Review

Code Under Construction: Neural Coding Over Development

Lilach Avitan[1] and Geoffrey J. Goodhill[1,2,*]

Developing animals must begin to interact with the world before their neural development is complete. This means they must build neural codes appropriate for turning sensory inputs into motor outputs adaptively as their neural hardware matures. We review some recent progress in the understanding of the relationship between neural coding and neural circuit development. We focus particularly on neural coding in the context of topographic maps and spontaneous activity, as well as receptive field and circuit development, drawing on examples from both mammalian visual cortex and fish optic tectum. Overall we suggest that neural coding strategies during development may be highly dynamic.

Building the Neural Code

Starting from a single cell, an organism must build not only a body but also a brain to sense the world and react appropriately. As any parent will appreciate, for humans this process of refinement takes many years. By contrast, for some animals the basic neural tools required for autonomous survival must be established much more rapidly. For instance, a larval zebrafish is able to use vision to hunt small, fast-moving prey such as *Paramecia* within 5 days of the fertilization of the egg [1,2]. During this brief period the fish must build neural computations which allow it to map noisy and unreliable sensory inputs into appropriate motor actions. Furthermore, these computations must be continuously updated as the animal grows and gains more experience of the world in which it must survive.

From this functional perspective, building a nervous system means constructing a neural code. **Neural encoding** (see Glossary) is the process of converting sensory stimuli into neural activity, which often takes the form of complex temporal patterns of action potentials across many neurons [3]. **Neural decoding** is the reverse process of extracting the information contained within these patterns. The neural code must satisfy many constraints, most importantly that of transmitting large amounts of information using small amounts of energy [4]. Crucially, the code that optimizes this constraint depends on the statistics of the environment [5]. Thus, through a combination of evolutionary experience encoded in genetic programs and dynamic plasticity of neural connections, an organism must adapt its nervous system to the world it inhabits.

In recent years great progress has been made in understanding nervous system development [6,7] and some of the key mathematical principles relevant to neural coding [8–11]. While some notable earlier works did explicitly address the links between the two (e.g., [12]), we suggest here that it is timely to consider these links again in the light of this new information. We first introduce some of the key mathematical principles underlying neural coding. We then discuss their relevance to some specific issues in neural development, namely topographic maps,

Highlights

A combination of theory and experiment has revealed some key principles of neural coding.

Several aspects of neural development can be interpreted in terms of optimizing neural coding.

Spontaneous activity may represent prior distributions, and prior distributions may be updated over development to facilitate Bayesian computations.

In addition to simply refining, coding strategies may qualitatively change over development.

[1]Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia
[2]School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia

*Correspondence:
g.goodhill@uq.edu.au (G.J. Goodhill).

CrossMark

spontaneous activity, and receptive field and neural circuit development, with a particular focus on the mammalian visual cortex and fish optic tectum.

Neural Encoding and Decoding

Neural activity is generally probabilistic: the same stimulus usually evokes slightly different responses each time it is presented. Probability theory therefore plays a central role in the mathematics of neural coding. The key equation relating encoding and decoding is Bayes' theorem. This gives the posterior probability $P(s|r)$ that the stimulus is s given that response r is observed (i.e., the decoding of response r) in terms of the probability $P(r|s)$ that stimulus s generates response r (also called the likelihood), the probability $P(r)$ of observing r over all stimuli, and the probability $P(s)$ of observing s, also called the prior probability of s (mathematical details are given in Box 1; Figure 1 gives schematic illustrations of the principles discussed in this section). Given substantial evidence that the brain utilizes Bayesian principles (e.g., [13–15]), it is thus apparent that some knowledge of $P(s)$, namely how likely each different type of stimulus is in the world, is important for decoding. While philosophers have argued for centuries regarding the extent to which the brain is a *tabula rasa*, it is clear that an organism's knowledge of $P(s)$ is at least modified by its interaction with the environment. Thus a key part of neural development can be seen as learning about $P(s)$.

Advances in experimental techniques now allow the activity of many neurons to be recorded simultaneously. This has led to much theoretical interest in population coding, in other words how neurons conspire in groups to represent information [16]. One case is that neurons act independently, in other words the probability of a neuron firing over the stimulus ensemble is independent of the probability of any other neuron firing (sometimes called **factorial coding**). A

Box 1. Encoding, Decoding and Bayes Theorem

The **prior** $P(s)$, likelihood $P(r|s)$ (encoding) and **posterior** $P(s|r)$ (decoding) **distributions** defined in the main text are related via **Bayes theorem**:

$$P(s|r) = \frac{P(r|s)P(s)}{P(r)} \qquad [I]$$

The prior over r, $P(r)$ can be determined by:

$$P(r) = \int_s P(r|s)P(s)ds. \qquad [II]$$

There are three main approaches for decoding s. The simplest to calculate is the maximum likelihood (ML) estimate, which is the stimulus which maximizes $P(r|s)$. This does not require knowledge of $P(s)$, but can produce misleading results when some stimuli are much more likely than others. This problem is solved by the **maximum a posteriori (MAP) decoding** estimate, which is the stimulus that maximizes $P(s|r)$, but at the cost of requiring knowledge of $P(s)$. However, a problem with both the ML and MAP estimates is that they pick the peak of their respective distributions, which may be misleading if these distributions are not well-summarized in this way [for instance if the mode (peak) of the distribution is a long way from the mean]. In Bayesian inference (an imperfect name, because the MAP estimate is also Bayesian) the estimate that minimizes a particular loss function is chosen. When this function is the squared difference between the estimate s_{bayes} and the true value, the estimate is given by the mean $P(s|r)$:

$$s_{bayes} = \int sP(s|r)ds \qquad [III]$$

However, this requires greater computational load than the ML or MAP estimates.

The above distributions clearly change over neural development. For instance, a gradual reduction in response variability leads to more precise likelihood functions and thus better decoding, regardless which specific method is used. However, whether there is a change in decoding strategies over development, for instance by moving from simpler to more sophisticated estimates, is unknown. In addition, while decoding can be a very useful method for analyzing the information content of a particular stage of neural processing, note that the real system may not explicitly perform decoding in mapping between sensory inputs and motor outputs.

Glossary

Efficient coding: an efficient neural code is one that minimizes the number of spikes (or other measure of energy cost) necessary to convey a particular set of signals. For instance, visual images are highly redundant (neighboring pixels tend to have similar values), and thus pixel-based codes are generally not efficient.

Factorial code: a code where the probability of each neuron firing over the stimulus ensemble is statistically independent of the probability of any other neuron firing. A weaker case is a decorrelation code where second-order correlations are zero but high-order correlations may not be.

Independent components analysis (ICA): the separation of a signal into components which are statistically independent.

Maximum likelihood (ML) and maximum a posterior (MAP) decoding: In a neural context, ML decoding usually means choosing the stimulus that is most likely to have generated an observed response, while MAP decoding means choosing instead the stimulus that is most likely given the observed response. ML and MAP decoding may give different answers when not all stimuli are equally likely.

Neural encoding and decoding: encoding refers to the mapping from a stimulus to a neural response, while decoding refers to the mapping fr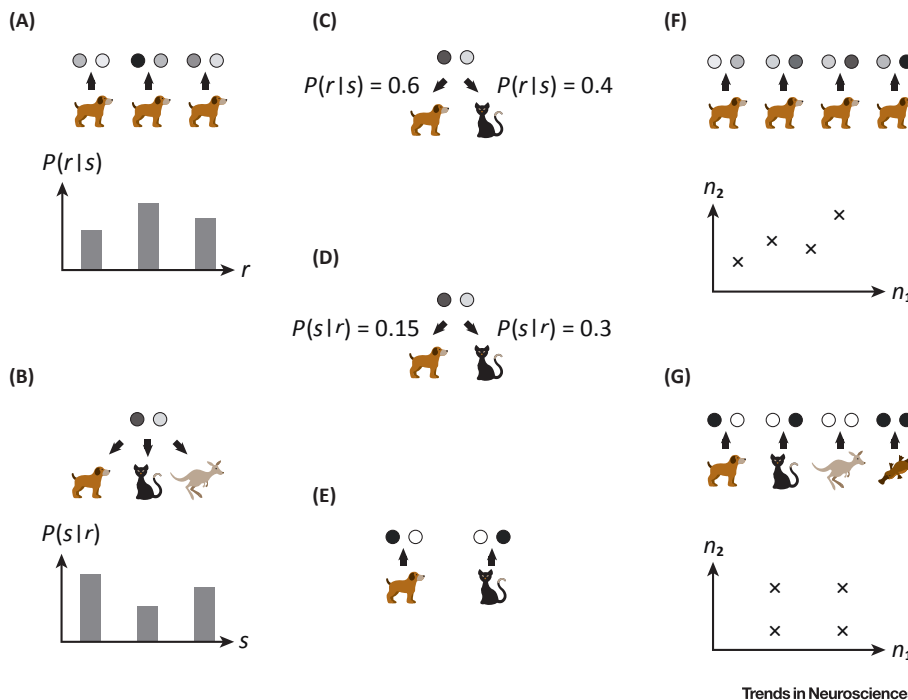om a neural response to a stimulus. In general these mappings are probabilistic: each stimulus can generate a range of responses, and each response could have been generated by a range of stimuli.

Noise correlations: two neurons are noise-correlated if their responses to the same stimulus are correlated. This indicates that a process causing the response of each neuron to vary is shared between the two neurons.

Optimal decoder: a decoder which optimizes some measure of the quality of the decoding, for instance its variance.

Principal components analysis (PCA): the projection of data from a high-dimensional space to a low-dimensional space which preserves the maximum amount of variance in the data.

Posterior distribution: the probability distribution of a variable (e.g., stimuli) taking into account specific evidence (e.g., an evoked neural response) relating to that variable.
Prior distribution: the probability distribution of a variable (e.g., stimuli) before specific evidence (e.g., an evoked neural response) is available to constrain the distribution.
Sparse coding: a type of coding where only a small proportion of neurons are active at any moment (population sparseness), or alternatively each neuron is activated only rarely (lifetime sparseness).

Trends in Neurosciences

Figure 1. Principles of Neural Coding. (A) In encoding we consider the probability distribution over responses for each stimulus. Two neurons are represented by grey circles, with their levels of activity being denoted by the shade of grey. The probability distribution $P(r|s)$ gives the likelihood of each response (where 'response' means the activity of both neurons) generated by a particular stimulus, in this case a dog. The arrows point upwards to represent encoding. (B) In decoding we consider the probability distribution over stimuli for each response. Given the particular response of the two neurons, $P(s|r)$ gives the probability of each of the three stimuli. The arrows point downwards to represent decoding. (C) In ML decoding we ask which stimulus has the highest probability of generating the observed response pattern. We imagine here that when the dog stimulus is presented it generates the observed response 60% of the time [$P(r|dog) = 0.6$], as compared to only 40% of the time for the cat stimulus [$P(r|cat) = 0.4$], hence the dog is the ML estimate for the stimulus given the response. (D) In MAP decoding we ask which stimulus was most likely to have generated the response, taking into account the relative frequencies of the stimuli. We imagine here that the probabilities of response given stimulus $P(r|s)$ are the same as in panel (C), but that the cat stimulus occurs in the world threefold more frequently than the dog stimulus. Now, when this particular response occurs it is more likely to have been a cat, because the cat was more frequent than the dog (by Bayes theorem $p(cat|r) = \frac{0.4 \times 0.75}{0.4+0.6}$, whereas $p(dog|r) = \frac{0.6 \times 0.25}{0.4+0.6}$). The cat is therefore the MAP estimate for the stimulus given the output. (E) In sparse coding (here population sparseness is shown) only a small proportion of output neurons from the population (in this case only one) are active in response to each stimulus. (F) Two output neurons are noise-correlated when fluctuations in their activity in response to the same stimulus are correlated. This is represented here by four repetitions of the dog stimulus, eliciting four different but correlated responses from the output neurons. This correlation is seen in the graph of the response of neuron 1 versus the response of neuron 2. (G) In factorial coding the activity of each of the output neurons is statistically independent from other output neurons across the set of stimuli, represented here by four stimuli generating four responses such that there is no correlation in the activity of the two neurons.

slightly weaker case is a decorrelation code, where second-order (but not necessarily higher-order) correlations are zero. Neurons in the real brain are often **noise-correlated**, meaning that stochastic fluctuations around the mean response are correlated between neurons. Such correlations can both increase or decrease the amount of information encoded compared to when noise correlations are not present [9], and affect the form of the **optimal decoder** [17,18]. Another important concept in population coding is **sparse coding**. The term carries two different meanings: the first is that only a few neurons are activated by each stimulus (population sparseness [19]), whereas the second is that each neuron is activated only rarely (lifetime sparseness [20]). As discussed below, explicitly optimizing particular measures of

coding sparseness for sensory stimuli can lead to receptive fields matching those seen biologically, such as orientation-selective cells in V1. Factorial and sparse coding are particular types of **efficient coding** in which the overarching goal is to form codes that minimize some energy cost, such as the number of spikes required.

Another important distinction is that spike trains can encode information (e.g., the level of activation of a sensory receptor, or the desired activation of a particular muscle) either through a rate code or a temporal code [21]. In the former, information is carried by the number of spikes that occur in some time-window, while in the latter the precise times at which the spikes occur is important. Temporal codes can carry many-fold more information per spike than rate codes [22], but are less robust than rate codes because they require more precise time measurements. While there is evidence for temporal as well as rate codes in adult animals [23,24], whether both exist in the developing nervous system remains unclear.

Topographic Maps

A crucial part of neural development is the formation of topographic maps between brain areas, the best-studied example being the map between the eye and the optic tectum/superior colliculus. These maps are driven initially by molecular cues, and then are later refined by neural activity. What role do these maps play in neural coding? A simple idea, which we refer to as topographic coding, is that stimulus information is represented by the spatial position of neural activity in the target structure, and is then read out by appropriately topographic maps projecting to downstream structures. For instance, the position of a spot in the visual field could be coded by a corresponding position of localized activity in the tectum. One of the attractions of topographic coding is that it can be established by molecular cues and/or spontaneous activity without sensory experience, in other words be ready to go early in development. The size of retinal projective fields refine over development [25], but counter-intuitively this does not strongly affect the quality of topographic coding [26] (Figure 2A). Instead, an important limiting factor is the degree of topographic order in the map, and early in development this is usually fairly crude. Decoding experimentally measured population activity in the larval zebrafish tectum in response to stimuli consisting of small stationary spots demonstrated that the decoded spot positions were often confused with each other [26].

However, other coding schemes are possible. The scheme that is under some circumstances statistically optimal is **maximum likelihood (ML) decoding** (Box 1). In this context, this means gathering statistics on the likelihood of each neuron's level of activity given each stimulus, and then decoding a pattern of activity over the population of neurons as the stimulus that was most likely to have produced that pattern (usually for simplicity assuming each neuron acts independently). Using ML decoding for the same patterns of population activity in the larval zebrafish tectum, as mentioned above, showed almost no confusion of spot positions [26]. This demonstrates that the tectal activity contained all the information needed to accurately decode spot position, but that topographic decoding was not a particularly effective way of extracting that information (Figure 2B). Under some conditions topographic decoding is statistically equivalent to ML decoding, but these conditions are almost certainly violated in the developing nervous system. ML decoding does not explicitly use the spatial positions of the neurons in the tectum, but it does require statistical information about the mapping between stimuli and responses, which must be obtained from experience. Together these observations raise the possibility that, unless topography sufficiently refines over development, a developmental switch could occur between topographic and more statistically optimal types of decoding as the animal gains more experience of the world.
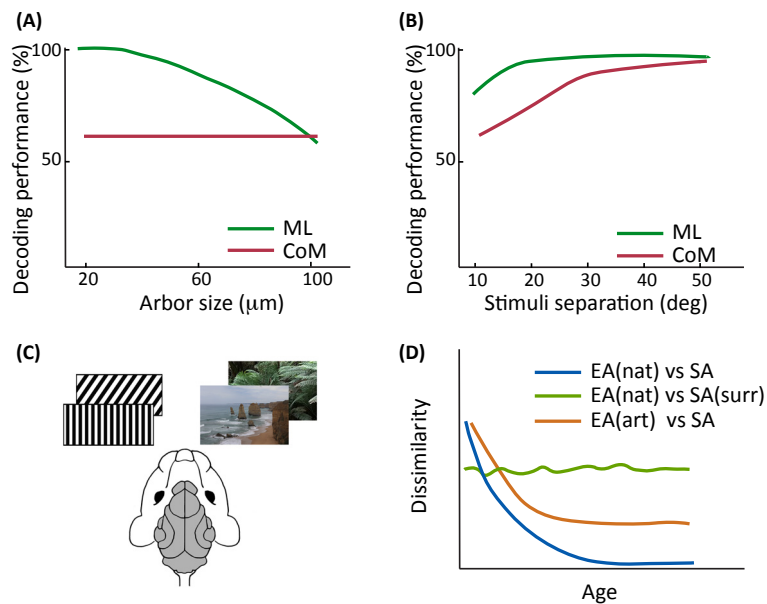
Trends in Neurosciences

Figure 2. Analyses of Topography and Spontaneous Activity in Neural Coding. (A,B) The dependence of decoding performance on system parameters in a simple computational model of the zebrafish retinotectal system (redrawn, with permission, from [26]). The performance of a topography-based decoder (center of mass, CoM) and a maximum likelihood (ML) decoder was compared. (A) While the performance of the ML decoder decreased as retinal arbor size increased, CoM performance was unaffected. The bluemenkhol mutant fish has wider retinal arbors and decreased visual acuity compared to wild-type fish [81], and this is more consistent with ML than CoM decoding. (B) Performance of the CoM decoder increased more slowly with stimulus separation compared to the ML decoder (deg, degrees). (C,D) Multiunit activity was recorded in the visual cortex of awake ferrets over development while viewing natural (nat) and artificial (art) stimuli (evoked activity, EA), or in the absence of visual cues (spontaneous activity, SA). The dissimilarity between evoked and spontaneous activity decreased with age, but more so for natural than for artificial stimuli. However dissimilarity did not decrease between evoked and spontaneous activity when temporal correlations in the latter were removed (surr, surrogate data) (adapted, with permission, from [46]).

A particularly challenging aspect of retinotectal map development in fish and amphibians is that the retina adds new neurons at the margin nearest the lens and expands radially, while the tectum expands by adding new cells linearly along its periventricular proliferative zone. To maintain a uniform topographic map, retinal axons must therefore continually shift their retinal termination zones [27]. While newly added tectal neurons take a few days to functionally integrate within tectal circuits [28], how they begin to contribute to coding from a quantitative perspective is unknown.

Spontaneous Activity

Spontaneous neural firing begins very early in nervous system development. One of the best-studied examples is spontaneous retinal waves. These are a robust phenomenon that is observed across many species [29–32]. These waves occur before light transduction by photoreceptors, and disappear around the time of eye-opening in mammals, regardless of visual experience [33]. From a coding perspective, retinal waves have an obvious interpretation: they are providing central structures with sensory statistics similar to those the animal will encounter once its sensory apparatus is better developed [34]. This allows activity-dependent learning rules to begin to adjust synaptic strengths in an appropriate direction [34,35]. Indeed this can also provide a counterbalance to unusual sensory experience early in life. For instance, when ferrets were raised seeing only spots of light (rather than the oriented edges that are

ubiquitous in normal visual experience), they still developed oriented receptive fields, suggesting that the development of such receptive fields is not dependent on experience [36]. Of note, in these experiments spontaneous retinal waves were not blocked, and it was subsequently shown using a sparse-coding model that only a small proportion of normal retinal wave activity – relative to direct visual experience of spots only – was required for oriented rather than spot-like receptive fields to develop [37].

Spontaneous activity also occurs in more central structures such as the cortex [38–40]. Input from sensory receptors plays an important role in driving some of this activity, but some of it is endogenously generated. For instance, when both eyes were removed during the first day of zebrafish development, spontaneous activity still emerged over the next few days in the optic tectum, sharing statistical similarities with tectal spontaneous activity in normal zebrafish [41,42]. Spontaneous patterns in central structures have been suggested to play a role in the encoding process, where stimuli recruit spontaneously generated patterns to represent stimulus features [43]. Supporting this idea is the observation that the spatiotemporal correlation structure in ferret primary visual cortex was only mildly modulated by sensory inputs [44], and could predict the overall structure of visually-evoked orientation columns [83]. It has also been suggested that spontaneous activity in the cortex acts as a self-correcting mechanism, whereby the network modifies erroneous representations of input variables with additional spiking, hence minimizing coding error [45].

Another key conceptual idea for the role of spontaneous activity is that it provides a representation of a Bayesian prior over stimuli $P(s)$ (Box 1). Among the most direct supports for this idea is the demonstration that the distance between the probability distributions for evoked and spontaneous activity reduces over development in ferrets [46]. This suggests that the statistics of spontaneous activity come to match those of activity evoked by $P(s)$ (Figure 2C,D). Similarity between evoked and spontaneous patterns of neural activity has also been observed in mice and fish [43,47]. However, this match only occurs (at least in ferrets) for natural but not artificial stimuli [46]. In humans, spontaneous activity preceding evoked responses accounted for a large portion of both neuronal and behavioral variability [48], as well as influencing perception [49], providing further evidence for a role for these patterns in the decoding process.

Receptive Field and Circuit Development

The emergence over development of structured receptive fields that encode information beyond topography, such as orientation selectivity in the visual system, has been a key target for theories of neural coding [12]. Sensory space is very high-dimensional (e.g., an image of size 1000×1000 pixels can be thought of as a point in a 1000 000-dimensional space). Important goals are therefore to understand the underlying regularities of this space, how these can be usefully represented in a population of neurons, and how these representations could be learned during development. The obvious mechanism for implementing such plasticity is through Hebbian learning.

Principal components analysis (PCA) performs dimension reduction by projecting points in the high-d space into a low-d space, in a manner where the axes of the low-d space correspond to the directions which preserve the most variance in the high-d space. Neurons can find these principal components via simple Hebbian learning rules [50]. The first two principal components of natural images are oriented edges; however, they are not localized in space, unlike the oriented receptive fields seen in primary visual cortex. More successful in this regard is **independent components analysis** (ICA), which finds statistically independent (rather than only uncorrelated, as in PCA) dimensions of the input space [51]. In addition to

accounting for aspects of sensory maps under normal conditions, ICA also captures changes that occur in cortical receptive fields when visual input is disrupted, for instance through monocular deprivation, stripe rearing, and strabismus [52]. However, the biological learning rules that could find independent components are less clear, although several suggestions have been made [53,54].

ICA is an example of sparse coding: each neuron is not active very often, but when it is active it is highly active. The probability distribution of the activity of each neuron is therefore relatively sharply peaked and has heavier tails than a Gaussian distribution. These properties can be quantified using kurtosis (a measure related to the 4th moment of the distribution), and sparse coding as described above corresponds to relatively high kurtosis measures. Sparse coding provides an intermediate point between local codes, where each neuron uniquely represents only one possible input, and dense codes, where a large proportion of neurons are active for each input. Local codes are easy to decode but have limited representational power: the number of neurons required scales with the number of inputs. By contrast, dense codes are harder to decode, but the number of neurons required scales only as the log of the number of inputs [55]. Multiple lines of evidence support the idea that neural codes become sparser over development, suggesting that forming sparse codes is a key goal of neural development [56–58]. However, what this means for hierarchical representations is unclear. Clearly, the statistical structure of the stimuli an animal receives is a function not only of the world but also the animal's own behavior. An intriguing recent proposal is that a developing organism might adapt its behavioral and neural coding strategies in a coordinated manner to maximize coding efficiency – so-called active efficient coding [59]. This strategy, using both computer simulation and implementation on a real robot, was shown to be capable of learning smooth-pursuit eye movements [60].

In mammals there is a distinction between two phases of circuit development – an early, experience-independent major establishment phase and a subsequent experience-dependent phase of further intensive circuitry changes (e.g., [61]). In ferrets the first phase is characterized by frequent retinal and cortical waves [38], the formation of rough topographic maps, and the onset of a map-refinement process in parallel to changes in the proportions of inhibitory neurons [62–64] (Figure 3A). At this stage orientation maps are recognizable, although homogeneity in columnar structure, signal strength, and mature orientation selectivity is not achieved until 3 weeks after eye opening [36,65–67]. The onset of visual experience at around postnatal day 30 coincides with a dramatic increase in the density of cortical synapses in all cortical layers, and a comparable increase in the outgrowth of both short- and long-range axonal projections [68]. A key substrate for forming sparse codes is likely the maturation of lateral inhibitory connections. By mutually inhibiting the activity of neurons that would otherwise fire together, such connections can reduce correlations in neural firing and thus reduce the redundancy in the representation [69,70]. Because both excitatory and inhibitory circuits refine after eye opening [71–73], and are susceptible to changes in visual experience [74–76], the effectiveness of lateral inhibition likely changes during postnatal development. However, the extent to which this may enhance the processing of visual cues is unknown.

The roles of neuronal selectivity, variance, and noise correlations in neural coding have recently been investigated in the context of direction coding in the ferret visual cortex [58]. Over a brief period of time following eye opening, the active population of neurons underwent a striking transformation from a highly dense response with complex spatiotemporal wave-like dynamics to a sparse distribution of active neurons [56–58]. This transformation was characterized by a decline in response variance and pairwise noise correlations, occurring with the same time-course and in the same neuronal population as the rise of direction selectivity. Longitudinal
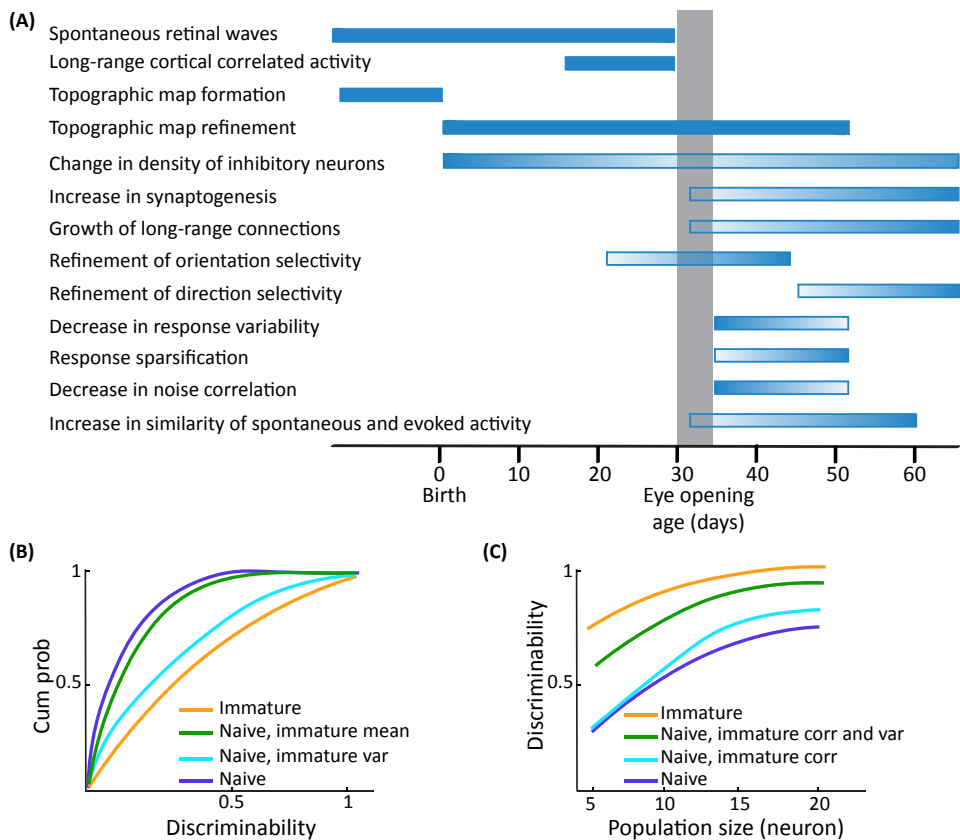
Trends in Neurosciences

Figure 3. The Development of the Neural Code in the Ferret Visual Circuit. (A) Schematic representation of some of the key processes related to neural coding occurring in the developing ferret visual system. The vertical grey bar indicates the age for eye opening. Horizontal blue gradient bars indicate a trend over development (white/low, blue/high). Pre-vision processes guide the establishment of the circuit [36,38,62–64]. Subsequent vision-guided processes [68,74,82] refine key properties of neural code [46,58,66,67]. (B,C) Reduction of variance is crucial for improved direction discrimination (redrawn, with permission, from [58]). Two-photon calcium imaging of visual cortex in response to moving grating stimuli in eight different directions was performed on naive (postnatal days P29–32) and immature (P33–36) ferrets. (B) Direction discriminability for single neurons increased from naive to immature animals (cum prob, cumulative probability). Combining naive levels of mean responses with immature levels of variance resulted in near-immature levels of discriminability. By contrast, combining naive levels of variance (var) with immature levels of mean response gave only a small improvement in discriminability over the naive case. This suggests that, over development, variance reduction plays a more important role in the improvement of discriminability than increases in selectivity. (C) Direction discriminability for the neural population increased from naive to immature animals. Combining naive direction selectivity with immature variance and correlation (corr) structure resulted in near-immature levels of discriminability. This effect was smaller when only immature correlation structure was used. This again suggests a crucial role for variance reduction over development in improving discriminability.

imaging in the same animals evaluated the contribution of each of these changes to the increase in discriminability between directions (Figure 3B,C). A reduction in response variance played a stronger role in improving discriminability than the increase in selectivity. High noise correlations present at eye opening limited discriminability, and removal of noise correlations (via data shuffling, hybrid datasets, or decreased noise correlation with visual experience) resulted in improved discriminability. Furthermore, the selectivities and structure of noise correlations were affected by visual experience, emphasizing the role of experience in shaping the neural code.

Concluding Remarks

Despite our increasing knowledge of the mathematical principles underlying neural coding, how coding emerges over development remains an open question (see Outstanding Questions). In the adult brain it has been suggested that different neuronal subpopulations may use different coding strategies, and additionally that these strategies may vary over time [10]. During development, neurons can change their information transmission capabilities, for instance by gain scaling [77]. Instead of thinking of development as simply elaborating and refining one specific coding strategy, it could be more meaningful to consider how a repertoire of coding strategies, and the temporal dynamics of the transitions between these, emerges in the face of the storm of anatomical and physiological changes occurring in both the brain and body during development.

In addition to being of obvious relevance to understanding neurodevelopmental disorders [78], deeper understanding of neural coding during development could also be important for the construction of future computing technologies. Recent developments in artificial intelligence (AI), although impressive (e.g., [79]), have typically occurred within a framework of fixed hardware built by human engineers, rather than via hardware that needed to construct itself while interacting productively with the world. In the same way as current AI algorithms such as deep learning originally took inspiration from neuroscience [80], our evolving understanding of neural development might ultimately lead to a new generation of breakthroughs in AI.

References

1. Budick, S. and O'Malley, D. (2000) Locomotor repertoire of the larval zebrafish: swimming, turning and prey capture. *J. Exp. Biol.* 203, 2565–2579
2. Gahtan, E. *et al.* (2005) Visual prey capture in larval zebrafish is controlled by identified reticulospinal neurons downstream of the tectum. *J. Neurosci.* 25, 9294–9303
3. Rolls, E.T. and Treves, A. (2011) The neuronal encoding of information in the brain. *Prog. Neurobiol.* 95, 448–490
4. Laughlin, S.B. (2001) Energy as a constraint on the coding and processing of sensory information. *Curr. Opin. Neurobiol.* 11, 475–480
5. Wark, B. *et al.* (2007) Sensory adaptation. *Curr. Opin. Neurobiol.* 17, 423–429
6. Ruthazer, E.S. and Aizenman, C.D. (2010) Learning to see: patterned visual activity and the development of visual function. *Trends Neurosci.* 33, 183–192
7. Seabrook, T.A. *et al.* (2017) Architecture, function, and assembly of the mouse visual system. *Ann. Rev. Neurosci.* 40, 499–538
8. Borst, A. and Theunissen, F.E. (1999) Information theory and neural coding. *Nat. Neurosci.* 2, 947–957
9. Averbeck, B.B. *et al.* (2006) Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366
10. Quiroga, R.Q. and Panzeri, S. (2013) *Principles of Neural Coding,* CRC Press
11. Panzeri, S. *et al.* (2017) Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* 93, 491–507
12. Linsker, R. (1988) Self-organization in a perceptual network. *Computer* 21, 105–117
13. Ma, W.J. *et al.* (2006) Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438
14. Series, P. and Seitz, A. (2013) Learning what to expect (in visual perception). *Front. Hum. Neurosci.* 7, 668
15. Pouget, A. *et al.* (2013) Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 16, 1170–1178
16. Yuste, R. (2015) From the neuron doctrine to neural networks. *Nat. Rev. Neurosci.* 16, 487–497
17. Wu, S. *et al.* (2001) Population coding with correlation and an unfaithful model. *Neural Comput.* 13, 775–797
18. Nirenberg, S. and Latham, P.E. (2003) Decoding neuronal spike trains: how important are correlations? *Proc. Natl. Acad. Sci.* 100, 7348–7353
19. Olshausen, B.A. and Field, D.J. (2004) Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487
20. Willmore, B. and Tolhurst, D. (2001) Characterising the sparseness of neural codes. *Netw. Comput. Neural Syst.* 12, 255–270
21. Usrey, W.M. and Reid, R.C. (1999) Synchronous activity in the visual system. *Annu. Rev. Physiol.* 61, 435–456
22. Rieke, F. *et al.* (1999) *Spikes: Exploring the Neural Code,* MIT Press
23. Hafting, T. *et al.* (2008) Hippocampus-independent phase precession in entorhinal grid cells. *Nature* 453, 1248–1252
24. Nemenman, I. *et al.* (2008) Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS Comput. Biol.* 4, e1000025
25. Dong, W. and Aizenman, C.D. (2012) A competition-based mechanism mediates developmental refinement of tectal neuron receptive fields. *J. Neurosci.* 32, 16872–16879
26. Avitan, L. *et al.* (2016) Limitations of neural map topography for decoding spatial information. *J. Neurosci.* 36, 5385–5396
27. Reh, T. and Constantine-Paton, M. (1984) Retinal ganglion cell terminals change their projection sites during larval development of *Rana pipiens*. *J. Neurosci.* 4, 442–457
28. Boulanger-Weill, J. *et al.* (2017) Functional interactions between newborn and mature neurons leading to integration into established neuronal circuits. *Curr. Biol.* 27, 1707–1720

Outstanding Questions

How does the quality of neural decoding change over development?

Is there only one neural code that is refined over development, or are there many codes that vary over space and time?

Do coding strategies switch during development?

How are the probability distributions required for neural decoding maintained and updated over development?

How strongly is the development of coding strategies influenced by sensory experience?

Does spontaneous activity during development play a key role in neural coding and decoding?

How does neural coding adapt to continually shifting connections, as occurs for instance in retinotectal mapping in fish and frogs?

Are there principles of neural development which can provide new inspiration for AI?

29. Ackman, J.B. and Crair, M.C. (2014) Role of emergent neural activity in visual map development. *Curr. Opin. Neurobiol.* 24, 166–175

30. Warland, D.K. *et al.* (2006) Dynamics of spontaneous activity in the fetal macaque retina during development of retinogeniculate pathways. *J. Neurosci.* 26, 5190–5197

31. Zhang, R.W. *et al.* (2016) Stereotyped initiation of retinal waves by bipolar cells via presynaptic NMDA autoreceptors. *Nat. Commun.* 7, 12650

32. Richter, L.M. and Gjorgjieva, J. (2017) Understanding neural circuit development through theory and models. *Curr. Opin. Neurobiol.* 46, 39–47

33. Demas, J. *et al.* (2003) Developmental loss of synchronous spontaneous activity in the mouse retina is independent of visual experience. *J. Neurosci.* 23, 2851–2860

34. Albert, M.V. *et al.* (2008) Innate visual learning through spontaneous activity patterns. *PLoS Comput. Biol.* 4, e1000137

35. Butts, D.A. *et al.* (2007) A burst-based 'Hebbian' learning rule at retinogeniculate synapses links retinal waves to activity-dependent refinement. *PLoS Biol.* 5, e61

36. Ohshiro, T. *et al.* (2011) Development of cortical orientation selectivity in the absence of visual experience with contour. *J. Neurophysiol.* 106, 1923–1932

37. Hunt, J.J. *et al.* (2012) Sparse coding on the spot: spontaneous retinal waves suffice for orientation selectivity. *Neural Comput.* 24, 2422–2433

38. Chiu, C. and Weliky, M. (2001) Spontaneous activity in developing ferret visual cortex in vivo. *J. Neurosci.* 21, 8906–8914

39. Ackman, J.B. *et al.* (2012) Retinal waves coordinate patterned activity throughout the developing visual system. *Nature* 490, 219–225

40. Shen, J. and Colonnese, M.T. (2016) Development of activity in the mouse visual cortex. *J. Neurosci.* 36, 12259–12275

41. Avitan, L. *et al.* (2017) Spontaneous activity in the zebrafish tectum reorganizes over development and is influenced by visual experience. *Curr. Biol.* 27, 2407–2419

42. Pietri, T. *et al.* (2017) The emergence of the spatial structure of tectal spontaneous activity is independent of visual inputs. *Cell Rep.* 19, 939–948

43. Miller, J.E. *et al.* (2014) Visual stimuli recruit intrinsically generated cortical ensembles. *Proc. Natl. Acad. Sci.* 111, E4053–E4061

44. Fiser, J. *et al.* (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* 431, 573–578

45. Koren, V. and Denève, S. (2017) Computational account of spontaneous activity as a signature of predictive coding. *PLoS Comput. Biol.* 13, e10053554

46. Berkes, P. *et al.* (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87

47. Romano, S.A. *et al.* (2015) Spontaneous neuronal network dynamics reveal circuit's functional adaptations for behavior. *Neuron* 85, 1070–1085

48. Schölvinck, M.L. *et al.* (2012) The influence of spontaneous activity on stimulus processing in primary visual cortex. *Neuroimage* 59, 2700–2708

49. Boly, M. *et al.* (2007) Baseline brain activity fluctuations predict somatosensory perception in humans. *Proc. Natl. Acad. Sci.* 104, 12187–12192

50. Dayan, P. and Abbott, L.F. (2001) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems,* MIT Press

51. Hyvrinen, A. *et al.* (2009) *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision,* Springer

52. Hunt, J.J. *et al.* (2013) Sparse coding can predict primary visual cortex receptive field changes induced by abnormal visual input. *PLoS Comput. Biol.* 9, e1003005

53. Savin, C. *et al.* (2010) Independent component analysis in spiking neurons. *PLoS Comput. Biol.* 6, e1000757

54. Isomura, T. and Toyoizumi, T. (2016) A local learning rule for independent component analysis. *Sci. Rep.* 6, 28073

55. Földiák, P. (2003) Sparse coding in the primate cortex. In *The Handbook of Brain Theory and Neural Networks* (Arbib, M.A., ed.), pp. 1064–1068, MIT Press

56. Rochefort, N.L. *et al.* (2009) Sparsification of neuronal activity in the visual cortex at eye-opening. *Proc. Natl. Acad. Sci.* 106, 15049–15054

57. Ikezoe, K. *et al.* (2012) Decorrelation of sensory-evoked neuronal responses in rat barrel cortex during postnatal development. *Neurosci. Res.* 73, 312–320

58. Smith, G.B. *et al.* (2015) The development of cortical circuits for motion discrimination. *Nat. Neurosci.* 18, 252–261

59. Zhao, Y. *et al.* (2012) A unified model of the joint development of disparity selectivity and vergence control. In *IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, IEEE

60. Teulière, C. (2015) Self-calibrating smooth pursuit through active efficient coding. *Robot. Auton. Syst.* 71, 3–12

61. Crair, M.C. *et al.* (1998) The role of visual experience in the development of columns in cat visual cortex. *Science* 279, 566–570

62. Huberman, A.D. *et al.* (2008) Mechanisms underlying development of visual maps and receptive fields. *Ann. Rev. Neurosci.* 31, 479–509

63. Tavazoie, S.F. and Clay Reid, R. (2000) Diverse receptive fields in the lateral geniculate nucleus during thalamocortical development. *Nat. Neurosci.* 3, 608–616

64. Gao, W.J. *et al.* (2000) Development of inhibitory circuitry in visual and auditory cortex of postnatal ferrets: immunocytochemical localization of calbindin- and parvalbumin-containing neurons. *J. Comp. Neurol.* 422, 140–157

65. Chapman, B. and Stryker, M. (1993) Development of orientation selectivity in ferret visual cortex and effects of deprivation. *J. Neurosci.* 13, 5251–5262

66. White, L.E. *et al.* (2001) The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature* 411, 1049–1052

67. Li, Y. *et al.* (2006) The development of direction selectivity in ferret visual cortex requires early visual experience. *Nat. Neurosci.* 9, 676–681

68. White, L.E. and Fitzpatrick, D. (2007) Vision and cortical map development. *Neuron* 56, 327–338

69. Srinivasan, M.V. *et al.* (1982) Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* 216, 427–459

70. Pecka, M. *et al.* (2014) Experience-dependent specialization of receptive field surround for selective coding of natural scenes. *Neuron* 84, 457–469

71. Katagiri, H. *et al.* (2007) Optimization of somatic inhibition at critical period onset in mouse visual cortex. *Neuron* 53, 805–812

72. Kuhlman, S.J. *et al.* (2011) Fast-spiking interneurons have an initial orientation bias that is lost with vision. *Nat. Neurosci.* 14, 1121–1123

73. Ko, H. *et al.* (2013) The emergence of functional microcircuits in visual cortex. *Nature* 496, 96–100

74. Ruthazer, E.S. and Stryker, M.P. (1996) The role of activity in the development of long-range horizontal connections in area 17 of the ferret. *J. Neurosci.* 16, 7253–7269

75. Maffei, A. *et al.* (2004) Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation. *Nat. Neurosci.* 7, 1353–1359

76. Chattopadhyaya, B. *et al.* (2004) Experience and activity-dependent maturation of perisomatic GABAergic innervation in primary visual cortex during a postnatal critical period. *J. Neurosci.* 24, 9598–9611

77. Gjorgjieva, J. *et al.* (2014) Intrinsic neuronal properties switch the mode of information transmission in networks. *PLoS Comput. Biol.* 10, e1003962

78. Rubenstein, J.L. (2011) Annual research review. Development of the cerebral cortex: implications for neurodevelopmental disorders. *J. Child Psychol. Psychiatry* 52, 339–355

79. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444

80. Hassabis, D. *et al.* (2017) Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258

81. Smear, M.C. *et al.* (2007) Vesicular glutamate transport at a central synapse limits the acuity of visual perception in zebrafish. *Neuron* 53, 65–77

82. Erisir, A. and Harris, J.L. (2003) Decline of the critical period of visual plasticity is concurrent with the reduction of NR2B subunit of the synaptic NMDA receptor in layer 4. *J. Neurosci.* 23, 5208–5218

83. Smith, G.B. *et al.* (2018) Long-range order from local interactions: organization and development of distributed cortical networks. *bioRxiv* Published online 2 June 2018. http://dx.doi.org/10.1101/335828