nature reviews neuroscience

https://doi.org/10.1038/s41583-026-01036-2

Perspective

Check for updates

Categorization is 'baked' into the brain

Lisa Feldman Barrett [1,2,3] ✉ & Earl K. Miller [4,5]

Abstract

Categorization, the grouping of objects, living organisms, actions or events into equivalence clusters, is fundamental to adaptive behaviour. Traditionally, it is assumed that categorization begins with feature detection and ends with assigning representations stored in memory. Here we review converging evidence from neuroanatomy, electrophysiology, brain imaging and cognitive science to suggest an alternative view: categorization is not the end stage of perception but occurs throughout signal processing, from the very beginning. It is a core computational strategy of the brain, implemented through a neural context created by predictive feedback signals that organize feedforward processing. Implications for theory, future research and neuropsychiatric disorders are discussed.

Sections

[1]Department of Psychology, Northeastern University, Boston, MA, USA. [2]Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, USA. [3]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [4]The Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: l.barrett@northeastern.edu

Introduction

A category is a group of objects, living organisms, actions or events that are similar enough to be treated as equivalent for some purpose or function. Categorization, the process of treating something as equivalent to something else, is fundamental to life. The ability to determine 'this is like that' allows animals to draw on past experiences to guide present actions. All animals categorize, generalizing from past experiences at temporal and spatial scales relevant to their biology and ecological niche. Categorization reduces metabolic costs by minimizing uncertainty in a constantly changing, partially predictable world.

In this Perspective, we discuss the possibility that categorization is not the pinnacle of brain processing, as is traditionally assumed. Instead, categorization is a fundamental operating principle that describes the functional consequences of signal processing throughout the brain. We integrate evidence from neuroanatomy, electrophysiology, brain imaging and cognitive science to suggest that category construction begins as patterns of predictive feedback signals for abstract motor plans. Feedback signals create a neural context that actively shapes the processing of feedforward sensory signals, organizing the reduction of their dimensionality to serve situated, functional goals. In this view, the brain categorizes sensory signals from the outset, rather than as a final-stage process following sensation, attention and perception.

We begin with a review of the main cytoarchitectural gradient that reduces the dimensionality of feedforward sensory inputs to the cerebral cortex. We then discuss the seemingly counterintuitive idea that reverse signal flow along this gradient expands dimensionality to generate prediction signals that shape the processing of and ultimately categorize those inputs to give them meaning. This universal continuous category construction starts at the limbic core of the brain, in areas that are at the core of the brain's regulation of the body, and supports metabolic efficiency by anticipating and preparing for energy needs before they arise, a condition known as allostasis[1,2]. In this way, continuous category construction and the resulting categorization satisfy a key biological constraint, energy optimization, which is an organizing principle of life and evolution[2–4].

The brain compresses sensory signals

A brain is a signal processor. It receives a wide array of signals from the sensory surfaces in the body, such as signals from the rods and cones of each retina, olfactory receptors embedded in the epithelium of each nasal cavity, glucose monitors in the intestines and muscle fibres, oxygen sensors in the carotid arteries and airways of the lungs, thermoreceptors in neurons and so on. With its own intrinsic signals, a brain continually transforms these incoming signals of very high dimensionality (many small, specific details) into fewer signals of lower dimensionality, a type of signal processing called signal compression. Redundancies (correlations in time and space) are reduced. Signal duration and intensity are modulated. Stochastic noise is reduced so that fewer dimensions have the same, or more, informational value. Noise can also be enhanced (for example, to sharpen faint or salient sensory signals, facilitate sensory discrimination and selection, improve motor control, and perhaps even improve control and coordination of the body; for example, refs. 5–10). Noise may also enable generative modelling of past states to construct compressed ensembles that could become possible future states[11].

In engineered systems, compression increases signal quality while decreasing processing cost. In a living animal, any immediate reduction in metabolic burden is a saving that can be invested elsewhere and, in the long run, improves health and the chances of reproductive success. The central nervous system of vertebrates has long been viewed as a system for compressing sensory signals[6,12–14] into more efficient, stable and relevant signals of lower dimensionality (for example, refs. 15–24). The brain's signal compression is lossy, meaning that high-dimensional details are not retained; when remembered later, they must be inferred. A substantial degree of compression takes place in the cerebral cortex itself, whose cytoarchitecture suggests that transforming the dimensionality of signals is intrinsic to its function.

A cortical architecture for dimensionality reduction

The morphology and structural arrangement of pyramidal neurons in the cerebral cortex, particularly in the superficial cortical layers, naturally form a dimensionality reduction gradient that runs from primary visual, auditory and somatosensory cortices (V1, A1 and S1, respectively) to multimodal (or heteromodal) limbic areas (depicted in Fig. 1a; for more detail and additional references, see refs. 25–27). We use the word 'limbic' in its original anatomical meaning, outlined by Broca, to denote the cortical areas that ring subcortical nuclei. Cortical limbic areas have monosynaptic connections to the subcortical nuclei that maintain allostasis and regulate metabolism via their control of the viscera and tissues of the body (for example, see refs. 28–31 and references therein).

Starting at the sensory edge of the gradient, pyramidal neurons gradually increase in size and decrease in density, forming a many-to-fewer gradient (Fig. 1b; also see ref. 32); correspondingly, there is a progressive increase in connectivity and in the diameter and myelination of pyramidal axons[27,33,34]. Consequently, high-dimensional signals arriving at V1, A1 and S1 are successively compressed into increasingly efficient, stable and relevant signals of lower dimensionality as they flow towards the limbic edge. Signals flow from many small, sparsely connected pyramidal neurons that map the receptive fields of the retina, cochlea and other sensory surfaces to fewer but increasingly large and more densely connected pyramidal neurons. They eventually arrive at the limbic edge, which contains some of the largest and most densely connected pyramidal cells in the cerebral cortex, including subgenual and pregenual sectors of the anterior cingulate cortex, posterior orbitofrontal cortex, ventral anterior insular cortex and medial entorhinal cortex. Importantly, several limbic cortices are also a member of the brain's 'rich club'[35] and function as a communication backbone that synchronizes signalling throughout the brain[36].

The many-to-fewer gradient of pyramidal neurons in the cerebral cortex is loosely coordinated with changes in the pattern of thalamocortical projections (Box 1). Most relevant here is the observed variation in the number and segregation of stellate or granule cells, which receive inputs from the thalamus, referred to as variation in cortical granularity. Structurally, limbic cortical regions are called agranular (or periallocortical) in their architecture, because their granule cells are not segregated into a well-defined layer 4 (or granular layer). Variations in cortical granularity, when combined with the topography of the pyramidal cells in layers 2 and 3, describe the degree of cortical lamination[34,37,38]. Feedforward signals flow from areas of more developed to less developed cortical lamination, whereas feedback signals flow in the other direction. Limbic cortices are the least laminated parts of the cerebral cortex, whereas V1, A1 and S1 have the most developed lamination, a point that we return to shortly when we discuss predictive category construction.

Further compression of feedforward signals occurs as they proceed from entorhinal cortex to the hippocampus (for example, refs. 39,40),

which is cortical (specifically, allocortical) in structure. Compression also occurs within the hippocampus itself, both along the transverse axis from the dentate through the CA fields to the subiculum and along the longitudinal axis from posterior to anterior (here we use the hippocampal terminology for a macaque brain; in a rodent brain, the longitudinal hippocampal gradient runs from dorsal to ventral).

Sensory signals for olfaction, gustation and interoception undergo more substantial compression before they reach their primary sensory

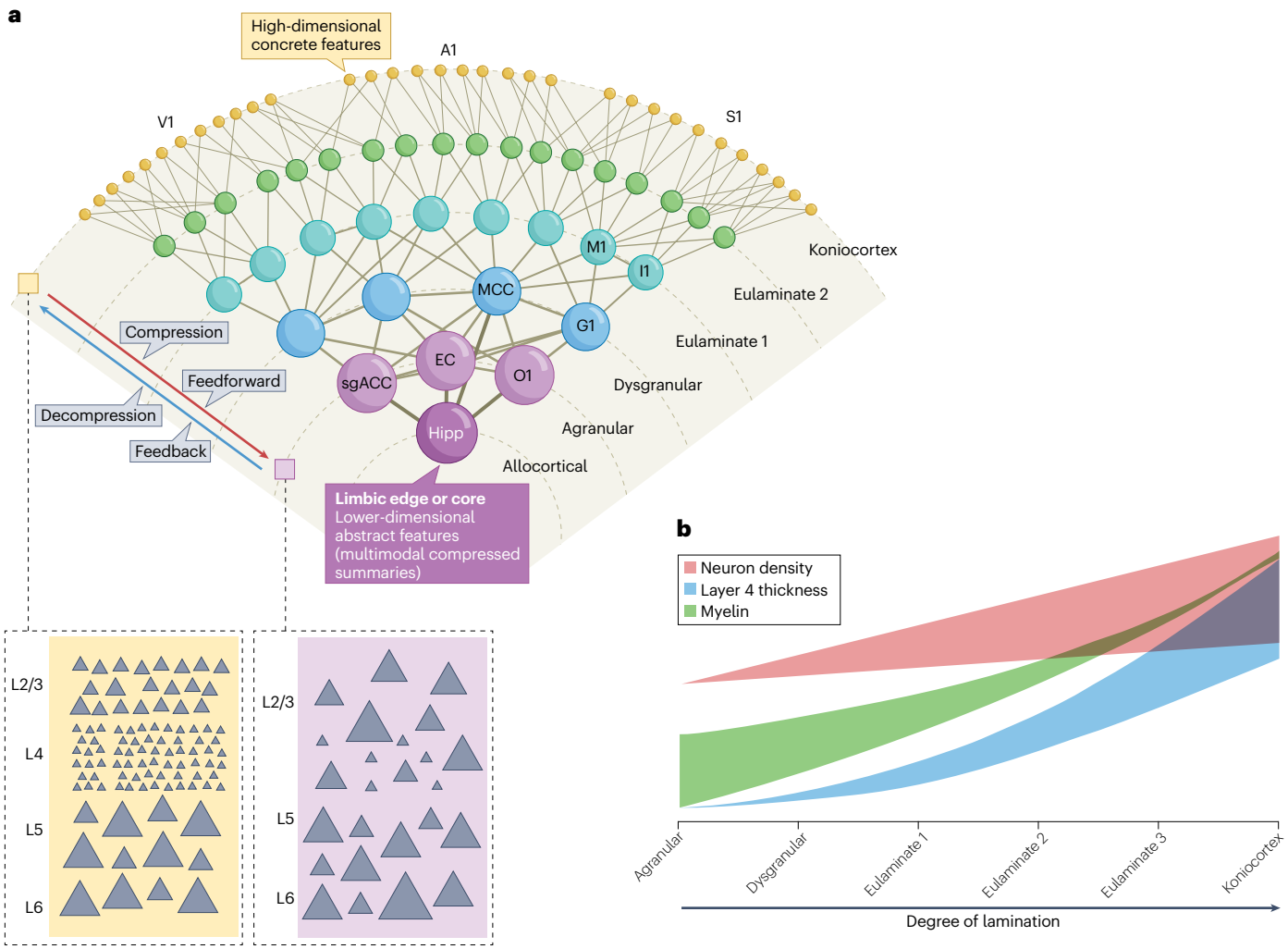


Fig. 1 | The limbic-to-sensory cytoarchitectural gradient of the cerebral cortex. a, A schematic depiction of the main cytoarchitectural gradient of pyramidal cells in layers 2 and 3 of the mammalian cerebral cortex, shown as concentric rows of coloured circles. Not all areas and connections are depicted. The primary sensory areas for distance senses (somatosensory, auditory and visual) are depicted as small yellow circles; these areas correspond to high-dimensional, concrete features. Areas that make up the limbic core of the cortex are depicted in violet and correspond to lower-dimensional, multimodal compressed summaries. Across this gradient, running from the limbic cortex to primary somatosensory cortex (S1), primary auditory cortex (A1) and primary visual cortex (V1), pyramidal neurons successively decrease in size, receptive field (not shown), connectivity (depicted by thickness of line) and intracortical myelination of individual axons (not shown), while at the same time increasing in density and cortical layering (segregation into different lamina). Each concentric dotted line depicts a 'cortical type', from areas of less laminar differentiation (bottom circles) to those of greater lamination (top circles), although in reality, the gradient is continuous and 'types' are not discrete. Cortical lamination is depicted in boxes at the bottom left, with the five-layered agranular (also called periallocortical) cortex in purple and six-layered koniocortical sensory areas in yellow. As feedback signals flow from limbic to sensory, they are decompressed so that their dimensionality is expanded. Feedforward signals flow in the other direction. Several primary sensory areas, such as primary olfactory cortex (O1), primary gustatory cortex (G1) and primary interoceptive cortex (I1) are topographically and cytoarchitecturally limbic or are close to the limbic edge. **b**, A schematic of cytoarchitectural shifts along the main cortical gradient. The *x* axis depicts increasing degrees of lamination. The *y* axis depicts the relative magnitudes of the features that contribute to laminar differentiation. For example, the shaded blue curve corresponds to the degree to which the stellate cells (also called granule cells) that receive thalamic inputs are segregated into a definable layer 4. Agranular areas have stellate cells, but they are not organized into a well-defined granular layer 4. The segregation of stellate cells increases in areas of cortex with greater definition in their cellular layers. Layer 4 size increases along the gradient, with koniocortices (V1, A1 and S1) having the largest, most defined layer 4. EC, entorhinal cortex; Hipp, hippocampus (dentate gyrus and CA); M1, primary motor cortex; MCC, mid cingulate cortex; sgACC, subgenual anterior cingulate cortex. Part **a** was adapted with permission from Fig. 3a in ref. 32, Elsevier. Part **b** was adapted from Fig. 9 in ref. 38, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Box 1 | Variations in thalamocortical projections along the main cortical gradient

There is a variety of types of thalamic input to the cerebral cortex (for a sample of relevant references, see refs. 100,106,252–256), and their projection patterns appear to vary with cortical lamination. The typical distinction is between so-called core projections, which are considered drivers of feedforward cortical signals, and so-called matrix projections, which are thought to modulate feedback cortical signals; the utility of this distinction has recently been challenged[257], but we use it here because it may be more familiar to readers. Core projections originate in the more lateral and posterior subnuclei of the thalamus, which receive a higher proportion of inputs from sensory surfaces. They arrive to the stellate cells in the middle layer of the cerebral cortex (layer 4) and are progressively denser in areas of greater cortical lamination towards the sensory edge of the cortical gradient. The more diffuse matrix projections, which originate from more medial and anterior thalamic nuclei, receive fewer inputs from sensory surfaces and relatively more multimodal, compressed inputs from the hippocampus as well as signals from nuclei involved in visceromotor and skeletomotor regulation, such as the hypothalamus and basal ganglia. Matrix projections arrive to all cortical layers other than layer 4. Agranular cortical regions at the limbic edge of the gradient have no well-defined layer 4, and their stellate cells primarily receive matrix projections. In fact, matrix projections preferentially target layers 1 through 3 of limbic cortices[106]. Cortical efferents to the thalamus from the deep layers of cortex (primarily layer 6) are thought to result in compressed, lower-dimensional patterns at the thalamus[107,258],which can then be relayed back to cortex via the medial nuclei to ensure efficient, coordinated brain-wide signal processing[109,110,259].

areas. All are within the insular cortex and are considered largely or partially limbic (Fig. 1a).

Several decades of high-resolution histological studies in macaque monkeys and other mammals, which reveal this many-to-few architectural arrangement, provide corroborating evidence for dimensionality reduction in feedforward signals along the cortical gradient (for example, refs. 34,37,38,41). Specifically, the majority of corticocortical feedforward signals originate in pyramidal cells within the deep layers of a cortical territory and terminate in the upper layers of cortical territories with fewer pyramidal cells (that is, in cortical territories with relatively less cortical lamination). The majority of corticocortical feedback signals flow in the other direction, along axons originating in the deep layers of relatively less-laminated cortical areas that project to pyramidal neurons in the upper layers of areas with relatively more cortical lamination. This pattern of connectivity is part of a larger signal processing motif referred to as the structural model of the cerebral cortex, described in Box 2, which we return to later in the paper when we discuss its relevance for categorization.

Electrophysiological recordings of single neurons in macaque monkeys similarly provide direct evidence of dimensionality reduction within this many-to-fewer, sensory-to-limbic, cytoarchitectural gradient. Feedforward sensory signals gradually become more abstract and behaviourally relevant with the parallel reduction in neural population dimensionality[42,43].

Brain imaging evidence of functional connectivity gradients within the cerebral cortex are similarly consistent with the hypothesis of signal compression in this many-to-fewer architectural gradient (for example, refs. 44–47). Evidence further suggests that cortical, hippocampus and cerebellar compression gradients align with one another (see ref. 48 for evidence and references therein).

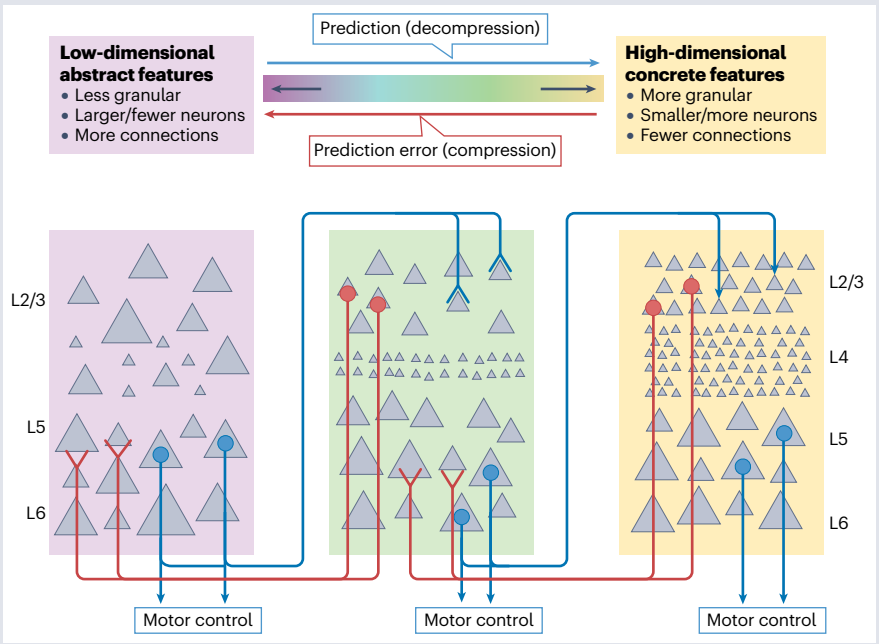
There is also a temporal aspect to the signal compression. Small pyramidal cells in the sensory end of the gradient change their firing at faster timescales (milliseconds), whereas the larger cells in the limbic edge change on slower timescales (seconds to minutes)[49–51].

One consequence of the spatial and temporal dimensionality reduction within the many-to-few convergent architecture is that the specificity of visual, auditory and somatosensory maps gradually disappears as feedforward signals move from their respective primary sensory cortices towards the limbic edge. The pyramidal neurons of V1, for example, create a high-dimensional, retinotopic map. This is called nearest-neighbour, array-to-array mapping because the spatiotemporal relationships of rods and cones in the retina correspond to the spatiotemporal relationships of receptive fields of neurons in primary visual cortex. Nearest-neighbour, array-to-array mapping also exists in A1 (for the cochlea) and S1 (for the body)[27,52]. As spatiotemporal redundancies are increasingly removed by lossy compression, nearest-neighbour mappings become increasingly less fine grained, both spatially and temporally, until the array-to-array signal maps are eventually lost. Mappings increasing reflect semantic or conceptual similarity[53]. The primary sensory areas for olfaction (O1), gustation (G1) and interoception (I1) do not contain fine-grained array-to-array spatial maps of their corresponding sensory surfaces, consistent with the evidence that olfactory, gustatory and interoceptive signals undergo substantial compression before they reach their respective primary sensory areas in the insula (again, see ref. 23 for evidence and references therein).

The anatomical arrangement and signal processing differences between proximal senses (olfaction, gustation and interoception) and distance senses (vision, audition and aspects of somatosensation) suggest that the former may condition and even gate the latter, a hypothesis that is supported by growing empirical evidence[54,55] (also see ref. 32 for discussion and additional references). More generally, accumulating evidence suggests that the proximal senses broadly shape brain-wide signalling through multiple pathways, consistent with their anatomical position in the limbic-to-sensory cortical gradient (for example, ref. 56; also see ref. 32 for discussion and additional references).

A second, related consequence of signal compression is multimodality. The compressed signals in the limbic edge of the gradient are statistically related to and contain some information about the signals of all sensory systems (for example, refs. 57–60), meaning that they share mutual information in the Shannon sense. Importantly, these multisensory compressed signal summaries at the limbic edge also function as abstract motor control signals that create action plans for both motor systems in the body: the visceromotor system, which controls the internal organs (or viscera) and other tissues of the body, and the skeletomotor system, which controls the voluntary muscles that move parts of the skeleton. Limbic cortical neurons, which create the most

Box 2 | The structural model of corticocortical architecture



The notion of a hierarchy is sometimes used to describe the flow of signal across the cerebral cortex. This is a point of considerable contention. Various signal flow organizations have been proposed[34], including parallel, overlapping streams of hierarchical signal flow (for example, ref. 260), task-dependent hierarchies (for example, refs. 261,262) and organizing schemes that dispense with the notion of hierarchy altogether (for example, refs. 263,264). At other times, the notion of a hierarchy is used to describe the architectural arrangement of neurons, which is then used to predict signal flow. We use this latter definition of hierarchy, relying on the structure model of corticocortical signalling by Barbas and colleagues (for example, refs. 37,38,41,265). This is a lamination-based organization of the cerebral cortex that describes its main gradient as a loose hierarchy of pyramidal cells that has been observed to predict the majority of cortical signal flow. This loose hierarchy, in conjunction with other cytoarchitectural features such as the landscape of inhibitory neurons and the patterns of re-entrant thalamocortical connections, allows us to make specific hypotheses about signal processing patterns that give rise to category construction and categorization.

Lamination-based hierarchies of the cerebral cortex and their ability to reliably predict something about signal flow remain an active area of research. The structural model is derived from more than three decades of anterograde and retrograde tract-tracing studies in primates and other mammals (for example, refs. 37,38,41,265). Other lamination-based hierarchical models of the cortex, such as the Felleman and van Essen model and the distance rule model, also have value, but do not predict signal flow across the expanse of the cerebral cortex as well as the structural model. In contrast to the structural model, hierarchy rules such as the distance rule do not hold in prefrontal areas[266].

According to the structural model, signal flow largely follows two particular motifs, one for feedback signals and the second for feedforward signals (see the Figure for a depiction). The purple box is a schematic depiction of agranular cortex in the limbic core (for example, posterior orbitofrontal cortex, which is a primary visceromotor cortex, or primary olfactory cortex). The green box is a schematic depiction of dysgranular cortex (for example, anterior mid cingulate cortex, which is a primary visceromotor cortex and an association region for the skeletomotor system, or dorsal mid insula, which is part of primary interoceptive cortex). The yellow box depicts a schematized granular cortex (for example, primary visual cortex or primary auditory cortex). The cytoarchitectural features of these cortical areas are printed in black (repeated from Fig. 2).

Feedback signals (depicted in blue) originate from neurons within the deep layers of relatively less-laminated cortical areas (layers 5 and 6) and terminate in the upper layers of similar or relatively more laminated cortex (layers 2 and 3; also see refs. 134,147).

The pyramidal cells in layers 5 and 6 of the cerebral cortex issue projections to subcortical and spinal neurons involved in control of the viscera and skeletal muscle (visceromotor and skeletomotor control signals, respectively), as well as to the thalamus. Efference copies of the motor control signals are sent to other cortical areas of comparable lamination (considered lateral pathways) and cortical areas of greater lamination as feedback or signals (in blue). These correspond to prediction signals[147] (Fig. 3). A single pyramidal cell projects simultaneously to multiple cortical and subcortical targets (for example, refs. 267–270 and also see ref. 139). Feedforward signals (in red) flow in the other direction. They originate from neurons within the upper layers of relatively more laminated cortical areas and terminate in the deep layers of similar or relatively less laminated cortex (again, see refs. 134,147).

The figure depicts heuristics for signal flow that apply to any portions of the cortical gradient that differ in laminar differentiation (and not just the areas depicted or named in the Figure). For example, primary interoceptive cortex (I1) has a laminar structure that is less differentiated than the structure in primary visual cortex (V1). Direct or indirect signals originating in the deep layers of I1 (layers 5 and 6) and terminating in the upper layers of V1 (layers 2 and 3) would function as feedback or visual prediction signals to V1. Signals flowing in the other direction would function as feedforward or prediction error signals (as discussed in the main text).

Both the feedback signal flow motif, with any single axon and its collaterals projecting to multiple neurons, and the feedforward motif, with a greater number of smaller neurons projecting to fewer, larger neurons, have been observed across the expanse of the cerebral cortex. Both motifs have functional implications for category construction, categorization and category learning, as discussed in the main text.

Figure adapted with permission from Fig. 3b in ref. 32, Elsevier.

compressed, multisensory cortical summaries, simultaneously serve as primary visceromotor control regions (for example, anterior cingulate cortices), skeletomotor association regions (for example, ventral premotor cortex) or both (for example, anterior mid cingulate[61,62]). The hippocampus, which is also considered limbic cortex (being allocortical), is multisensory and plays a substantial role in sensing and controlling the viscera (for example, refs. 63,64). This empirical evidence of sensory and motor multimodality in the cortical areas with the most signal compression is consistent with evidence that cognition is embodied (for example, ref. 65).

Motor control signals leave pyramidal cells in the deep layers of limbic cortices (particularly layer 5) as descending efferent signals to subcortical areas. This is the case throughout the entire expanse of the cerebral cortex, but here we are particularly concerned with the abstract motor control signals descending from the limbic edge of the gradient. Efferent signals from the deep layers of limbic cortices expand in dimensionality (they become increasingly particularized) as they move down the midbrain and brainstem to the spinal cord, where they eventually specify particular muscle contractions in the skeletomotor system[61,66]. In a similar manner, efferent signals decompress as they descend to the vagus nerve and spinal cord to control the autonomic nervous system, the immune system, the endocrine system and other systems of the body's internal milieu. Low-dimensional action plans can therefore be understood as multimodal action categories[25,67,68], action goals[61] or action maps[69] that correspond to ethologically relevant behaviours.

Summary

In summary, a many-to fewer convergent architectural gradient in the cerebral cortex is an anatomical organization that exploits the statistical structure of co-occurring feedforward signals, reducing their dimensionality by compressing them into shared lower-dimensional sensorimotor summaries. Compression is a signal processing substrate that contributes to categorization: momentary equivalences are created from differences for behavioural purposes, always involving both visceromotor and skeletomotor action plans. Many high-dimensional sensory signals become functionally similar to one another in terms of the lower-dimensional action plans that they contribute to. Categorization, entailed by compression, offers a powerful adaptive advantage. It reduces the massive sensory complexity of an animal's environmental niche to enable efficient and flexible action planning. This advantage becomes even more powerful when signal compression is shaped by past experience manifest as signal decompression, which we discuss in the next section.

The traditional view and the neural context view of categorization

Every animal's immediate environment (its niche) is a signal scape of exceptionally high dimensionality, rivalled only by the number of signals inside its body. As an animal moves, signals from the body's sensory surfaces are in constant flux. Animals must reduce this massive sensory complexity of their signal scape in a manner that makes action planning more efficient according to their own physical and temporal ecology. Creating momentary similarities from differences for the purpose of action planning is a necessary form of simplification to survive and thrive in the ever-changing and only partly predictable world.

Categories capture inferred relations beyond what raw sensory inputs explicitly provide. This can range from simple, deterministic groupings based on a single feature to more complex groupings that are probabilistic, conjunctive or disjunctive and depend on inferring abstract features not directly present in the external world (for a review, see ref. 70). As a result, many small sensory particulars that co-occur during a specific spatiotemporal event become functionally similar to one another in terms of the lower-dimensional signals they contribute to. We previously described function as 'action plans' or 'action concepts'. Psychologically speaking, the functions of these lower-dimensional signals can also be described as abstract mental features, such as 'threat', 'value', 'reward', 'pleasure', 'pain' and other so-called semantic features[29,53].

The utility of creating functional equivalences from sensory differences is amplified when an animal abstracts across different spatiotemporal events, lest they be stranded in an immediate present of novelty and uncertainty. Different high-dimensional sensory arrays that occur in different contexts are rendered functionally equivalent when they are compressed across numerous instances into a common low-dimensional action category. Any abstract action category can be decompressed into more than one pattern of motor particulars, as described above.

A human, for example, has the possibility of compressing multiple high-dimensional ensembles of feedforward metabolic, proprioceptive, olfactory, gustatory, visual, auditory and haptic signals into a common low-dimensional action category for 'threat' or for 'reward'. When compressed to share abstract features, prior events from different situations can be generalized to plan action for a new occurrence in the present, even when those past events looked different, sounded different, tasted different, required different specific muscle contractions and so on. In particular, words may serve as powerful features of equivalence that invite categorization (for example, refs. 71–74) because they prompt a search for some underlying sameness that transcends any noticeable differences (for example, refs. 75–78). The ability to transcend detailed physical differences by treating them as functionally equivalent supercharges a brain to flexibly generalize from past to present and to efficiently accumulate knowledge for use in the future (for a similar point, see refs. 15,17).

Categories are therefore an effective avenue of simplification and generalization both during a single signal-processing event and across longer spans of time. Different signal patterns with a dimensionality of x, that vary and in space and time, become equivalent or interchangeable for the purpose of planning action when they are compressed to share the same (or a similar) multimodal summary of fewer dimensions, n. A multimodal summary of n dimensions, therefore, can be thought of as the abstract features of equivalence for a category of signal ensembles of x dimensionality, where $n < x$. Any set of higher-dimensional signals that are compressed (and therefore become equivalent in terms of) some smaller set of lower-dimensional signals can be understood as a category. This is how categorization enables inference and permits generalizations from past category instances that share the same features of equivalence. The categories that result from lossy compression have been depicted both as neural geometry within a neural state space that includes both x and n dimensions[79] (as discussed in ref. 80) and as geometry within a mental feature space in which the dimensions correspond to the mental features that are thought to arise from the neural computations. In this latter option, n features are depicted as lower-dimensional manifolds within a higher x-dimensional geometric space[81].

This logic holds for signal compression across the entire cortical gradient. A feature of equivalence at one level of dimensionality reduction (for example, the neurons whose signals constitute the category RED) serves as a higher-dimensional feature for a more abstract

category of lower dimensionality (for example, 'red' is a feature of FLOWER), with features becoming relatively more abstract as signals flow towards the limbic edge. When viewed from this perspective, categorization is not about assigning instances to equivalence groupings. It goes beyond what is given (sensory signals) to create equivalences that exist in relation to the brain doing the signal processing. As a consequence, a brain is an information-gaining system that creates meaning. Once categorized, a high-dimensional signal ensemble and its causal impacts are explained[82].

Effective categorization involves choosing the optimal dimensionality for features of equivalence, tailoring them to immediate task demands (for example, refs. 15,20,39,83). A brain that relies primarily on abstract features of equivalence risks overgeneralization, context insensitivity and poor episodic memory (as occurs in major depressive disorder). Generalization becomes more difficult when relying on higher-dimensionality sensory features of equivalence to the exclusion of lower-dimensional, multimodal abstractions and results in insufficiently flexible and overly situation-specific actions (as occurs in some variants of autism spectrum disorder).

The traditional view of categorization

In the traditional view, depicted in Fig. 2, categorization begins with the bottom-up decoding of objects and scenes in the world, starting with feature detection. This traditional decoding account is assumed to be the process by which newly encountered objects, people and events are detected in the world and are assigned to existing categories, or by which newly acquired categories come to be represented by forming stable mental structures called concepts (for example, refs. 84–86). For example, while out for a walk, say you encounter a small furry black animal that purrs, with whiskers and pointy little ears. It is assumed that your brain registers low-level features of the animal, the street you are walking on and so on, detecting lines, edges and other features of high dimensionality within the receptive fields of primary visual cortex, muscle contractions and cutaneous tactile features in the receptive fields of somatosensory cortex, vibrations that become tones in the receptive fields of primary auditory cortex, odorants in primary olfactory cortex and so on. Feedforward sensory signals are thought to successively activate neurons as they move along the cytoarchitectural hierarchy of the cerebral cortex, successively binding features together until those signals, as compressed summaries, reach the territory of semantic memory. At this point, the signal patterns are supposedly evaluated and categorized, eventually resulting in an appropriate action plan, such as to reach down and pet the cat or swerve to avoid it. (Traditional accounts of categorization rarely, if ever, consider the visceromotor actions and other dynamics of the body's internal milieu that are required to support the skeletomotor movements involved.)

In this conventional description, categorization is almost exclusively entailed by signal compression. A high-dimensional signal array is simplified into a more manageable, lower-dimensional summary that can be categorized by matching it to a category prototype or prior instance (an exemplar) that is retrieved from semantic memory. So, as you encounter the cat, your brain 'decodes' the meaning of all the sensory signals and somehow binds them into a compressed summary. Only then does your brain retrieve a representation of the category CAT to categorize the compressed summary. The prototype, for example, is said to be a single instance that summarizes a graded distribution of similarity or family resemblance for all cats across time and space[87]. Every individual cat that has ever existed or will exist in the world shares a family resemblance because of their family resemblance with this prototype. Most laboratory studies of categories, in which objects are presented as stimuli to be categorized, are designed with the assumption that categories are pre-existing in the world and therefore stable across place and time[88]. Even exemplar-based theories of categorization, which allow for more context sensitivity

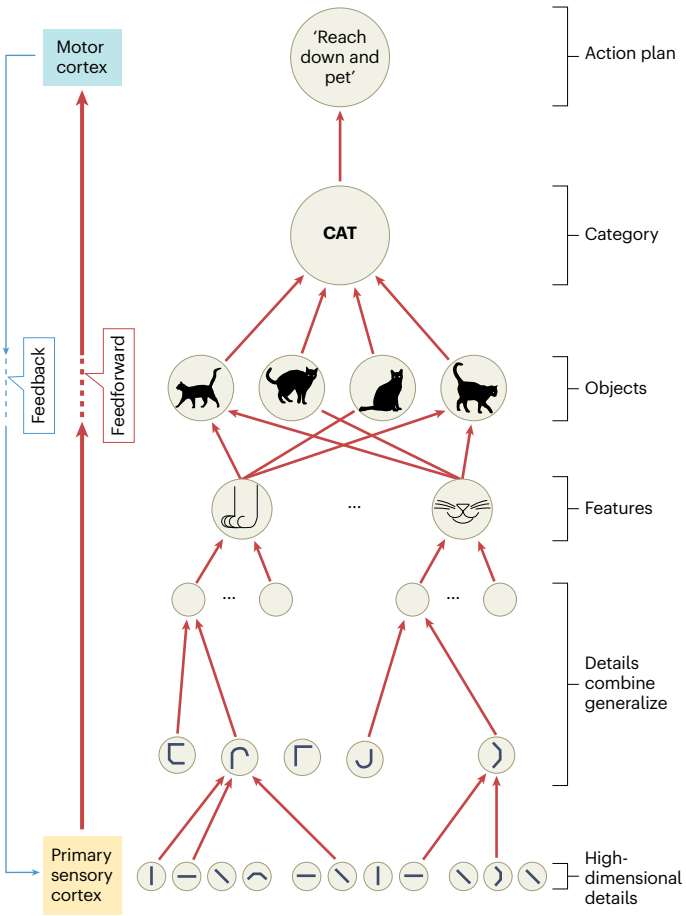


Fig. 2 | The traditional view of categorization. In the conventional feedforward conception, categorization is almost exclusively signal compression. A high-dimensional signal array of features in objects and scenes is decoded and simplified into a more manageable, lower-dimensional summary. The summary is categorized by matching it to a category prototype or exemplar that is retrieved from semantic memory. In this example (read from the bottom up), you detect lines, edges and other features of high dimensionality within the receptive fields of primary visual cortex, muscle contractions and cutaneous tactile features in the receptive fields of somatosensory cortex, vibrations that become tones in the receptive fields of primary auditory cortex, odorants in primary olfactory cortex and so on. Feedforward sensory signals successively activate neurons as they move along the cytoarchitectural hierarchy of the cerebral cortex, successively binding features together to create limbs, paws, a nose, whiskers and so on, until the brain has constructed a compressed representation of a cat. This representation is categorized by comparing it to the prototype or prior exemplar of a cat that is retrieved from semantic memory, resulting in an appropriate action plan to reach down and pet the cat. Traditional accounts of categorization rarely, if ever, consider the visceromotor actions and other dynamics of the body's internal milieu that are required to support the skeletomotor movements involved.

in categorization, assume that a brain stores and retrieves previously experienced instances (exemplars) of already existing categories, so as to categorize new objects or events in terms of their similarity to the retrieved category instances (for example, refs. 85,89).

Categories as dimensionality expansion of feedback signals

The traditional decoding account of categorization as the compression of feedforward signals is made implausible, however, by a growing number of empirical considerations. Together, this evidence suggests that categories first emerge as low-dimensional feedback signals (corresponding to abstract, multimodal features) whose dimensionality is then expanded (decompressed) as they propagate from the limbic edge to the sensory and motor periphery. These feedback signals of expanding dimensionality actively shape the compression of feedforward signals and therefore play a prominent (and perhaps dominant) role in how incoming sensory signals are categorized.

First is the repeated observation that feedback signals far outnumber feedforward signals in the cerebral cortex, suggesting that signals flowing from few (limbic) to many (sensory) projection neurons far exceed those that flow in the opposite direction. Precise ratios for the proportion of feedforward versus feedback synapses across the entire cortical gradient are not yet available and, given the complexity of the task, may not be available for some time[90], but anatomical studies of individual cortical areas are instructive. V1, constituting the sensory edge of the gradient, contains no projection neurons that feedback to other cortical areas[90]. The majority of synapses in V1 bring feedback signals from other areas of cortex; they are not core projections from the lateral geniculate nucleus of the thalamus that carry feedforward signals (on the order of 90% are feedback; for example, refs. 91–93). Even the lateral geniculate nucleus of the thalamus, which carries retinal signals to V1, receives more feedback from the cortex than feedforward signals from the retina[92]. Moreover, feedback connections are capable of generating their own receptive fields in V1 neurons, in addition to the traditional feedforward receptive fields, allowing a brain to infer feedforward signals that are missing[94]. Thus, categorization is unlikely to be solely or even primarily driven by feedforward dimensionality reduction. Even in primary sensory cortices, the signal flow is primarily efferent to the thalamus rather than afferent from it.

Second, as feedforward signals arrive to any given pyramidal neuron, they are met and shaped by the feedback signals that the neuron is already processing. Feedback signals oscillate more slowly than feedforward signals and therefore can provide a temporal organizing context for their processing (called cross-frequency phase coupling; for example, ref. 95). The cortical compression of feedforward signals can therefore be actively shaped by the many feedback signals that are decompressing in few-to-many divergent patterns.

Third, there is some question as to whether feedforward signals to the cerebral cortex are pure, bottom-up sensory signals in the first place. Afferent sensory signals are positioned to be shaped by efferent signals even before they reach the brain. For example, efferent signals originating in the brain shape processing in the olfactory bulb (for example, ref. 96) and in the retina (for example, ref. 97; also ref. 98). Efferent signals in the vagus nerve and the nodose ganglia have the potential to shape ascending viscerosensory signals (see ref. 23 for discussion and references). Once in the brain, there are many opportunities for afferent and efferent signals to influence one another before the afferent signals reach the cerebral cortex, for example, in brainstem nuclei, superior colliculus, hypothalamus and thalamus.

In addition, as discussed in Box 1, the entire cerebral cortex receives two types of thalamic projection, neither of which can be considered pure bottom-up signals. Core projections originate in the more lateral and posterior subnuclei of the thalamus and preferentially terminate in the stellate neurons in or around layer 4. These projections receive a higher proportion of inputs from sensory surfaces than matrix projections and are considered 'drivers' of feedforward sensory signals[99,100], but they are not unimodal. For example, the lateral geniculate nucleus receives signals from the retina as well as interoceptive inputs from the periaqueductal grey (for example, ref. 101), the parabrachial nucleus (for example, ref. 102) and the hypothalamus (for example, ref. 103); it also receives monosynaptic inputs from the limbic pregenual anterior cingulate (for example, ref. 104; for additional references, see refs. 29,105). The more diffuse matrix projections originate in more medial and anterior 'higher-order' thalamic subnuclei and arrive at all cortical layers other than layer 4, preferentially targeting layers 1 through 3 of limbic cortices[106]. They receive fewer inputs from sensory surfaces and relatively more inputs from the hippocampus, hypothalamus, basal ganglia and deep layers of cortex. These projections are thought to bring compressed, lower-dimensional signals[107,108] that shape feedback signals and ensure efficient, coordinated brain-wide signalling[109,110].

Fourth, even if the feedforward signals received by the cortex were pure bottom-up signals, object decoding could be achieved only if the receiving cortical neurons had fixed receptive fields so that their axons functioned as labelled lines with inherent meaning independent of context. Then, somehow these features would bind into a coherent percept. However, it is now well established that cortical receptive fields are highly sensitive to context. Neurons across the cerebral cortex show multifunctional 'mixed selectivity', meaning that they code many different features in a manner that changes with context (for example, refs. 111–113). This context-dependent variation is observed even at the sensory edge of the cortical gradient, such as in V1 (for example, refs. 114–120).

These four sources of evidence together suggest that a pyramidal neuron's response necessarily depends on the larger ensemble of neurons that it is communicating with momentarily, called its neural context (for discussions see refs. 111,121–125). Given the sheer number of feedback connections in cortex, the neural context for any cortical pyramidal neuron probably contains the many feedback signals it receives (relative to the fewer feedforward signals) in addition to inputs from inhibitory neurons. Inhibitory neurons, particularly those in the upper cortical layers, dynamically shape the transmission of signals (for example, refs. 126,127). A neuron's action potentials therefore have relational meaning that is situated in and changes with its neural context. That an action potential's meaning depends on the relation between sending and receiving neurons is further reinforced by the signal flow motif described in Box 2.

Even if feedforward signals were purely bottom-up with inherent meanings, their decoding would be unlikely to proceed to categorization in a purely feedforward fashion. Feedforward signals are inherently uncertain and ambiguous. Afferent sensory signals are woefully incomplete[128]. For example, synapses in the retina are noisy, and up to half the signal information is lost as it makes its way along the optic nerve to the brain[2]. Afferent sensory signals are also effects that can have more than a single cause. Signals from the body's sensory surfaces are the outcomes of changes in the body and in the world. Any given signal corresponding to an edge, a tone, a glucose or oxygen concentration, and so on, can result from more than one plausible change. Such

ambiguities are more fervent in sensory modalities with substantial spatial compression at the sensory surfaces (such as olfaction and some viscerosensory modalities). A brain never has access to the events that initiate the sensory signals, resulting in an inverse problem of massive scale. This inverse problem is made more challenging by the fact that the incoming signals dynamically change, carrying different degrees of noise and varying in temporal scale. A brain must solve this inverse problem, from outcomes to causes, in a metabolically efficient way for an animal to plan and execute action in the moment and, ultimately, to survive and reproduce.

Summary

The dominance of feedback signals in the cortex and the ambiguity of feedforward signals makes the traditional account of categorization unlikely. Instead, what emerges is an inverse problem: a brain must always attempt to infer from feedforward signals the causal factors that produced them. A predictive processing account of categorization offers a counterintuitive solution to the brain's inverse problem: the lower-dimensional, feedback signals that dominate the cerebral cortex and determine how a neuron responds to higher-dimensional feedforward signals are realized in advance of the feedforward signals. An ensemble of feedback signals can be understood as the features of equivalence for a momentarily constructed category of signals that anticipate feedforward signals. Decompressing feedback signals are possible futures that ultimately give feedforward signals meaning in relation to the momentary internal state and ecology of the animal. Ultimately, any conjunction of expanding feedback signals and compressing feedforward signals eventually results in the categorization of feedforward signals in terms of their metabolic and motor consequences. In the next section, we describe the details of how a predictive processing approach, rooted in the structural model of corticocortical connections, describes categorization as emerging from whole-brain feedback and feedforward signal dynamics. Neurons at the limbic edge of the cerebral cortex are the source of prediction signals, powerfully shaping category construction and categorization from the outset.

Prediction and category construction

Predictive processing has emerged as a powerful neurocomputational research tradition of many research programmes that account for diverse psychological phenomena subserved by a brain. A core hypothesis is that motor control signals, and the sensory prediction signals they give rise to, begin as compressed multimodal feedback signals in the brain, fashioned from reassembled past experiences (that is, they are remembered). Ample evidence exists for such memory reconstructions (for example, refs. 129–132), with the assembly of low-dimensional features proceeding faster than for high-dimensional features (for example, ref. 133). Feedback signals are continually tested against feedforward signals, correcting when necessary. A variety of specific computational proposals abound, but they are united by the idea that cognition, in the most general sense (including motor control, emotion, perception and so on), can be described within a common set of operating principles[134]. In our account of predictive processing, one of these operating principles is categorization[25]. Feedback signals, as groups of prediction signals, solve the inverse problem and give feedforward signals meaning in terms of the action plans (or the visceromotor and skeletomotor control signals) required to deal with them. This predictive account of categorization integrates the several decades of neuroanatomical evidence discussed previously (Box 2) with evidence that supports a predictive account of brain function (for a review of evidence, see refs. 134–139, and for specific examples, see refs. 140–145).

We hypothesize that category construction begins as compressed multimodal summaries in the hippocampus and in the deep layers of agranular and some dysgranular cortices that make up the limbic edge. They are reinstated (remembered) ensembles of low-dimensional signals that correspond to abstract features. These low-dimensional signals expand as efferent signals that cascade down to successively more numerous, smaller neurons in the hypothalamus, midbrain, brainstem, vagus nerve and spinal cord. In effect, these signals decompress as probabilistic inferences of higher dimensionality for controlling the autonomic nervous system, the immune and endocrine systems, and the skeletomotor system. At the same time, copies of these efferent motor control signals (called efference copies[146]) leave the deep layers of limbic cortices and flow along the fewer-to-many cortical gradient to the superficial layers of increasingly granular cortices (in which superficial pyramidal neurons are increasingly smaller with increasingly fewer connections). These signals decompress (expand) as probabilistic inferences of successively higher dimensionality, eventually serving as low-level sensory prediction signals when they reach the more granular primary sensory areas of cortex.

This account of category construction is heavily influenced by the decades of empirical evidence that support the structural model: feedback (or prediction) signals flow along the fewer-to-many cortical gradient in a manner that is organized by patterns of relative cortical lamination. Feedback signals arise from pyramidal neurons in the deep layers of the relatively less-laminated areas. Areas of the limbic edge, being the least laminated of all cortical areas, are therefore at the top of this hierarchy[26,147]. They send but do not receive feedback signals. This is consistent with the idea that limbic cortical neurons function as the most powerful source of feedback signals in the brain[148] and inform our hypothesis that category construction in the cerebral cortex begins at the limbic edge.

Category construction therefore involves signal decompression in two directions at once (Fig. 3a). Lower-dimensional signals descend subcortically as efferent, motor control signals destined for the vagus nerve and spinal cord. These signals carry ever-more particularized (decompressed) commands to coordinate and regulate the organs, tissues and skeletal muscles of the body. At the same time, efference copies (for example, ref. 149) of these motor control signals function as cortical feedback signals that are increasingly decompressed into probabilistic inferences of higher dimensionality as they cascade towards the primary sensory cortices. Efference copies are the predicted consequences of the motor movements that will result from the motor control signals. The implication is that all categories involve motor plans and movements, even though movements will not always be perceptible to the naked eye.

A category can therefore be described as an ensemble of low-dimensional motor control signals and the feedback signals they give rise to, fashioned from past experiences via Hebbian learning principles[150] (Fig. 3b). A category is predictive because it is assembled before the arrival of the expected feedforward signals. A category is abstract because the signals originating in limbic cortices are necessarily low-dimensional, multimodal summaries, as already reviewed. These signals propagate from the limbic edge and are increasingly decompressed into probabilistic inferences of higher dimensionality. At every neuron, incoming signals of n dimensions expand into an ensemble of signals of x dimensions (recalling that $n < x$). Each n dimension

therefore functions as a more abstract feature of equivalence (with some likelihood of occurrence) for the resulting higher-dimensional *x* signals.

A category, in this view, is a dynamic event, not a fixed, stable cognitive structure. It is constructed in the moment, not retrieved. In a specific situation, a momentary ensemble of lower-dimensional

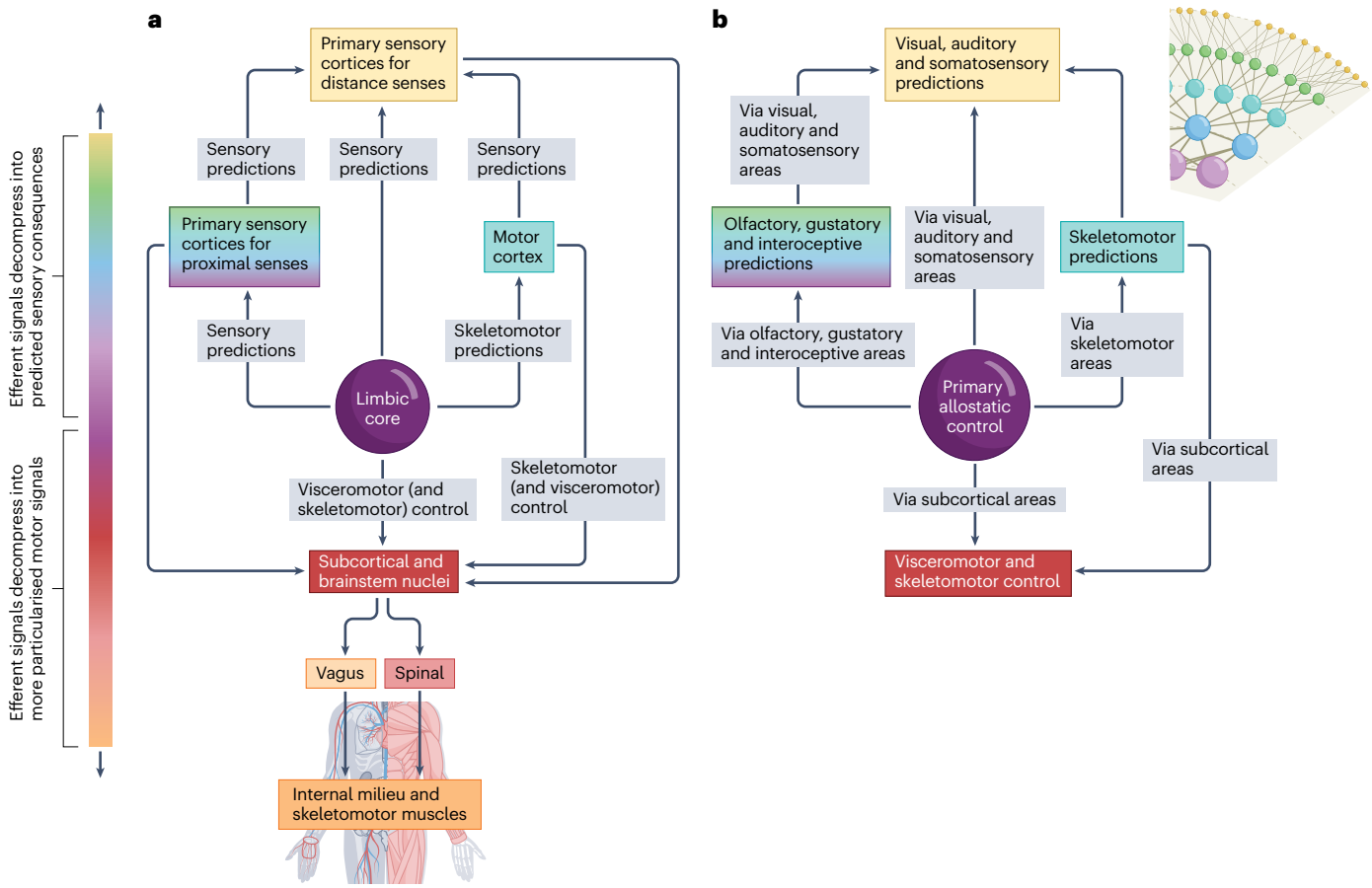


Fig. 3 | Prediction signal flow in category construction. a, A schematic summary of hypothesized prediction signal flow. The coloured shapes depict anatomical areas within the mammalian nervous system and the arrows depict signal flow. The cortical gradient is organized to propagate progressively decompressed signals in two directions simultaneously, as indicated in the long vertical arrow on the left side of the figure running from yellow (top) to orange (bottom). The purple to orange colours indicate the progressive expansion of dimensionality in motor control signals. The purple to yellow colours indicate the progressive expansion of dimensionality in sensory prediction signals. Motor control signals originate in the primary visceromotor regions in the limbic core of the cortex and become increasingly particularized as they descend through subcortical nuclei to the vagus and spinal cord, where they coordinate and control the internal milieu of the body (depicted by the left side of the body) and the skeletal muscles (depicted by the right side of the body), respectively. At the same time, efference copies of motor control signals function as prediction signals when they reach the upper layers of cortical areas of greater laminar differentiation. They progressively decompress (and become increasingly particularized) as they propagate towards the sensory edge of the cortical gradient. Efference copies of visceromotor control signals are received by the upper layers of primary motor cortex (M1) as skeletomotor prediction signals. Efference copies that project to the upper layers of primary sensory areas function as sensory prediction signals. This same signal motif exists throughout the cortical gradient. The pyramidal neurons in layer 5 of M1 send skeletomotor control signals that are progressively expanded in dimensionality as they descend through subcortical nuclei to the vagus and spinal cord. At the same time, efference copies of skeletomotor control signals function as another set of sensory prediction signals that reach the upper layers of primary sensory areas of relatively greater laminar differentiation (that is, primary sensory areas for the distance senses: primary somatosensory cortex (S1), primary auditory cortex (A1) and primary visual cortex (V1)). Pyramidal neurons in layer 5 of the proximal sensory modalities (olfaction (O1), gustation (G1) and interoception (I1)) send signals to subcortical areas (depicted as unlabelled arrows, because their functions have not yet been clarified in the predictive processing framework). Their efference copies are sensory prediction signals that arrive to the upper layers of S1, A1, and V1. **b**, The same figure is relabelled to depict the types of signal involved in category construction during feedback-driven predictive signal flow. The coloured shapes now depict signals and the arrows depict anatomical features. Category construction begins as multimodal, compressed signal summaries that originate in the hippocampus and deep layers of the limbic core of the cerebral cortex. These summaries can be described as abstract (lower-dimensional) features such as reward, threat and so on. A single multimodal summary is decompressed into a grouping of possible future motor movements and their resulting sensations in a specific situation or context, which includes the momentary energetic state of the body. Each possible future has some prior probability of functionality (for example, ref. 250), yet all are equivalent in terms of the abstract (low-dimensional) action concept, goal or plan that they originate from. In this view, a category is a dynamic signal event, constructed as ensemble of low-dimensional motor control signals and the possible allostatic futures they give rise to, fashioned from past experiences. Parts **a** and **b** were modified with permission from Fig. 4 in ref. 32, Elsevier. Similar schematics have been published in refs. 26,251.

signals corresponds to possible future motor movements and their resulting sensations in a specific context – a group of feedback signals that in given situation are functionally equivalent in terms of the even lower-dimensional abstract features from which those signals were inferred. High-dimensional signal arrays, on the other hand, are momentarily equivalent by their relation to signals for lower-dimensional motor control plans in the next moment (that is, an action concept or action tendency) and the corresponding abstract mental features (for example, threat, reward, value and so on). This description holds for so-called concrete categories that are perceptually similar, such as the category CAT. The features of equivalence for the category CAT that make for an ideal house pet will differ from those that make for an ideal mouser or an animal to be admired in a zoo. The description also holds for more abstract versions of the category CAT, such as one that gives rise to a person with a cool demeanour who plays jazz (for example, 'a cool cat'). A category of the same name, such as CAT, can vary with each use in the same manner as constructed memories. Following evidence from the structural model, each category begins as a low-dimensional, abstract visceromotor plan that will support particular skeletomotor movements, such as bending to pet the cat (for example, the mental feature 'approach'), swerving to avoid a cat (for example, the mental feature 'avoid'), saccading to look around in a cage (for example, the mental feature 'explore') or tapping a finger on table while listening to some cool jazz (for example, the mental feature 'enjoy'). Even sitting still requires a motor plan for coordinating visceral organs, blood flow, ATP production, gene expression for potential viruses versus bacteria, contracting some skeletal muscles while relaxing others and so on (for discussion, see ref. 32).

This way of thinking about categories is consistent with decades of research in cognitive science on their contextual nature (for example, refs. 85,151–154), as well as evidence that categories are constructed in the moment and are goal-based, meaning they are extemporaneous groupings created for a particular function in a particular context (reviewed in refs. 25,155–157). A category is something a brain does, not something a brain has.

To summarize thus far, category construction is a brain-wide ensemble of signals corresponding to a pattern of lower-dimensional, abstract features that decompress into multiple patterns of higher-dimensional, particularized features. The features of equivalence for a momentary situated category are grounded not in high-dimensional sensory features but in more compressed summaries. These summaries correspond to the function or the goal that the category serves in that particular situation, rather than a stable function that remains the same across individuals and situations.

Accordingly, any given category construction event can be described as a context-dependent probability distribution of possible instances. It is an exemplar account of categories, of sorts, but without the assumption that the brain is retrieving a distribution of instances from an already existing category. The distribution's prototype, if there is one, would correspond to the ideal instance of the category in that particular situation[158], even if the instance never existed in reality and must therefore be inferred (for example, ref. 159). The pattern of neural signals that construct a category, which are distributed across the brain and give rise to these varying features, can themselves vary from situation to situation, instance to instance, a phenomenon known as degeneracy or multiple realizability[160,161]. Category knowledge about cats, for example, would not be described as a single distribution of instances with a family resemblance to a single prototypical instance, but as a distribution of spatiotemporally varying CAT categories. Each momentary CAT category will have its own prototype, depending on the goal that CAT is serving in a given situation (as a pet, a mouser, a zoo animal or a person with a cool demeanour); in effect, a distribution of situated CAT prototypes (rather than a single, unchanging prototype). Across a person's lifetime, they will construct a population of possible CAT categories (or a distribution of categories), each with graded similarity in their features. Think of this as a vocabulary of CAT categories. Further, in principle, the vocabulary of momentary CAT categories (with graded distributions) can vary across people, particularly if those people come from different backgrounds with different opportunities to culturally inherit variable knowledge about cats (for example, in some cultures, cats are food to be eaten, whereas in others they are deities to be worshipped).

In category learning, signals are hypothesized to flow (and compress) in the feedforward direction. For example, dimensionality reduction has been directly observed with category learning, as cortical networks shift from higher-dimensional sensory signals to lower-dimensional, task-related ensembles[162]. Neurons that are relatively closer to the limbic edge of the gradient begin to show low-dimensionality selectivity for categories, with less sensitivity to high-dimensional sensory details[163–167]. Dimensionality reduction can also manifest as 'stretching' of task-relevant dimensions to expand their representational space[168]. Notably, this effect is more prominent in areas that are closer to the limbic end of the cortical gradient (Zhang et al. sampled various locations across the cerebral cortex, including lateral prefrontal cortex, V4, MT, frontal eye fields, lateral parietal cortex and inferotemporal cortex; limbic cortical areas were not explicitly studied. The most prominent stretching was observed in lateral prefrontal cortex, where the amount of dimensionality reduction is considerable[168]).

As in other predictive processing accounts, feedforward signals are hypothesized to function as teaching signals that reinforce or modify future motor control plans. A brain learns to more optimally issue motor commands to control future sensory events (including, most importantly, the sensory events that inform on metabolic conditions of the body, as we discuss in the next section). For any given neuron, feedforward signals confirm or constrain feedback signals. This account is assumed to hold for unexpected feedforward signals (so-called positive prediction errors) as well as for expected feedback signals that fail to materialize (so-called negative prediction errors; for example, refs. 169–171). Any difference between a compressed summary at a given point along the cortical gradient and the neural context at that point would pass back towards the limbic edge as a positive prediction error (presence of feedback signals that are not confirmed). In effect, feedforward signals help to select one momentary category exemplar from many.

Another source of selection comes from attentional signals, also called precision signals, that are hypothesized to adjust the strength and reliability of prediction signals and prediction error signals (refs. 172,173). Prediction signals are thought to be weighted according to their anticipated value to explain incoming sensory signals, which is equivalent to their prior probability[25,172]. Prediction error signals are weighted by their salience, which is equivalent to the predicted value of the allostatic information those errors will provide[25,174]. Via this selection, one of many possible sensory and motor futures is transformed into the present.

Importantly, feedforward signals that continue to propagate towards the limbic edge of the cortex (as prediction errors) remain uncategorized and have the potential to modify the originating multimodal control signals. Hence our hypothesis that some feedforward

signals (those that are prediction errors) are important for category learning. Prediction errors can result from a local mismatch between feedback (expectation) and feedforward (input) at each neuron (for example, ref. 175) or can be computed in relatively less-laminated cortical areas and broadcast to relatively more laminated cortical areas (for example, ref. 176; notably, in this study, conceptual representations (word associations) were used to generate sensory predictions). Consistent with this account of prediction error, top-down feedback signals act as inhibitory filters[177,178] that suppress excitatory feedforward signals that match them[140,145]. Prediction errors, therefore, result from a mismatch between the excitatory (feedforward) and inhibitory (feedback) signals. This account is also consistent with recent computational evidence suggesting that predictive learning (prediction signals supervising the encoding of prediction errors) creates low-dimensional representations[179].

The inhibitory function of feedback signals comes from direct evidence of cortical rhythms that is consistent with our predictive account of categorization. Cortical rhythms are oscillations or repetitive patterns of electrical activity produced by neurons corresponding to cycles of depolarization and hyperpolarization driven by electric field fluctuations. They work through a non-synaptic form of neural communication, called ephaptic coupling, in which the electrical fields surrounding neurons influence the degree of excitability of other neurons[180]. Dendrites may play an important role in generating these electrical fields (and may also take advantage of them for their own electrical communication). This more direct electrical communication helps to entrain neuronal spiking (and the chemical communication between neurons at synapses). As a result, millions of neurons synchronize and coordinate their activity in a manner that helps to achieve changes in dimensionality (expansion and reduction)[80,181–185]. Gamma oscillations (>30 Hz) are faster, higher dimensional and strongest in the upper layers of cortex that originate feedforward prediction error signals[140,178]. They are closely linked to sensory-driven spiking and exert fine-scale (microscale) influence. By contrast, slower alpha/beta (12–30 Hz) oscillations have a broader (mesoscale) influence and support low-dimensional organization[177,186]. They tend to be strongest in the deep layers of cortex that originate feedback signals[140,178,187], which are tied to current context and goals[178] that we have described as category construction. It has been hypothesized that alpha and/or beta oscillations function as inhibitory patterns that constrain gamma-linked spiking, acting like 'stencils' that permit local expression of content where alpha/beta is suppressed[177]. Each stencil corresponds to a distinct task operation. This dynamic mirrors observed associations: alpha and/or beta oscillations are associated with feedback signalling that supports predictive processing as a functional consequence of dimensionality reduction, whereas gamma oscillations are associated with feedforward signalling[178,188]. More generally, neural oscillations are thought to play a role in the dimensionality expansion and reduction that we are describing as categorization (for example, ref. 80).

Our predictive account of categorization may also explain the mixed selectivity of neurons whose responses vary with context. All neurons show some contextual variation in their receptive fields, as noted above. Alpha and/or beta oscillatory signals, whose power increases with top-down demand[167], may create the neural context that sculpts spiking accordingly[42,111,113,189].

Furthermore, the landscape of inhibitory cortical neurons is consistent with these rhythm-related observations and with our predictive account of categorization more generally. The faster-spiking, parvalbumin-expressing inhibitory neurons that contribute relatively more to gamma oscillations[94,190,191] are relatively more concentrated at the sensory end of the fewer-to-many cortical gradient[32]. These cells are cytoarchitecturally positioned to control the timing and frequency of action potentials in pyramidal cells (located near cell bodies and spike initiation zones (ref. 192 and references therein); they also modulate the stellate cells that receive thalamic signals. Somatostatin-expressing inhibitory neurons, by contrast, are slower spiking and are relatively more concentrated at the limbic end of the gradient. They preferentially form synapses on dendrites and are positioned to gate and modulate incoming excitatory inputs (ref. 192 and references therein) in a manner that shapes the neural context of pyramidal and stellate activity. They are not as driven by feedforward input when compared with parvalbumin-expressing inhibitory neurons (for example, ref. 94). This relative difference in inhibitory neuron distribution is associated with overall differences in excitatory–inhibitory balance along the main cortical gradient. Limbic areas have relatively more excitatory receptors and show relatively more excitation (prone to a default level of excitation), whereas sensory areas have relatively more inhibitory receptors and show relatively more inhibition[193]. In these ways, predictive category signals shape the processing of incoming signals to the cortex.

There are many consequences of considering all categories to be functional and goal-based events, constructed in the moment. They range from the practical to the scientific and philosophical. Most immediately, this view helps to explain how different ensembles of physical signals can have the same meaning within a context and across contexts, boosting generalizability and conferring considerable contextual flexibility in responding. In everyday life, for example, cats are encountered in situations, never in isolation. A category of cats that make ideal pets might be soft, cuddly and purr softly when stroked. Cats who are successful mousers will be agile and quick, perhaps a bit aggressive. Cats who draw visitors to a zoo will be large and majestic. Cool cats, regardless of their biological species, will be aloof, a touch dismissive and maybe great musicians; and so on. This view also explains how a single physical signal or pattern of signals can have different meanings in different situations (for example, being soft and cuddly, purring, agile, aggressive and so on is not specific to cats). The value of this approach is perhaps most apparent when attempting to understand how people learn, construct, and use categories that have tremendous within-group variation and across-group similarity that cannot be handled by traditional views of categorization, such as emotion categories (for discussion and review of evidence, see refs. 81,194).

These considerations suggests that a single prototype for CAT is actually more like a stereotype that masks substantial structured variation in the meaning and function of the category, a position that is consistent with the population thinking introduced by Charles Darwin in *On the Origin of Species* (see ref. 195). It also implies that the category CAT is not something 'out there' in some mind-independent world, referring to all possible past and future instances of the word 'cat'. Instead, the human brain creates a category for CAT in relation to a specific context. The features of similarity are linked to the function the category serves in that specific instance. Function and features of equivalence are linked: the features of equivalence change depending on the requirements of the situation, because the function of the category changes. The most important similarities that form any specific CAT category may be not physical but functional.

Summary

The account offered thus far overturns the conventional understanding of categorization. The cerebral cortex does not reduce dimensionality after the fact to match feedforward signals to the memory of a stored category representation. Instead, the cortex uses low-dimensional temporal dynamics as anticipatory categories that shape the neural context for processing feedforward signals. Categorization begins with ad hoc groupings of possible futures and ends when predictive feedback constrains feedforward signals to produce specific actions and perceptions.

When predictions fail, residual feedforward signals drive learning, compressing towards the limbic edge of the cerebral cortex. A key insight is that visceromotor preparation, and the skeletomotor preparation that it supports, precede perception and are integral to the categories being constructed. This anatomical reality places allostasis — predictive regulation of the body — at the core of the neural context and feedback control. Cortical architecture may support abstract allostatic plans as equivalence features, enabling metabolically efficient generalization and action planning.

Allostasis and energy optimization at the core of categorization

We have established that feedforward sensory signals, when predicted and constrained by feedback signals, are categorized and become meaningful in relation to those signals. Meaning, therefore, is not defined as what a signal is but by what it does when interacting with other signals. The primary meaning of incoming signals, as well as the actions and features of sensation and perception that they contribute to, is rooted in their interaction with prediction signals that ultimately arise in limbic cortices (Fig. 3b and Box 2). This places allostasis and efficient energy regulation at the functional core of categorization and the meaning that derives from it.

Allostasis is not a reactive condition of the body, but a phenomenon in which the brain actively coordinates and regulates the systems of the body according to energetic costs and benefits. The "core task of all brains… is to regulate the organism's internal milieu… by anticipating needs and preparing to satisfy them before they arise."[2]. Survival, growth, reproduction and the energetic demands of living a life (not to mention information transmission across generations via genes and cultural inheritance) require an animal to manage the continual intake and expenditure of biological resources[196]. Metabolic and other energetic expenditures are required to acquire resources in the first place, to learn about where resources are and how to get them, and to plan and execute the physical movements necessary (while protecting against threats and dangers along the way). An animal thrives when it has sufficient resources to explore the world. Experience consolidated within the brain's dendritic and synaptic connections makes those experiences available to predictively guide later decisions about future energy acquisition and expenditures.

Consider that every animal body has multiple internal systems to coordinate and regulate: systems to shuttle nutrients and eliminate waste products, systems to regulate microbiota and systems for regulating adaptive immunity, as well as systems to sense the world surrounding the body and secure resources from it. At the same time, animals must make efficient use of those resources, which requires that they balance trade offs[196]. The brain's allostatic efforts achieve this balancing act: maintaining vital functions (for example, ion gradients in cells, cell division, telomere repair and cell apoptosis), both during moments of rest and during moments when large metabolic outlays are anticipated (that is, during 'stress'[197]). The brain's allostatic efforts ensure that resources are used as efficiently as possible (at the level of the whole organism) regardless of the level of metabolic output required (for example, when sleeping, running, learning and so on). Individual biological systems may use the more familiar homeostatic regulation (reactive regulation to perturbations that return a system back to an optimal set point or range of points around which a system operates). As a collective, however, the many systems that make up a body are coordinated and regulated by the broader allostatic efforts of the brain.

Brain imaging evidence has identified an intrinsic, distributed allostatic network in the brain, anchored in the cortical areas along the limbic edge plus their subcortical connections (Fig. 4). This network largely replicates in humans the tract-tracing 'ground-truth' connectivity identified in non-human animal studies[29,105] and reliably includes both cortical regions and subcortical nuclei. The topography of this network overlaps with the central autonomic network, the skeletomotor system, the salience and/or cingulo-opercular network, the somato-cognitive network and the default network[32]. Areas in this network are routinely implicated in a range of perceptual, cognitive and emotional phenomena[29]. Atrophy and dysfunction within this broader allostatic system have been documented for mental and physical illness[198–202], involving profound metabolic disruptions, including in obesity, diabetes, depression, anxiety, addiction, chronic pain and chronic stress (for discussion and references therein, see ref. 32).

In addition, it is now well established that allostasis-related signals broadly shape what was formerly considered brain-wide 'intrinsic' or 'spontaneous' signalling[56,203] (for discussion, see ref. 32). Evidence suggests that these signals are responsible for the most prominent spatiotemporal patterns observed in the 'resting state' blood oxygen level-dependent (BOLD) signal, such as the now well-known intrinsic networks and functional connectivity gradients[204–206]. Such signalling can be understood as related to the feedback (predictive category construction) signals that provide the neural context for processing feedforward signals (for example, refs. 141,207).

With this allostatic foundation in mind, our predictive processing account of categorization builds on traditional predictive processing work in three ways. First, the goal of prediction may not be error minimization per se (for example, refs. 136,208), but optimizing the metabolic cost of living (including signal processing within the brain)[209]. As a general rule, a system optimizes cost by predicting and correcting instead of reacting (for example, refs. 2,210–213). Networks trained to anticipate incoming signals develop predictive architectures[214] and algorithms to predict incoming signals minimize the cost of signal processing[142].

Categorization, therefore, can be understood as a means by which a brain maintains allostasis, or how it anticipates the energetic needs of its body in a specific spatiotemporal context by generalizing from allostatically similar past events and preparing to meet those energetic needs before they arise. To this end, the brain also predictively models the sensory conditions in the body, a process called interoception[215], which is part of allostasis because it supports the motor control of the body's organs and tissues (visceromotor control) in the same manner that somatosensation supports skeletomotor control.

Second, other predictive processing accounts often rely on the familiar concept of homeostasis (regulation in response to perturbations that return a system to its optimal set point or range of points in which it operates) or define allostasis in a way that is similar to homeostasis. In our view, allostasis, which is characterized as 'stability through

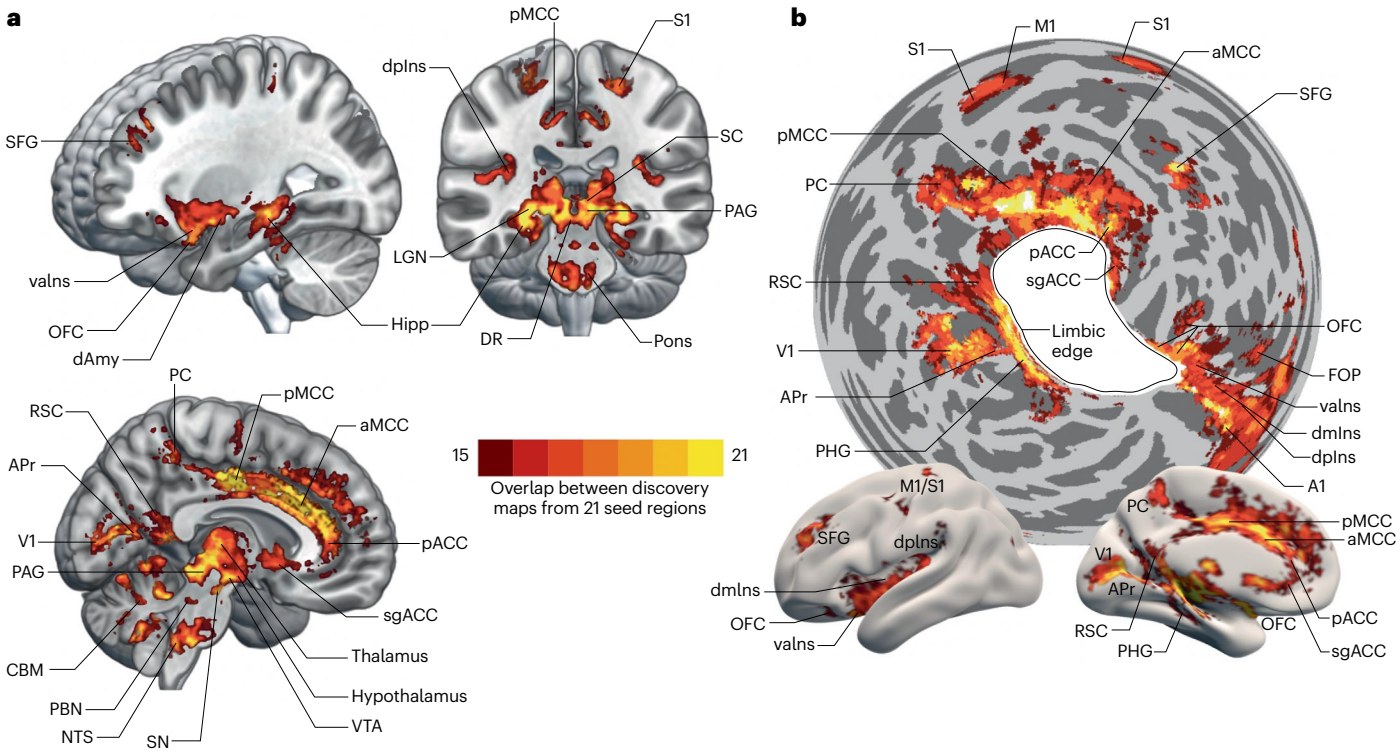


Fig. 4 | An intrinsic brain network for allostasis. a, The topography of the distributed, intrinsic allostatic network depicted in volumetric space. This is a conjunction map (a combination of multiple, independently produced statistical maps that are superimposed on one another). The colours depict the number of binarized whole-brain connectivity maps from each seed region of interest (ROI) ($p < 0.05$), which shared overlapping regions (ranging from 15 to 21, reflecting the total number of cortical and subcortical ROI seeds used in the analysis). A bootstrapping strategy (1,000 samples) was used to identify weak but reliable correlations, avoiding type II errors caused by stringent statistical thresholds; this was important particularly for identifying connections involving subcortical ROIs. Intrinsic functional connectivity was estimated using ultrahigh-resolution brain imaging at 7 Tesla ($N = 90$), which replicated an earlier study at 3 Tesla ($N = 550$) (ref. 29) but with improved subcortical nucleus localization afforded by higher spatial resolution (1.1 mm isotropic), better signal-to-noise-ratio, and the use of in vivo brainstem and diencephalic nuclei atlases. **b**, The allostatic system projected onto an azimuthal flat map representation of the cortical surface (left hemisphere). A1, primary auditory cortex; aMCC, anterior mid cingulate cortex; APr, area prostriata; CBM, cerebellum; daIns, dorsal anterior insula; dAmy, dorsal amygdala; dmIns, dorsal mid insula; dpIns, dorsal posterior insula; DR, dorsal raphe; FOP, frontal operculum; hippo, hippocampus (dentate gyrus and cornu ammonis); LGN, lateral geniculate nucleus; M1, primary motor cortex; mdThal, mediodorsal thalamus; MGN, medial geniculate nucleus; mvAIns, medial ventral anterior insula; NAcc, nucleus accumbens; NTS, nucleus tractus solitarius; OFC, orbital frontal cortex; pACC, pregenual anterior cingulate cortex; PAG, periaqueductal grey; PBN, parabrachial nucleus; PC, precuneus; PHG, parahippocampal gyrus; Pf, parafascicular nucleus; pMCC, posterior mid cingulate cortex; RSC, retrosplenial cortex; S1, primary somatosensory cortex; SC, superior colliculus; SFG, superior frontal gyrus; sgACC, subgenual anterior cingulate cortex; SN, substantia nigra; V1, primary visual cortex; vaIns, ventral anterior insula; VTA, ventral tegmental area. Part **a** was adapted with permission from Fig. 1c in ref. 32, Elsevier. Part **b** was adapted with permission from Fig. 1e in ref. 32, Elsevier.

change' is computationally distinct from resisting perturbations (homeostasis). In addition, homeostasis is usually reactive, whereas allostasis is predictive.

Third, traditional views of predictive processing suggest that sensory prediction signals are efference copies of skeletomotor control signals (for example, refs. 139,216,217). Others have touched on the importance of visceromotor control, but our proposal belongs to a handful of efforts that emphasize it as central (for example, refs. 218–220). Furthermore, we consider the inter-relations between visceromotor and skeletomotor systems rather than treating them in isolation. Visceromotor control is necessary for and biologically intertwined with the control of skeletomotor action. Collateral axons from deep pyramidal neurons, which we referred to earlier as efference copies of the motor control signals, project to the pyramidal neurons in the superficial layers of more laminated cortical columns as feedback or prediction signals (Box 2). One implication of this architecture is that primary motor cortex (M1), which has a more developed laminar structure than limbic cortices, receives skeletomotor prediction signals as efference copies of visceromotor control signals from the limbic edge of the cortical gradient (Figs. 1 and 3; also refs. 26,147 and references therein). These signals may help to explain the maps of the adrenal glands and gut that have been identified in M1 of the primate brain[221,222]. A reasonable avenue for further study is the hypothesis that allostatic control signals are part of the ethological action maps that have been identified in M1 when neurons are stimulated at more ecologically valid durations[69]. Another implication is that the two motor systems must be coordinated. Anything otherwise would be a metabolic drag on the system, which could result in disease. For example, depression is a disorder associated with profound metabolic disruptions that may result, in part, from a relative

Glossary

Abstract
As used here, to 'abstract' means to generalize across different high-dimensional patterns of sensory features and motor movements to the same lower-dimensional mental feature (for example, 'threat' or 'reward'). Abstract features are efficient summaries of many more signals from many modalities (multimodal compressed summaries). They are not without modality (amodal). An abstract feature is never directly sensed and is not directly measurable by physical means; it is always created from physical features that are themselves sensed and can be measured (that is, from concrete mental features).

Binding
A hypothetical process of integrating separate, distributed features into a coherent whole.

Category
A grouping of individual occurrences, such as events or objects, that are similar enough to be interchangeable or equivalent for some function or use in some context.

Category construction
The process of creating extemporaneous groupings of occurrences, such as patterns of prediction signals, based on perceived similarities or uses in a given context, ignoring differences that are irrelevant.

Cytoarchitectural gradient
A gradual, continuous transition in the structure, density and connectivity of cells.

Dimensionality
The number of independent features, properties or variables needed to describe something, such as an object or event.

Distance senses
Sensory modalities that report on conditions of the world that are external to and some distance from the body (vision, audition and aspects of somatosensation that derive from the same lateral-line system as audition, and function as a distance sense in the water where vertebrates first evolved). The primary sensory cortices for distance senses display nearest-neighbour, array-to-array mapping (for example, the neurons in V1 correspond topographically to receptors in the retina). A considerable amount of dimensionality reduction in the distance senses occurs in the cerebral cortex after those signals are received by primary sensory cortices (when compared with the more proximal senses, whose signals arrive at their primary sensory areas in the cerebral cortex already largely compressed).

Engineered systems
Combinations of components deliberately and purposefully designed by humans that together perform specific, useful functions.

Feature of equivalence
A feature that renders a group of instances similar enough to be equivalent for some function or use in some context.

Feedback signals
Signals that propagate away from the limbic edge of the cerebral cortex towards the sensory edge (known as centrifugal, or 'moving away from the centre'). Also known as 'top-down' signals. The original meaning of 'feedback' comes from engineering and systems theory to mean signals that merely regulate or modify a process.

Feedforward signals
Signals that flow towards the limbic edge of the cerebral cortex (known as centripetal, or 'moving towards the centre'). Also known as 'bottom-up' signals, although, strictly speaking, signals arriving at the cerebral cortex from subcortical areas have already been integrated with or affected by 'top-down' signals originating in the cortex (for example, afferent signals from the sensory surfaces meet efferent motor control signals in the vagus nerve, nuclei within the brainstem, the superior colliculus, the hypothalamus and the thalamus, meaning that feedforward signals, when they arrive at cortex, are not purely bottom-up).

Internal milieu
A term, originally coined by the physiologist Claude Bernard in 1878, that refers to blood and other tissues that maintain a constant internal environment inside the body, even in the face of external perturbations. In modern usage, 'internal milieu' is a general term that refers to organs and tissues of the viscera that are coordinated and regulated by the autonomic nervous system, the immune system and the endocrine system.

Labelled line
A hypothesized dedicated neural pathway between a sensory receptor and a neuron in the brain, such that the receptor fixes the neuron's receptive field. Activity in the pathway would have an intrinsic meaning that corresponds to one specific quality or feature coded by the receptor.

Limbic core
See 'Limbic edge'.

Limbic edge
The word 'limbic' in Latin means 'border', 'edge' or 'hem'. The term was originally used by the anatomist Paul Broca in 1878 ('le grand lobe limbique') to refer to the parts of the cerebral cortex that form a border or ring around subcortical nuclei. In this Perspective, 'limbic edge' refers to the hippocampus, subiculum, presubiculum, parasubiculum, medial entorhinal cortex, posterior orbitofrontal cortex, ventral anterior insula, and portions of cingulate, retrosplenial, parahippocampal and perirhinal cortices. These regions are architecturally allocortical, agranular (periallocortical) or dysgranular (proisocortical) and have monosynaptic connections to midbrain and brainstem nuclei that control the viscera and tissues of the body. This is not to be confused with the now discredited 'limbic system' concept, which hypothesizes a set of brain structures devoted to emotion and motivation.

Mixed selectivity
A term used to describe individual neurons that have variable receptive fields and fire in relation to different signals, corresponding to different features. The meaning of any action potential in these neurons is relational and varies with context.

Precision signals
Signals that adjust the strength and reliability of prediction and prediction error signals. Precision signals function as attention signals.

Prediction signals
Memory-based patterns of feedback signals that forecast upcoming motor movements or anticipated feedforward signals.

Prediction error signals
Discrepancies between expected (feedback) signals and incoming (feedforward) signals that function as teaching signals in the brain.

Glossary (continued)

Proximal senses
Sensory modalities that report on conditions in or near the body (for example, olfaction, gustation and interoception). The primary sensory areas for proximal senses are located in or near the limbic edge of the cerebral cortex. For example, primary olfactory cortex (or piriform cortex, O1) is located in the agranular anterior insula, primary gustatory cortex (G1) is located in the dysgranular mid insula, and the dorsal mid portion of primary interoceptive cortex (I1) is in the dysgranular (the other portion of I1, which resides in dorsal posterior insula, is fully granular in structure). Unlike the primary sensory cortices for distance senses (vision, audition and aspects of somatosensation), O1, G1 and I1 do not display nearest-neighbour, array-to-array mapping (for example, primary olfactory cortex does not contain a spatial map of olfactory receptors in the nasal epithelium). Much of the spatial dimensionality reduction and some temporal dimensionality reduction in these more proximal senses occurs before those signals reach their respective primary sensory areas in the cerebral cortex. Some aspects of somatosensation, such as proprioception (joint, limb or body position), may in fact be a proximal sense.

Rich club
A set of highly interconnected nodes with a high degree of mutual connectivity forming a central backbone of the connectome that is important for brain-wide neural synchrony.

Stochastic noise
Random fluctuations governed by a probability distribution (for example, Gaussian or Poisson). Outcomes vary even under the exact same conditions.

Visceromotor
The body's organs and tissues are referred to as 'viscera'. Control of these tissues is called visceromotor control. Visceromotor control areas in the cerebral cortex traditionally include the subgenual and pregenual anterior cingulate cortex, anterior mid cingulate cortex, posterior orbitofrontal cortex and ventral anterior insula. Other areas not typically considered important to visceromotor control, but that have direct monosynaptic connections to midbrain and brainstem nuclei that are involved in regulating the viscera, include the ventral portions of premotor cortex, entorhinal cortex, the hippocampus (dentate gyrus and CA) and subiculum.

'disconnection' between the two motor systems. Intractable depression is associated with atrophy of the cingulum bundle, which connects primary visceromotor regions in the anterior cingulate with premotor regions in the mid and posterior cingulate (for example, ref. 223). Deep-brain stimulation just anterior to the subgenual anterior cingulate results in repair of the cingulum bundle by oligodendrocytes[224].

Correspondingly, the abstract 'action concepts', as discussed above, can in fact be understood as representing the causal relationship[225] between visceromotor and skeletomotor actions and their sensory consequences. In effect, feedback prediction signals, as candidates for categorizing incoming feedforward inputs, make those inputs meaningful in terms of energetics — in terms of the visceromotor movements and the skeletomotor movements that they support, as well as any acquisition of new resources (for example, 'rewards'). On this interpretation, prediction error salience is a function of predicted allostatic value. A brain establishes whether and how to spend energy resources to learn any unanticipated sensory inputs (sensory prediction errors) to improve its predictive efficacy according to their anticipated future impact on energy regulation.

Summary

Metabolism is meaning. High-dimensional feedforward sensory signals are rendered equivalent in terms of energy regulation and action. The basic operating principle of the brain — categorization — grounds meaning in allostatic control and efficient energy regulation across different levels of metabolic output. A 'goal' is a future (predicted) allostatic state of the system. 'Motivation' is any energy expended to achieve that goal. Categorization, by way of predictive categories, constrains and reduces the complexity and uncertainty of incoming sensory signals, thereby reducing costly uncertainty while supporting metabolic efficiency in action planning. A brain learns to construct categories, not to reduce prediction error per se, but to more optimally use motor commands to control future sensory events according to their allostatic cost, which the brain continually predicts and compares to the incoming viscerosensory signals that report on the energetic conditions of the body. Consequently, categorization allows an animal to continually act on and modify its niche — the part of the world that is relevant to energy regulation and survival[226]. Metabolic efficiency, by way of effective categorization, translates into physical health and mental health, whereas inefficiency translates into illness. Metabolic efficiency also increases evolutionary fitness, both by increasing surplus energy available for mating, reproduction and caring for offspring and by decreasing the frequency with which an organism must seek nutrients and expose itself to predators[4,209,227].

Conclusion

All animals categorize[228]. Even single-cell organisms can be said to categorize as they generalize from a varied past to predict and act in the service of allostasis in the spatiotemporal present[229]. However, the world does not come pre-sorted into categories. Categorizing may not involve detecting regularities post hoc but instead involves constructing similarities from differences. In this Perspective, we have integrated phenomena at different levels of analysis and timescales to offer several provocative proposals: category construction, categorization and category learning arise from different combinations of signal compression and decompression in a brain. Category construction occurs predictively at a slower timescale and involves the decompression of feedback signals to generate the motor control signals and sensory predictions signals that guide inference. Categorization occurs at a faster timescale when feedforward signals encounter the neural context created by category construction. Category learning occurs at an even slower timescale as prediction errors are consolidated and made available for later category construction.

A category, therefore, may not be a representation that an animal has, but a signal processing event that an animal does, predictively, to constrain the meaning of a high-dimensional ensemble of incoming signals in a particular situation. Categorization renders these signals meaningful — similar to one another and to past allostatic events — in terms of some goal or function. Categories guide allostasis and action by anticipating incoming sensory signals and giving them meaning in terms of energy expenditure in the process of creating lived experience (whatever that means for the animal in question).

We placed allostasis, the predictive coordination and regulation of the body, at the core of categorization, on the basis of the anatomical,

physiological and brain imaging evidence we reviewed. This conceptual manoeuvre suggests that, fundamentally, a brain groups feedforward signals into equivalence classes and gives them meaning for the purpose of optimal energy regulation at the organismic level.

Taken together, these hypotheses and the inferences they give rise to offer the beginnings of a coherent neurocomputational framework, in line with a whole-brain approach to brain mind mapping (for example, ref. 230). Our proposal is consistent with broad swaths of empirical evidence at different levels of analysis from different disciplines, some of which we covered in this review. Like any novel framework, however, our proposal is not definitive. Further experimentation is required to test, refine or reject the various hypotheses outlined here (see Box 3 for examples of outstanding questions and future directions). Two simple but important course corrections in the literature would be (1) to extend investigations of the cerebral cortex all the way to the limbic edge (rather than stopping at lateral prefrontal cortex) and (2) to routinely include visceromotor measurements in studies of the mind, including studies that focus on the skeletomotor system (and vice versa). The mark of any new scientific idea is not necessarily which existing questions it answers but which new questions it offers.

More generally, our proposal suggests that sensation, perception, memory, action, decision making, goal-directed behaviour and so on may not arise from their own dedicated signal processing motifs but instead may be facets or mental features of each category event, arising from a common signal processing framework with allostasis at its core[32]. This broader view of categorization as a basic operating principle of the brain also has the potential to unite different scientific constructs for meaning making, including decision making, cognitive mapping, explanation and appraisal. Categories, constructed as cascades of decompressing feedback or prediction signals, can describe reference signals (for example, ref. 231), forward models or simulators (in a control theoretic sense) for the signals of relatively higher dimensionality that code more particularized features (for example, ref. 232). Predictively constructed categories, as outlined here, can also describe what it means for a brain to have generative, internal models (for example, ref. 233); to create perceptual simulations or perceptual inferences (for example, ref. 234), top-down inferences (for example, ref. 235), Bayesian inferences (for example, ref. 213), causal inferences (for example, ref. 225), latent cause inferences (for example, ref. 236) or policies (in reinforcement learning); to perform conceptual combination (for example, ref. 237); or simply to remember (supported by Hebbian learning, not declarative memory specifically[238,239]). Integrating these concepts with one another and with their metabolic costs and consequences could result in increased inferential power and usable (justified) scientific knowledge from what are currently separate scientific domains.

Box 3 | A sample of outstanding questions and future directions

1. One of the core postulates is that category construction originates in limbic cortices, initiating a neural context for the compression and ultimate categorization of feedforward signals. For example, existing anatomical studies indicate that there are long-range connections between limbic cortices, such as the anterior cingulate cortex (ACC), and primary sensory regions, such as primary visual cortex (V1)[271], as well as evidence that the ACC sends prediction signals to V1[272]. Given this evidence, and the fact that the ACC is a primary visceromotor region and a premotor association region in the skeletomotor system, we would hypothesize that the ACC routinely sends visual prediction signals to V1, and this does appear to be the case. These signals appear to be the source of neural firing in V1 after retinal lesions and subsequent visual deprivation[273]. More studies like this, with greater spatial and temporal specificity at the single-neuron level, as well as at populational levels of analysis, are required. It would also be important to examine how such dynamics allow for rapid categorization anywhere along the cortical hierarchy, reconciling evidence for rapid (~100 ms) category-selective responses in ventral temporal cortex (for example, as discussed in ref. 274).
2. Recent evidence suggests that feedback signals tend to synapse at the tufts or apical ends of cortical pyramidal dendrites (primarily in the upper layers and layer 5), whereas feedforward signals tend to synapse closer to the cell body (on basal dendrites)[275]. The apical dendrites have been observed to generate electrical spikes that control the responsiveness of basal dendrites and the cell body[276]. If this dynamic continues to be observed, it may imply that feedback signals are in a better spatial position to influence feedforward signals than vice versa, via the complex signal processing that occurs at dendrites (also see ref. 277).
3. Neurons in the superior colliculus, a multimodal subcortical structure that is important for vision, somatosensation and interoception, also contain signals for abstract sensory categories, underscoring how widespread categorization is in the brain[278]. How do such findings enrich or challenge the proposal outlined here?
4. Is spontaneous signalling in the brain synonymous with category construction? Existing evidence suggests so[218,219]. A more detailed analysis is warranted.
5. The idea that continuous category construction is a fundamental operating principle of the brain would benefit from more explicit mathematical testing. For example, would a computational model of continuous category construction be consistent with the notion of recursive neural programs (for example, ref. 279)?
6. If limbic hubs initiate category construction and play a substantive role in shaping the categorization of feedforward signals, then characteristic patterns of impairment, such as reduced abstraction, reduced granularity, diminished flexibility or increased reliance on sensory details, may correspond to damage or structural variation in these regions. This broad hypothesis is consistent with existing evidence but requires more specific testing.
7. How might computational or network-based models of categorization be useful in testing the hypotheses offered here? For example, could a network-based modelling approach empirically compare our neural context hypothesis of category construction against more traditional prototype or exemplar-based theories? Might our prediction error-based hypothesis for category learning be compared with dual-process theories of categorization?

In addition, it is possible to speculate that neuropsychiatric disorders, neurodegeneration, other brain dysfunctions and even neurodiversity could be understood in terms of the abstractness, granularity or situatedness of a person's categories, with consequences for the flexibility and generativity of categorization and therefore for action, experience and energy efficiency. As we suggested earlier in this Perspective, a brain that primarily relies on lower-dimensionality, abstract features of equivalence such as 'negativity', 'threat' or 'reward', for example, may be a brain that risks metabolic deficits from overgeneralization, context insensitivity, experiences that are low in granularity and poor episodic memory. This describes major depressive illness[240,241]. By contrast, an over-reliance on higher-dimensionality sensory and motor features of equivalence to the exclusion of lower-dimensional, multimodal abstractions risks metabolic deficits from an inability to sufficiently generalize. The result will be too much sensory complexity, persistent uncertainty and overly granular, inflexible responses. Such may be the case in some cases of neurodivergence, such as autism spectrum disorder (discussed in ref. 242). Both major depressive disorder and autism are associated with profound problems with energy regulation.

The few-to-many cortical gradient discussed here also aligns with various evolutionary and developmental changes in the vertebrate brain, including cortical expansion, allometric scaling across species, increased cerebral metabolism in the upper layers of the cerebral cortex and modified excitatory/inhibitory balance (see refs. 27,243 for discussion and references therein). As a consequence, what may have changed across evolution and what may differ in vertebrate species is the degree of abstraction that a brain is capable of constructing because of general brain-scaling functions, metabolic adaptations in the brain and, correspondingly, the signals available in an animal's body and its niche (see refs. 67,242 and references therein). Compared with other vertebrates, humans live in an extended spatiotemporal world, making allostasis more challenging. The human brain's capacity for category construction must be equal to the task. It can create highly abstract, functional similarities from broad arrays of spatiotemporal differences. Accordingly, the human brain has expanded association cortices in the frontal lobes, parietal cortex and inferotemporal cortex when compared with other primates, including other great apes (for example, ref. 244). There are also changes in metabolic costs of brain function (for example, ref. 245), particularly in the upper layers of the cerebral cortex (for example, refs. 246–248). Brain expansion and metabolic change potentially enable the increased signal compression and dimensionality reduction afforded in the human brain, creating the opportunity for categories of greater abstraction. Together, this expansion and the enhanced abstraction create the possibilities for human activities not available to other animals, such as the creation of social reality[249] and the practice of science (as discussed in ref. 81). Via these activities, human brains collectively act on the world as a way of predictively regulating their own bodies and the bodies of others and then experience the world as it is relevant for that process. Evolution, in this sense, has conferred the capacity for humans to spatially and temporally expand the energetic control of the human body ever further into the world.

Published online: 13 April 2026

References

1. Sterling, P. Allostasis: a model of predictive regulation. *Physiol. Behav.* **106**, 5–15 (2012).
2. Sterling, P. & Laughlin, S. *Principles of Neural Design* (MIT Press, 2015).
3. Niven, J. E. & Laughlin, S. B. Energy limitation as a selective pressure on the evolution of sensory systems. *J. Exp. Biol.* **211**, 1792–1804 (2008).
4. Pontzer, H. Energy expenditure in humans and other primates: a new synthesis. *Annu. Rev. Anthropol.* **44**, 169–187 (2015).
5. White, O., Babič, J., Trenado, C., Johannsen, L. & Goswami, N. The promise of stochastic resonance in falls prevention. *Front. Physiol.* **9**, 1865 (2019).
6. Adar, O., Shakargy, J. D. & Ilan, Y. The constrained disorder principle: beyond biological allostasis. *Biology* **14**, 339 (2025).
7. Mendez-Balbuena, I. et al. Improved sensorimotor performance via stochastic resonance. *J. Neurosci.* **32**, 12612–12618 (2012).
8. Krauss, P., Tziridis, K., Schilling, A. & Schulze, H. Cross-modal stochastic resonance as a universal principle to enhance sensory processing. *Front. Neurosci.* **12**, 12:578 (2018).
9. Nobusako, S. et al. Stochastic resonance improves visuomotor temporal integration in healthy young adults. *PLoS ONE* **13**, e0209382 (2018).
10. Vázquez-Rodríguez, B. et al. Stochastic resonance at criticality in a network model of the human cortex. *Sci. Rep.* **7**, 13020 (2017).
11. Ghosh, A., Rho, Y., McIntosh, A. R., Kötter, R. & Jirsa, V. K. Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS Comput. Biol.* **4**, e1000196 (2008).
12. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
13. Barlow, H. B. in *Sensory Communication* (ed. Rosenblith, W. A.) Ch. 13 (MIT Press, 1961).
14. Shannon, C. & Weaver, W. *The Mathematical Theory of Communication* (Univ. Illinois Press, 1964).
15. Badre, D., Bhandari, A., Keglovits, H. & Kikumoto, A. The dimensionality of neural representations for control. *Curr. Opin. Behav. Sci.* **38**, 20–28 (2021).
16. Bates, C. J. & Jacobs, R. A. Efficient data compression in perception and perceptual memory. *Psychol. Rev.* **127**, 891–917 (2020).
17. Bernardi, S. et al. The geometry of abstraction in hippocampus and prefrontal cortex. *Cell* **183**, 954–967.e21 (2020).
18. Guell, X., Schmahmann, J. D., Gabrieli, J. D. & Ghosh, S. S. Functional gradients of the cerebellum. *eLife* **7**, e36652 (2018).
19. Kharabian Masouleh, S., Plachti, A., Hoffstaedter, F., Eickhoff, S. & Genon, S. Characterizing the gradients of structural covariance in the human hippocampus. *NeuroImage* **218**, 116972 (2020).
20. Mack, M. L., Preston, A. R. & Love, B. C. Ventromedial prefrontal cortex compression during concept learning. *Nat. Commun.* **11**, 46 (2020).
21. Przeździk, I., Faber, M., Fernández, G., Beckmann, C. F. & Haak, K. V. The functional organisation of the hippocampus along its long axis is gradual and predicts recollection. *Cortex* **119**, 324–335 (2019).
22. Reber, T. P. et al. Representation of abstract semantic knowledge in populations of human single neurons in the medial temporal lobe. *PLoS Biol.* **17**, e3000290 (2019).
23. Shaffer, C., Barrett, L. F. & Quigley, K. S. Signal processing in the vagus nerve: hypotheses based on new genetic and anatomical evidence. *Biol. Psychol.* **182**, 108626 (2023).
24. Straub, I. et al. Gradients in the mammalian cerebellar cortex enable Fourier-like transformation and improve storing capacity. *eLife* **9**, e51771 (2020).
25. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cognit. Affect. Neurosci.* **12**, 1–23 (2017).
26. Chanes, L. & Barrett, L. F. Redefining the role of limbic areas in cortical processing. *Trends Cognit. Sci.* **20**, 96–106 (2016).
27. Finlay, B. L. & Uchiyama, R. Developmental mechanisms channeling cortical evolution. *Trends Neurosci.* **38**, 69–76 (2015).
28. Öngür, D. & Price, J. L. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10**, 206–219 (2000).
29. Kleckner, I. R. et al. Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nat. Hum. Behav.* **1**, 0069 (2017).
30. Girn, M., Setton, R., Turner, G. R. & Spreng, R. N. The "limbic network," comprising orbitofrontal and anterior temporal cortex, is part of an extended default network: evidence from multi-echo fMRI. *Netw. Neurosci.* **8**, 860–882 (2024).
31. Zhang, J. et al. Cortical and subcortical mapping of the human allostatic–interoceptive system using 7 Tesla fMRI. *Nat. Neurosci.* **28**, 2380–2391 (2025).
32. Theriault, J. E. et al. It's not the thought that counts: allostasis at the core of brain function. *Neuron* **113**, 4107–4133 (2025).
33. Glasser, M. F., Goyal, M. S., Preuss, T. M., Raichle, M. E. & Van Essen, D. C. Trends and properties of human cerebral cortex: correlations with cortical myelin content. *NeuroImage* **93**, 165–175 (2014).
34. Hilgetag, C. C. & Goulas, A. 'Hierarchy' in the organization of brain networks. *Phil. Trans. R. Soc. B* **375**, 20190319 (2020).
35. Zhang, J. et al. Topography impacts topology: anatomically central areas exhibit a "high-level connector" profile in the human cortex. *Cereb. Cortex* **30**, 1357–1365 (2020).
36. van den Heuvel, M. P. & Sporns, O. Rich-club organization of the human connectome. *J. Neurosci.* **31**, 15775–15786 (2011).
37. Barbas, H. General cortical and special prefrontal connections: principles from structure to function. *Annu. Rev. Neurosci.* **38**, 269–289 (2015).
38. John, Y. J., Zikopoulos, B., García-Cabezas, M. Á & Barbas, H. The cortical spectrum: a robust structural continuum in primate cerebral cortex revealed by histological staining and magnetic resonance imaging. *Front. Neuroanat.* **16**, 897237 (2022).
39. Benna, M. K. & Fusi, S. Place cells may simply be memory cells: memory compression leads to spatial tuning and history dependence. *Proc. Natl Acad. Sci. USA* **118**, e2018422118 (2021).

40. Gluck, M. A. & Myers, C. E. Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus* **3**, 491–516 (1993).
41. Barbas, H. & Rempel-Clower, N. Cortical structure predicts the pattern of corticocortical connections. *Cereb. Cortex* **7**, 635–646 (1997).
42. Brincat, S. L., Siegel, M., Von Nicolai, C. & Miller, E. K. Gradual progression from sensory to task-related processing in cerebral cortex. *Proc. Natl Acad. Sci. USA* **115**, E7202–E7211 (2018).
43. Siegle, J. H. et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).
44. Bernhardt, B. C., Smallwood, J., Keilholz, S. & Margulies, D. S. Gradients in brain organization. *NeuroImage* **251**, 118987 (2022).
45. Haueis, P. Multiscale modeling of cortical gradients: the role of mesoscale circuits for linking macro- and microscale gradients of cortical organization and hierarchical information processing. *NeuroImage* **232**, 117846 (2021).
46. Paquola, C. et al. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS Biol.* **17**, e3000284 (2019).
47. Zhou, D. et al. Compression supports low-dimensional representations of behavior across neural circuits. Preprint at *bioRxiv* https://doi.org/10.1101/2022.11.29.518415 (2022).
48. Katsumi, Y. et al. Correspondence of functional connectivity gradients across human isocortex, cerebellum, and hippocampus. *Commun. Biol.* **6**, 401 (2023).
49. Raut, R. V., Snyder, A. Z. & Raichle, M. E. Hierarchical dynamics as a macroscopic organizing principle of the human brain. *Proc. Natl Acad. Sci. USA* **117**, 20890–20897 (2020).
50. Shafiei, G. et al. Topographic gradients of intrinsic dynamics across neocortex. *eLife* **9**, e62116 (2020).
51. Wang, X.-J. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nat. Rev. Neurosci.* **21**, 169–178 (2020).
52. MacIver, M. A. & Finlay, B. L. The neuroecology of the water-to-land transition and the evolution of the vertebrate brain. *Phil. Trans. R. Soc.* **B 377**, 20200523 (2022).
53. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796 (2009).
54. Azzalini, D., Rebollo, I. & Tallon-Baudry, C. Visceral signals shape brain dynamics and cognition. *Trends Cognit. Sci.* **23**, 488–509 (2019).
55. Engelen, T., Solcà, M. & Tallon-Baudry, C. Interoceptive rhythms in the brain. *Nat. Neurosci.* **26**, 1670–1684 (2023).
56. Tort, A. B. L., Brankačk, J. & Draguhn, A. Respiration-entrained brain rhythms are global but often overlooked. *Trends Neurosci.* **41**, 186–197 (2018).
57. Braga, R. M., Sharp, D. J., Leeson, C., Wise, R. J. S. & Leech, R. Echoes of the brain within default mode, association, and heteromodal cortices. *J. Neurosci.* **33**, 14031–14039 (2013).
58. Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H. & Johnson, K. A. Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. *J. Neurosci.* **32**, 10649–10661 (2012).
59. Szinte, M. & Knapen, T. Visual organization of the default network. *Cereb. Cortex* **30**, 3518–3527 (2020).
60. Wei, W. et al. A function-based mapping of sensory integration along the cortical hierarchy. *Commun. Biol.* **7**, 1–14 (2024).
61. Rizzolatti, G. et al. in *Principles of Neural Science* (eds. Kandel, E. R. et al.) 412–425 (McGraw-Hill, 2013).
62. Vogt, B. A. in *Cingulate Neurobiology and Disease* (ed. Vogt, B. A.) 65–94 (Oxford Univ. Press, 2009).
63. Barnaveli, I., Viganò, S., Reznik, D., Haggard, P. & Doeller, C. F. Hippocampal-entorhinal cognitive maps and cortical motor system represent action plans and their outcomes. *Nat. Commun.* **16**, 4139 (2025).
64. Lathe, R., Singadia, S., Jordan, C. & Riedel, G. The interoceptive hippocampus: mouse brain endocrine receptor expression highlights a dentate gyrus (DG)–cornu ammonis (CA) challenge-sufficiency axis. *PLoS ONE* **15**, e0227575 (2020).
65. Barsalou, L. W. Grounded cognition: past, present, and future. *Top. Cognit. Sci.* **2**, 716–724 (2010).
66. Kalaska, J. et al. in *Principles of Neural Science* (eds. Kandel, E. R. et al.) Ch. 37 (McGraw-Hill, 2013).
67. Barrett, L. F. & Finlay, B. L. Concepts, goals and the control of survival-related behaviors. *Curr. Opin. Behav. Sci.* **24**, 172–179 (2018).
68. Hickok, G. *The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition* 292 (W. W. Norton, 2014).
69. Graziano, M. S. A. in *Shared Representations* (eds Obhi, S. S. & Cross, E. S.) 38–58 (Cambridge Univ. Press, 2016).
70. Seger, C. A. & Miller, E. K. Category learning in the brain. *Annu. Rev. Neurosci.* **33**, 203–219 (2010).
71. Giffin, C., Wilkenfeld, D. & Lombrozo, T. The explanatory effect of a label: explanations with named categories are more satisfying. *Cognition* **168**, 357–369 (2017).
72. Vouloumanos, A. & Waxman, S. R. Listen up! Speech is for thinking during infancy. *Trends Cognit. Sci.* **18**, 642–646 (2014).
73. Waxman, S. R. & Gelman, S. A. in *The Making of Human Concepts* (eds Mareschal, D., Quinn, P. C. & Lea, S. E. G.) 99–130 (Oxford Univ. Press, 2010).
74. Waxman, S. R. & Markow, D. B. Words as invitations to form categories: evidence from 12- to 13-month-old infants. *Cognit. Psychol.* **29**, 257–302 (1995).
75. Booth, A. E. & Waxman, S. Object names and object functions serve as cues to categories for infants. *Dev. Psychol.* **38**, 948–957 (2002).
76. Graham, S. A., Kilbreath, C. S. & Welder, A. N. Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Dev.* **75**, 409–427 (2004).
77. Nazzi, T. & Gopnik, A. Linguistic and cognitive abilities in infancy: when does language become a tool for categorization? *Cognition* **80**, B11–B20 (2001).
78. Welder, A. N. & Graham, S. A. The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Dev.* **72**, 1653–1673 (2001).
79. Chung, S. & Abbott, L. F. Neural population geometry: an approach for understanding biological and artificial neural networks. *Curr. Opin. Neurobiol.* **70**, 137–144 (2021).
80. Kerrén, C., Reznik, D., Doeller, C. F. & Griffiths, B. J. Exploring the role of dimensionality transformation in episodic memory. *Trends Cognit. Sci.* **29**, 614–626 (2025).
81. Barrett, L. F. & Theriault, J. in *Handbook of Social Psychology 6th Edition* (eds Gilbert, D., Fiske, S., Finkel, E. & Mendes, W.) https://doi.org/10.70400/BPQW3358 (Situational Press, 2025).
82. Lombrozo, T. Explanation and categorization: how 'why?' informs 'what?'. *Cognition* **110**, 248–253 (2009).
83. Muhle-Karbe, P. S. et al. Goal-seeking compresses neural codes for space in the human hippocampus and orbitofrontal cortex. *Neuron* **111**, 3885–3899.e6 (2023).
84. Estes, W. K. *Classification and Cognition* (Oxford Univ. Press, 1994).
85. Medin, D. L. & Schaffer, M. M. Context theory of classification learning. *Psychol. Rev.* **85**, 207–238 (1978).
86. Smith, E. E. in *Foundations of Cognitive Science* (ed. Posner, M.) 501–526 (MIT Press, 1989).
87. Rosch, E. & Mervis, C. B. Family resemblances: studies in the internal structure of categories. *Cognit. Psychol.* **7**, 573–605 (1975).
88. Ashby, F. G. & Valentin, V. V. in *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (ed. Wixted, J. T.) https://doi.org/10.1002/9781119170174.epcn508 (Wiley, 2018).
89. Nosofsky, R. M. Choice, similarity, and the context theory of classification. *J. Exp. Psychol.* **10**, 104–114 (1984).
90. Rockland, K. S. Notes on visual cortical feedback and feedforward connections. *Front. Syst. Neurosci.* **16**, 784310 (2022).
91. Markov, N. T. et al. Weight consistency specifies regularities of macaque cortical networks. *Cereb. Cortex* **21**, 1254–1272 (2011).
92. Sherman, S. M. & Guillery, R. W. The role of the thalamus in the flow of information to the cortex. *Phil. Trans. R. Soc. Lond. B* **357**, 1695–1708 (2002).
93. Sporns, O. *Networks of the Brain* (MIT Press, 2011).
94. Keller, A. J., Roth, M. M. & Scanziani, M. Feedback generates a second receptive field in neurons of the visual cortex. *Nature* **582**, 545–549 (2020).
95. Aru, J. et al. Untangling cross-frequency coupling in neuroscience. *Curr. Opin. Neurobiol.* **31**, 51–61 (2015).
96. Boyd, A. M., Kato, H. K., Komiyama, T. & Isaacson, J. S. Broadcasting of cortical activity to the olfactory bulb. *Cell Rep.* **10**, 1032–1039 (2015).
97. Warwick, R. A. et al. Top-down modulation of the retinal code via histaminergic neurons of the hypothalamus. *Sci. Adv.* **10**, eadk4062 (2024).
98. Schröder, S. et al. Arousal modulates retinal output. *Neuron* **107**, 487–495.e9 (2020).
99. Halassa, M. M. & Sherman, S. M. Thalamocortical circuit motifs: a general framework. *Neuron* **103**, 762–770 (2019).
100. Sherman, S. M. & Guillery, R. W. *Exploring the Thalamus and its Role in Cortical Function* 253–286 (MIT Press, 2006).
101. Beitz, A. J. The organization of afferent projections to the midbrain periaqueductal gray of the rat. *Neuroscience* **7**, 133–159 (1982).
102. Uhlrich, D. J., Cucchiaro, J. B. & Sherman, S. M. The projection of individual axons from the parabrachial region of the brain stem to the dorsal lateral geniculate nucleus in the cat. *J. Neurosci.* **8**, 4565–4575 (1988).
103. Card, J. P. & Moore, R. Y. Organization of lateral geniculate-hypothalamic connections in the rat. *J. Comp. Neurol.* **284**, 135–147 (1989).
104. Fillinger, C., Yalcin, I., Barrot, M. & Veinante, P. Efferents of anterior cingulate areas 24a and 24b and midcingulate areas 24a' and 24b' in the mouse. *Brain Struct. Funct.* **223**, 1747–1778 (2018).
105. Zhang, J. et al. Cortical and subcortical mapping of the allostatic-interoceptive system in the human brain using 7 Tesla fMRI. *Nat. Neurosci.* **28**, 2380–2391 (2025).
106. Müller, E. J. et al. Core and matrix thalamic sub-populations relate to spatio-temporal cortical connectivity gradients. *NeuroImage* **222**, 117224 (2020).
107. Phillips, J. M. et al. Primate thalamic nuclei select abstract rules and shape prefrontal dynamics. *Neuron* **113**, 2014–2027.e12 (2025).
108. Shine, J. M. et al. Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat. Neurosci.* **22**, 289–296 (2019).
109. Halassa, M. M. & Saalmann, Y. B. in *The Cerebral Cortex and Thalamus* (eds Usrey, W. M. & Sherman, S. M.) Ch. 46 (Oxford Univ. Press, 2024).
110. Shine, J. M. The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics. *Prog. Neurobiol.* **199**, 101951 (2021).
111. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
112. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).

113. Tye, K. M. et al. Mixed selectivity: cellular computations for complexity. *Neuron* **112**, 2289–2303 (2024).
114. Olshausen, B. A. & Field, D. J. in *23 Problems in Systems Neuroscience* (eds van Hemmen, J. L. & Sejnowski, T. J.) 182–212 (Oxford Univ. Press, 2006).
115. Albright, T. D. & Stoner, G. R. Contextual influences on visual processing. *Annu. Rev. Neurosci.* **25**, 339–379 (2002).
116. Basole, A., White, L. E. & Fitzpatrick, D. Mapping multiple features in the population response of visual cortex. *Nature* **423**, 986–990 (2003).
117. David, S. V., Vinje, W. E. & Gallant, J. L. Natural stimulus statistics alter the receptive field structure of V1 neurons. *J. Neurosci.* **24**, 6991–7006 (2004).
118. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).
119. Parker, P. R. L., Brown, M. A., Smear, M. C. & Niell, C. M. Movement-related signals in sensory areas: roles in natural behavior. *Trends Neurosci.* **43**, 581–595 (2020).
120. Spillmann, L., Dresp-Langley, B. & Tseng, C. Beyond the classical receptive field: the effect of contextual stimuli. *J. Vis.* **15**, 7 (2015).
121. Bressler, S. L. & McIntosh, A. R. in *Handbook of Brain Connectivity* (eds Jirsa, V. K. & McIntosh, A.) 403–419 (Springer, 2007).
122. Gjorgjieva, J., Drion, G. & Marder, E. Computational implications of biophysical diversity and multiple timescales in neurons and synapses for circuit performance. *Curr. Opin. Neurobiol.* **37**, 44–52 (2016).
123. Heald, J. B., Wolpert, D. M. & Lengyel, M. The computational and neural bases of context-dependent learning. *Annu. Rev. Neurosci.* **46**, 233–258 (2023).
124. Saxena, S. & Cunningham, J. P. Towards the neural population doctrine. *Curr. Opin. Neurobiol.* **55**, 103–111 (2019).
125. Willems, R. M. & Peelen, M. V. How context changes the neural basis of perception and language. *iScience* **24**, 102392 (2021).
126. Denève, S. & Machens, C. K. Efficient codes and balanced networks. *Nat. Neurosci.* **19**, 375–382 (2016).
127. Sillito, A. M. The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *J. Physiol.* **250**, 305–329 (1975).
128. Balasubramanian, V. Heterogeneity and efficiency in the brain. *Proc. IEEE* **103**, 1346–1358 (2015).
129. Favila, S. E., Samide, R., Sweigart, S. C. & Kuhl, B. A. Parietal representations of stimulus features are amplified during memory retrieval and flexibly aligned with top-down goals. *J. Neurosci.* **38**, 7809–7821 (2018).
130. Lifanov-Carr, J. et al. Reconstructing spatiotemporal trajectories of visual object memories in the human brain. *eNeuro* **11**, ENEURO.0091–24.2024 (2024).
131. Linde-Domingo, J., Treder, M. S., Kerrén, C. & Wimber, M. Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nat. Commun.* **10**, 179 (2019).
132. Pinotsis, D. A., Buschman, T. J. & Miller, E. K. Working memory load modulates neuronal coupling. *Cereb. Cortex* **29**, 1670–1681 (2019).
133. Lifanov, J., Linde-Domingo, J. & Wimber, M. Feature-specific reaction times reveal a semanticisation of memories over time and with repeated remembering. *Nat. Commun.* **12**, 3177 (2021).
134. Hutchinson, J. B. & Barrett, L. F. The power of predictions: an emerging paradigm for psychological research. *Curr. Dir. Psychol. Sci.* **28**, 280–291 (2019).
135. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 1–24 (2013).
136. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
137. Gabhart, K. M., Xiong, Y. S. & Bastos, A. M. Predictive coding: a more cognitive process than we thought? *Trends Cognit. Sci.* **29**, 627–640 (2025).
138. Keller, G. B. & Mrsic-Flogel, T. D. Predictive processing: a canonical cortical computation. *Neuron* **100**, 424–435 (2018).
139. Rao, R. P. N. A sensory–motor theory of the neocortex. *Nat. Neurosci.* **27**, 1221–1235 (2024).
140. Bastos, A. M., Lundqvist, M., Waite, A. S., Kopell, N. & Miller, E. K. Layer and rhythm specificity for predictive routing. *Proc. Natl Acad. Sci. USA* **117**, 31459–31469 (2020).
141. Koren, V. & Denève, S. Computational account of spontaneous activity as a signature of predictive coding. *PLoS Comput. Biol.* **13**, e1005355 (2017).
142. Luczak, A., McNaughton, B. L. & Kubo, Y. Neurons learn by predicting future activity. *Nat. Mach. Intell.* **4**, 62–72 (2022).
143. Pinotsis, D. A., Loonis, R., Bastos, A. M., Miller, E. K. & Friston, K. J. Bayesian modelling of induced responses and neuronal rhythms. *Brain Topogr.* **32**, 569–582 (2019).
144. Pinotsis, D. A. et al. Linking canonical microcircuits and neuronal activity: dynamic causal modelling of laminar recordings. *Neuroimage* **146**, 355–366 (2017).
145. Xiong, Y. S. et al. Propofol-mediated loss of consciousness disrupts predictive routing and local field phase modulation of neural activity. *Proc. Natl Acad. Sci. USA* **121**, e2315160121 (2024).
146. Straka, H., Simmers, J. & Chagnaud, B. P. A new perspective on predictive motor signaling. *Curr. Biol.* **28**, R232–R243 (2018).
147. Barrett, L. F. & Simmons, W. K. Interoceptive predictions in the brain. *Nat. Rev. Neurosci.* **16**, 419–429 (2015).
148. Joyce, M. K. P. & Barbas, H. Cortical connections position primate area 25 as a keystone for interoception, emotion, and memory. *J. Neurosci.* **38**, 1677–1698 (2018).
149. Wolpert, D. M. & Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Netw.* **11**, 1317–1329 (1998).
150. Pinotsis, D. A., Siegel, M. & Miller, E. K. Sensory processing and categorization in cortical and deep neural networks. *Neuroimage* **202**, 116118 (2019).
151. Hoffman, P., McClelland, J. L. & Lambon Ralph, M. A. Concepts, control, and context: a connectionist account of normal and disordered semantic cognition. *Psychol. Rev.* **125**, 293 (2018).
152. Schyns, P. G., Goldstone, R. L. & Thibaut, J. P. The development of features in object concepts. *Behav. Brain Sci.* **21**, 1–17 (1998).
153. Wilson-Mendenhall, C. D., Barrett, L. F. & Barsalou, L. W. Variety in emotional life: within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Soc. Cogn. Affect. Neurosci.* **10**, 62–71 (2015).
154. Yee, E. & Thompson-Schill, S. L. Putting concepts into context. *Psychon. Bull. Rev.* **23**, 1015–1027 (2016).
155. Barsalou, L. W. in *Building Categories in Interaction: Linguistic Resources at Work* (eds. Mauri, C., Fiorentini, I. & Goria, E.) 35–72 (John Benjamins, 2021).
156. Casasanto, D. & Lupyan, G. in *The Conceptual Mind: New Directions in the Study of Concepts* (eds Margolis, E. & Laurence, S.) 543–566 (MIT Press, 2015).
157. Coraci, D. A unified model of ad hoc concepts in conceptual spaces. *Minds Mach.* **32**, 289–309 (2022).
158. Voorspoels, W., Storms, G. & Vanpaemel, W. Idealness and similarity in goal-derived categories: a computational examination. *Mem. Cogn.* **41**, 312–327 (2013).
159. Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. *J. Exp. Psychol.* **77**, 353–363 (1968).
160. Edelman, G. M. & Gally, J. A. Degeneracy and complexity in biological systems. *Proc. Natl Acad. Sci. USA* **98**, 13763–13768 (2001).
161. Marder, E. & Taylor, A. L. Multiple models to capture the variability in biological neurons and networks. *Nat. Neurosci.* **14**, 133–138 (2011).
162. Brincat, S. L. & Miller, E. K. Prefrontal cortex networks shift from external to internal modes during learning. *J. Neurosci.* **36**, 9739–9754 (2016).
163. Freedman, D. J. & Assad, J. A. Experience-dependent representation of visual categories in parietal cortex. *Nature* **443**, 85–88 (2006).
164. Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).
165. Wallis, J. D., Anderson, K. C. & Miller, E. K. Single neurons in prefrontal cortex encode abstract rules. *Nature* **411**, 953–956 (2001).
166. Antzoulatos, E. G. & Miller, E. K. Differences between neural activity in prefrontal cortex and striatum during learning of novel, abstract categories. *Neuron* **71**, 243–249 (2011).
167. Wutz, A., Loonis, R., Roy, J. E., Donoghue, J. A. & Miller, E. K. Different levels of category abstraction by different dynamics in different prefrontal areas. *Neuron* **97**, 716–726.e8 (2018).
168. Zhang, X.-Y. et al. Adaptive stretching of representations across brain regions and deep learning model layers. *Nat. Commun.* **16**, 10302 (2025).
169. Yaron, A. et al. Auditory cortex neurons that encode negative prediction errors respond to omissions of sounds in a predictable sequence. *PLoS Biol.* **23**, e3003242 (2025).
170. Hoy, C. W. et al. Asymmetric coding of reward prediction errors in human insula and dorsomedial prefrontal cortex. *Nat. Commun.* **14**, 8520 (2023).
171. Braga, A. & Schönwiesner, M. Neural substrates and models of omission responses and predictive processes. *Front. Neural Circuits* **16**, 799581 (2022).
172. Adams, R. A., Bauer, M., Pinotsis, D. & Friston, K. J. Dynamic causal modelling of eye movements during pursuit: confirming precision-encoding in V1 using MEG. *Neuroimage* **132**, 175–189 (2016).
173. Feldman, H. & Friston, K. J. Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* **4**, 215 (2010).
174. Parr, T. & Friston, K. J. Attention or salience? *Curr. Opin. Psychol.* **29**, 1–5 (2019).
175. Deneve, S. Bayesian spiking neurons I: inference. *Neural Comput.* **20**, 91–117 (2008).
176. Yan, C., de Lange, F. P. & Richter, D. Conceptual associations generate sensory predictions. *J. Neurosci.* **43**, 3733–3742 (2023).
177. Lundqvist, M. et al. Working memory control dynamics follow principles of spatial computing. *Nat. Commun.* **14**, 1429 (2023).
178. Miller, E. K., Lundqvist, M. & Bastos, A. M. Working memory 2.0. *Neuron* **100**, 463–475 (2018).
179. Recanatesi, S. et al. Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nat. Commun.* **12**, 1417 (2021).
180. Pinotsis, D. A., Fridman, G. & Miller, E. K. Cytoelectric coupling: electric fields sculpt neural activity and "tune" the brain's infrastructure. *Prog. Neurobiol.* **226**, 102465 (2023).
181. Buzsáki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).
182. Jadi, M. P. & Sejnowski, T. J. Cortical oscillations arise from contextual interactions that regulate sparse coding. *Proc. Natl Acad. Sci. USA* **111**, 6780–6785 (2014).
183. Buzsáki, G. & Vöröslakos, M. Brain rhythms have come of age. *Neuron* **111**, 922–926 (2023).
184. Fröhlich, F. & McCormick, D. A. Endogenous electric fields may guide neocortical network activity. *Neuron* **67**, 129–143 (2010).
185. Anastassiou, C. A., Perin, R., Markram, H. & Koch, C. Ephaptic coupling of cortical neurons. *Nat. Neurosci.* **14**, 217–223 (2011).
186. Lundqvist, M., Miller, E. K., Nordmark, J., Liljefors, J. & Herman, P. Beta: bursts of cognition. *Trends Cognit. Sci.* **28**, 662–676 (2024).
187. Pinotsis, D. A. & Miller, E. K. Differences in visually induced MEG oscillations reflect differences in deep cortical layer activity. *Commun. Biol.* **3**, 707 (2020).

188. Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L. & Miller, E. K. Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* **9**, 394 (2018).
189. Pinotsis, D. A. & Miller, E. K. Beyond dimension reduction: stable electric fields emerge from and allow representational drift. *NeuroImage* **253**, 119058 (2022).
190. Bartos, M., Vida, I. & Jonas, P. Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nat. Rev. Neurosci.* **8**, 45–56 (2007).
191. Buzsáki, G. & Wang, X.-J. Mechanisms of gamma oscillations. *Annu. Rev. Neurosci.* **35**, 203–225 (2012).
192. Miao, C., Cao, Q., Moser, M.-B. & Moser, E. I. Parvalbumin and somatostatin interneurons control different space-coding networks in the medial entorhinal cortex. *Cell* **171**, 507–521.e17 (2017).
193. van den Heuvel, M. P. et al. Multimodal analysis of cortical chemoarchitecture and macroscale fMRI resting-state functional connectivity. *Hum. Brain Mapp.* **37**, 3103–3113 (2016).
194. Barrett, L. F. et al. The theory of constructed emotion: more than a feeling. *Perspect. Psychol. Sci.* **20**, 392–420 (2025).
195. Mayr, E. *What Makes Biology Unique?: Considerations on the Autonomy of a Scientific Discipline* (Cambridge Univ. Press, 2004).
196. Picard, M., Kempes, C., Pontzer, H., Behnke, A. & Shaulson, E. D. Energy constraints on human health. Preprint at https://doi.org/10.31219/osf.io/ne3qg_v1 (2025).
197. McEwen, B. S. Stress, adaptation, and disease: allostasis and allostatic load. *Ann. N. Y. Acad. Sci.* **840**, 33–44 (1998).
198. Crossley, N. A. et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395 (2014).
199. de Lange, S. C. et al. Shared vulnerability for connectome alterations across psychiatric and neurological brain disorders. *Nat. Hum. Behav.* **3**, 988–998 (2019).
200. Goodkind, M. et al. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305 (2015).
201. Sprooten, E. et al. Addressing reverse inference in psychiatric neuroimaging: meta-analyses of task-related brain activation in common mental disorders. *Hum. Brain Mapp.* **38**, 1846–1864 (2017).
202. Stam, C. J. Hub overload and failure as a final common pathway in neurological brain network disorders. *Netw. Neurosci.* **8**, 1–23 (2024).
203. Bolt, T. et al. Autonomic physiological coupling of the global fMRI signal. *Nat. Neurosci.* **28**, 1327–1335 (2025).
204. Bolt, T. et al. A parsimonious description of global functional brain organization in three spatiotemporal patterns. *Nat. Neurosci.* **25**, 1093–1103 (2022).
205. Gu, Y. et al. Brain activity fluctuations propagate as waves traversing the cortical hierarchy. *Cereb. Cortex* **31**, 3986–4005 (2021).
206. Raut, R. V. et al. Global waves synchronize the brain's functional systems with fluctuating arousal. *Sci. Adv.* **7**, eabf2709 (2021).
207. Luczak, A., Bartho, P. & Harris, K. D. Gating of sensory input by spontaneous cortical activity. *J. Neurosci.* **33**, 1684–1695 (2013).
208. Parr, T., Pezzulo, G. & Friston, K. J. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* (MIT Press, 2022).
209. Theriault, J. E. et al. A functional account of stimulation-based aerobic glycolysis and its role in interpreting BOLD signal intensity increases in neuroimaging experiments. *Neurosci. Biobehav. Rev.* **153**, 105373 (2023).
210. Chalk, M., Marre, O. & Tkačik, G. Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl Acad. Sci. USA* **115**, 186–191 (2018).
211. Hechler, A., De Lange, F. P. & Riedl, V. The energy metabolic footprint of predictive processing in the human brain. Preprint at *bioRxiv* https://doi.org/10.1101/2023.12.08.570804 (2023).
212. Manookin, M. B. & Rieke, F. Two sides of the same coin: efficient and predictive neural coding. *Annu. Rev. Vis. Sci.* **9**, 293–311 (2023).
213. Sengupta, B., Stemmler, M. B. & Friston, K. J. Information and efficiency in the nervous system — a synthesis. *PLoS Comput. Biol.* **9**, e1003157 (2013).
214. Ali, A., Ahmad, N., de Groot, E., Johannes van Gerven, M. A. & Kietzmann, T. C. Predictive coding is a consequence of energy efficiency in recurrent neural networks. *Patterns* **3**, 100639 (2022).
215. Quigley, K. S., Kanoski, S., Grill, W. M., Barrett, L. F. & Tsakiris, M. Functions of interoception: from energy regulation to experience of the self. *Trends Neurosci.* **44**, 29–38 (2021).
216. Bastos, A. M. et al. Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
217. Shipp, S., Adams, R. A. & Friston, K. J. Reflections on agranular architecture: predictive coding in the motor cortex. *Trends Neurosci.* **36**, 706–716 (2013).
218. Pezzulo, G., Zorzi, M. & Corbetta, M. The secret life of predictive brains: what's spontaneous activity for? *Trends Cognit. Sci.* **25**, 730–743 (2021).
219. Dimakou, A., Pezzulo, G., Zangrossi, A. & Corbetta, M. The predictive nature of spontaneous brain activity across scales and species. *Neuron* **113**, 1310–1332 (2025).
220. Corcoran, A. W., Pezzulo, G. & Hohwy, J. From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition. *Biol. Philos.* **35**, 32 (2020).
221. Levinthal, D. J. & Strick, P. L. The motor cortex communicates with the kidney. *J. Neurosci.* **32**, 6726–6731 (2012).
222. Levinthal, D. J. & Strick, P. L. Multiple areas of the cerebral cortex influence the stomach. *Proc. Natl Acad. Sci. USA* **117**, 13078–13083 (2020).
223. Alagapan, S. et al. Cingulate dynamics track depression recovery with deep brain stimulation. *Nature* **622**, 130–138 (2023).
224. Fujimoto, S. et al. Deep brain stimulation induces white matter remodeling and functional changes to brain-wide networks. *Brain Stimul.* **18**, 242–243 (2025).
225. Lochmann, T. & Deneve, S. Neural processing as causal inference. *Curr. Opin. Neurobiol.* **21**, 774–781 (2011).
226. Laland, K., Matthews, B. & Feldman, M. W. An introduction to niche construction theory. *Evol. Ecol.* **30**, 191–202 (2016).
227. Simpson, S. J. & Raubenheimer, D. *The Nature of Nutrition: A Unifying Framework from Animal Adaptation to Human Obesity* (Princeton Univ. Press, 2012).
228. Mareschal, D., Quinn, P. C. & Lea, S. E. G. (eds) *The Making of Human Concepts* (Oxford Univ. Press, 2010).
229. Freddolino, P. L. & Tavazoie, S. Beyond homeostasis: a predictive-dynamic framework for understanding cellular behavior. *Annu. Rev. Cell Dev. Biol.* **28**, 363–384 (2012).
230. Westlin, C. et al. Improving the study of brain–behavior relationships by revisiting basic assumptions. *Trends Cognit. Sci.* **27**, 246–257 (2023).
231. Sennesh, E. et al. Interoception as modeling, allostasis as control. *Biol. Psychol.* **167**, 108242 (2022).
232. Blakemore, S. J., Goodbody, S. J. & Wolpert, D. M. Predicting the consequences of our own actions: the role of sensorimotor context estimation. *J. Neurosci.* **18**, 7511–7518 (1998).
233. Berkes, P., Orbán, G., Lengyel, M. & Fiser, J. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331**, 83–87 (2011).
234. Barsalou, L. W. Grounded cognition. *Annu. Rev. Psychol.* **59**, 617–645 (2008).
235. McMains, S. & Kastner, S. Interactions of top-down and bottom-up mechanisms in human visual cortex. *J. Neurosci.* **31**, 587–597 (2011).
236. Gershman, S. J., Blei, D. M. & Niv, Y. Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010).
237. Frankland, S. M. & Greene, J. D. Concepts and compositionality: in search of the brain's language of thought. *Annu. Rev. Psychol.* **71**, 273–303 (2020).
238. Isomura, T., Shimazaki, H. & Friston, K. J. Canonical neural networks perform active inference. *Commun. Biol.* **5**, 55 (2022).
239. Johansen, J. P. et al. Hebbian and neuromodulatory mechanisms interact to trigger associative memory formation. *Proc. Natl Acad. Sci. USA* **111**, E5584–E5592 (2014).
240. Barrett, L. F., Quigley, K. S. & Hamilton, P. An active inference theory of allostasis and interoception in depression. *Philos. Trans. R. Soc. Lond. B* **371**, 20160011 (2016).
241. Shaffer, C., Westlin, C., Quigley, K. S., Whitfield-Gabrieli, S. & Barrett, L. F. Allostasis, action, and affect in depression: insights from the theory of constructed emotion. *Annu. Rev. Clin. Psychol.* **18**, 553–580 (2022).
242. Barrett, L. F. *How Emotions Are Made: The Secret Life of the Brain* (Pan Macmillan, 2017).
243. Sydnor, V. J. et al. Neurodevelopment of the association cortices: patterns, mechanisms, and implications for psychopathology. *Neuron* **109**, 2820–2846 (2021).
244. Sherwood, C. C., Bauernfeind, A. L., Verendeev, A., Raghanti, M. A. & Hof, P. R. in *Evolution of Nervous Systems* 2nd edn (ed. Kaas, J. H.) 121–139 (Academic, 2017).
245. Kuzawa, C. W. et al. Metabolic costs and evolutionary implications of human brain development. *Proc. Natl Acad. Sci. USA* **111**, 13010–13015 (2014).
246. Krienen, F. M., Yeo, B. T. T., Ge, T., Buckner, R. L. & Sherwood, C. C. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proc. Natl Acad. Sci. USA* **113**, E469–E478 (2016).
247. Sherwood, C. C. & Gómez-Robles, A. Brain plasticity and human evolution. *Annu. Rev. Anthropol.* **46**, 399–419 (2017).
248. Wei, Y. et al. Genetic mapping and evolutionary analysis of human-expanded cognitive networks. *Nat. Commun.* **10**, 4839 (2019).
249. Barrett, L. F. *Seven and a Half Lessons About the Brain* (HarperCollins, 2020).
250. Gallivan, J. P., Bowman, N. A. R., Chapman, C. S., Wolpert, D. M. & Flanagan, J. R. The sequential encoding of competing action goals involves dynamic restructuring of motor plans in working memory. *J. Neurophysiol.* **115**, 3113–3122 (2016).
251. Mesulam, M. M. From sensation to cognition. *Brain* **121**, 1013–1052 (1998).
252. Jones, E. G. Synchrony in the interconnected circuitry of the thalamus and cerebral cortex. *Ann. N. Y. Acad. Sci.* **1157**, 10–23 (2009).
253. Jones, E. G. The thalamic matrix and thalamocortical synchrony. *Trends Neurosci.* **24**, 595–601 (2001).
254. Sherman, S. M. & Guillery, R. W. Distinct functions for direct and transthalamic corticocortical connections. *J. Neurophysiol.* **106**, 1068–1077 (2011).
255. Sherman, S. M. Thalamus plays a central role in ongoing cortical functioning. *Nat. Neurosci.* **19**, 533–541 (2016).
256. Usrey, W. M. & Sherman, S. M. Corticofugal circuits: communication lines from the cortex to the rest of the brain. *J. Comp. Neurol.* **527**, 640–650 (2019).
257. Sherman, S. M. & Usrey, W. M. A reconsideration of the core and matrix classification of thalamocortical projections. *J. Neurosci.* **44**, e0163242024 (2024).
258. Shine, J. M. et al. The low-dimensional neural architecture of cognitive complexity is related to activity in medial thalamic nuclei. *Neuron* **104**, 849–855.e3 (2019).
259. Cappe, C., Morel, A., Barone, P. & Rouiller, E. M. The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cereb. Cortex* **19**, 2025–2037 (2009).
260. Lamme, V. A. F. & Roelfsema, P. R. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* **23**, 571–579 (2000).
261. Zeki, S. Multiple asynchronous stimulus- and task-dependent hierarchies (STDH) within the visual brain's parallel processing systems. *Eur. J. Neurosci.* **44**, 2515–2527 (2016).

262. Zeki, S. "Multiplexing" cells of the visual cortex and the timing enigma of the binding problem. *Eur. J. Neurosci.* **52**, 4684–4694 (2020).
263. Demirtaş, M. et al. Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron* **101**, 1181–1194.e13 (2019).
264. Tognoli, E. & Kelso, J. A. S. The metastable brain. *Neuron* **81**, 35–48 (2014).
265. García-Cabezas, M. Á, Zikopoulos, B. & Barbas, H. The structural model: a theory linking connections, plasticity, pathology, development and evolution of the cerebral cortex. *Brain Struct. Funct.* **224**, 985–1008 (2019).
266. Markov, N. T. et al. Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* **522**, 225–259 (2013).
267. Lin, H.-M. et al. Reconstruction of intratelencephalic neurons in the mouse secondary motor cortex reveals the diverse projection patterns of single neurons. *Front. Neuroanat.* **12**, 86 (2018).
268. Parent, M. & Parent, A. Single-axon tracing study of corticostriatal projections arising from primary motor cortex in primates. *J. Comp. Neurol.* **496**, 202–213 (2006).
269. Rockland, K. S. & Drash, G. W. Collateralized divergent feedback connections that target multiple cortical areas. *J. Comp. Neurol.* **373**, 529–548 (1996).
270. Weisenhorn, D. M. V., Ilung, R. B. & Spatz, W. B. Morphology and connections of neurons in area 17 projecting to the extrastriate areas mt and 19DM and to the superior colliculus in the monkey *Callithrix jacchus*. *J. Comp. Neurol.* **362**, 233–255 (1995).
271. Zhang, S. et al. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* **345**, 660–665 (2014).
272. Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A. & Keller, G. B. A sensorimotor circuit in mouse cortex for visual flow predictions. *Neuron* **95**, 1420–1432.e5 (2017).
273. Keck, T. et al. Synaptic scaling and homeostatic plasticity in the mouse visual cortex in vivo. *Neuron* **80**, 327–334 (2013).
274. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).
275. Larkum, M. A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* **36**, 141–151 (2013).
276. Takahashi, N., Oertner, T. G., Hegemann, P. & Larkum, M. E. Active cortical dendrites modulate perception. *Science* **354**, 1587–1590 (2016).
277. Larkum, M. E. Are dendrites conceptually useful? *Neuroscience* **489**, 4–14 (2022).
278. Peysakhovich, B. et al. Primate superior colliculus is causally engaged in abstract higher-order cognition. *Nat. Neurosci.* **27**, 1999–2008 (2024).
279. Fisher, A. & Rao, R. P. N. Recursive neural programs: a differentiable framework for learning compositional part-whole hierarchies and image grammars. *Proc. Natl Acad. Sci. USA Nexus* **2**, pgad337 (2023).

Acknowledgements
The authors thank H. Reimann, J. Zhang, J. Theriault and J. Rodriguez for their contributions to the original figures that were adapted herein and A. Mahoney for editorial assistance in compiling references and proofreading. H. Reimann also contributed to our discussion of the inhibitory landscape. Preparation of this article was supported by grants from the National Institute on Aging (R01AG071173), the US Army Research Institute for the Behavioral and Social Sciences (W911NF-16-1-019), the Unlikely Collaborators Foundation, The Freedom Together Foundation, The Picower Institute for Learning and Memory, the Simon Center for the Social Brain, Office of Naval Research MURI N00014-23-1-2768 and Army Research Office MURI W911NF2410228. The views, opinions and/or findings contained in this review are those of the authors and shall not be construed as an official Department of the Army position, policy or decision, unless so designated by other documents. Nor do they necessarily reflect the views of the Unlikely Collaborators Foundation.

Author contributions
Both authors contributed equally to the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Peer review information *Nature Reviews Neuroscience* thanks Stefan Mihalis, Giovanni Pezzulo and H. Steven Scholte for their contribution to the peer review of this work.