c). Adapted from [66]. (d,e) Generating multiple tasks from the same meta-task. Starting with a common meta-task (d), many tasks can be instantiated (e).

Tasks with a common meta-task A classic meta-task example is the Harlow task [29]. In this meta-task, an animal/agent learns to choose between two objects (Figure 2d), one rewarding, and the other not. For each instance of this meta-task, a new set of two objects is used. Within a task, the objects are shown at different spatial locations in each trial. Critically, each task instance only lasts for a few trials (Figure 2e), so the animal/agent needs to learn efficiently within a task to maximize reward. Here each concrete task requires rapid learning, so the meta-task can be described as learning-to-learn [61]. Similarly, learning to categorize using a small number of examples can be considered a meta-task, where each task instance would involve several examples from new categories [60,36].

Many other tasks can be conceptualized as instances of corresponding meta-tasks. The task of navigating a particular maze can be an instance of the more general maze navigation task [57]. A 2-arm bandit task with a particular reward probability for each arm is a special case of the general 2-arm bandit task [61], which itself is an instance of the n-arm bandit task. Starting from a generic enough meta-task, we can generate many, even infinite, interrelated tasks.

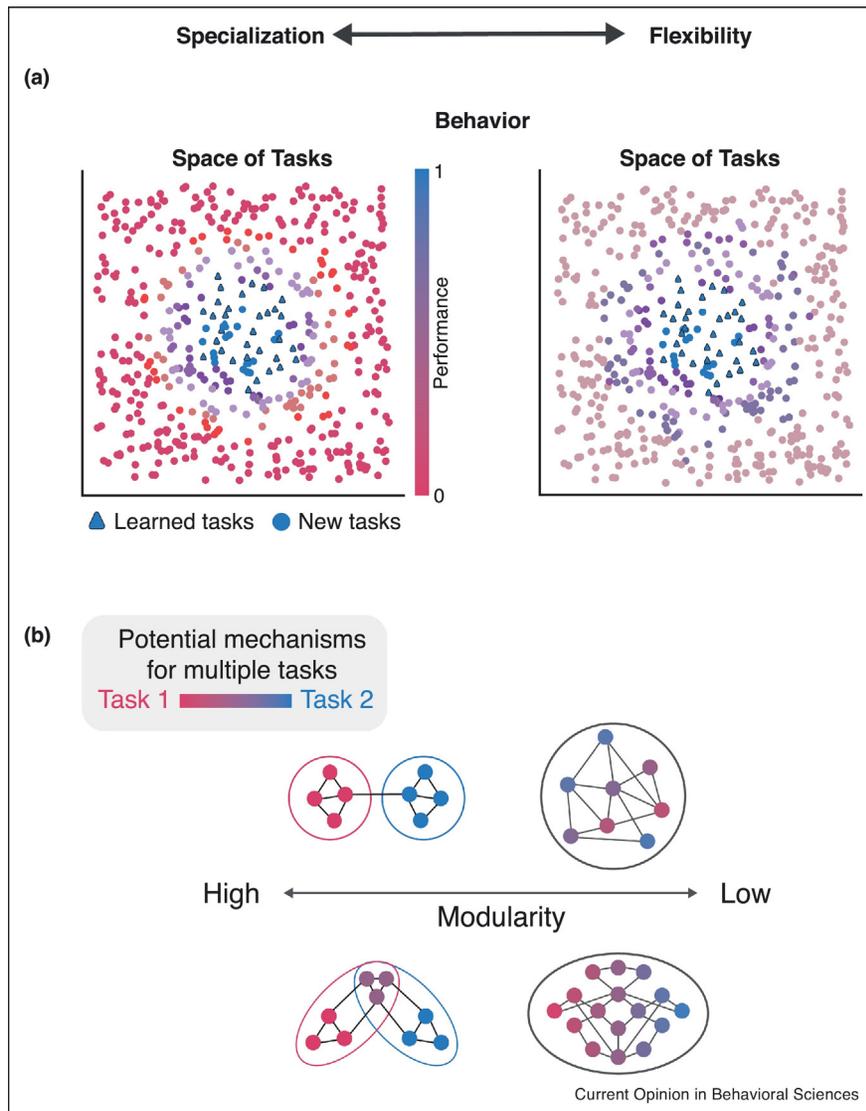
The benefit of constructing a large set of tasks with a shared meta-task is that the difficulty of individual task

can be kept the same. This can, for example, allow us to probe whether a model is getting faster at learning new tasks. In addition, studying tasks from a common meta-task allows us to investigate whether networks have acquired the abstract structure of the meta-task, and how the task structure is represented.

The specialization-flexibility continuum

It is clear that there is no single neural mechanism for multiple tasks. Instead, there are many potential neural mechanisms, depending on the collection of tasks, the method that a biological or artificial system uses to acquire the tasks, and (in the case of biological systems) the brain areas studied. Overall, little is known about how any of these aspects influence the neural mechanism. Here we propose that even though there are practically infinite possible ways to choose a set of tasks and to train networks, the resulting neural mechanisms usually live along a specialization-flexibility continuum (Figure 3a) [28]. Solutions occupying different places on this continuum would lead to different neural mechanisms, and demand different types of training paradigms to reach. At the extremes of the specialization-flexibility continuum are two distinct types of solutions for a set of acquired/trained tasks, the specialized and the flexible solution (Figure 3a). In the case of an animal/agent that has learned to perform multiple tasks, the two types of

Figure 3



(a) (left) For a set of learned tasks (triangle), the specialized solution leads to high performance for learned tasks, but low performance when learning tasks far from the learned tasks. (right) The flexible solution improves expected performance over a wide range of tasks at the expense of a lower performance for the learned tasks. (b) Schematic showing potential neuronal-level mechanisms for multiple tasks, left: modularity, right: mixed-selectivity. Color indicates the level each unit is engaged in tasks 1 and 2. Red: unit only engaged in task 1, blue: unit only engaged in task 2, purple: unit engaged in both tasks. Adapted from [67**].

solutions will differ in the degree they specialize to the set of learned tasks.

Consider an agent that has already learned a set of tasks $S=(A, B, C, D, \dots)$, and is about to learn a new task X . A specialized solution is characterized by the agent's high performance or efficiency for the set of learned tasks S , but relative difficulty in learning the new task X , if X is *dissimilar to tasks in S* . Here task X is similar to the set of tasks S if it shares many subtasks or a meta-task with the tasks in S (such that shared neural representations/processes are used). Because the organization of tasks is agent-dependent as argued before,

the distance between tasks would have to be as well. In comparison, a flexible solution to the set of tasks S may not achieve as high performance as the specialized solution, but it would allow for better learning when X is dissimilar to tasks in S . This difference between specialized and flexible solutions are illustrated in Figure 3a. We emphasize that when a new task X is similar to S , both the specialized and flexible solutions can learn it rapidly, or even perform it without learning (i.e., without connectivity weight changes).

This continuum from specialization to flexibility is conceptually connected to several other contrasting concepts.

Perhaps most relevant to this distinction, decades of cognitive psychology (and cognitive neuroscience) research has established that controlled versus automatic processing is a fundamental distinction in human cognition [51,13]. Controlled processing is characterized as being especially flexible but capacity limited and inefficient. In contrast, automatic processing is characterized as being highly inflexible but high capacity and efficient. Controlled processing occurs when a task is novel (decreasing with practice) and when there is conflict between representational mappings (e.g., from stimulus to response) that needs to be resolved. Automatic processing occurs in all other cases, consisting of consistent mappings between representations from extensive practice (e.g., walking, driving a familiar route, etc.). Given that these modes of processing map directly onto the specialized-flexible continuum, it appears that the human brain deals with this computational trade-off by switching from one mode of processing to the other as necessary – a form of meta-optimization [15,7]. Specifically, it appears that the human brain uses flexible cognitive control brain systems early in learning [16,17,12] and in the face of conflict [6,37], switching to specialized systems (when possible due to low conflict) to implement automatic processing after extensive practice [12,50]. It will be important for future work to explore the relationship between this computational trade-off generally (e.g., in computational models) and the particular manner in which human (and other animal) brains deal with this trade-off.

From modularity to mixed-selectivity

Here we describe two neural mechanisms that may correspond to specialized and flexible solutions, respectively. In particular, we will mainly focus on mechanisms at the neuronal level, namely how neurons are engaged in each task, and how the group of neurons engaged in one task is related to the group of neurons engaged in another task.

The first neural mechanism is modularity (Figure 3b). A neural circuit capable of performing multiple tasks can potentially consist of multiple groups of neurons, or modules. A particular subset of modules will be engaged when the network is performing each task. The second neural mechanism is mixed-selectivity (Figure 3b). In a neural circuit exhibiting mixed selectivity [25], neurons do not belong to fixed functional modules, unlike the modular mechanism. Mixed selective neurons are characterized as being nonlinearly selective to many task variables (e.g., sensory stimulus, action). Furthermore, the selectivity is task-dependent. Collectively, these neurons form representations that are high-dimensional, supporting readout of many combinations of task variables [47].

We argue that in the brain, modularity is typically the result of specialization. Highly evolved and stereotypical

computations are usually supported by modular neural circuits. Neuronal-level modularity is evident in highly-specialized early sensory processing areas. Mammalian retina consists of more than 60 cell types [41], with at least 30 functionally distinct types of output cells [2]. Mouse olfactory system consists of more than 1000 types of olfactory receptor neurons [10]. Modularity is also apparent at the neural system level. Mammalian cortex is made up of about a dozen modular brain systems [31] consisting of many areas (almost 400 in humans [26]), some of which are highly specialized such as areas dedicated to face processing in primates [59].

We have previously described that even highly specialized networks can appear flexible and rapidly learn many new tasks as long as the new tasks are close to the learned tasks S . Here we explain how this could be achieved in a modular circuit. Consider a set of tasks generated from a common set of subtasks. A highly specialized network can support each subtask with a module (a group of neurons). The entire task can be performed by activating the corresponding modules. Such a network can be flexible in the sense that it can generalize to new tasks that use the same subtasks. But it may have difficulty learning new tasks that involve new subtasks, as that would require breaking existing modules. Further, there would likely be difficulty learning and coordinating the correct combination of modules, given that more than one combination is possible among three or more modules.

While specialization can drive modularity, flexible solutions demand mixed-selectivity. Neurons are usually mixed-selective in higher-order brain areas critical for flexible behavior, such as prefrontal cortex and hippocampus. In the prefrontal cortex (which is part of the frontoparietal system), for example, many tasks engage a significant proportion of neurons [25]. In the hippocampus, spatial and non-spatial information are nonlinearly mixed [1,27].

From a state-space perspective, a population of neurons with mixed selectivity can support high-dimensional representation of sensory information. A higher-dimensional representation can lead to read out of more combinations of inputs, supporting faster learning of new tasks [56]. In contrast, specialized solutions should favor lower-dimensional representations where the network only represents the combinations of sensory inputs useful to the learned tasks.

Even though we described modularity and mixed-selectivity as two opposing mechanisms, they can clearly co-exist. Both mechanisms are observed in the brain, after all. A brain area that is mixed-selective can be itself one module of the larger modular neural system. Further, there is evidence that the frontoparietal system (which consists of neurons with high mixed selectivity)

coordinates specialized modules to facilitate transfer of previous learning to novel task contexts [16,17,30].

How to train neural networks to be specialized or flexible

In neuroscience, it is increasingly common to compare biological neural circuits with artificial ones [65,64,40,53,54]. With the exponential growth of the deep learning field, there are many varieties of training methods available for artificial neural networks. Here we discuss how training methods used for artificial neural networks can influence whether the solution developed is more specialized or flexible.

Overall, we predict that conventional training methods that rely on a large amount of training data will likely lead to specialized solutions. These methods are the standard ones in machine learning and many neuroscience applications. The best models for matching neural activity in higher-order visual areas are deep convolutional networks [64] trained on the ImageNet dataset [18], containing more than 1 million images. The ImageNet task of object classification is a general-enough task that many related visual tasks benefit from using backbone networks trained on ImageNet [39,22]. These results again demonstrate that specialized solutions can allow rapid learning on tasks close to learned tasks.

In the previous section, we showed that modularity is commonly observed in specialized brain areas. In artificial neural networks the causal link from specialization to modularity can be studied more readily than in biological neural systems. A recurrent neural network trained to perform 20 interrelated cognitive tasks developed specialized neural populations, each serving a common computation behind multiple tasks [67**]. In this work, the emergence of functionally specialized modules is not a result of regularization that sparsifies activity or connectivity, instead, it appears simply under the pressure to perform all tasks well.

A notable case of specialized solutions is when each task has a dedicated output network following an input network shared across all tasks [11,49]. The system is modular in the sense that each output network is only engaged in a single task. The optimal size of each output module relative to the size of the shared input network depends on how similar the tasks are [33*,68]. The advantage of such systems is that multiple tasks can be performed simultaneously. However, learning a new task involves training a separate output network, which can be difficult when learning a large number of tasks.

When a network is trained on a large number of task instances from one meta-task, it can develop a specialized solution that still generalizes to new task instances from the same meta-task. Following an example mentioned

above, a network can be trained to perform many specific 2-arm bandit task instances, each with particular reward probabilities and a limited amount of training data available. Once a network masters the 2-arm bandit meta-task, it can quickly learn to perform new task instances. This method of training a network to learn a meta-task so it can quickly learn task instances has been subject to a large body of machine learning work under the topic of learning-to-learn or meta-learning. A neural network can be trained using meta-learning methods to flexibly categorize [60,52,23], adapt to new reinforcement-learning tasks [61**,62,21,5], and imitate behaviors [20]. Networks trained this way can develop powerful specialized solutions to the meta-task. Little is known about the neuronal-level neural mechanism in networks trained this way. It would be interesting to know whether these networks also develop modular solutions.

How to train a network to stay flexible to new tasks that do not share subtasks or a meta-task with previously learned tasks? We suggest that a network can stay flexible for many tasks by explicitly preventing specialization to the tasks currently being learned. This can be achieved in artificial neural networks through various continual learning methods [34,69,4] that discourage large changes to existing connection weights during training. A related strategy is to train only a small subset of connections for any given task [55,46,42]. These methods are originally proposed to prevent neural networks from forgetting previously learned tasks when learning new tasks, however, we argue that these methods can also help neural networks learn new tasks far from the set of learned tasks. The neural network can be initialized with random connectivity, which will endow it with mixed selectivity [48]. When learning a new task, mixed selectivity can be preserved as long as learning does not strongly alter the random connectivity [67**,55].

Concluding remarks

Having the same system solve multiple tasks can provide strong constraints on both conceptual and computational models of the brain. It prevents us from building theories and models that overfit the tasks being studied at hand. Having a multi-task system further opens up many new questions, especially when systematically-generated task collections are used. Training non-human animals to perform multiple tasks can be relatively difficult. The use of artificial neural networks in neuroscience and cognitive science can alleviate this problem by offering a complimentary model system where multiple tasks are more easily trained. However, depending on the particular training methods, profoundly different solutions can arise. Thus, modelers should choose training techniques based on the type of solutions (specialized or flexible) they intend to build.

We have discussed two neuronal-level mechanisms for multiple tasks, modularity and mixed-selectivity. Of course, much remains to be learned about each mechanism. Another line of intriguing questions is to better connect mechanisms at the neuronal-, state-space-, and behavioral-level. For example, what happens at the neuronal level when an agent or animal had a eureka moment (at the behavioral level) that several tasks all belong to the same meta-task or share a common subtask? Addressing these questions requires neural circuit/network models that are versatile enough to perform multiple tasks, yet simple enough to facilitate analysis and understanding.

Acknowledgements

The authors acknowledge support by Simons Society of Fellows Junior Fellowship, National Science Foundation Grant Number 1707398, and the Gatsby Charitable Foundation to G.R.Y., the US National Institutes of Health under awards R01 AG055556 and R01 MH109520 to MWC.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Aronov Dmitriy, Nevers Rhino, Tank David W: **Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit.** *Nature* 2017, **543**:719.
 2. Baden Tom, Berens Philipp, Franke Katrin, Rosón Miroslav Román, Bethge Matthias, Euler Thomas: **The functional diversity of retinal ganglion cells in the mouse.** *Nature* 2016, **529**:345.
 3. Badre David, Nee Derek Evan: **Frontal cortex and the hierarchical control of behavior.** *Trends Cognit Sci* 2018, **22**:170-188.
 4. Benna Marcus K, Fusi Stefano: **Computational principles of synaptic memory consolidation.** *Nature Neurosci* 2016, **19**:1697.
 5. Botvinick Mathew, Ritter Sam, Wang Jane X, Kurth-Nelson Zeb, Blundell Charles, Hassabis Demis: **Reinforcement learning, fast and slow.** *Trends Cognit Sci* 2019.
 6. Botvinick Matthew M, Braver Todd S, Barch Deanna M, Carter Cameron S, Cohen Jonathan D: **Conflict monitoring and cognitive control.** *Psychol Rev* 2001, **108**:624.
 7. Boureau Y-Lan, Sokol-Hessner Peter, Daw Nathaniel D: **Deciding how to decide: Self-control and meta-decision making.** *Trends Cognit Sci* 2015, **19**:700-710.
 8. Britten Kenneth H, Shadlen Michael N, William T, Newsome J, Movshon Anthony: **The analysis of visual motion: a comparison of neuronal and psychophysical performance.** *J Neurosci* 1992, **12**:4745-4765.
 9. Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba Openai gym. arXiv preprint arXiv:1606.01540, 2016.
 10. Buck Linda, Axel Richard: **A novel multigene family may encode odorant receptors: a molecular basis for odor recognition.** *Cell* 1991, **65**:175-187.
 11. Caruana Rich: **Multitask learning.** *Mach Learn* 1997, **28**:41-75.
 12. Chein Jason M, Schneider Walter: **Neuroimaging studies of practice-related change: fmri and meta-analytic evidence of a domain-general control network for learning.** *Cognit Brain Res* 2005, **25**:607-623.
 13. Chein Jason M, Schneider Walter: **The brain's learning and control architecture.** *Curr Directions Psychol Sci* 2012, **21**:78-84.
 14. Cole Michael W, Bagic Anto, Kass Robert, Schneider Walter: **Prefrontal dynamics underlying rapid instructed task learning reverse with practice.** *J Neurosci* 2010, **30**:14245-14254.
 15. Cole Michael W, Braver Todd S, Meiran Nachshon: **The task novelty paradox: Flexible control of inflexible neural pathways during rapid instructed task learning.** *Neurosci Biobehav Rev* 2017, **81**:4-15.
 16. Cole Michael W, Laurent Patryk, Stocco Andrea: **Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control.** *Cognit Affect Behav Neurosci* 2013, **13**:1-22.
 17. Cole Michael W, Reynolds Jeremy R, Power Jonathan D, Repovs Grega, Anticevic Alan, Braver Todd S: **Multi-task connectivity reveals flexible hubs for adaptive task control.** *Nature Neurosci* 2013, **16**:1348.
 18. Deng Jia, Dong Wei, Socher Richard, Li Li-Jia, Li Kai, Fei-Fei Li: **Imagenet: A large-scale hierarchical image database.** In *2009 IEEE conference on computer vision and pattern recognition* 2009:248-255. Ieee.
 19. Driscoll Laura N, Pettit Noah L, Minderer Matthias, Chettih Selmaan N, Harvey Christopher D: **Dynamic reorganization of neuronal activity patterns in parietal cortex.** *Cell* 2017, **170**:986-999.
 20. Duan Yan, Andrychowicz Marcin, Bradly Stadie OpenAI, Ho Jonathan, Schneider Jonas, Sutskever Ilya, Abbeel Pieter, Zaremba Wojciech: **One-shot imitation learning.** *Adv Neural Inform Process Syst* 2017:1087-1098.
 21. Duan Yan, Schulman John, Chen Xi, Bartlett Peter L, Sutskever Ilya, Abbeel Pieter: **RL 2: Fast reinforcement learning via slow reinforcement learning.** arXiv preprint arXiv:1611.02779. 2016.
 22. Esteva Andre, Kuprel Brett, Novoa Roberto A, Ko Justin, Swetter Susan M, Blau Helen M, Thrun Sebastian: **Dermatologist-level classification of skin cancer with deep neural networks.** *Nature* 2017, **542**:115.
 23. Finn Chelsea, Abbeel Pieter, Levine Sergey: **Model-agnostic meta-learning for fast adaptation of deep networks.** In *Proceedings of the 34th International Conference on Machine Learning-Volume* 2017, **70**:1126-1135 JMLR. org.
 24. Freedman David J, Assad John A: **Experience-dependent representation of visual categories in parietal cortex.** *Nature* 2006, **443**:85.
 25. Fusi Stefano, Miller Earl K, Rigotti Mattia: **Why neurons mix: high dimensionality for higher cognition.** *Curr Opin Neurobiol* 2016, **37**:66-74.
 26. Glasser Matthew F, Coalson Timothy S, Robinson Emma C, Hacker Carl D, Harwell John, Yacoub Essa, Ugurbil Kamil, Andersson Jesper, Beckmann Christian F, Jenkinson Mark *et al.*: **A multi-modal parcellation of human cerebral cortex.** *Nature* 2016, **536**:171.
 27. Gulli Roberto Adamo, Duong Lyndon, Corrigan Benjamin Whelehan, Doucet Guillaume, Williams Sylvain, Fusi Stefano, Martinez-Trujillo Julio Cesar: **Flexible coding of memory and space in the primate hippocampus during virtual navigation.** *bioRxiv* 2018:295774.
 28. Hardcastle Kiah, Ganguli Surya, Giocomo Lisa M: **Cell types for our sense of location: where we are and where we are going.** *Nature Neurosci* 2017, **20**:1474.
- This review discusses the distinction between specialist and generalist circuits, two concepts closely related to specialization and flexibility discussed here.
29. Harlow Harry F: **The formation of learning sets.** *Psychol Rev* 1949, **56**:51.
 30. Ito Takuya, Kulkarni Kaustubh R, Schultz Douglas H, Mill Ravi D, Chen Richard H, Solomyak Levi I, Cole Michael W: **Cognitive task information is transferred between brain regions via resting-state network topology.** *Nature Commun* 2017, **8**:1027.
 31. Ji Jie Lisa, Spronk Marjolein, Kulkarni Kaustubh, Repovš Grega, Anticevic Alan, Cole Michael W: **Mapping the human brain's**

cortical-subcortical functional network organization. *NeuroImage* 2019, **185**:35-57.

32. Johnson Justin, Hariharan Bharath, van der Maaten Laurens, Li Fei-Fei C, Zitnick Lawrence, Girshick Ross: **Clevr: A diagnostic dataset for compositional language and elementary visual reasoning.** *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017:2901-2910.
- The proposed CLEVR dataset shows how to construct a large number of tasks using a small set of subtasks as building blocks.
33. Kell Alexander JE, Yamins Daniel LK, Shook Erica N, Norman-Haignere Sam V, McDermott Josh H: **A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy.** *Neuron* 2018, **98**:630-644.
- This paper studied how neural network models solve speech and music recognition tasks. It systematically varied the degree of shared versus separate processing.
34. Kirkpatrick James, Pascanu Razvan, Rabinowitz Neil, Veness Joel, Desjardins Guillaume, Rusu Andrei A, Milan Kieran, Quan John, Ramalho Tiago, Grabska-Barwinska Agnieszka *et al.*: **Overcoming catastrophic forgetting in neural networks.** *Proc Natl Acad Sci* 2017, **114**:3521-3526.
35. Lake Brenden M, Baroni Marco: **Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks.** *arXiv preprint arXiv:1711.00350.* 2017.
36. Lake Brenden M, Salakhutdinov Ruslan, Tenenbaum Joshua B: **Human-level concept learning through probabilistic program induction.** *Science* 2015, **350**:1332-1338.
37. Li Qi, Yang Guochun, Li Zhenghan, Qi Yanyan, Cole Michael W, Liu Xun: **Conflict detection and resolution rely on a combination of common and distinct cognitive control networks.** *Neurosci Biobehav Rev* 2017, **83**:123-131.
38. Lin Tsung-Yi, Maire Michael, Belongie Serge, Hays James, Perona Pietro, Ramanan Deva, Piotr Dollár C, Zitnick Lawrence: **Microsoft coco: Common objects in context.** *In European conference on computer vision* 2014:740-755. Springer.
39. Long Jonathan, Shelhamer Evan, Darrell Trevor: **Fully convolutional networks for semantic segmentation.** *In Proceedings of the IEEE conference on computer vision and pattern recognition* 2015:3431-3440.
40. Mante Valerio, Sussillo David, Shenoy Krishna V, Newsome William T: **Context-dependent computation by recurrent dynamics in prefrontal cortex.** *nature* 2013, **503**:78.
41. Masland Richard H: **The neuronal organization of the retina.** *Neuron* 2012, **76**:266-280.
42. Masse Nicolas Y, Grant Gregory D, Freedman David J: **Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization.** *Proc Natl Acad Sci* 2018, **115**:E10467-E10475.
43. Murray John D, Bernacchia Alberto, Roy Nicholas A, Constantinidis Christos, Romo Ranulfo, Wang Xiao-Jing: **Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex.** *Proc Natl Acad Sci* 2017, **114**:394-399.
44. Emin Orhan A, Ma Wei Ji: **A diverse range of factors affect the nature of neural representations underlying short-term memory.** *Nature Neurosci* 2019:1.
45. Pan Sinno Jialin, Yang Qiang: **A survey on transfer learning.** *IEEE Trans Knowledge Data Eng* 2009, **22**:1345-1359.
46. Rajan Kanaka, Harvey Christopher D, Tank David W: **Recurrent network models of sequence generation and memory.** *Neuron* 2016, **90**:128-142.
47. Rigotti Mattia, Barak Omri, Warden Melissa R, Wang Xiao-Jing, Daw Nathaniel D, Miller Earl K, Fusi Stefano: **The importance of mixed selectivity in complex cognitive tasks.** *Nature* 2013, **497**:585.
48. Rigotti Mattia, Daniel D, Rubin Ben Dayan, Wang Xiao-Jing, Fusi Stefano: **Internal representation of task rules by recurrent**

dynamics: the importance of the diversity of neural responses. *Front Comput Neurosci* 2010, **4**:24.

49. Sebastian Ruder: **An overview of multi-task learning in deep neural networks.** *arXiv preprint arXiv:1706.05098* 2017.
50. Schneider Walter, Chein Jason M: **Controlled & automatic processing: behavior, theory, and biological mechanisms.** *Cognit Sci* 2003, **27**:525-559.
51. Shiffrin Richard M, Schneider Walter: **Controlled and automatic human information processing: II. perceptual learning, automatic attending and a general theory.** *Psychol Rev* 1977, **84**:127.
52. Snell Jake, Swersky Kevin, Zemel Richard: **Prototypical networks for few-shot learning.** *In Advances in Neural Information Processing Systems* 2017:4077-4087.
53. Francis Song H, Yang Guangyu R, Wang Xiao-Jing: **Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework.** *PLoS Comput Biol* 2016, **12**:e1004792.
54. Francis Song H, Yang Guangyu R, Wang Xiao-Jing: **Reward-based training of recurrent neural networks for cognitive and value-based tasks.** *Elife* 2017, **6**:e21492.
55. Sussillo David, Abbott Larry F: **Generating coherent patterns of activity from chaotic neural networks.** *Neuron* 2009, **63**:544-557.
56. Tang Evelyn, Mattar Marcelo G, Giusti Chad, Lydon-Staley David M, Thompson-Schill Sharon L, Bassett Danielle S: **Effective learning is accompanied by high-dimensional and efficient representations of neural activity.** *Nature Neurosci* 2019:1.
57. Teh Yee, Bapst Victor, Czarnecki Wojciech M, Quan John, Kirkpatrick James, Hadsell Raia, Heess Nicolas, Pascanu Razvan: **Distral: Robust multitask reinforcement learning.** *In Advances in Neural Information Processing Systems* 2017:4496-4506.
58. Todorov Emanuel, Erez Tom, Tassa Yuval: **Mujoco: A physics engine for model-based control.** *In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* 2012:5026-5033. IEEE.
59. Tsao Doris Y, Freiwald Winrich A, Tootell Roger BH, Livingstone Margaret S: **A cortical region consisting entirely of face-selective cells.** *Science* 2006, **311**:670-674.
60. Vinyals Oriol, Blundell Charles, Lillicrap Timothy, Wierstra Daan *et al.*: **Matching networks for one shot learning.** *In Advances in neural information processing systems* 2016:3630-3638.
61. Wang Jane X, Kurth-Nelson Zeb, Kumaran Dharshan, Tirumala Dhruva, Soyer Hubert, Leibo Joel Z, Hassabis Demis, Botvinick Matthew: **Prefrontal cortex as a meta-reinforcement learning system.** *Nature Neurosci* 2018, **21**:860.
- This study demonstrated that recurrent neural networks that receive past reward and action as inputs can be trained to rapidly solve new reinforcement learning tasks drawn from a meta-task.
62. Wang Jane X, Kurth-Nelson Zeb, Tirumala Dhruva, Soyer Hubert, Leibo Joel Z, Munos Remi, Blundell Charles, Kumaran Dharshan, Botvinick Matt: **Learning to reinforcement learn.** *arXiv preprint arXiv:1611.05763.* 2016.
63. Weston Jason, Bordes Antoine, Chopra Sumit, Rush Alexander M, Merriënboer Bart van, Joulin Armand, Mikolov Tomas: **Towards ai-complete question answering: A set of prerequisite toy tasks.** *arXiv preprint arXiv:1502.05698.* 2015.
64. Yamins Daniel LK, DiCarlo James J: **Using goal-driven deep learning models to understand sensory cortex.** *Nature Neurosci* 2016, **19**:356.
65. Yamins Daniel LK, Hong Ha, Cadieu Charles F, Solomon Ethan A, Seibert Darren, DiCarlo James J: **Performance-optimized hierarchical models predict neural responses in higher visual cortex.** *Proc Natl Acad Sci* 2014, **111**:8619-8624.
66. Yang Guangyu Robert, Ganichev Igor, Wang Xiao-Jing, Shlens Jonathon, Sussillo David: **A dataset and architecture for visual reasoning with a working memory.** *In European Conference on Computer Vision* 2018:729-745. Springer.

This paper shows how to construct a large set of cognitive tasks using the same method as the CLEVR dataset.

- 67. Yang Guangyu Robert, Madhura R, Joglekar H, Francis Song, Newsome William T, Wang Xiao-Jing: **Task representations in neural networks trained to perform many cognitive tasks.** *Nature Neurosci* 2019:1.

This paper trained a recurrent neural network to perform 20 interrelated cognitive tasks. It shows how modular solutions arise from specialization.

- 68. Yosinski Jason, Clune Jeff, Bengio Yoshua, Lipson Hod: **How transferable are features in deep neural networks?** *Adv Neural Inform Processing Syst* 2014:3320-3328.
- 69. Zenke Friedemann, Poole Ben, Ganguli Surya: **Continual learning through synaptic intelligence.** *In Proceedings of the 34th International Conference on Machine Learning-Volume 70* 2017:3987-3995. JMLR. org.