Chapter 4

Why let networks grow?

Thomas R. Shultz, Shreesh P. Mysore,
and Steven R. Quartz

> 'Let it grow, Let it grow.'
>
> Grateful Dead, 1973, Eric Clapton, 1974,
> Boogie Brown Band, 2003

Introduction

Beyond the truism that to develop is to change, the process of developmental change itself has been relatively neglected in developmental psychology. Methodologically, most behavioural studies have traditionally utilized a cross-sectional approach. This has revealed a great deal about how certain behavioural and cognitive abilities differ at various points in development, but it tends to reveal less about the developmental processes that operate to transform those cognitive and behavioural abilities. A variety of factors account for the lack of explanations of developmental change, ranging from the methodological challenges their study entails to principled, learning-theoretic arguments against the very existence of such developmental processes (Macnamara, 1982; Pinker, 1984).

Among the most influential arguments against developmental change was Chomsky's (1980) instantaneity argument that starts with the null hypothesis that children are qualitatively similar learners to adults and supposes that development is instantaneous, in the sense that there are no time dependencies in development. Departing from this view and supposing that children are initially more restricted learners only results in weakening their acquisition properties. The upshot is that *less is less*, and development only reduces learning capability. Thus, as a sort of charity principle, one ought to at least start with the hypothesis that children do not differ substantially from adults, because explaining development is difficult enough without having to do so with greatly diminished learning ability.

These sorts of arguments led to the widespread assumption that there is little theoretical insight to be gained from studying processes of developmental change. Indeed even with the advent of connectionist models and their adoption in developmental psychology, most of the models that were used were qualitatively similar to models of adult learning. That is, most models used a fixed feedforward architecture in which the main free parameters were connection strengths. Therefore such models implicitly adopted Chomsky's argument and began with the assumption that the immature and mature states are qualitatively identical.

Yet how well-founded is this assumption? Although Chomsky's approach is often considered to be a biological one, it is actually quite unbiological in the sense that it is not developmental. Genuine biological approaches invariably emphasize the importance of development (Gilbert, 2003). Viewed from the perspective of developmental cognitive neuroscience, it is now well established that the structural features of neural circuits undergo substantial alterations throughout development (Quartz and Sejnowski, 1997). Do such changes add anything of interest to the explanation of developmental change? If so, does a better understanding of the processes of developmental change undermine Chomsky's argument, and thereby demonstrate that developmental change is crucial for understanding the nature of cognitive development?

In this chapter we ask the question, why let networks grow? We begin by reviewing the wealth of accumulating data from neuroscience that network growth appears to be a much more central feature of learning than is traditionally assumed. Indeed it appears that the assumption that the main modifiable parameters underlying learning are changes in connection strength in an otherwise fixed network may need revision. We discuss this question with an eye toward evidence for learning-directed architectural changes in the brain throughout the lifespan. Next, we consider the computational implications of learning-directed growth. There we consider whether less really is less, or whether by breaking down the traditional distinction between intrinsic maturation and cognitive processes of learning, the learning mechanism underlying cognitive development thereby becomes substantially different, and more powerful, in its learning properties than a fixed architecture. We then present a case in which activity-dependent network growth underlies important acquisition properties in a model system, auditory localization in the barn owl. This non-human system is very well characterized in terms of biological underpinnings, which provide important clues into the likely biological mechanisms underlying human cognitive development, and which give rise to general computational principles of activity-dependent network growth. Then we present evidence from modelling cognitive development in children. There, we explore a number of computational simulations utilizing the cascade-correlation (CC) algorithm. In contrast to the more commonly used back-propagation algorithm, CC starts with a minimal network and adds new units as a function of learning. CC can be viewed as an abstract rule for neural growth, and thus as a means of exploring the computational and learning properties of such developmental processes. In the final section we present the main conclusions from this work and point to areas of future research and open research questions.

Experience-dependent architectural plasticity

Two broad categories of neural plasticity can be distinguished based on the nature of expression and encoding of change in the nervous system. They are 'synaptic efficacy change' and 'architectural plasticity'. The former, as the name suggests, refers to changes to existing, functional synapses in a network whose topology is kept fixed, while the latter refers to changes in the underlying topology of the network. Whereas synaptic efficacy change has been the predominantly studied form of plasticity (Martin and Morris, 2002), neuroscience research also provides evidence for activity-dependent plasticity. Structural mechanisms largely implement architectural plasticity, and they include dendritic spine motility, spinogenesis (the induction of new spines), synaptogenesis (the formation of new synapses), and neurogenesis (the growth of new neurons). We provide a brief sketch of the research literature in these areas.

Most excitatory synapses in the brain are formed on small protrusions from neuronal dendrites called spines. Spines are extremely important postsynaptic structures containing several neurotransmitter receptors, and various biochemical signaling mechanisms, among other elements. Based on the synaptic input they receive, spines trigger different signaling pathways that can result in short-term or long-term synaptic changes. Rapid morphological plasticity of spines, spine movement and their growth and retraction are collectively referred to as spine motility. Motility was found to occur spontaneously in hippocampal neurons in slice cultures from rats (Dailey and Smith, 1996; Fischer *et al.*, 1998). Since its discovery, one of the main functions attributed to motility is that of exploring the extracellular space in search of presynaptic partners. This view is supported by the finding that formation of synapses stabilizes previously mobile dendritic protrusions (Dailey and Smith, 1996). Motility has also been found in response to electrical stimulation of neurons (Maletic-Savatic *et al.*, 1999), thus suggesting that experience may play a role in regulating motility. Direct evidence for this possibility stems from the finding that trimming the whiskers of rats, thereby causing significant sensory deprivation, resulted in a 40 per cent drop in spine motility in corresponding whisker barrels in the rodent brain (Lendvai *et al.*, 2000).

While the above studies looked directly at movement and morphological changes as they occurred, other studies have looked at the effects of motility indirectly by comparing the density of spines and their morphological distribution between control and learning conditions. In a trace eye-blink conditioning task in rats, it was found that the density of spines in the basal CA1 dendrites of the hippocampus—an area that is known to be important for encoding such memories—was greater after conditioning (Leuner *et al.*, 2003). Interestingly, when the acquisition of this association was blocked using a pharmacological agent that interferes with the formation of long-term memories, this spine density increase was blocked as well. Spatial training of rats, which is known to produce an increase in their subsequent ability to learn in spatial tasks, also produces an increase in the density of spines in the hippocampus (Moser *et al.*, 1997). Other tasks like odour discrimination training (Knafo *et al.*, 2001), visual stimulation, and even space flight (Yuste and Bonhoeffer, 2001), have been found to produce spine density increases in the appropriate brain regions in rats. A case can thus be made that experience-dependent plasticity in dendritic spines may facilitate architectural reorganization of neural circuits in response to functional demands.

In addition to the above studies that examined spine changes, other studies have looked directly at changes in synapse numbers in response to learning. For instance, Chang and colleagues (1991) showed that long-term potentiation increased synaptic numbers in two-year-old rats *in vitro*. Other studies report an increase in the cerebellar synapse density in rats following classical eye-blink conditioning experiments (Black *et al.*, 1990; Kleim *et al.*, 2002), and synaptogenesis in the hippocampus following spatial training of rats in the Morris water maze task, a classic spatial learning experiment (Ramirez-Amaya *et al.*, 1999). Such results are not restricted to rodents (for instance, see Stewart and Rusakov, 1995, for similar effects in chicks). Experimental evidence for direct sensory-stimulus-dependent synapse formation in adult animals was first reported in rodent whisker barrels, where it was shown that localized increases in both synapse number (35 per cent) and spine density (25 per cent) followed specific whisker stimulation (Knott *et al.*, 2002; Zito and Svoboda, 2002). Thus, several lines of evidence establish the occurrence of synaptogenesis in adult neuronal tissue following physiologically relevant stimuli and learning paradigms.

Finally, we turn to neurogenesis in adult animals. In traditional neuroscience, the idea that neurons may be added continually after birth was not considered seriously (Gross, 2000). Since the early 1960s, this view gradually began to change when Altman and colleagues (Altman, 1962; Altman and Das, 1965) showed evidence for new neurons in several brain regions of adult rats (neocortex, olfactory bulb, dentate gyrus), and cats (neocortex). Since then, evidence for neurogenesis has accumulated in adult songbirds (Nottebohm, 1985; Goldman and Nottebohm, 1983; Paton and Nottebohm, 1984), dentate gyrus of rats (Stanfield and Trice, 1988), adult mouse hippocampus (Kempermann *et al.*, 1997), and dentate gyrus and olfactory bulb of macaque monkeys (Kornack and Rakic, 1999). Recent studies have even reported the addition of new neurons in the neocortex of macaque monkeys throughout adulthood (Gould *et al.*, 1999b). We note that the results in the neocortex are still controversial as some researchers have expressed doubt about them based on objections to the techniques used to establish neurogenesis (Nowakowski and Hayes, 2000; Rakic, 2002). In humans, it has been found that new neurons are added in the dentate gyrus (Eriksson *et al.*, 1998), but not to the olfactory bulb (Sanai *et al.*, 2004). The latter study points to the presence of a substantial number of adult neural stem cells—cells that can potentially generate neurons—in a region of the forebrain called the subventricular zone, that intriguingly appear not to produce neurons in the adult human brain.

The above evidence indicates that neurons are indeed added to adult mammalian neuronal circuits, although this addition appears to be less prevalent in humans. An important question that remains is whether this neurogenesis is functionally important or merely a 'vestige of development' (Gross, 2000; Kempermann, 2002; Nottebohm, 2002), especially since the functional integration of new neurons into well-developed circuits may be difficult (Rakic, 2004). The following considerations suggest that neurogenesis does, in fact, mediate function. New neurons are added to structures that are important for learning and memory, such as the hippocampus, lateral prefrontal cortex, inferior temporal cortex and posterior parietal cortex. Neurogenesis, like synaptogenesis and spine density changes, has been correlated with hippocampally dependent learning experiences. For instance, trace eye-blink conditioning and spatial learning in animals lead to an increase in the number of neurons through an extension of neuronal lifetimes (Gould *et al.*, 1999a). There appears to be a critical period following cell production such that learning occurring in this period increases neuronal lifespan. In parallel to this evidence, stressful experiences that result in a downregulation of cell proliferation in the dentate gyrus (Gould and Gross, 2002) are implicated in lower performance in hippocampally dependent learning tasks, suggesting a causal link between the two. Several conditions that increase neurogenesis (enriched environments, increased oestrogen levels, wheel running, etc.) in mice and rats also enhance performance (Gould and Gross, 2002). Similarly, increases in social complexity have been found to enhance the survival of new neurons in birds (Lipkind *et al.*, 2002). Finally, it has been found that the physiological properties of adult-generated granule cells in the dentate gyrus of the hippocampus resemble those of granule cells in young rats (Overstreet-Wadiche *et al.*, 2006). This suggests that adult-generated neurons may share some properties with embryonic and early postnatal neurons in their ability to extend axons, form new connections more readily and to make more synapses. These characteristics may make adult neurogenesis an attractive 'feature', rather than a developmental 'bug' in neuronal circuits.

In summary, there is now a large body of evidence that suggests that spine motility, spinogenesis, synaptogenesis, and more recently, neurogenesis are all active mechanisms for implementing architectural plasticity in neuronal circuits in response to real learning and memory needs faced by organisms.

Computational importance of learning-directed growth

Formal learning theory (FLT) deals with the ability of a learner to arrive at a target concept based on examples of the concept. Three typical features of FLT models are that the learning algorithm searches through a predefined space of candidate hypotheses (concepts), it is expected to learn the concept exactly, and no restrictions are placed on the actual time taken by the learner to arrive at the target concept. FLT is therefore interested in 'exact' and 'in-principle' learnability, and the expectation is that generalization—a measure of performance on novel examples—is achieved. The classical formulation of FLT is discussed in the context of language learning by Gold (1967). Perhaps the key insight from formal work on language learning is that the learner must utilize a highly restricted set of all possible concepts in order to have even the possibility of generalizing. In other words, far from employing a general learner, from this perspective a language learning system must be meticulously tailored to the problem at hand, either by the designer in the case of artificial systems, or presumably by evolution in the case of biological learners.

In the early 1980s, formal learning theory underwent a substantial change from Gold's limit-based framework to a model of probably-approximately-correct (PAC) learning (Valiant, 1984). The PAC model relaxes two of the three key requirements of formal learning theory. The impractical assumption of infinite time horizons is eliminated and the stringent restriction of exact learning relaxed. In short, PAC learning deals not with in-principle learnability, but with learnability in polynomial time, and not with exact learning, but with approximate learning. Formally, a concept is PAC-learnable if the learner arrives with probability 1-d at a hypothesis that classifies all the examples correctly with probability 1-e, for arbitrarily small e and d. Nevertheless, the hypothesis space is still fixed a priori. A fixed hypothesis space yields such problematic theoretical issues as Fodor's paradox (Fodor, 1980), which states that nothing can be learned that is not already known; and hence nothing is really learned. The idea here is that no hypothesis that is more complex than the ones in the given hypothesis space can be evaluated and hence learned. Therefore, all concepts need to be available in the hypothesis space before the search begins.

Constructive learning (Piaget, 1970, 1980; Quartz, 1993; Quartz and Sejnowski, 1997) addresses this issue of a fixed hypothesis space. The central idea of Piagetian constructivism is the construction of more complex hypotheses from simpler ones. This issue is dealt with more formally by Quartz (Quartz, 1993; Quartz and Sejnowski, 1997). Constructivist learning models deal directly with the issue of increasing hypothesis complexity as learning progresses (Shultz *et al.*, 1995; Westermann, 2000). Activity-dependent structural plasticity is viewed as the mechanism that implements constructivist learning. Constructive neural networks offer a clear way of viewing learning and development as constituting a 'plasticity continuum'. Synaptic weight change may be a form of plasticity that occurs at fast timescales, whereas architectural changes occur on slower timescales. Further, it is also possible, as we saw earlier, that the developmental processes of structural plasticity underlie learning even in mature animals.

Baum (1988) showed that networks with the ability to add structure during the process of learning are capable of learning in polynomial time any learning problem that can be solved in polynomial time by any algorithm whatsoever (Quartz and Sejnowski, 1997), thus conferring a computational universality to this paradigm. Interestingly, the bias-variance trade-off can be broken by constructivist learning, by adding hypotheses incrementally to the space in such a way as to keep variance low while reducing bias. Further, in the context of neurobiology, the burden of innate knowledge is relaxed. Given a basic set of primitives (in the form of mechanisms and physical substrate), construction of a network occurs under the guidance of experience within genetic constraints. Finally, computational arguments show that it is unlikely that evolution has prepared brain networks in human children for all of the various learning problems to which they might eventually be exposed (Sirois and Shultz, 1999). It is far more likely that brain networks will need to be constructed and their architectures modified as novel and unexpected learning problems arise.

Architectural plasticity and auditory localization in barn owls

The auditory localization system (ALS) of the barn owl is an excellent example of representational change that is produced by experience-driven architectural plasticity. Beacause developmental mechanisms and programmes appear to be highly conserved across many species and phyla, systems like this may reveal important general principles of neurobiological self-organization and mechanisms by which instructive environmental signals play a role in the construction of representational structures that also most likely operate in human development. Whereas many animals can localize sounds soon after birth (Field *et al.*, 1980) indicating the hard-wiring of at least a part of this system (Brainard and Knudsen, 1998; Knudsen *et al.*, 1991), experience, in addition to intrinsic programmes, plays an important role in shaping and modifying the auditory localization pathway in barn owls.

Visual displacement experiments in barn owls

Barn owls localize sound in the horizontal direction primarily using an auditory cue called the interaural time difference (or ITD). This is defined as the delay or the time difference between the arrival of sound from a source at one ear versus the other. Unless the sound source is positioned symmetrically with respect to the two ears, the signal has to travel different distances to reach the near versus the far ear, resulting in a non-zero ITD.

One of the pathways in which ITD information is processed in barn owls is the midbrain pathway (see Figure 4.1). Here, cue information arrives in a frequency-dependent manner at the central nucleus of the inferior colliculus (ICC), and flows through the external nucleus of the inferior colliculus (ICX) to the optic tectum (OT). The ICC has a tonotopic, or frequency-dependent, organization, and cue information from various frequency channels of the ICC is combined (shown as thick arrows) to form a topographic, map-like representation of auditory space in the ICX. The ICX then conveys this topographic information about auditory space to the OT (Knudsen, 2002). Interestingly, the OT has a similar topographic map of *visual* space mediated by visual input from the retina and forebrain. It has been found that the auditory and visual maps of space are aligned and integrated in the OT (Knudsen and Brainard, 1995).

Visual displacement experiments have explored the nature of information processing in the midbrain pathway and have shed light on the mechanisms by which plasticity occurs in response to external demands. These experiments involve the use of prismatic spectacles to disrupt auditory-location

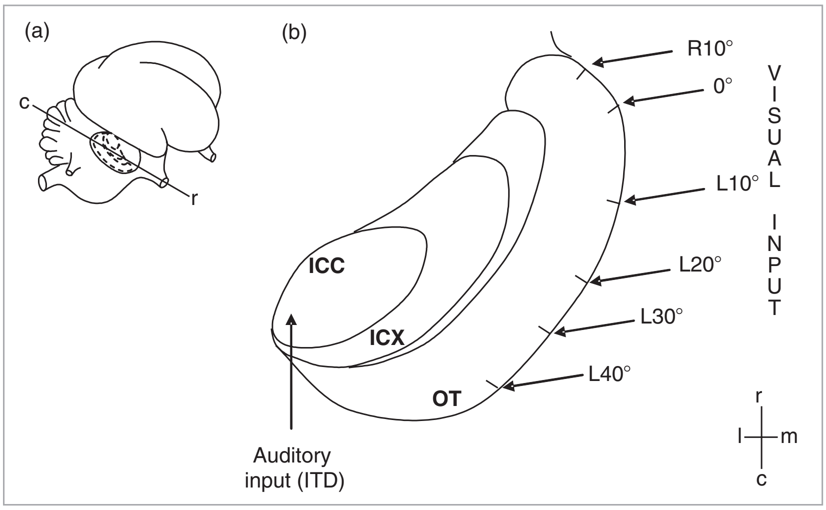

Fig. 4.1 Midbrain auditory pathway in barn owls. (a) Cartoon showing lateral view of a barn owl's brain. The straight line represents the plane of section through the midbrain. (b) Schematic representation of the sectioned surface of the *right* midbrain. Shown are the three important processing areas in the midbrain localization pathway: ICC—central nucleus of the inferior colliculus, ICX—external nucleus of the inferior colliculus, and OT—optic tectum. The arrows on the left and the right show respectively the flow of auditory and visual information into the midbrain. The topographic nature of the visual input into the OT is emphasized by individual arrows which indicate some of the spatial locations from which visual information arrives at the OT. Here, R10° (L10°) represents the location in the visual field that is 10° to the right (left) of the central azimuthal position. R: rostral, c: caudal, m: medial, l: lateral. Adapted from (DeBello *et al.*, 2001).

associations by displacing the visual input along the azimuth by a predetermined amount (Brainard and Knudsen, 1998). For instance, when an owl fitted with spectacles that produce a 10° rightward shift looks straight ahead, it receives visual information centred at 10° to the left, rather than at 0°, or straight ahead. As a result, whereas a normal owl can visually localize a sound-source placed at 0° by looking straight ahead, one fitted with a right-shifting prism has to learn to rotate its head to the right by the appropriate amount, in order to localize the same source.

It has been found that owls less than 200 days old display a remarkable ability to behaviourally adapt to prism experience—up to a 23° shift—in about seven weeks. Removal of prisms after adaptation results in gradual recovery of normal orienting behaviour. On the other hand, adult owls adapt poorly to prism experience (Knudsen, 1998). Interestingly, shifted-back juvenile owls that are subsequently fitted with prisms as adults do show adaptability, but only up

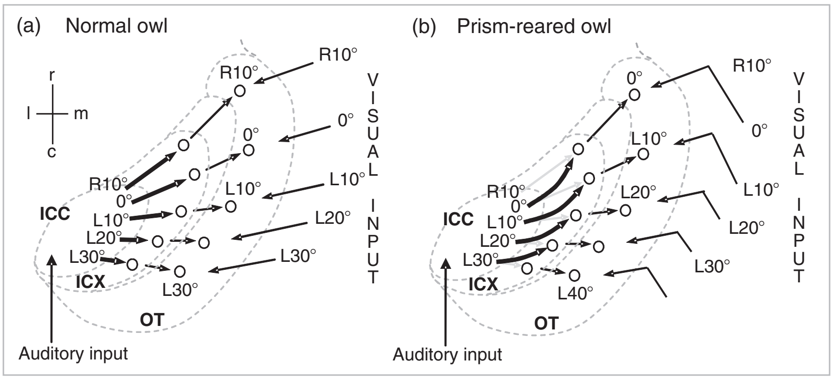

Fig. 4.2 Auditory localization plasticity in prism-reared owls. Adaptation to prism experience in the auditory localization behaviour of juvenile owls has been found to occur through activity-dependent architectural modifications of the topographic representation of auditory space. It has been found that axonal growth and synaptogenesis occur only between the ICC and ICX layers in the midbrain (DeBello *et al.*, 2001; Feldman and Knudsen, 1997). As a result, while the auditory tuning of neurons in the ICC remains unchanged, that of the ICX (and consequently the OT) neurons shifts by the amount of visual displacement induced by the prisms (Brainard and Knudsen, 1993, 1995). (a) Schematic representation of a section through the right midbrain of a normal barn owl showing the default anatomical projections (arrows) between different layers. Auditory (ITD) information arriving at the ICC and visual information arriving at the OT are shown. The open circles represent individual neurons. The text above each circle represents the azimuthal centre of the auditory receptor field of that neuron. The text adjacent to the arrows represents the centre of the spatial location from which auditory or visual information arrives to the midbrain. The thick arrows between ICC and ICX represent the integration of tonotopic information across various frequency channels of the ICC. (b) Schematic representation of a section through the right midbrain of a prism-reared barn owl. The auditory input remains unchanged, while the visual input is shifted as a result of prism fitting (depicted with arrows that are bent). New anatomical projections between the ICC and ICX that develop in response to prism-rearing are shown with thick curved arrows. Text above the neurons in the OT highlights the shift in the auditory tuning of these neurons. For instance, the tuning of the topmost neuron in the OT shifts from R10° to 0°. Here, R10° (L10°, represents the location in the visual field that is 10° to the right (left) of the central azimuthal position. R: rostral, c: caudal, m: medial, l: lateral. Adapted from DeBello *et al.* (2001).

to the adjustments they had originally displayed (Knudsen, 1998). Recent evidence (Linkenhoker and Knudsen, 2002) also shows that the capacity for behavioural plasticity in adult owls can be increased with incremental training, that is, where the adults experience prismatic shifts in small increments.

Detailed investigations into the mechanisms mediating this behavioural adaptation have yielded several insights. Particularly interesting is the fact that learning is mediated by structural changes—the growth of new axons and the formation of new synapses—in the neural architecture of prism-fitted birds (see Figure 4.2).

We are developing a model of this system to understand the computational mechanisms through which different aspects of the observed architectural adaptation can be implemented in a biologically feasible manner. Box 4.1 discusses the model, and preliminary results show that a Hebb-like learning rule can account for many of the experimental details. The model makes predictions about the site of an important (inhibitory) component of plasticity, and about the biological signals that can trigger structural change.

Box 4.1 Modelling plasticity in the auditory localization system (ALS) of barn owls

We implemented a simple connectionist model of the auditory localization network in barn owls using firing-rate coding neurons. Figure 4.3 shows the architecture of the model based on published details of the anatomical connections in the midbrain localization pathway (Hyde and Knudsen 2000; Knudsen 2002). The different layers are the ICC, ICX, OTD (deep layers of the OT), OTM (medial layers of the OT), and OTS (superficial layers of the OT). Neurons in the ICX are either excitatory (ICXE) or inhibitory (ICXI). Individual neurons are shown as circles, and their point-to-point projections are indicated by arrows. We assume without loss of generality that every degree in the visual (and auditory) field is coded for by a separate neuron in any given layer. The output of each neuron is a nonlinear function of the weighted sum of its inputs. The driving assumption of our modelling effort is that all the default structures and pathways already exist and that the phenomenon of interest is the plasticity between the ICC and ICX layers in response to prismatic shifts in the visual field. The *ab initio* development of ITD maps has been modeled in literature (for instance, see Kempster *et al.*, 2001), but we do not focus on this. For simplicity, we omit the frequency dependence of ICC neuronal tuning in this model, and assume without loss of generality that the ITD tuning of neurons is all-or-none corresponding to the best ITD.

Past modelling efforts have either used a foveation-based error signal to drive learning (Gelfland *et al.*, 1988; Pouget *et al.*, 1995; Rucci *et al.*, 1997) which has since been shown to be incorrect, or have primarily relied

Box 4.1 *(continued)*

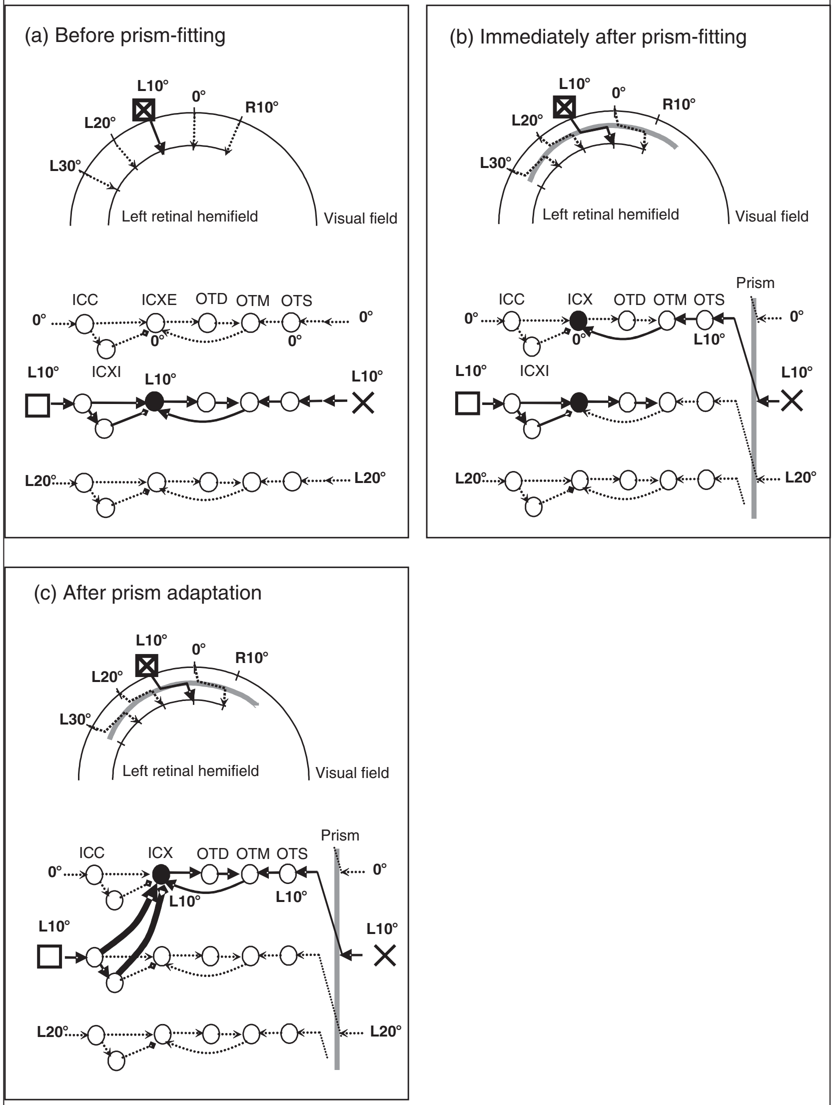

Fig. 4.3 Model of the midbrain auditory localization system in barn owls. Shown are the three stages in the adaptation of the system to prism experience. In the top half of each panel, the outer semi-circle is a cartoon representation of the audiovisual field experienced by an owl. The inner arc depicts the left retinal hemifield. Representative locations in the audiovisual field are represented in degrees. An audio stimulus is represented as an open square, a visual stimulus as an 'X', and a combined audiovisual stimulus as a square and an 'X'. In all cases, an audio-visual stimulus is presented to the owl at L10° (i.e., the location in space that is at 10° to the left of the central azimuthal position).

Box 4.1 *(continued)*

Fig. 4.3 Caption Continued The bottom half of each panel shows the topographic nature of the information flow (dotted and continuous arrows) through different layers in a model of the midbrain. Each column of circles represents a different layer of neurons in the midbrain, and each row represents the point-to-point pathway that processes information from one spatial location. The inhibitory neurons in the ICX project to the excitatory neurons in the ICX and this inhibitory connection is shown by an arrow with a diamond-shaped head between an ICXI and an ICXE neuron. Auditory input enters the model at the left through ICC neurons, and visual input at the right, through OT neurons. (a) Here, a colocalized audiovisual input is presented to a normal owl at L10° (shown in the top half). The default anatomical connections and flow of information are shown in the bottom half. We see that audio input arrives at the neuron encoding for L10° in the ICX (filled circle), and the visual input arrives at the same ICX neuron via a back projection from the corresponding OT neuron. The pathway activated by this input is represented with continuous arrows, while non-active pathways are shown with dotted arrows. As mentioned in the main text, such a stimulus produces a firing response in the ICX neuron about 10ms after its presentation. (b) In this panel, the same audio-visual input as in (a) is presented to a juvenile owl immediately after prism-fitting. The prism is schematically shown in the top half of the panel as a thick arc, which refracts all incoming visual input, and shifts it to the right by 10°. The bottom half shows auditory input arriving as in a normal owl, and being processed by the ICX neuron in the middle row (filled circle). However, due to prism-fitting visual information from L10° is now processed by the neuron in the OTS that had previously processed visual information from 0° (compare with panel a). The prism is shown here as a thick line. The point-to-point anatomical projections imply that the downstream neurons that subsequently process this visual information are also different from those in a normal owl. We see that the visual signal eventually gets back-projected to the ICX neuron in the top row (shown as a filled circle), whose auditory tuning corresponds to 0°. Thus, there is a mismatch between the processing of auditory and visual information at the ICX. It has been found that in this case, the ICX neuron in the top row, which receives only the visual component of the bimodal stimulus, fires approximately 90 ms after stimulus presentation (Gutfreund *et al.*, 2002). This delayed firing of the ICX neuron constitutes the error signal that subsequently drives plasticity in this system. (c) In this panel, the same audio-visual stimulus as before is presented to the prism-fitted juvenile owl after it has displayed behavioral adaptation to prism experience. New anatomical projections that were observed between the ICC and ICX (DeBello *et al.*, 2001; Feldman and Knudsen, 1997) are shown in the bottom half of this panel by thick arrows. These new projections have shifted the auditory tuning of the ICX neurons (but not that of the ICC neurons), and as a result, the mismatch in the audiovisual information processing in the ICX has been eliminated. The ICX neuron that now processes the audiovisual input is in the top row (shown by a filled circle), and its new auditory tuning corresponds to L10°.

Box 4.1 *(continued)*

on reward-based learning schemes (Pouget *et al.*, 1995; Rucci *et al.*, 1997) about which little is known in the owl's auditory localization system. Although Rosen *et al.* (1994) did not use a reinforcement learning scheme, they implemented learning via the back-propagation algorithm. The learning rule we use is inspired by the classical Hebbian scheme. We move away from reward-based methods and a global error signal, and implement learning using a template-based error signal. This is based on recent evidence showing that the teaching or error signal that drives plasticity in this system is a visually based topographic template signal (Hyde and Knudsen, 2001). It has also been recently established that visual input from the OT does indeed project to the ICX and that this back-projecting error signal which is responsible for the calibration of the map is encoded in the selective and delayed firing of ICXE neurons in response to the presentation of a co-localized audiovisual stimulus to a prism-shifted system (Gutfreund *et al.*, 2002). In such a set-up, ICXE neurons receive either the visual or the auditory component of the input (see Figure 4.3). Under normal circumstances, a co-localized bimodal stimulus produces a response in the corresponding ICX neuron about 10 ms after the stimulus is presented, but significantly before (about 80 ms before) the response in the prism-shifted case (Gutfreund *et al.*, 2002). In our model, we detect the error signal by checking for ICX activity 90 ms after the presentation of an audiovisual stimulus. Weights are reduced between ICC neurons that do not fire and ICXE (and ICXI) neurons that do fire as a result of the back-projecting error signal from the OT. If the error occurs repeatedly for a sufficiently long period of time (implying a systematic error in encoding information), new synapses are formed, that is, the population of the ICC-ICXI and ICC-ICXE weight matrices is increased. Once this happens, the learning rule increases weights between ICC neurons that fire and ICXE (and ICXI) neurons that also fire soon after (indicating the arrival of auditory input). The time course of weight evolution is implemented to be sigmoidal, and this prevents an indefinite increase in weights caused by the classical Hebbian rule. The results of this scheme match those from experiments in that the ITD tuning curves of the ICX neurons shift in the adaptive direction after training, and we are able to capture the essence of the learning exhibited by the barn owls. Current work is looking into relaxing some of the assumptions made at the beginning, and at incorporating more details into the model.

Structural plasticity and representation construction

The initial topographic organization of the auditory localization system suggests that structural plasticity may be necessary for the owl to adapt to experimentally induced visual displacement. Indeed, a temporally restricted developmental window for structural plasticity may be the reason adult owls are unable to adapt beyond a few degrees of shift (Brainard and Knudsen, 1998). An interesting question in this context is, what are the implications of this capacity for architectural change to the representational properties of the network in the context of the highly constrained neurobiological framework?

Investigating the representational consequences of architectural plasticity necessitates the use of appropriate quantifiers of representational complexity. One way to quantify the 'ability' or 'complexity' of a network is in terms of its memory capacity (Poirazi and Mel, 2001), defined using a function-counting approach. We apply a similar approach to the midbrain auditory localization network in the barn owl by treating the auditory and visual inputs as the *input* to the network, and the activities of the ICX neurons as its *output*. If we assume for simplicity that every neuron codes uniquely for one degree in the visual (and auditory) field, and that each neuron in one layer projects uniquely to the corresponding neuron in another, then a simple calculation yields that the initial network architecture implements one input–output function, whereas the post-learning architecture (that results after a one degree prism shift) implements a qualitatively different function. That is, structural plasticity has permitted an increase in the repertoire of localization functions that the network can compute from one to two, while satisfying topographic projection constraints. According to this calculation scheme, the increase in memory capacity (or function complexity) possible in response to an n-degree prismatic shift in juvenile owls is n. We note here that the ability to grow introduces extra degrees of freedom in the input–output transformations in a manner that is consistent with the shift in the resulting output. That is, if the output units are considered to be linear for simplicity, then the input–output function between a pair of layers is a linear transformation (or a matrix), and the point-to-point nature of the anatomical projections constrain all off-diagonal elements of this matrix to be zero. Synaptic weight modifications just scale the diagonal elements, whereas growth permits the population of appropriate off-diagonal elements resulting in qualitatively different functions. No amount of diagonal scaling can reproduce the latter effect. As a result, if there is no growth but only synaptic efficacy change, the memory capacity of the network and hence its representational complexity remain constant. This observation is consistent with the hypothesis that a lack of the ability to grow in adult circuits may be the reason adult owls (>200 days old) are unable to adapt to large visual shifts.

The general lesson from this model system is that environmentally derived instructive signals help regulate architectural plasticity and that this form of plasticity underlies abrupt qualitative change in the representational ability of a network. Further, this representational change typically increases the functional complexity of a network. Viewed from the perspective of learning theory, the initial network represents the hypothesis space that is explored while learning from examples and architectural plasticity allows the developing owl to go beyond this original hypothesis space to learn novel functions that are originally outside of this space. In contrast, the mature owl appears not to possess the same degree of architectural plasticity, and so is limited in terms of learning capacity. This highlights the fundamental role of architectural plasticity as a form of constructive learning and illustrates why it is fundamentally different from synaptic weight change, which only implements a search within a given hypothesis space. Hence architectural plasticity is a powerful and indispensable ally in representation construction and it is likely a general theme of neurobiological development in many species, including humans, a species to which we now turn.

Evidence from modelling cognitive development in children

One way to assess the importance of constructive growth in cognitive development is to compare models whose networks are allowed to grow to models containing static networks. This comparison is facilitated by the fact that the principal difference between two of the most popular neural algorithms for modelling cognitive development is that one of them incorporates growth and the other does not (Shultz, 2003). Both back-propagation (BP) and cascade-correlation (CC) involve multilayered feedforward networks that learn by adjusting their connection weights to reduce error, defined as the discrepancy between actual and target outputs. BP networks are typically designed by a programmer and retain their initial structure throughout learning. In contrast, CC networks begin with a minimal topology containing only inputs and outputs defined by a programmer, and then proceed to add hidden units as needed to master the training problem.

Box 4.2 describes the CC algorithm, which has been successfully applied to a wide variety of phenomena in cognitive development in children (Shultz, 2003). There are presently three domains of cognitive development to which both BP and CC networks have been applied. Here we briefly review the simulations in these domains in order to examine the effects of growth on simulation success. The three relevant domains are the balance scale, the integration of velocity, time, and distance cues for moving objects, and age changes in sensitivity to correlations vs. features in infants' category learning. Note that in none of these domains do we know about the actual brain circuits involved. Nor are the BP and CC algorithms considered to be realistic implementations of the details of real brain circuits. They are instead highly abstracted, functional algorithms, albeit ones that are inspired by some general neurophysiological constraints (Shultz, 2003).

Box 4.2 The cascade-correlation (CC) algorithm

There are two phases in CC learning—the input phase and the output phase. Training typically begins in output phase by using the so-called delta rule to adjust the weights entering output units in order to reduce network error. When error reduction stagnates for a set number of epochs or when the problem has not been mastered within another, larger certain number of epochs, CC switches to input phase. An epoch is a pass through the training patterns.

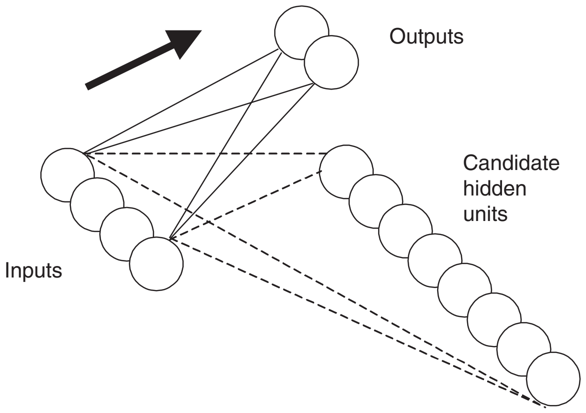

Fig. 4.4 The first input phase of training in a hypothetical CC network. Solid lines represent frozen connection weights, and dashed lines represent trainable connection weights (Shultz, 2003, adapted with permission from MIT Press).

In the input phase, the task is no longer to reduce error, but rather to reconceptualize the problem by recruiting a useful hidden unit downstream of the existing units but upstream of the output units. A pool of typically eight candidate hidden units, usually equipped with sigmoid

Box 4.2 *(continued)*

activation functions, is evaluated for recruitment (see Fig. 4.4). This evaluation consists of training randomly initialized weights from the inputs and any existing hidden units to these candidates, with the goal of maximizing a modified correlation between candidate activation and network error. In other words, the CC algorithm in input phase is looking for a new hidden unit whose activation patterns track the network's current error, becoming either very active or very inactive with error fluctuations—in short, a candidate unit which is sensitive to the major difficulties that the network is currently having. The same delta rule is used here in input phase, but instead of minimizing error as in output phase, it is used to maximize correlations.

When these correlations stagnate over a certain number of input-phase epochs, the candidate with the highest absolute correlation with network error is selected and installed into the network, and the less successful candidates are discarded. This implements a kind of proliferation and pruning of units and connections that is characteristic of selection models of brain development (Changeux and Danchin, 1976). The new recruit is initialized with randomized output weights of positive values if the correlation with error was negative or with negative values if the correlation with error was positive.

At this point, CC returns to output phase to determine how to best utilize the new conceptualization of the problem afforded by its latest recruit. In this way, new and more complex interpretations build on top of older, simpler interpretations. Cycling between input and output phases continues until the training problem is mastered. CC can be interpreted as implementing the processes of synaptogenesis and neurogenesis that our review shows to be at least partly under the control of learning. CC also implements the hierarchical, cascaded progression now thought to be characteristic of brain development (Thivierge, Rivest and Shultz, 2003).

Balance-scale stages

The balance-scale problem involves presenting a child with a rigid beam balanced on a fulcrum (Siegler, 1976, 1981). The beam is equipped with several pegs spaced at regular intervals to the left and right of the fulcrum. The experimenter typically places some number of identical weights on one peg on the left side and some number of weights on one peg on the right side. While supporting blocks prevent the beam from moving, the child is asked to predict which side of the beam will descend, or whether the scale will remain balanced, once the supporting blocks are removed.

The 'rules' that children use to make balance-scale predictions can be diagnosed by presentation of six different types of problems. Three problems are relatively simple in the sense that one cue (either weight or distance from the fulcrum) predicts the outcome because the other cue does not differ from one side of the scale to the other. Three other problems are relatively complex in that the two relevant cues conflict with each other, with weight predicting one outcome and distance predicting the opposite outcome. For simple or complex types, there are three possible outcomes in which the direction of the tip is predicted by weight information or by distance information or the scale remains balanced. The pattern of a child's predictions across these six different problem types can be used to diagnose the 'rule' the child is apparently using to make predictions. Of course, in a connectionist network a rule is not implemented as a symbolic if-then structure, but is rather an emerging epiphenomenon of the network's topology and connection-weight values.

There is a consensus in the psychological literature that children progress through four different stages of balance-scale performance (Siegler, 1976, 1981). In stage 1, children use weight information only, predicting that the side with greater weight will descend or that the scale will balance when the two sides have equal weights. In stage 2, children continue to use weight information but begin to use distance information when the weights are equal on each side. In stage 3, weight and distance information are used about equally, but the child guesses when weight and distance information conflict on the more complex problems. Finally in stage 4, children respond correctly on all types of problems, whether simple or complex.

In one of the first connectionist simulations of cognitive development McClelland and Jenkins found that a static BP network with two groups of hidden units segregated for either weight or distance information progressed through the first three stages of the balance scale and even into the fourth stage (McClelland and Jenkins, 1991). However, the network never settled into stage 4 performance, instead cycling between stage 3 and stage 4 functioning for as long as the programmer's patience lasted. Further simulations indicated that BP networks could settle into stage 4, but only at the cost of missing stages 1 and 2 (Schmidt and Shultz, 1991). There seems to be no way for BP networks to capture both consistently mature stage 4 performance and earlier progressions through stages 1 and 2. In contrast, the first CC model of cognitive development naturally captured all four stages of balance-scale performance, and did so without requiring any hand-designed segregation of hidden units (Shultz *et al.*, 1994). Capturing balance-scale stages is not a trivial matter, as witnessed by the shortcomings of earlier symbolic-rule-learning models (Langley, 1987; Newell, 1990).

Both BP and CC models are able to capture naturally the other psychological regularity of balance-scale development, the torque-difference effect. This is the tendency for problems with large absolute torque differences (from one side of the scale to the other) to be easier for children to solve, regardless of their current stage (Ferretti and Butterfield, 1986). Connectionist feed-forward networks produce perceptual effects like torque difference whenever they learn to convert quantitative input differences into a binary output judgment. The bigger the input differences—in this case, from one side of the fulcrum to the other side—the clearer the hidden-unit activations, and the more decisive the output decisions. Symbolic rule-based models are unable to naturally capture the torque-difference effect because symbolic rules typically don't care about the amount of differences, only the direction of differences.

The ability of neural-network models to capture stages 1 and 2 on the balance scale is due to a bias towards equal-distance problems in the training patterns. These are problems in which the weights are placed equally distant from the fulcrum on each side. This bias forces the network to first emphasize weight information at the expense of distance information because weight information is much more relevant to reducing prediction errors. Once weight information is successfully dealt with, then the network can turn its attention to distance information, particularly on those problems where distance varies from one side of the fulcrum to the other. In effect, the network must find the particular region of connection-weight space that allows it to emphasize the numbers of weights on the scale and then move to another region of weight space that focuses on the multiplication of equally important weight and distance information. It requires a powerful learning algorithm to make this move in connection-weight space. Apparently a static BP network, once committed to using one source of information in stage 1, cannot easily find its way in weight space to the stage 4 region merely by continuing to reduce error. A constructive algorithm, such as CC, has an easier time with this move because each newly recruited hidden unit effectively changes the shape of connection-weight space by adding a new dimension. This new dimension affords a path on which to move towards the stage-4 region. More colloquially, BP cannot have its cake and eat it too, but CC, due to its dynamic increases in computational power, can both have and eat its cake, essentially by continuing to bake.

Integration of velocity, time, and distance cues

In Newtonian physics, the velocity of a moving object is the ratio of distance traveled to the time of the journey: velocity = distance/time. Algebraically then, distance = velocity x time, and time = distance/velocity. Some of the cleanest evidence on children's acquisition of these relations was collected by Wilkening (1981), who asked children to predict one dimension (e.g., time) from knowledge of the other two (e.g., velocity and distance). For example, three levels of velocity information were represented by the locomotion of a turtle, a guinea pig and a cat. These three animals were described as fleeing from a barking dog, and a child participant was asked to imagine these animals running while the dog barked. The child's task was to infer how far an animal would run given the length of time the dog barked, an example of inferring distance from velocity and time cues. CC networks learning similar tasks typically progress through an equivalence stage (e.g., velocity = distance), followed by an additive stage (e.g., velocity = distance − time), and finally the correct multiplicative stage (e.g., velocity = distance/time) (Buckingham and Shultz, 2000). Some of these stages had been previously found with children, and others were subsequently confirmed as predictions of our CC simulations. As with children in psychological experiments, our CC networks learned to predict the value of one dimension from knowledge of values on the other two dimensions.

We diagnosed rules based on correlations between network outputs and the various algebraic rules observed in children. To be diagnosed, an algebraic rule had to correlate positively with network responses, account for most of the variance in network responses, and account for more variance than any other rules did. As with rules for solving balance-scale problems, these rules are epiphenomena emerging from patterns of network connection weights. For velocity and time inferences, CC networks first acquired an equivalence rule, followed by a difference rule, followed in turn by the correct ratio rule. Results were similar for distance inferences, except that there was no equivalence rule for distance inferences. In making distance inferences, there is no reason a network should favour either velocity or time information because both velocity and time vary proportionally with distance.

A shift from linear to non-linear performance occurred with continued recruitment of hidden units. Linear rules include equivalence (e.g., time = distance), sum (e.g., distance = velocity + time), and difference (e.g., velocity = distance − time) rules, whereas non-linear rules include product (e.g., distance = velocity x time) and ratio (e.g., time = distance/velocity) rules.

Because the sum and difference rules in the second stage are linear, one might wonder why they require a hidden unit. The reason is that networks without a hidden unit are unable to simultaneously encode the relations among the three dimensions for all three inference types. In distance inferences, distance varies directly with both velocity and time. But in velocity inferences, distance and time vary inversely, and in time inferences, distance and velocity also vary inversely. Networks without hidden units are unable to encode these different relations. The first-recruited hidden unit differentiates distance information from velocity and time information, essentially by learning connection weights with one sign (positive or negative) from the former input and opposite signs from the latter inputs. This weight pattern enables the network to consolidate the different directions of relations across the different inference types.

In contrast to constructive CC networks, static BP networks seem unable to capture these stage sequences (Buckingham and Shultz, 1996). If a static BP network has too few hidden units, it fails to reach the correct multiplicative rules. If a static BP network has too many hidden units, it fails to capture the intermediate additive stages on velocity and time inferences. Our extensive exploration of a variety of network topologies and variation in critical learning parameters suggests that there is no static BP network topology that can capture all three types of stages in this domain. Even the use of cross-connections that bypass hidden layers, a standard feature of CC, failed to improve the stage performance of BP networks. There is no apparent way to get BP to cover all three stages because the difference between underpowered and overpowered BP networks is a single hidden unit. Thus, we conclude that the ability to grow in computational power is essential in simulating stages in the integration of velocity, time, and distance cues.

Age changes in infant category learning

Our final simulation comparison concerns age changes in sensitivity to correlations vs. features in category learning by infants. Using a stimulus-familiarization-and-recovery procedure to study categorization, Younger and Cohen found that 4-month-olds process information about independent features of visual stimuli, whereas 10-month-olds are able to abstract relations among those features (Younger and Cohen, 1983, 1986). Such findings relate to a long-standing controversy about the extent to which perceptual development involves integration (Hebb, 1949) or differentiation (Gibson, 1969) of stimulus information. A developing ability to understand relations among features suggests that perceptual development involves information integration, a view compatible with constructivism.

Infants are assumed to construct representational categories for repeated stimuli, ignoring novel stimuli that are consistent with a category, while

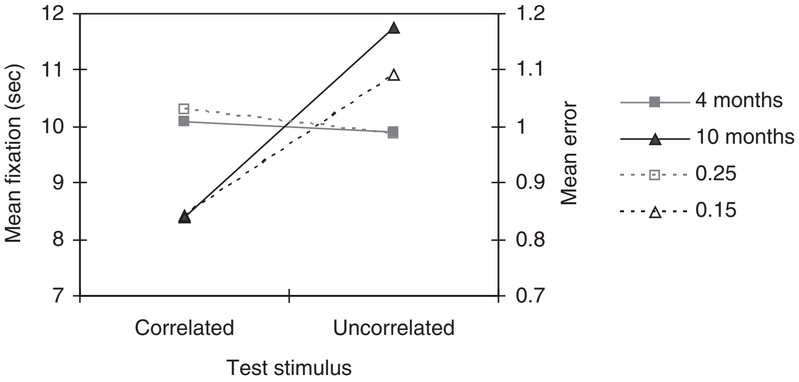

Fig. 4.5 Mean fixation of key test stimuli in infants (solid lines) from Younger and Cohen (Younger and Cohen, 1986, reprinted with permission of the Society for Research in Child Development), and mean CC network error at two levels of score-threshold (dashed lines) from Shultz and Cohen (Shultz and Cohen, 2004, adapted with permission).

concentrating on novel stimuli that are not members of a category (Cohen and Arthur, 1983; Younger and Cohen, 1985). After repeated presentation of visual stimuli with correlated features, 4-month-olds recovered attention to stimuli with novel features more than to stimuli with either correlated or uncorrelated familiar features (Younger and Cohen, 1983, 1986). In contrast, 10-month-olds recovered attention to both stimuli with novel features and familiar uncorrelated features more than to stimuli with familiar correlated features. Mean fixation time in seconds is shown for the key interaction between age and correlated vs. uncorrelated test stimuli, plotted in solid lines in Figure 4.5 against the left-hand y-axis. This pattern of recovery of attention indicates that young infants learned about the individual stimulus features, but not about the relationships among features, whereas older infants in addition learned about how these features correlate with one another.

These infant experiments were recently simulated with CC encoder networks (Shultz and Cohen, 2004). Encoder networks have the same number of input units as output units, and their job is to reproduce their input values on their output units. They learn to do this reproduction by encoding an input representation onto their hidden units and then decoding that representation onto output units. In this fashion, encoder networks develop a recognition memory for the stimuli they are exposed to. Network error can be used as an index of stimulus novelty. Both BP and CC encoders were applied to the infant data (Shultz and Cohen, 2004). In CC networks, age was implemented by varying the score-threshold parameter: 0.25 for 4-month-olds and 0.15 for

10-month-olds. This parameter governs how much learning occurs because learning stops only when all network outputs are within score-threshold of their target values for all training patterns. It was assumed that older infants would learn more from the same exposure time than would younger infants. Mean network error for the key interaction between age and correlated vs. uncorrelated test stimuli is plotted with dashed lines in Figure 4.5 against the right-hand y-axis. On the assumption that network error reflects stimulus novelty (and thus infant interest), the CC networks closely matched the infant data—more error to the uncorrelated test stimulus than to the correlated test stimulus only at the smaller score-threshold of 0.15.

In sharp contrast, a wide range of static BP networks could not capture this key interaction. Our BP simulator was modified to use a score-threshold parameter to decide on learning success, just as in CC. The same coding scheme used in the CC simulations was also used here. A wide variety of score-threshold values were explored in a systematic attempt to cover the infant data. BP networks were equipped with three hidden units, the typical number recruited by CC networks. BP network topology was varied to explore the possible roles of network depth and the presence of cross connections in simulation success. Both flat and deep BP networks were tried, both with and without the cross-connection weights native to CC. There were six input and six output units in BP networks just as there were in CC networks. In the BP networks, there was never lower error on correlated than on uncorrelated test items, the signature of effective correlation detection. It is difficult to prove that any algorithm cannot in principle cover a set of phenomena, but we certainly gave BP a fair shake, running 9 networks in each of 80 simulations (2 familiarization sets × 10 score-threshold levels × 4 network topologies).

Unlike previous psychological explanations that postulated unspecified qualitative shifts in processing with age, our computational explanation based on CC networks focused on quantitatively deeper learning with increasing age, a principle with considerable empirical support over a wide range of ages and experiments (e.g., Case *et al.*, 1982). CC networks also generated a crossover prediction, with deep learning showing a correlation effect, and even more superficial learning showing the opposite—a similarity effect, in the form of more error to (or interest in) the correlated test item than to the uncorrelated test item. This is called a *similarity* effect because the uncorrelated test item was most similar to those in the familiarization set. This simulation predicted that with a single, suitable age group, say 10-month-olds, there would be a correlation effect under optimal learning conditions and a similarity effect with less than optimal familiarization learning. Tests of this prediction found that 10-month-olds who habituated to the training stimuli looked longer at the uncorrelated than the correlated test stimulus, but those who did not habituate did just the opposite, looking longer at the correlated than the uncorrelated test stimulus (Cohen and Arthur, unpublished).

In addition to covering the essential age by correlation interaction, CC networks covered two other important features of the infant data: discrimination of the habituated from the correlated test stimulus via the presence of two additional non-correlated features, and more recovery to test stimuli having novel features than to test stimuli having familiar features.

Although the BP simulations did not capture the infant data, they were useful as control conditions in establishing why CC networks were successful. Namely, the growth of cascaded networks with cross-connection weights seems critical to capturing detection of correlations between features and the developmental shift to this ability from earlier feature-value learning.

Because CC is sufficiently abstracted from the low-level details of actual neural circuits, it is not clear whether its recruitment of hidden units corresponds to synapto- or neurogenesis. This issue concerns whether candidate units already exist somewhere else in the system or whether they are created fresh before recruitment. From a sufficiently abstract computational view, it seems not to matter. But if CC does implement synaptogenesis, then it assumes establishment of fresh synapses rather than adding duplicate synaptic connections between units.

Conclusions

The chief biological constraint implemented in our models is that of letting networks grow while they learn. We allowed our networks to grow by forming new connections or by recruiting new hidden units. The modeling revealed a number of computational advantages for network growth, all of which made for more realistic coverage of both animal and human development.

The work reviewed here indicates that brain networks in a variety of species clearly grow as they learn. There is good evidence of synaptogenesis in birds and mammals in the cerebellum, hippocampal formation, and sensory and motor cortices. There is consensus that neurogenesis occurs in the olfactory bulb and dentate gyrus of a variety of mammals including monkeys, but published evidence of cortical neurogenesis is still somewhat controversial. Evidence indicates that both synapto- and neurogenesis increase with pressures and opportunities to learn, and that they play an important functional role in learning.

Computational analysis shows that constructive learning algorithms offer several advantages over static learning algorithms. Constructive algorithms can learn any learnable problem in realistic time by reducing bias while keeping variance low, thus avoiding the bias-variance trade-off that plagues static learners. Unlike static learners, constructive algorithms can also escape from Fodor's paradox about not being able to learn anything genuinely new. New representations constructed by growing networks can be qualitatively different than anything they possessed before. Furthermore neural-network algorithms that grow offer a precise and convincing way to implement constructivist accounts of cognitive development wherein new knowledge builds on earlier knowledge. Finally computational arguments show it is unlikely that evolution has prepared brain networks in human children for all of the learning problems to which they might eventually be exposed. It is more likely that brain networks are constructed as novel problems arise.

Recent investigations of sound localization in the barn owl provide a clear and compelling example of learning-directed brain growth, complete with evidence of synaptogenesis in the auditory topographic map in the inferior colliculus when juvenile barn owls are fitted with optical prisms that shift the visual field across the azimuth. This qualitative change in neural circuitry and sound localization was modeled with a Hebbian-like learning rule that was allowed to adjust old synapses and form new ones in response to the mismatch between auditory and visual information.

Implementing network growth in simulations of three different domains of psychological development in children produced superior data coverage when contrasted with comparable static networks. Network growth offers a number of computational advantages in covering these phenomena including the building of more complex knowledge on top of simpler knowledge, superior ability to escape from local minima during error reduction, and deeper learning of training problems.

The simulation of infant category learning predicted a correlation effect under optimal learning conditions and a similarity effect with less than optimal learning. There is already some confirmation of this prediction in that infants who habituated to training stimuli showed a correlation effect, but those who did not habituate showed the opposite, similarity effect.

An interesting general hypothesis raised by our computer simulations is that the function of learning-driven neural growth is to increase computational power so that learning tasks can be mastered. Precisely how this hypothesis might be tested in future brain research remains to be worked out.

It is noteworthy that these computational constraints can operate in a variety of learning paradigms, including both Hebbian learning and multilayered feedforward CC networks. As used here, some of the differences between these learning paradigms became somewhat blurred. Typically considered as an unsupervised learner, the Hebb rule was here modified to use the discrepancy between visual and auditory information to produce an error signal. Typically considered as a supervised learner, CC can often be construed as not requiring a human teacher to supply target outputs. For example, CC networks learn about balance-scale problems and about velocity, time and distance relations by computing the discrepancy between what they predict will happen and what actually happens (observations of the natural environment). Likewise, CC encoder networks (such as those used in our simulations of category learning in infants) use the stimulus itself as the target output, compute error as the discrepancy between input and output activations, and in effect implement an unsupervised learning scheme.

Although we emphasized network growth in this chapter, it is also clear from neuroscience research that brains discard little-used neurons and synapses. In a drive towards even greater biological realism, it would be worth experimenting with networks that both grow and prune. Some preliminary work along these lines has found that pruning of CC networks during growth and learning improves learning speed, generalization, and network interpretability (Thivierge *et al.*, 2003). This would appear to be yet another case of a biological constraint offering computational advantages.

The bottom line is that there is a coherent alternative to the instantaneity hypothesis that we described in the Introduction. This constructivist alternative takes both development and learning seriously, and embraces the non-instantaneous features of development as fundamental to the understanding of development. As we have illustrated in this chapter, this developmental trajectory of passing through increasingly complex representational and processing capacities is both biologically realistic and computationally attractive in terms of the acquisition capacities it confers on a developing system. Models of psychological development that fail to grow come up short on all these counts. Letting networks grow and providing a role for environmentally-derived instructive signals to participate in this growth is a major departure in the study of cognitive development.

Acknowledgements

The research discussed here was supported by NSF grant 0093757 (Steven R. Quartz and Shreesh P. Mysore), and a grant from the Natural Sciences and Engineering Research Council of Canada (Thomas R. Shultz). We are grateful to Evan Balaban, Frederic Dandurand, Karim Nader, Yoshio Takane and J. P. Thivierge for helpful comments on earlier drafts.

Glossary

Basal CA1 dendrites The basal dendrites of neurons in a region called CA1 that is part of the hippocampus.

Bias-variance trade-off In formal learning theory, a quantification of the generalization ability of a learner, or equivalently, of its generalization error, is achieved statistically by defining it to be the sum of two components—bias squared and variance. Bias is defined as the distance of the best hypothesis in the space from the target concept, whereas variance is the distance of the current hypothesis from the target. Restricting the hypothesis space tends to reduce variance while increasing bias (the target concept may be far away from the few hypotheses in the space). On the other hand, expanding the hypothesis space increases variance but also increases the likelihood that the target hypothesis is included in the space (low bias). The trade-off that occurs when trying to minimize the generalization error is called the bias-variance trade-off.

Delay or classical eye-blink conditioning This is similar to trace conditioning in that the animal is trained to learn the association between an unconditioned stimulus and an aversive stimulus. The main difference is that the aversive stimulus comes on while the unconditioned stimulus is still on.

Neocortex Evolutionarily, this is the most recent part of the brain, and it is associated with higher cognitive functions in humans. Hence evidence of neurogenesis in this area in macaques has come as an unexpected finding, with potentially far-reaching implications to cognitive development, and to stem-cell research.

Polynomial time The complexity of a computational problem is one of the indices used to characterize and classify it. The time taken to solve the problem as a function of its size is a commonly used measure to quantify problem complexity. If the time taken to compute or solve a problem scales polynomially with its size, then it is said to be a polynomial time problem, and is considered to be tractable.

Score threshold A parameter in CC learning that governs depth of learning; learning proceeds until all output-unit activations are within score threshold of their target values.

Spatial training Typically, this involves having rats perform challenging spatial tasks on a day-to-day basis in their cages. Here, this was implemented by adding several movable ‘floors’ connected by ladders to the rats’ cages, and by having them look for food placed at different locations on these floors. Doing this on a regular basis positively affects their performance in subsequent spatial tasks. The hippocampus is known to play in important role in spatial navigation in rats.

Torque The product of weight and distance on one side of a balance scale.

Trace eye-blink conditioning The animal is trained to learn the role of an unconditioned stimulus, like an auditory tone, in predicting an aversive stimulus that follows it. In eye-blink conditioning, the aversive stimulus is a puff of air in the eyes which causes the animal to blink. The word 'trace' refers to the fact that the aversive stimulus onsets with a delay after the unconditioned stimulus ends. The idea here is that encoding this associative relationship requires a memory 'trace' in the hippocampus.

Tuning curve (auditory or visual) of a neuron Characterization of the strength of the response of a neuron as the stimulus varies. For instance, the auditory tuning curve of an ICX neuron is a curve that represents the strength of its response as a function of the location of a sound stimulus in space. Usually, the response of a neuron peaks when the sound source is at a particular spatial location and this peak determines the 'best' stimulus location for the neuron.

Whisker barrels Clusters of neurons in layer IV in the somatosensory cortex of rodents that process information from specific whiskers. These barrels are organized topographically, in the sense that neighbouring whiskers transmit information to neighbouring barrels. Also, each whisker is associated with a unique barrel.

References

Altman J (1962). Are new neurons formed in the brains of adult mammals? *Science*, 135, 1127–1128.

Altman J and Das GD (1965). Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *Journal of Comparative Neurology*, 124, 319–335.

Baum EB (1988). Complete representations for learning from examples. In Y Abu-Mostafa, ed. *Complexity in information theory*, pp. 77–98. Springer-Verlag.

Black JE, Isaacs KR, Anderson BJ, Alcantara AA, and Greenough WT (1990). Learning causes synaptogenesis, whereas motor activity causes angiogenesis, in cerebellar cortex of adult rats. *Proceedings of the National Academy of Sciences*, 87(14), 5568–5572.

Brainard MS and Knudsen EI (1993). Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the reward representation of auditory space in the barn owl. *Journal of Neuroscience*, 13, 4589–4608.

Brainard MS and Knudsen EI (1995). Dynamics of visually guided auditory plasticity in the optic tectum of the barn owl. *Journal of Neurophysiology*, 73, 595–614.

Brainard MS and Knudsen EI (1998). Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *Journal of Neuroscience*, 18, 3929–3942.

Buckingham D and Shultz TR (1996). Computational power and realistic cognitive development. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, 507–511, Mahwah, NJ: Erlbaum.

Buckingham D and Shultz TR (2000). The developmental course of distance, time, and velocity concepts: A generative connectionist model. *Journal of Cognition and Development*, 1, 305–345.

Case R, Kurland DM, and Goldberg J (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33, 386–404.

Chang PL, Isaacs KR, and Greenough WT (1991). Synapse formation occurs in association with the induction of long-term potentiation in two-year-old rat hippocampus *in vitro*. *Neurobiology of Aging*, 12, 517–522.

Changeux JP and Danchin A (1976). Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. *Nature*, 264, 705–712.

Chomsky N (1980). Rules and representations. *Behavioral and Brain Sciences*, 3, 1–61.

Cohen LB and Arthur AE (1983). Perceptual categorization in the infant. In E Scholnick (ed.) *New trends in conceptual representation*, pp. 197–220. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen LB and Arthur AE (unpublished). The role of habituation in 10-month-olds' categorization.

Dailey ME and Smith SJ (1996). The dynamics of dendritic structure in developing hippocampal slices. *Journal of Neuroscience*, 16(9), 2983–2994.

DeBello WM, Feldman DE, and Knudsen EI (2001). Adaptive axonal remodeling in the midbrain auditory space map. *Journal of Neuroscience*, 21(9), 3161–3172.

Eriksson PS, Perfilieva E, Bjork-Eriksson T *et al.* (1998). Neurogenesis in the adult human hippocampus. *Nature Medicine*, 4(11), 1313–1317.

Feldman DE and Knudsen EI (1997). An anatomical basis for visual calibration of the auditory space map in the barn owl's optic tectum. *Journal of Neuroscience*, 17, 6820–6837.

Ferretti RP and Butterfield EC (1986). Are children's rule-assessment classifications invariant across instances of problem types? *Child Development*, 57(6), 1419–1428.

Field J, Nuir D, Pilon R, Sinclair M, and Dodwell P (1980). Infants' orientation to lateral sounds from birth to three months. *Child Development*, 51, 295–298.

Fischer M, Kaech S, Knutti D, and Matus A (1998). Rapid actin-based plasticity in dendritic spines. *Neuron*, 20(5), 847–854.

Fodor J (1980). On the impossibility of learning 'more powerful' structures. In M Piatelli-Palmarini, ed. *The debate between Jean Piaget and Noam Chomsky*, pp. 142–152. Routledge & Kegan Paul.

Gelfland JJ, Pearson JC, Spence CD, and Sullivan WE (1988). Multisensor integration in biological systems. *3rd IEEE International Symposium on Intelligent Control*, 3, 147–153, Arlington, VA.

Gibson EJ (1969). *Principles of perceptual learning and development*. New York, Appleton-Century-Crofts.

Gilbert SF (2003). *Developmental biology*. Sunderland, MA, Sinauer Associates.

Gold M (1967). Language identification in the limit. *Information and Control*, 10, 447–474.

Goldman SA and Nottebohm F (1983). Neuronal production, migration, and differentiation in a vocal control nucleus of the adult female canary brain. *Proceedings of the National Academy of Sciences*, 80(8), 2390–2394.

Gould E, Beylin A, Tanapat P, Reeves A, and Shors TJ (1999a). Learning enhances adult neurogenesis in the hippocampal formation. *Nature Neuroscience*, 2(3), 260–265.

Gould E and Gross CG (2002). Neurogenesis in adult mammals: some progress and problems. *Journal of Neuroscience*, 22(3), 619–623.

Gould E, Reeves AJ, Graziano MSA, and Gross CG (1999b). Neurogenesis in the neocortex of adult primates. *Science*, 286, 548–552.

Gross CG (2000). Neurogenesis in the adult brain: death of a dogma. *Nature Reviews Neuroscience*, 1, 67–73.

Gutfreund Y, Zheng W, and Knudsen EI (2002). Gated visual input to the central auditory system. *Science*, 297, 1556–1559.

Hebb DO (1949). *The organization of behavior*. New York, Wiley.

Hyde PS and Knudsen EI (2000). Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. *Journal of Comparative Neurology*, 421, 146–160.

Hyde PS and Knudsen EI (2001). A topographic instructive signal guides the adjustment of the auditory space map in the optic tectum. *Journal of Neuroscience*, 21(21), 8586–8593.

Kempermann G (2002). Why new neurons? Possible functions for adult hippocampal neurogenesis. *Journal of Neuroscience*, 22(3), 635–638.

Kempermann G, Kuhn HG, and Gage FH (1997). More hippocampal neurons in adult mice living in an enriched environment. *Nature*, 386(6624), 493–495.

Kempter R, Leibols C, Wagner H, and Hemmen JLV (2001). Formation of temporal-feature maps by axonal propagation of synaptic learning. *Proceedings of the National Academy of Sciences, USA*, 98(7), 4166–4171.

Kleim JA, Freeman JH, Jr., Bruneau R *et al.* (2002). Synapse formation is associated with memory storage in the cerebellum. *Proceedings of the National Academy of Sciences*, 99(20), 13228–13231.

Knafo S, Grossman Y, Barkai E, and Benshalom G (2001). Olfactory learning is associated with increased spine density along apical dendrites of pyramidal neurons in the rat piriform cortex. *European Journal of Neuroscience*, 13(3), 633–638.

Knott GW, Quairiaux C, Genoud C, and Welker E (2002). Formation of dendritic spines with GABAergic synapses induced by whisker stimulation in adult mice. *Neuron*, 34(2), 265–273.

Knudsen EI (1998). Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science*, 279, 1531–1533.

Knudsen EI (2002). Instructed learning in the auditory localization pathway of the barn owl. *Nature*, 417, 322–328.

Knudsen EI and Brainard MS (1995). Creating a unified representation of visual and auditory space in the brain. *Annual Review of Neuroscience*, 18, 19–43.

Knudsen EI, Esterly SD, and Lac SD (1991). Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls: acoustic basis and behavioral correlates. *Journal of Neuroscience*, 11, 1727–1747.

Kornack D and Rakic P (1999). Continuation of neurogenesis in the hippocampus of the adult macaque monkey. *Proceedings of the National Academy of Sciences*, 96, 5768–5773.

Langley P (1987). A general theory of discrimination learning. In D Klahr, P Langley and R Neches, eds. *Production systems models of learning and development*, pp. 99–161. Cambridge, MA: MIT Press.

Lendvai B, Stern EA, Chen B, and Svoboda K (2000). Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex *in vivo*. *Nature*, 404, 876–881.

Leuner B, Falduto J, and Shors TJ (2003). Associative memory formation increases the observation of dendritic spines in the hippocampus. *Journal of Neuroscience*, 23(2), 659–665.

Linkenhoker BA and Knudsen EI (2002). Incremental training increases the plasticity of the auditory space map in adult barn owls. *Nature*, 419, 293–296.

Lipkind D, Nottebohm F, Rado R, and Barnea A (2002). Social change affects the survival of new neurons in the forebrain of adult songbirds. *Behavioral Brain Research*, 133(1), 31–43.

Macnamara J (1982). *Names for things: A study of child language*. MIT Press.

Maletic-Savatic M, Malinow R, and Svoboda K (1999). Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science*, 283(5409), 1923–1927.

Martin SJ and Morris RG (2002). New life in an old idea: the synaptic plasticity and memory hypothesis revisited. *Hippocampus*, 12(5), 609–636.

McClelland JL and Jenkins E (1991). Nature, nurture, and connections: Implications of connectionist models for cognitive development. *Architectures for Intelligence: The Twenty-second Symposium on Cognition*, 41–73, Hillsdale, NJ: Erlbaum.

Moser MB, Trommald M, Egeland T, and Andersen P (1997). Spatial training in a complex environment and isolation alter the spine distribution differently in rat CA1 pyramidal cells. *Journal of Comparative Neurology*, 380(3), 373–381.

Newell A (1990). *Unified theories of cognition*. Cambridge, MA, Harvard University Press.

Nottebohm F (1985). Neuronal replacement in adulthood. *Annals of the New York Academy of Sciences*, 457, 143–161.

Nottebohm F (2002). Why are some neurons replaced in adult brain? *Journal of Neuroscience*, 22(3), 624–628.

Nowakowski RS and Hayes NL (2000). New neurons: extraordinary evidence or extraordinary conclusions. *Science*, 288, 771.

Overstreet-Wadiche LS, Bensen AL, and Westbrook GL (2006). Delayed development of adult-generated granule cells in dentate gyrus. *Journal of Neuroscience*, 26(8), 2326–2334.

Paton JA and Nottebohm FN (1984). Neurons generated in the adult brain are recruited into functional circuits. *Science*, 225(4666), 1046–1048.

Piaget J (1970). *Genetic epistemology*. Columbia University Press.

Piaget J (1980). *Adaptation and intelligence: Organic selection and phenocopy*. University of Chicago Press.

Pinker S (1984). *Language learnability and language development*. Harvard University Press.

Poirazi P and Mel BW (2001). Impact of active dendrites and structural plasticity on the memory capacity of neural tissues. *Neuron*, 29, 779–796.

Pouget A, Deffayet C, and Sejnowski TJ (1995). Reinforcement learning predicts the site of plasticity for auditory remapping in the barn owl. In G Tesauro, D Touretzky and T Leen (eds) *Advances in Neural Information Processing Systems 7*, pp. 125–132. MIT Press.

Quartz SR (1993). Neural networks, nativism and the plausibility of constructivism. *Cognition*, 48, 223–242.

Quartz SR and Sejnowski TJ (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, 20, 537–596.

Rakic P (2002). Neurogenesis in the adult primate neocortex: an evaluation of the evidence. *Nature Reviews Neuroscience*, 3, 65–71.

Rakic P (2004). Neuroscience: immigration denied. *Nature*, 427(6976), 685–686.

Ramirez-Amaya V, Escobar ML, Chao V, and Bermudez-Rattoni F (1999). Synaptogenesis of mossy fibers induced by spatial water maze overtraining. *Hippocampus*, 9(6), 631–636.

Rosen DJ, Rumelhart DE, and Knudsen EI (1994). A connectionist model of the owl's sound localization system. In J Cowan, G Tesauro and J Alspector (eds) *Advances in neural information processing systems* 6, pp. 606–614. San Francisco, CA: Morgan-Kauffman.

Rucci M, Tononi G, and Edelman GM (1997). Registration of neural maps through value-dependent learning: modeling the alignment of auditory and visual maps in the barn owl's optic tectum. *Journal of Neuroscience*, 17(1), 334–352.

Sanai N, Tramontin AD, Quinones-Hinojosa A *et al.* (2004). Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration. *Nature*, 427(6976), 740–744.

Schmidt WC and Shultz TR (1991). *A replication of McClelland's balance scale model*. McGill University, Montreal: McGill Cognitive Science Centre.

Shultz TR (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Shultz TR and Cohen LB (2004). Modeling age differences in infant category learning. *Infancy*, 5, 153–171.

Shultz TR, Mareschal D, and Schmidt WC (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning*, 16, 57–86.

Shultz TR, Schmidt WC, Buckingham D, and Mareschal D (1995). Modeling cognitive development with a generative connectionist algorithm. In TJ Simon and GS Halford, eds. *Developing cognitive competence: New approaches to process modeling*, pp. 205–261. Hillsdale, NJ: Erlbaum.

Siegler RS (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8, 481–520.

Siegler RS (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, 46, Serial No. 189.

Sirois S and Shultz TR (1999). Learning, development, and nativism: Connectionist implications. *Twenty-first Annual Conference of the Cognitive Science Society*, 689–694, Mahwah, NJ: Erlbaum.

Stanfield BB and Trice JE (1988). Evidence that granule cells generated in the dentate gyrus of adult rats extend axonal projections. *Experimental Brain Research*, 72, 399–406.

Stewart MG and Rusakov DA (1995). Morphological changes associated with stages of memory formation in the chick following passive avoidance training. *Behavioral Brain Research*, 66(1–2), 21–28.

Thivierge JP, Rivest F, and Shultz TR (2003). A dual-phase technique for pruning constructive networks. *IEEE International Joint Conference on Neural Networks*, 559–564.

Valiant LG (1984). A theory of the learnable. *Communications of the ACM*, 27, 1134–1142.

Westermann G (2000). Constructivist neural networks models of cognitive development. *Division of Informatics.* University of Edinburgh, Edinburgh.

Wilkening F (1981). Integrating velocity, time, and distance information: a developmental study. *Cognitive Psychology*, 13(2), 231–247.

Younger BA and Cohen LB (1983). Infant perception of correlations among attributes. *Child Development*, 54(4), 858–867.

Younger BA and Cohen LB (1985). How infants form categories. In G Bower, ed. *The psychology of learning and motivation: Advances in research and theory*, pp. 211–247. New York: Academic Press.

Younger BA and Cohen LB (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, 57(3), 803–815.

Yuste R and Bonhoeffer T (2001). Morphological changes in dendritic spines associated with long-term synaptic plasticity. *Annual Review of Neuroscience*, 24, 1071–1089.

Zito K and Svoboda K (2002). Activity-dependent synaptogenesis in the adult mammalian cortex. *Neuron*, 35, 1015–1017.