Atypical representational change: Conditions for the emergence of atypical modularity

Michael S. C. Thomas and Fiona M. Richardson

School of Psychology, Birkbeck College, London

Thomas, M. S. C., & Richardson, F. (in press). Atypical representational change: Conditions for the emergence of atypical modularity. In. M. Johnson and Y. Munakata (Eds.) *Attention and Performance XXI.* Oxford: Oxford University Press.

Address for correspondence:

Dr. Michael Thomas
School of Psychology
Birkbeck College, University of London
Malet Street, Bloomsbury
London WC1E 7HX, UK
Email: m.thomas@bbk.ac.uk
Tel.: +44 (0)20 7631 6386
Fax: +44 (0)20 7631 6312

Introduction

There are a number of developmental disorders that display uneven cognitive profiles in their developmental endstates, exhibiting areas of relative strength and relative weakness. For example, in Specific Language Impairment and developmental dyslexia, a relative weakness is observed in various aspects of language compared with relative strength in non-verbal abilities. Disorders that show differential performance in numerical cognition, face recognition, and motor co-ordination have also been identified. Neurogenetic developmental disorders can display more complex patterns of uneven performance affecting multiple domains, such as the cases of Williams syndrome (relatively stronger language, face recognition, and social cognition, relatively weaker visuospatial cognition, numerical cognition, and problem solving, against a background of low IQ) and autism (a central triad of deficits in communication, imagination, and socialisation). However, the theoretical implication of these uneven cognitive profiles remains a matter for debate. One of the central issues concerns the origin of cognitive modules that are specialised for functions such as language, visuospatial cognition and face recognition. Where do these modules come from, and can they be selectively disrupted in developmental disorders?

One theoretical standpoint, which we will call *innate modularity*, argues that evidence of uneven cognitive profiles in genetic developmental disorders points to an innate basis for functional specialisation. High-level cognitive structures are taken to be pre-specified during normal development, preceding the influence of experience. Developmental disorders represent a case of differential perturbations to different innate modules. Such a proposal need not invoke innate knowledge within the modules, since a developmental process could serve to put in place the content of each component. If an innate module were atypical, this would lead to a differential deficit in the adult endstate for that domain, while initially normal components would lead to domains with normal endstate performance. However, initially normal components might be compromised by attempting to compensate across development for the faulty one(s). The uneven cognitive profile is then explained with reference to the functional (modular) structure of the normal adult system (since this is pre-specified). Proponents of this position do not rule out the possibility that qualitatively atypical functional structures could occur in developmental disorders. Instead they argue that the empirical evidence has not supported it. Thus, Temple and Clahsen (2002, p.770) argued on behavioural grounds that "there remains no empirical evidence in any developmental disorder that the ultimate functional architecture has fundamentally different organisation from normal, rather than merely lacking or having reduced development of components of normal functional architecture". Tager-Flusberg (2000, p.33) commented that "despite some variation in size (either smaller or larger) and other surface features, in fact, across a wide range of disorders it is actually quite remarkable how similar the brains of different populations are to one another and to normally developing children… To be sure, there is some functional variation, but not much beyond the degree that is observed in normal people… We need to view the brain as a dynamic system that develops along flexible but fairly bounded and directed pathways". More recently, Tager-Flusberg et al. (2003, p.22) added that "there is much less deviance in the developmental processes and neurocognitive organisation in people with genetically based disorders than has been portrayed in the literature".

A second theoretical standpoint called *neuroconstructivism* argues that assumptions of innate modularity are inconsistent with what is known about early brain and cognitive development (Elman et al., 1996; Karmiloff-Smith, 1998). For example, Karmiloff-Smith (1998) pointed out that current evidence indicates that there is no region-specific gene expression in the areas of cortex that come to underlie higher cognitive functions in adults. Moreover, brain imaging of infants suggests that modular structure may be emergent, in that it is a product of the developmental process rather than a precursor to it. For example, both localisation and specialisation of ERP waveforms increase in response to faces and spoken words across the first two years of life (De Haan, 2001; Mills, Coffey-Corina, and Neville, 1997). In Karmiloff-Smith's view, the uneven cognitive profiles found in adults with developmental disorders are due to subtle differences in the neurocomputational properties of neonate brain, constraints that are both less detailed and less domain-specific than the processing structures involved in higher-level cognitive functions. A cascade of developmental processes then attenuates or exaggerates these initial differences, so that the process of modularisation may be disrupted. The result could be an atypical modular structure in which even the cognitive processes underlying the relative cognitive *strengths* are atypical. Karmiloff-Smith (1998) suggested that the absence of overt evidence for atypical modularity stems from the poor sensitivity of standardised cognitive tests used to verify normal performance, and the restricted research attention paid to areas of strength in disorders. Reviewing evidence for face processing and language development in Williams syndrome, both areas of relative strength, she argued that behavioural and brain evidence are consistent with atypical processing underlying performance. Her conclusion was that abnormal cognitive phenotypes should not automatically be described with reference to normal adult functional structure because the structure itself may be atypical.

In assessing the relative merits of these two positions, evidence from functional brain development in infancy does appear to support the idea that modularity is emergent rather than pre-specified, with functional specialisation of brain areas increasing with age and expertise. Proponents of innate modularity typically de-emphasise the developmental process in their explanatory models, and therefore do not offer accounts of existing infant data in terms of pre-specified structures. However, one could defend the innate modularity position in at least two ways. First, one could express scepticism that we know enough about brain development to rule out the possibility that specialisation occurs via intrinsic factors (representing the maturation of pre-specified modules), or argue that brain-imaging evidence as it stands holds little relevance to the development of cognitive structures per se. Thus Dudek (2001, p.146) predicted that "a unique gene will be found for each and every distinct cortical area"; and Fodor (1998, p.130) argued that "nobody knows whether the infant's brain is plastic *in respects that affect cognitive architecture*" (italics in original). Second, one could claim that the modules present in the infant are less abstract than those found in the adult, and that development serves to glue these together into higher-level modules. Thus Baron-Cohen (1999) proposes a 'minimalist' innate modularity to explain theory-of-mind deficits in autism.

However, the neuroconstructivist and innate modularity positions could converge if it were the case that the emergence of modularity was difficult to disrupt. That is, modularity could be a product of development as neuroconstructivism suggests, but the neurocomputational properties that guide the emergence of large-scale functional structure might not be altered by the kinds of genetic mutations found in developmental disorders so that normal patterns of specialisation emerge. By contrast, if the constraints that shape the properties of the eventual functional components are irrevocably tied to the constraints that drive the emergence of modularity, then neuroconstructivism and innate modularity must represent opposing and empirically distinguishable theories.

The aim of this article, then, is to address whether there can be common ground between the two approaches, given what we currently know about the principles that guide the emergence of functionally specialised neurocomputational structures. Ultimately, we will introduce a set of computational simulations to investigate the constraints that would disrupt or preserve the emergence of functional architecture. However, to get to that point we need to do some groundwork. First, we need to unpack the theoretical claims of neuroconstructivism regarding the emergence of modularity. We will illustrate the relevant issues using the example of the lateralisation of language in the brain. Second, we need to review current computational approaches to the emergence of modularity in order to identify the constraints that guide specialisation in the normal case. Third, we need to identify a set of architectures and sample cognitive domains through which we can investigate the effects of disruption to these constraints. Specifically, to explain uneven cognitive profiles in an emergent framework, we need to ask whether different parameters guide the emergence of specialised functional components versus the computational properties within the eventual components themselves.

The emergence of modularity

The proposal that functional structure is emergent (i.e., formed as a product of development) sits between two more extreme positions that contribute to current theories of functional modularity (Bates & Roe, 2001). The first is *equipotentiality*, which proposes that all areas of cortex are equally able to perform all cognitive functions at birth. The second is *innate modularity* (which Bates and Roe refer to as 'irreversible determinism'). This states that areas of cortex are innately and irreversibly specialised for certain cognitive functions. The emergentist position seeks to reconcile two empirical facts: (1) at a broad scale, there is reasonable uniformity of outcome in the assignment of cortical areas to functions in normal adults; (2) there appears to be flexibility after early brain damage: following focal lesions, otherwise healthy children often show recovery to within the normal range of cognitive abilities. However, the exact extent both of uniformity and the completeness of recovery are still matters for debate. The constraints that guide the emergence of specialised structure must be strong enough to explain the uniformity but weak enough to accommodate the recovery. Once flexibility is added, it is also necessary to explain why the outcome should be an array of specialised systems rather than multiple redundant systems: if a component can aid in the recovery of a function, why did it not take on this function in the first place?

The emergentist proposal is that two factors explain specialisation: (1) *domain-relevance* and (2) *competition between areas for functional specialisation* (e.g., Elman et al., 1996; Karmiloff-Smith et al., 1998). Domain-relevance means that some brain areas are more suited to carrying out the computations for a given cognitive domain than others, without encoding any specific details of that domain. An area will bear differing degrees of computational relevance to a range of possible domains. Competition refers to a process where the activity of one component tends to increase at the expense of other components. To the extent that representational change is activity dependent, such change will occur differentially in the 'winning' component.

Specialisation then occurs as follows. The initial substrate of the cortex is computationally heterogeneous. Different areas are more or less able to perform the computations required for different cognitive domains. These areas compete with each other to acquire the various cognitive domains, a competition biased by the information to which various areas are initially exposed by global connectivity. (For example, auditory areas would tend to beat visual areas to compute audition since they are biased by the rich connectivity that delivers auditory input to this area. This is independent of the relative ability of these areas to perform the computations.) The winners of the competition come to specialise in a given domain, modifying their structures to represent the regularities of each domain and thereby becoming domain specific. Importantly, areas may be able to process alternative domains but less efficiently. Following damage, the less-suited areas are then able to acquire, or develop a partial specialisation for, an alternative domain, so long as they can access the relevant information. This explains recovery. Overall, the account produces two candidate constraints to guide specialisation: (1) the set of domain relevancies present in the initial state (that is, the default set of mechanisms present at the onset of development), and (2) the (biased) competition that drives eventual specialisation.

The example of language

The domain of language, and in particular the phenomenon of language lateralisation, serves to illustrate the type of data that support theories of specialised functional structure. In adults, the processing of syntax and semantics is mostly left lateralised, implying that language is specialised to structures in this hemisphere. However, this conclusion emerges more strongly from lesion data (where left-hemisphere damage produces aphasia but right-hemisphere damage does not) than from brain-imaging data (where homologous areas of the right hemisphere show activation in some comprehension and production tasks, albeit at lower levels; see Price, 2003). The left-dominance of syntax and semantics exhibits *uniformity of outcome*. It is found in around 95% of adults, irrespective of handedness (Bates & Roe, 2001). Brain-imaging data suggest that left lateralisation for word recognition emerges in infancy. Mills, Coffey-Corina, and Neville (1997) found bilateral ERP patterns in response to single words in 13-17 month olds, but left lateralised and more focal patterns in 20 month olds. The restriction to the left hemisphere was more closely associated with comprehension ability than chronological age, arguing against a maturational effect strictly linked to age.

The parameters that drive this emergent specialisation are as yet unknown. There are anatomical differences in brain structure between the hemispheres, for instance a larger left temporal plane, but this was found in only 65% of individuals, somewhat short of the 95% that exhibit left lateralisation of language (Reggia & Schulz, 2002). Indeed, structural hemispheric differences have been argued to dissociate from functional differences, since rare individuals with *situs inversus* (left-right reversed internal organs) nevertheless still show left-lateralised language (Kennedy et al., 1999; Walker, 2003). Bates and Roe (2001) argued for a functional difference that might pull language development to the left side, specifically that left temporal areas are better than right temporal areas in the fine perceptual discrimination required to recognise (and later produce) speech sounds. Huttenlocher (2002) noted that the development of synapses in the auditory cortex precedes that in Wernicke's area, which in turn precedes synaptogenesis in Broca's area. This reflects the sequence of functional development found in these regions of cerebral cortex (perception of speech sounds precedes language comprehension which precedes language production), implying a possible role for plasticity in mediating specialisation.

Despite evidence of left lateralisation at 20 months, the *flexibility* of the system is illustrated by the fact that if unilateral brain damage occurs prior to the age of 5-7 years, children recover to acquire language in the normal range. Although the side of damage predicts the profile of recovery, eventually, as adults, there is no obvious effect of side of damage, or a disadvantage for initial left-sided damage that can only be revealed by subtle psycholinguistic measures (Bates & Roe, 2001; Huttenlocher, 2002). Following early left-sided damage, then, the greater plasticity of the child's brain may permit the right hemisphere to acquire language. To do so, it is possible that the right hemisphere constructs a *de novo* language system. However, it is also possible that language development exploits derelict right-hemisphere structures remaining from the earlier phase of bilateral processing (Huttenlocher, 2002); or even that recovery takes advantage of existing *redundant* right-hemisphere systems that are suppressed during normal performance. Certainly in normal adults, homologous right-hemisphere areas appear to be inhibited by the left hemisphere during language production, and released from that inhibition after left hemisphere damage (for left and right pars opercularis: Blank et al., 2003). A similar effect has been observed in the perception of same-species vocalisations in rhesus monkeys: right-hemisphere inhibition in the temporal pole in the normal adult, followed by greater right hemisphere activation after disconnection from the left hemisphere (Poremba et al., 2004). Thus the brain's position on a scale between *specialisation* and *redundancy* is not settled.

Moreover, it has recently been claimed that specialisation is not fixed in adulthood, with the balance of left vs. right contributions changing across life span. Szaflarski et al. (2004) used fMRI to assess regions of interest in 121 right-handed children and adults between 5 and 63 years of age in a verb generation task. They reported that language became more left lateralised with age in children and young adults, reaching a maximum lateralisation between 25 and 35, and then started to become more bilateral again. These researchers suggested that the increase in specialisation reflects improved linguistic skills, maturation of the central nervous system, and pruning of synaptic connections, while the later reduction reflects compensation for age-related loss of functional capacity.

Although specialisation has often been associated with expertise, the functional significance of left-sided language specialisation is not clear-cut. It has been suggested that women demonstrate more bilateral patterns of activation than men (e.g., in a reading task: Shaywitz et al., 1995), while performing better on language-related tasks (Kimura, 1992) and exhibiting better recovery from unilateral left-sided damage (Strauss, Wada, & Hunter, 1992). Atypical language lateralisation is not necessarily associated with language pathology (Knecht et al., 2001), suggesting that different individuals may utilise different patterns of cortical organisation to process the same information and to produce normal behaviour (weakening claims of uniformity of outcome). Specialisation to a single hemisphere is therefore not necessarily associated with better performance and may be associated with greater vulnerability to damage.

If emergent language lateralisation can stand as a proxy for specialisation, answers to the following questions remain unclear: (1) What are the neurocognitive factors that drive specialisation? (2) When does specialisation fully occur (if at all)? (3) To what extent does specialisation become fixed with age or experience? (4) Is recovery better explained by reorganisation or by bilateral redundancy? And (5) does specialisation necessarily convey an (externally observable) behavioural advantage? To investigate these questions further, the process of specialisation must be specified in greater detail, and for this we turn to computational modelling.

Computational approaches to emergent modularity

Three principal types of computational account have been put forward to account for emergent functional specialisation (Jacobs, 1999). In the *mixture-of-experts* approach (Jacobs, 1997; Jacobs et al., 1991), the initial system is comprised of components that have different computational properties. A separate mechanism gates the contribution of these components to the output. When the overall system is presented with a task, the gating mechanism mediates a competition between the set of components, allowing the most successful component for each training pattern both to drive output performance and to update its weights to become better at that pattern. Across training, certain mechanisms come to specialise on sets of patterns, by virtue of having an initial (perhaps small) advantage in processing those patterns. In *the neural selectionism* or *parcellation* approach, the initial computational system has a surplus of connections. However, during learning, many of these connections are weeded out (pruned), whereas others are stabilised depending on usage. In addition, a locality constraint favours the stabilisation of connections between nearby processing units. The result is that nearby units communicate with each other and come to perform the same functions, whereas those far apart do not communicate and come to specialise in different functions (Jacobs & Jordan, 1992; Johnson & Karmiloff-Smith, 1992). In the *wave of plasticity* approach, the initial computational system experiences differential responsiveness to learning, both spatially and temporally. Plasticity is reduced over time across a sheet of computational units, so that one side of the sheet loses its plasticity earlier than the other. The result is that later maturing units can employ the functions computed by the earlier maturing units as input, thereby computing more abstract functions from them. The later maturing units effectively specialise in more abstract aspects of the problem domain (Shrager & Johnson, 1996).

To date, the majority of computational modelling approaches to atypical development have focused on processing anomalies within pre-specified functional modules. For example, parameter variations have been used to explain characteristics of autism in categorisation networks, characteristics of dyslexia in reading networks, or characteristics of Williams syndrome and Specific Language Impairment in inflectional morphology networks (see Thomas & Karmiloff-Smith, 2002a, for a review). One model has offered a preliminary insight to atypical specialisation. Oliver et al. (2000) explored the parameters that affect the successful emergence of topographic maps in a self-organising network. In this model, the network was presented with four bars on an input retina and had to learn a map with four regions, one specialised to recognise each bar. Oliver et al. found that several computational parameters disrupted the organisation of the map, including reducing the length of lateral inhibitory connections in the output layer and altering the relative similarity of the four input patterns. They suggested that the simulations might offer a model of ‘encapsulation’ and the emergence of information-processing modules. The parameter variations could therefore represent atypical modularisation via a disrupted parcellation process. However, the functional significance of the disrupted networks was hard to evaluate given the simple problem domain.

Three other computational models are also relevant in that they establish some of the conditions under which specialisation can occur. Dailey and Cottrell (1999) used a mixture-of-experts model to capture the emergence of specialised structures for face recognition in a system also trained to recognise images of books, cups, and cans. The model had two components that competed to classify the images. In one version of the model, there was no processing difference between the two components other than their initial random connection weights. Partial specialisation of the faces to one mechanism did occur but the effect was not particularly strong. Nevertheless, this condition demonstrated that with very low domain relevance, competition is sufficient to produce specialisation. In a second version, one component was fed high spatial frequency information from the visual input while the other was fed low spatial frequency information. The result was much more reliable emergent specialisation for face recognition to the low spatial frequency component.

Monaghan and Shillcock (2004) employed a similar approach to capture hemispheric asymmetries in a model of unilateral visual neglect. The hidden layer of a three-layer network was split so that the left side had gaussian units with narrow receptive fields, while the right side had gaussian units with wide overlapping receptive fields. This manipulation implemented the assumption that the two hemispheres have different spatial scales at which they prefer to operate. The model was trained to recognise the location and length of lines presented on an input retina. When the network was given a unilateral lesion after training, its performance on a line bissection task replicated a number of asymmetries found after cases of human brain damage. These included larger displacements of the centre of the line following damage to the coarse-coded right side than to the left side of the hidden layer, and faster recovery after left damage than right damage. In this model, there was no explicit competition process, yet domain relevance led to specialisation.

Finally, Reggia and Schulz (2002) reviewed a number of their models designed to explore the computational conditions under which two cerebral regions that communicate via a simulated corpus callosum produce emergent specialisation between the two hemispheres. Using both self-organising and backpropagation networks, they examined the effect of using excitatory vs. inhibitory connections between the hemispheres, as well as a range of other parameter variations. Their results demonstrated that specialisation can occur in the absence of competition between the hemispheres or even in the presence of excitation, so long as the parameter settings of the two hemispheres are different enough. However, specialisation effects were stronger with inhibitory connections implementing a competitive process.

The current simulations

In the following sections, we report the results of new computational work designed to investigate possible disruptions to the emergence of functionally specialised structures. We focused on four issues:

- What are the computational parameters that affect emergent specialisation in associative (task-driven) systems?
- What are the relative contributions of domain-relevance and competition, the two features of the emergentist approach, in driving emergent specialisation?
- Are there computational parameters that determine whether a system with multiple components will exhibit pre-specified modularity, emergent modularity, or redundancy?
- When a normally developed system with specialised components experiences damage, how do parameter settings alter the patterns of recovery? In particular, under what conditions do side-of-lesion effects persist after recovery (consistent with innate modularity) and under what conditions do side-of-lesion effects eventually disappear (consistent with emergentism), as in the case of language development after early unilateral brain damage (Bates & Roe, 2001)?

Architectures, cognitive domains, and parameters: Previous modelling work has drawn a distinction between two types of neurocomputational learning systems (O'Reilly, 1998). In *self-organising* systems, the emergence of functional structure is based upon patterns of similarity in the input. Self-organising systems are normally used to learn categories within input sets. In *error-driven* systems, the emergence of functional structure is guided through an error signal. The error signal is normally used to allow the system to learn input-output transformations. It is possible that the key parameters guiding specialisation differ between the two types of system, and we therefore consider them separately. While the models are greatly simplified from neurobiological and behavioural reality, they are nevertheless based on some basic principles of biological neural elements, circuits and synaptic plasticity (Reggia & Schulz, 2002). We take two exemplar cognitive domains, one appropriate to a self-organising system, one to an error-driven system, both of which demonstrate emergent specialisation of functional structure in our chosen architectures. The two cognitive domains are relatively rich, facilitating evaluation of possible decrements on behavioural task performance caused by different parameter settings. The self-organising domain is a feature-based set of semantic representations. The error-driven domain is English past tense. As in previous simulations of developmental deficits, parameter manipulations are carried out prior to training the various models, and subsequent developmental trajectories and endstate deficits are then assessed (Thomas & Karmiloff-Smith, 2002b). The parameters we consider are the amount of processing resources for both self-organising and error-driven systems, plasticity and the nature of the learning rule in the error-driven system, competition mechanisms in both systems, and changes in input frequency in both systems.

Simulations

Simulation 1: Specialisation in self-organising systems

In this simulation, we consider the effects of parameter manipulations on the emergence of specialised structures in a self-organising feature map (Kohonen, 1995), one example of a self-organising system. This architecture is relevant to basic sensory development, e.g. of topographic maps. Our model is similar to that employed by Oliver et al. (2000) described previously, in that the relationship between inputs is reflected in the topography of the map formed on the cortical layer. The architecture we used had simpler activation dynamics but was exposed to a richer cognitive domain, permitting a consideration of the quality of the learned representations for driving categorisation. The example categories were drawn from semantics and included humans, animals, vegetables, fruit, dairy produce, kitchen utensils, tools, and vehicles, defined over a set of features. Our interest was to explore two constraints in the model: (1) a domain-relevant constraint – the level of resources available in the output layer (that is, the number output units available to represent the categories); and (2) the competition process occurring in the network. To alter conditions of competition, we either reduced the neighbourhood size of the units that would update their weights for a given pattern (i.e., the size of the region of units affected when a given unit became associated with a given pattern), or we altered the balance of the input set to favour some categories over others. The final self-organising maps were evaluated with regard to whether they maintained the same regions of specialisation, and whether they permitted the same degree of discrimination within each semantic category.

Simulation details

Architecture: We employed the Self-organising feature map (SOFM) available in the Matlab Neural Networks toolbox (Demuth & Beale, 2002). The SOFM is a 2-layer network with full connectivity between the layers. When a pattern is presented to the network, the output unit with the weight vector closest to the input vector becomes the winner. The weights for this unit and for those in a given neighbourhood surrounding it are then updated. The result is a topological arrangement of the SOFM, with neighbouring regions of the input space coming to be represented by neighbouring units on the map. The network was trained in two phases. The first 'ordering' phase defined the broad topology (with a large but reducing neighbourhood size and large but reducing learning rate). The second 'tuning' phase refined the topology (smaller fixed neighbourhood size and smaller and slowly declining learning rate). Six parameters defined training: the number of presentations of the training set in the ordering phase, the total number of presentations, the ordering phase learning rate, the tuning phase learning rate, and the final neighbourhood size. In the normal condition, the following values were used, respectively: 1000 presentations, 3000 presentations, 0.9, 0.02, and a size of 1. Two atypical parameter sets were used to disrupt competition: we either set the neighbourhood size to zero throughout training (the 'No neighbourhood' condition); or we allowed a brief ordering phase of ten presentations of the training set before setting the neighbourhood to zero (the 'Fast commitment' condition). The output layer used a hexagonal topology, and a box-distance function to calculate the distance between two units. In the normal condition, a 14x14 output layer was used. These resources were progressively reduced to 12x12, 10x10, and 7x7 maps. The input layer comprised 154 units. Each network was run three times with different random seeds for the initial weights.

Training set: We exposed the network to a training set comprising semantic representations for 58 concepts. The concept set was based on that used by Small et al. (1996), who constructed an item set from concepts employed in neuropsychological tests of semantic deficits. Our training set comprised 8 major categories: humans, animals, vegetables, fruit, dairy produce, kitchen utensils, tools, and vehicles. Concepts were defined over 154 meaningful semantic features (such as "has legs", "is food"). Each concept activated on average 19 of these features. An individual input pattern consisted of 154 binary digits, indicating the presence or absence of each semantic feature. From the set of 58 prototypes, multiple exemplars (variations on a given prototype) were generated from each prototype to produce a final training set of 185 items (for example, if "apple" was the prototype, "green apple" and "red apple" might be two exemplars). A single random order of the items was created and this was used in all conditions, to ensure that any variability stemmed from internal rather than external constraints. To investigate the effect of altering the balance of inputs from different categories, two Altered Input Competition conditions were created. One combined the exemplars for Living concepts with the prototypes of Non-Living (A) for a training set of 143 items. The other combined the prototypes for Living concepts with the exemplars for Non-living concepts (B) for a set of 100 items.

Results

The maps produced in each condition are presented in Figure 1 (a colour version of this diagram is available at http://www.psyc.bbk.ac.uk/people/academic/thomas_m/). To plot these diagrams, we used a cluster analysis of the similarity structure in the training set to assign a greyscale to each pattern. This colour was then assigned to the unit(s) that became activated on presentation of this pattern to the trained network. If several patterns activated the same unit, the colours were averaged. The number of patterns activating each unit determined the size of the circle representing that unit. Fig. 1(a) depicts the normal condition for the 14x14 map. The systematic change of shading across the map indicates that the network formed separate areas representing the major semantic categories. Of the network's 196 output units, on average 78 (40%) were activated by one or more inputs. Figs. 1(b) to 1(d) illustrate the effect of reducing the map size. Three points are of note. First, for smaller maps, a similar organisation of specialised areas still emerged. Replications revealed some instability in whether fruit/vegetables or people/animals were to the right of vehicles, but Living patterns were always separate from Non-living. Second, for the smaller maps, the circles became progressively larger. That is, each unit came to respond to more input patterns. As a result, these maps were less able to discriminate between items *within* each category. Third, as the maps became smaller, they progressively filled up: on average the 12x12 activated 70 of its 144 units (48%), the 10x10 activated 58 of its 100 units (58%), and the 7x7 36 of its 49 units (72%).

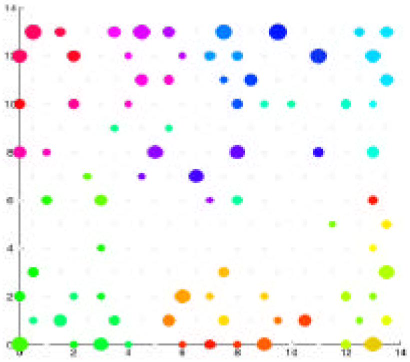
Figs. 1(e) and (f) demonstrate the effect of altering the balance of input patterns from each category. These networks demonstrated comparable use of units and broadly similar organisation of the maps, although the location of areas could be different. The effects of altering the inputs were sometimes subtle, with expansion of areas for over-represented categories and reduction of areas for under-represented categories, but little change in the discrimination that the maps offered within categories (at least with reference to the altered training sets). In the normal condition, Living occupied 60% of the active map space and Non-living 40%. For Competition A, where the Living category was considerably larger, Living items occupied 73% of the active map space and Non-living 27%. For Competition B, where the proportion of Non-living items was larger, Non-living items occupied 72% of the active map space and Living 28%. Fig. 1(g) demonstrates the effect of eliminating neighbourhoods. The result is a very disordered map, sparsely filled (16 of 196 units), an absence of specialised regions defined by similarity, and very poor discrimination. Fig. 1(h) demonstrates the result of quickly shrinking the neighbourhoods. The result is an organisation that reflects the similarity structure of the input, but again, a sparsely filled network (9/196) and very poor (indeed absent) discrimination within categories.

Discussion

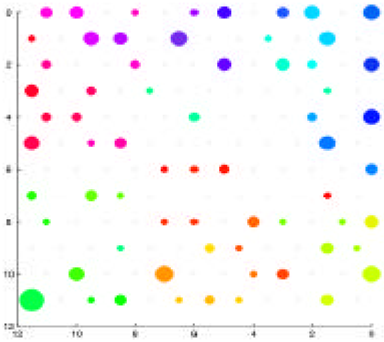
When the domain-relevant parameter of resource level was varied in the SOFM, the functional specialisation of regions was retained but within-category discrimination was lost. When competition was changed at the input level, regions for each category changed their relative sizes whilst retaining many of the features of overall organisation. When competition was altered at output (represented by changes in the neighbourhood function), both organisation and discrimination were lost. This finding reflects observations from the empirical literature. Huttenlocher (2002) notes that healthy infants suffering either focal or diffuse damage to the cerebral cortex tend to exhibit an impairment in the overall efficiency of cortical functions (reflected in a decrease in IQ) rather than patterns of differential cognitive deficits. There are also parallels with much lower level functioning in the research on sensory map formation in the animal literature. In comparative studies of cortical field development in marsupials (the short-tailed opossum), it was found that reducing the size (processing resources) of the cortical neuroepithelial sheet unilaterally at an early stage of development nevertheless led to the normal spatial relationships between visual, somatosensory, and auditory cortical fields in the reduced cortical sheet, but an increase in neurons responding to multiple inputs (Huffman et al., 1999). By contrast, changing the competition between inputs by peripheral innervation of vision early in development led to a subsequent alteration in the organisation of adult cortical areas, with auditory and somatosensory systems expanding to capture the usual visual area (Kahn & Krubitzer, 2002). In sum, this simulation suggests that the level of processing resources is a parameter that can be disrupted (and lead to performance decrements) without compromising functional specialisation in self-organising systems. By contrast, functional specialisation is more readily altered when the parameter of competition is atypical.

Figure 1. Self-organising maps for variations in map size (a to d), variations in input (e to f) and variations in competition within the output layer (g to h)

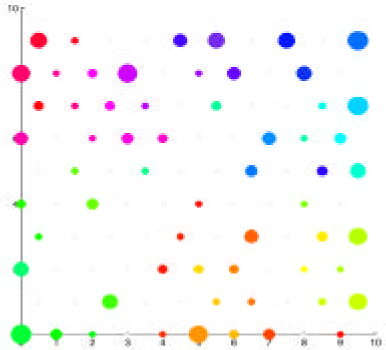
(a) Normal (14x14)



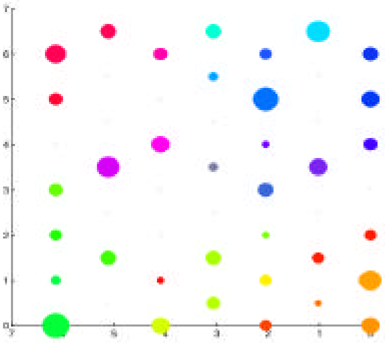
(b) Reduced resources (12x12)



(c) Reduced resources (10x10)



(d) Reduced resources (7x7)



🟥 Tools

🟪 Utensils

🟩 Vehicles

🟦 Vegetables

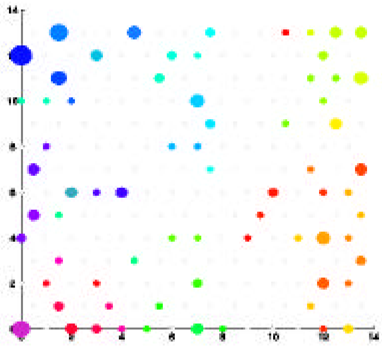
🟪 Dairy produce

🟦 Fruit

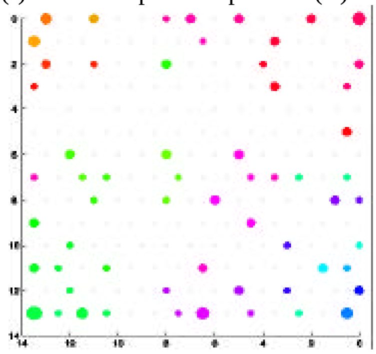
🟧 Animals

🟨 Humans

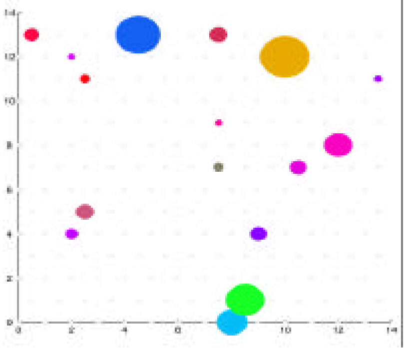
(e) Altered Input Competition (A)



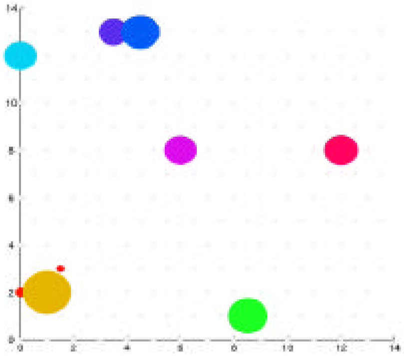
(f) Altered Input Competition (B)



(g) Altered Output layer competition (No neighbourhood)



(h) Altered Output layer competition (Fast commitment)



Simulation 2: The emergence of functional specialisation in Error-driven systems

In this simulation, we consider the effects of parameter manipulations on the emergence of specialised structures in an error-driven system. We employed a base model that was required to learn the transformations for a quasi-regular domain, which is characterised by a predominant rule along with a smaller set of exceptions to the rule. The problem domain was drawn from work on language development, specifically the acquisition of English past tense within inflectional morphology. In this task, the model is required produce the past tense form of a word when presented with its present tense at input (for example, regular: "part-parted", exception: "go-went"). This domain is useful for two reasons. First it has a bipartite organisation of regular versus exception mappings. We have previously shown that in a type of a mixture-of-experts model, those two classes of mapping can show emergent specialisation to two processing mechanisms (Thomas & Karmiloff-Smith, 2002b). Second, there has been an extensive debate within the field of language development on whether the cognitive system deploys *a priori* separate mechanisms to learn the two parts of the past tense domain (Pinker, 1994, 1999), or whether acquisition proceeds via a single undifferentiated system (Rumelhart & McClelland, 1986). Researchers supporting the former theory have already speculated on the competitive processes necessary to control the two pre-specified mechanisms during development, which will become relevant in Simulation 3. In the meantime, this training set again provides a relatively rich cognitive domain against which we can assess both performance and functional specialisation. In this simulation, we took the base model of Thomas and Karmiloff-Smith (2002b) and explored the effect on emergent specialisation of varying four computational parameters determining the domain-relevant computational properties of the system. (We consider variations in competition in the following simulation). The four parameters were: (1) processing resources, (2) plasticity, (3) the nature of the associative learning rule, and (4) input frequency.

Simulation details

Architecture: The base model was a backpropagation network in which input and output layers are connected by two routes: either by direct connections (the Direct route) or via a set of hidden units (the Indirect route). The normal condition of the model was trained using the backpropagation algorithm with a cross entropy error measure, learning rate of 0.1 and momentum of 0, for 500 presentations of the training set (random order without replacement). The network had 90 input units and 100 output units, with 20 hidden units in the Indirect route. Processing resources were varied by including 100 hidden units in the Indirect route. Plasticity was varied by multiplying the learning rate by a factor of 4 either in the Direct route (the '41' conditions) or the Indirect route (the '14' condition). The learning rule was varied by changing the backpropagation to employ RMS error to the target (BP-RMS), producing a network more vulnerable to entrenchment (Thomas & Karmiloff-Smith, 2003). Six replications of each network were run using different random seeds. In all figures, error bars portray the standard error of the means across the six replications.

Training: The training set comprised 508 training items, with a further set of 410 test items assessing regular generalisation. Performance was assessed on five categories of items: Regular mappings (410 items within the training set), generalisation of the regular Rule to 410 novel items similar to the regular items in the training set, and three types of exception patterns. Exception patterns varied in their level of inconsistency with regular items, which might alter the extent to which they are driven to use alternate processing mechanisms. EP1 exceptions (20) were most consistent with the regular training items. EP2 exceptions (68) were less consistent with the regular training items. EP3 exception items (10) were most inconsistent, sharing input similarity with regular items but requiring a qualitatively different transformation. Training items were split into high and low frequency groups. To ensure the acquisition of the EP3 items, these were given a higher token frequency that all other patterns, hence they are labelled EP3f. This training set therefore permits assessment of the effects of mapping type on specialisation, including (1) a continuum of consistency and (2) the effect of token frequency. Performance was measured at 1, 2, 5, 10, 25, 50, 100, 200, and 500 epochs of training (full details can be found in Thomas and Karmiloff-Smith, 2002b).

Results

All parameter conditions led to successful acquisition of this quasi-regular domain, albeit with developmental trajectories that could be accelerated or delayed, and relative rates of acquisition of regulars and exceptions that could be altered. Figure 2(a) demonstrates the developmental trajectories for the five pattern types in the base model and includes the effect of token frequency on acquisition. The base model contains two processing routes (Direct and Indirect) and the problem domain contains two types of problem (Regular / Rule vs. Exceptions). Figure 2(b) demonstrates the specialisation of function of the problem types to the two routes. Specialisation was assessed by measuring the differential impairment of a unilateral lesion to each route, at a level that did not cause performance to stay at ceiling or hit floor (this corresponded to a loss of 50% of connections). Using the standard logic of neuropsychology, if a pattern type was impaired more by damage to the Direct mechanism than to the Indirect mechanism, it was assumed to be more specialised to the Direct mechanism. The Fig.2(b) indicates that partial specialisation of this system emerged across training, with Regulars and Rule-based generalisation preferring the Direct mechanism, EP1 relying equally on both routes, and both the more inconsistent EP2 and EP3f patterns preferring the Indirect route. Technically, this partial specialisation occurs because the model requires hidden units in order to learn exception patterns, since the mapping problem is linearly inseparable and cannot be solved with only one layer of weights. More broadly, the exceptions form the harder part of the problem that requires the power of the hidden units to solve. However, the one layer of weights in the Direct route is more plastic than the two layers of weights in the Indirect route, so most mappings are initially acquired by the Direct route and then progressively move over to the Indirect route with further training. Importantly, Fig.2 demonstrates that the token frequency of mappings modulates the pattern of specialisation, tending to accelerate the shift from Direct to Indirect routes – that is, the relatively frequency of items in the training set itself is sufficient to alter patterns of specialisation.

Figure 3(a) depicts the developmental trajectories for the increased resources, BP-RMS and differential plasticity conditions, while Fig. 3(b) illustrates the emergent specialisation for these conditions. Altering the learning algorithm has a subtle effect on specialisation, changing the ability of the Direct route to accommodate both EP1 and EP3f patterns early in training, and driving EP1 across to the Indirect route. However, both resource changes and plasticity changes have marked effects on specialisation. Provision of extra resources in the hidden layer pulls all functions across to this route. The relative plasticity of the two routes is able to override structure-function correspondences (i.e., how well the two routes are suited to computing the two parts of the problem domain) and impose functional specialisation by a method that might be called the 'who gets there first' approach. However, all conditions achieved only partial rather than full specialisation by the end of training.

Figure 2. (a) Acquisition profile of the dual-route network, including the impact of token frequency. (b) Specialisation of the patterns to each route.

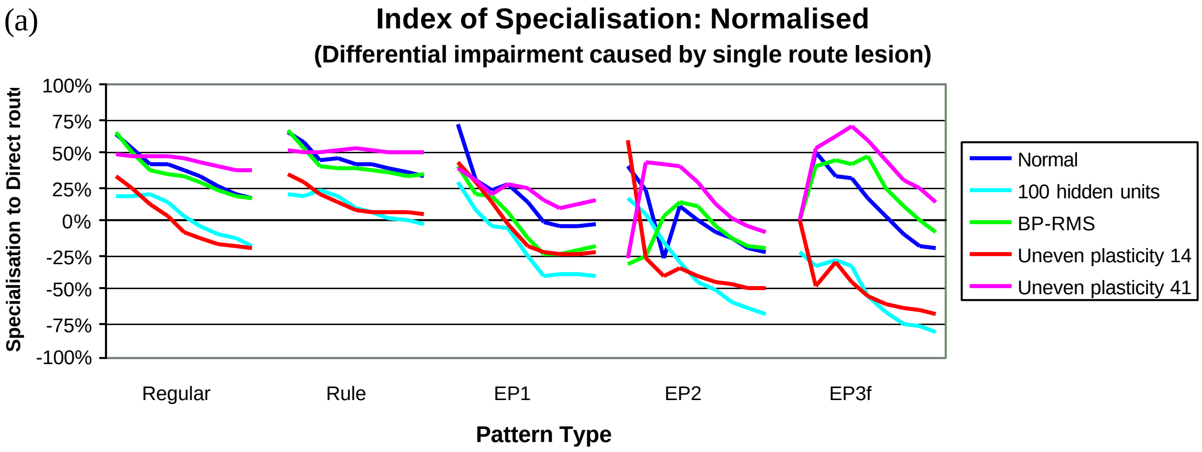
Discussion

This model demonstrates that structure-function correspondences can lead to emergent specialisation of function in an error-driven system, but that resources and relative plasticity of the processing routes play a significant role in driving specialisation. More subtle effects were produced by modifications to the learning rule. Even the token frequency of the patterns could modulate trajectories of specialisation. The influential role of plasticity reproduces similar findings by Reggia and Schulz (2002) in their simulations of hemispheric asymmetry. It is also consistent with proposals by Huttenlocher (2002) that the differential rates of plasticity in different cortical areas (indexed by synaptogenesis) may play a role in driving functional specialisation. However, differential plasticity can only have an effect when the relation between mechanisms and domains is 'relevant' rather than specific; that is, each mechanism must have at least some ability to process each domain.

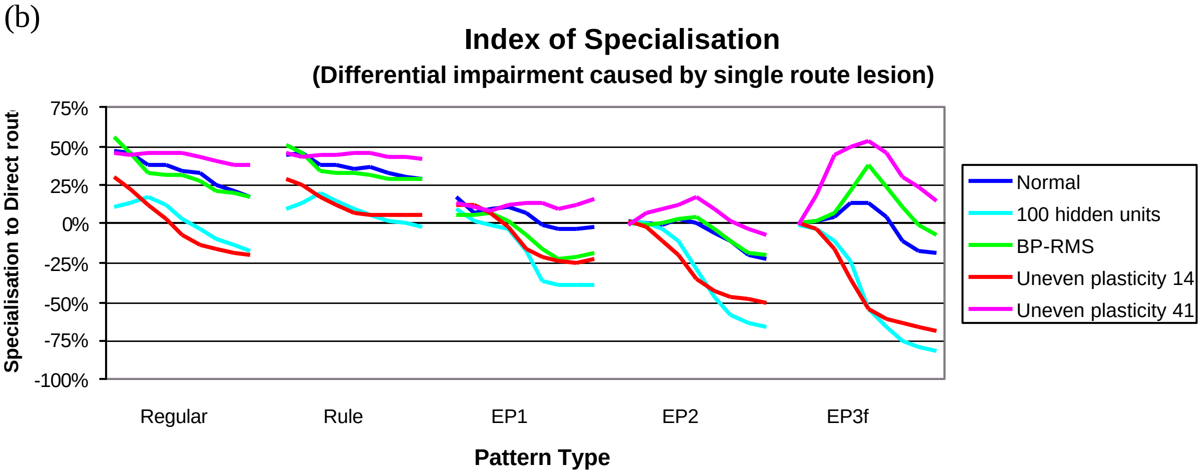
Returning to the broader picture, inappropriate levels of resources or region-specific changes of plasticity appear as constraints that would alter emergent functional structure in error-driven systems. However, importantly, the outward behavioural consequences of these alterations were subtle, producing little more than uneven or delayed development. On the other hand, the results *did not* point to any parameters that would allow the emergence of normal functional structure while showing circumscribed performance decrements. If anything, they pointed in the opposite direction: performance similar to normal could be achieved by an atypical underlying functional structure.

Figure 3. (a) The effect of parameter changes on acquisition profiles. (b) The effect of parameter changes on emergent specialisation.

(a)



(b)



Simulation 3: Varieties of competition

In this simulation, we consider variations to the competition process in error-driven systems. To foreshadow the results, it turned out that competition alone could mediate between a developmental system with *pre-specified modularity*, one with *emergent modularity*, and one with multiple *redundant* components. To understand this, we need to take a brief step backwards: A question one might ask of the preceding simulation is why specialisation occurred *at all*. While the architecture included two processing mechanisms like a traditional mixture-of-experts model, there was no gating mechanism to force the two routes to compete. Why, then, did they specialise? Why didn't each route attempt to compute all the patterns to the best level it could, thereby producing redundancy of function?

Further analysis revealed that a different form of competition was operating in this network, one that we will call 'Update' competition. For each input, both routes generate a contribution to the output layer. The difference between this output and the target leads to an error signal that allows weights to be changed in both routes. However, if one route is producing the correct answer before the other route has figured out its contribution, there will be no disparity between output and target, and therefore no error signal to drive further weight change in the non-contributing route. This form of competition is sufficient to drive specialisation. However, Update competition does not prevent weaker routes from making a contribution per se, it merely freezes the contribution when the error at output has been eliminated. As a result, Update competition encourages *co-operation and partial emergent specialisation*. Monaghan and Shillcock's (2004) model of hemispheric asymmetries in unilateral visual neglect provides an example of specialisation through update competition.

Update competition contrasts with two other forms of competition: (1) 'Input' competition and (2) 'Output' competition. In 'Input' competition, each mechanism is only presented with the patterns that it must learn. Because Input competition can ensure that each component is exposed only to patterns from a single cognitive domain, it is a way to implement *pre-specified modularity*. One might envisage at least three ways in to implement Input competition: (i) it might stem from the initial pre-wiring of the system, what Elman et al. (1996) call the global architecture. Certain areas of the cortex receive certain inputs and not others by virtue of their location. (ii) It might be the outcome of a self-organising process, whereby connections from certain inputs may be pruned as a function of learning. For example, when the self-organising map learns, strengthening the weights from the input layer to the winning area of the map means that connections to other areas are weakened. Eventually, areas distant from the winning location will simply stop receiving the signal for a given input and therefore can no longer compete to be activated by it. (iii) Input competition might be achieved by some kind of intelligent 'gatekeeper' that directs the input patterns to various mechanisms depending on their identity. Fodor (2000) argues that the gatekeeper to support this form of Input competition would have to be a (rather powerful) domain-general processing mechanism. Such a gatekeeper figures in a recent proposal by Pinker (1999) for a dual-component cognitive system for acquiring the English past tense. Construed in terms of our base model, Pinker's proposal amounts to training the Direct route only on Regular patterns. A gatekeeper would need to identify these from exception mappings, even though the two are fairly similar at input.

The third form of competition is 'Output' competition. In this case, all mechanisms are allowed to compute an answer for a given input. However, only the 'best' output will drive behaviour, while the other mechanisms are either ignored or potentially inhibited. For instance, in the example of language, we saw earlier that activation of left-sided language areas causes inhibition of homologous right-sided areas, consistent with some bilaterally of language function that is silenced by Output competition (Blank et al., 2003). Pinker's (1994) model of past tense formation also includes Output competition, whereby (in terms of our current architecture) the Indirect route would overrule the Direct mechanism and drive output under certain circumstances. One complication with Output competition is how to decide which mechanism is providing the 'best' output, particularly if the overall system is presented with a novel pattern where neither mechanism necessarily has a 'correct' answer. One possibility is to take something like the highest activation level. Finally, the use of Output competition is consistent with producing *multiple redundant systems* where all components attempt to learn all parts of the problem.

Although we have identified three different types of competition, combinations of these three types are possible. For example, within this scheme the traditional mixture-of-experts model is a combination of Update and Output competition. In this simulation, we explore the implications of all combinations of Input, Update, and Output competition on the emergence of specialised or redundant structures, as well as on the external developmental trajectories exhibited by each type of overall system. The domain-relevant computational properties of the system are held constant.

Simulation details

Architecture and training: We employed the same architecture as in Simulation 2. For simplicity, competition types were treated dichotomously, as present or absent. Update competition was implemented by training both Direct and Indirect routes in tandem, so that both contributed simultaneously to the output and error was propagated back to both routes. Input competition was implemented by training the Direct route in isolation on the Regular patterns, and the Indirect route in isolation on the Exception patterns. It was thus implemented in an absolute form. The output layer was common to both and therefore the routes were constrained to use the same threshold settings on the output units. Output competition was implemented by training both routes in isolation as above but now on the whole training set. During testing, the output activations were computed separately for each route. To determine the 'best' output, a thresholded version of each was created, with values set to 1 if a unit was activated above 0.5, and 0 if it was activated below 0.5. The Euclidean distance between each actual output vector and its thresholded version was then calculated. The route with the smallest distance reflected the most 'binary' output. Since all targets in the training set were 0 or 1, a more binarised output could be judge a more confident response. The most binary output vector from the two routes was assigned the winner and therefore the output from the whole system. Again, this is an absolute implementation of Output competition. Note that traditional the mixture-of-experts architecture permits weighted combinations of each route (see Dailey & Cottrell, 1999).

Input, Update, and Output competition could each be employed in the network, providing 8 combinations. However, Update competition is meaningless if both routes are not being supplied with the input (i.e., if Input competition is on), since a route without input cannot contribute to the output during training. This leaves 6 combinations. The network was trained using the parameters of Simulation 2 for these six combinations. When the Indirect route was trained in isolation, its 20 hidden units risked making it underpowered to learn a given set of mappings. An additional condition using 100 hidden units in the Indirect route was therefore also assessed. Since generalisation to novel inputs was one of the performance metrics, networks trained with Input competition were tested by presenting the input to both routes, as in Pinker's (1994) instantiation of Input competition.

Results

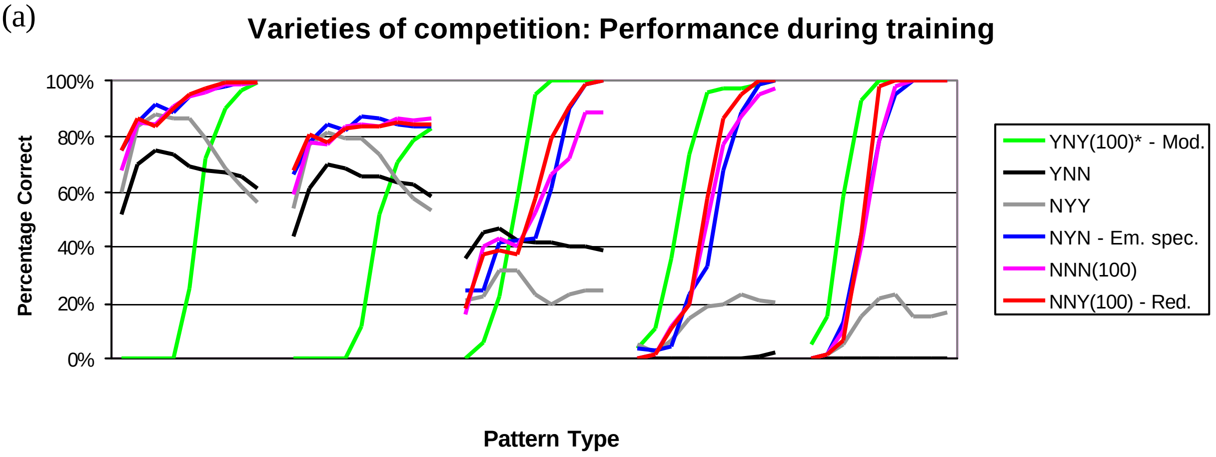
Figure 4 shows the developmental trajectories for the six conditions. Where acquisition was only successful with 100 hidden units in the Indirect route, only this trajectory is plotted. Two of the combinations were unsuccessful in acquiring the quasi-regular domain. Otherwise, competition decisions tended to modulate developmental trajectories, sometimes differentially across regulars and exceptions.

The two unsuccessful combinations failed for the following reasons. The network trained without Input competition but with Update and Output competition (NYY) was unsuccessful because the mechanisms that had co-operated in reducing the error on the output layer were now required to function in isolation and in competition. The division of labour meant neither had enough labour on its own. The network trained with Input competition but without Update and Output competition (YNN) was unsuccessful because two mechanisms trained in isolation were not co-ordinated at output, and therefore interfered with each other's responses.

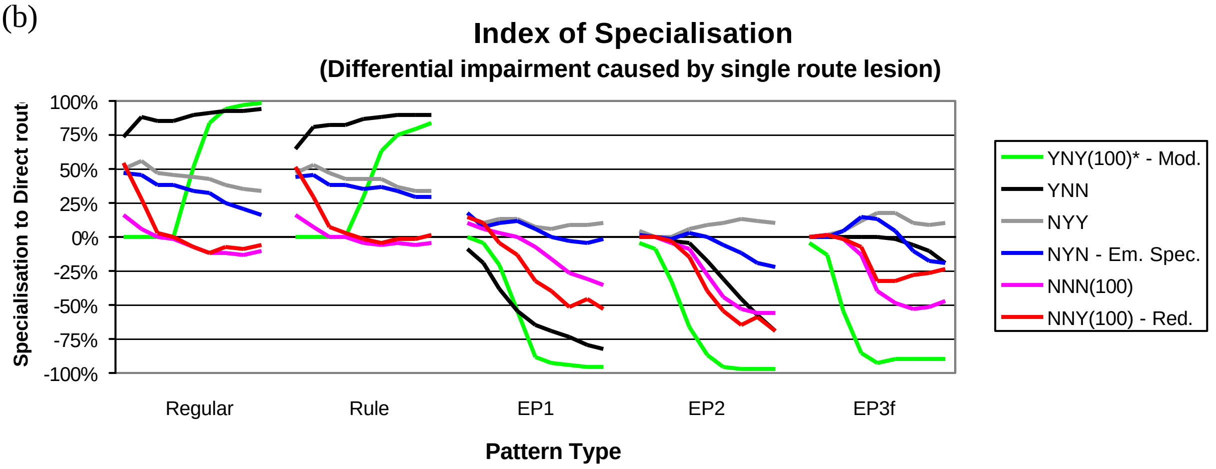
The network with Input competition, Output competition, but no Update competition (YNY100) captures the combination proposed by Pinker (1994) for how the child acquires this domain of grammar (a modification of this model proposed by Pinker, 1999, was also explored, however the results are not reported here). Two things are notable here. First, this combination produced a *pre-specified modular* system that successfully acquired the domain. Second, the conditions under which acquisition was successful were rather circumscribed. The hidden layer of the Indirect route required 100 hidden units. But even in this case, performance (as selected by Output competition) was mostly driven by the Direct route. This is because the Exception patterns were learnt by a mechanism with two layers of weights rather than one. The two layers take more training to produce an equivalent level of binarisation of output values. Thus, even when the Indirect route had the correct answer, the Direct route was more confident of the incorrect answer. Only by biasing the output competition (increasing the ‘confidence’ measure of the Indirect route by a factor of 200, a value determined via a parameter search) was the Indirect route successfully able to drive the output for its set of patterns. It is possible that such a calibration could have been acquired by learning, e.g., by gradually biasing the Output competition each time a route produced the correct answer but did not win the output competition. Nevertheless, under the conditions used here, pre-specified modularity to drive performance at a common output was not a robust solution for acquisition.

Figure 4. (a) Acquisition profiles for different combinations of competition. (b) Specialisation profiles. The first letter indicates the presence of Input competition (Y or N), the second letter indicates the presence of Update competition (Y or N), and the third Output competition (Y or N). 100 indicates the use of 100 hidden units in the Indirect layer. Mod. = Pre-specified modularity, Em. Spec. = Emergent specialisation, Red. = Redundancy. * = Indirect route had to be biased during Output competition to permit successful acquisition.

(a)



(b)



As we saw in Simulation 2, the sole use of Update competition produced *emergent specialisation* (NYN, NYN100). Successful acquisition was achieved with only 20 hidden units, implying that emergent specialisation represents an efficient use of resources. Use of additional resources (Fig.3b) also produced partial specialisation, but now with the heavily resourced route playing the dominant role.

When neither Input nor Update competition was used (NNN100, NNY100), acquisition was successful with or without Output competition, so long as additional resources were used in the Indirect route. As we shall see in the next simulation, these combinations produced *redundant* systems. The system performed better with Output competition, since the routes did not have to shout over each other to drive performance – the most confident route could produce behaviour. At the end of training, 90% of Regulars were produced by the Indirect route, 100% of Exceptions, and 50% of Rule generalisations. The main contribution of the Direct device was therefore in generalisation (see Taatgen & Anderson, 2002, for a similar result with an ACT-R hybrid symbolic-associative model acquiring the past tense domain). A summary of the outcomes of competition combinations can be found in Table 1.

Table 1. The effect of varying types of competition on specialisation in a dual route network

Input Competition	Update Competition	Output Competition	
		Yes	No
Yes	Yes	N/A	N/A
	No	Pre-specified modularity	Unsuccessful acquisition
No	Yes	Unsuccessful acquisition	Emergent specialisation
	No	Redundancy	Redundancy

Discussion

This simulation demonstrated that in error-driven systems, competition is exceedingly important in driving functional specialisation because, for a given set of domain-relevancies / structure-function correspondences, it can differentiate between pre-specified modularity, emergent specialisation and redundancy. For example, one could have an otherwise equipotential system segregated by Input and Output competition into a pre-specified modular system. In this case, the equipotentiality could be demonstrated by taking a processing element from, say, the vision component and placing it in the audition component, where it would be equally at home and start to be conditioned by the input/output mappings of that domain. Such equipotentiality of function would, however, be fully consistent with pre-specified modularity. Therefore the theoretical distinction between innate modularity and equipotentiality that we encountered in the Introduction is not necessarily a dichotomy – in this case the two exist as different settings of a competition parameter.

Although the conceptual outcome of the combinations of competition might have been anticipated in advance, implementation demonstrated that (1) emergence was a resource-efficient form of acquisition, (2) competition decisions nevertheless had implications for developmental trajectories even when acquisition was successful, and (3) pre-specified modularity required a more delicate balance of parameters than the other combinations to learn this particular problem domain. Overall, the implication of this simulation is that if the conditions of competition are not perturbed, modular architecture should not be disrupted.

Simulation 4: Recovery patterns of systems with different functional structure

In the Introduction, we noted the challenge of trying to reconcile uniformity of outcome in functional specialisation with the flexibility implied by recovery after damage. This simulation sought to assess the implications of the specialisation induced by competition for recovery after damage. It did so with particular reference to (1) the possibility of redundant systems aiding recovery (e.g., for language, see Blank et al., 2003); and (2) whether systems produced side-of-damage effects, either during recovery or once recovery was complete (e.g., for language, see Bates & Roe, 2001). The combinations of competition from Simulation 3 that successfully acquired the problem domain were given either a bilateral lesion to both routes, unilateral damage to the Direct route, or unilateral damage to the Indirect route, and their patterns of recovery observed. Note that these models assume unchanging plasticity with age.

Simulation details

Architecture and training: For the successful networks in Simulation 3, a probabilistic lesion of 75% of network connections was applied either to both routes, to the Direct route alone, or to the Indirect alone. Networks were then retrained for 500 epochs, using same parameter settings as in Simulation 3. Performance was measured after 1, 2, 5, 10, 25, 50, 100, 200, and 500 epochs of training after damage.

Results

Figure 5 demonstrates the recovery profiles following a *bilateral* lesion, for the *pre-specified modular* system, for the *emergent specialisation* systems with 20 and 100 hidden units, and for the *redundant* system with 100 hidden units and Output competition. For Regular and Rule patterns, only the pre-specified modular system failed to show strong recovery. For Exception patterns, recovery was weaker in the emergent system with limited resources and the redundant system, and stronger in the rich emergent and modular networks. Rich resources aided recovery. Interestingly, the rich emergent system was more successful in recovering from overall damage than the redundant system. This was because it could use its remaining resources co-operatively. In most cases, recovery was aided by frequency, either the higher type frequency of Regulars or the higher token frequency of the EB3f exceptions.

Figure 5. Recovery profiles for different dual route systems following a lesion to both routes.

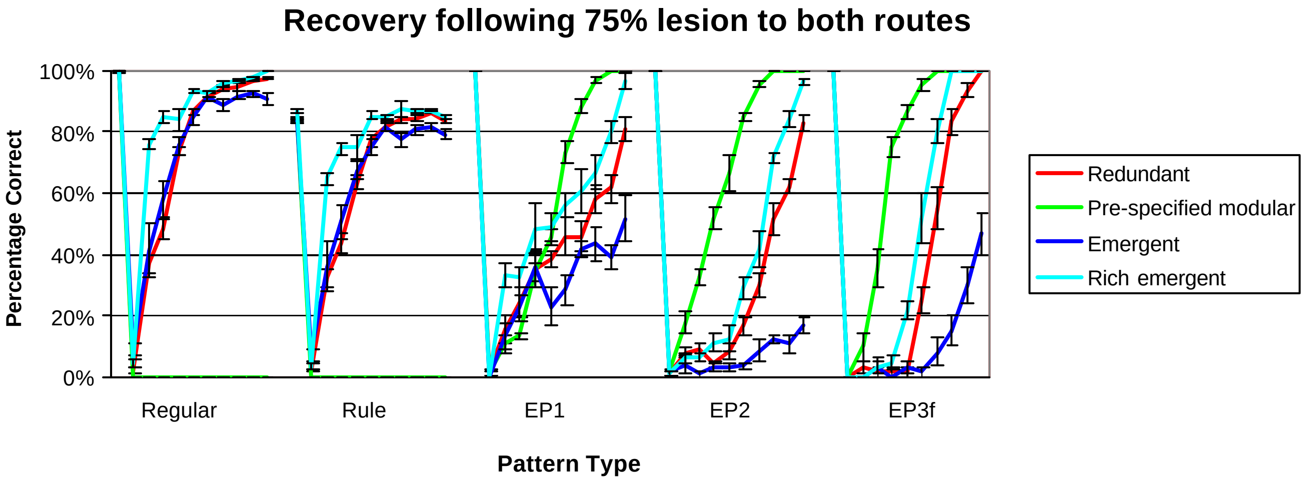


Figure 6 (a) to (c) contrast recovery patterns following *unilateral* lesions to each route for the pre-specified modular, emergent, and redundant systems. For the two conditions with Output competition (modular and redundant), Fig. 6(d) indicates which route was driving performance across recovery.

The emergent systems (Fig.6a) exhibited differential vulnerability for the two types of damage. Direct damage caused a decrement across the board and Indirect damage targeting Exceptions. Recovery after Direct damage was fast for the resource rich version, and slow and incomplete for the resource poor version. Recovery after Indirect damage was slower for both versions and complete only for the resource rich version. Overall, the resource rich emergent system, with co-operation between its routes, demonstrated the strongest recovery. The trajectories of recovery differed depending on the side of damage, but notably, following recovery, no there was *no evidence of the initial side of damage*. By contrast, the pre-specified modular system (Fig.6b) revealed extreme side of lesion effects and poor recovery. Damage to the Direct route produced complete and irrecoverable loss of Regular and Rule patterns, because the Indirect route now dominated the output competition. Damage to the Indirect route produced Exception impairments with slow and incomplete recovery, as the Direct route now tended to dominate the output. The redundant system (Fig.6b) was robust to Direct route damage but showed vulnerability of Exception mappings to Indirect damage and incomplete recovery.

Finally Fig.6c demonstrates how the presence of Output competition causes a system to shift to driving behaviour from an intact mechanism following unilateral damage. This is adaptive only if the intact mechanism possesses the appropriate knowledge for correct performance. The recovering mechanism may later start to influence behaviour once more. In the current case the recovering mechanism primarily drove the (limited) recovery of Exception performance in the modular and redundant systems.

Discussion

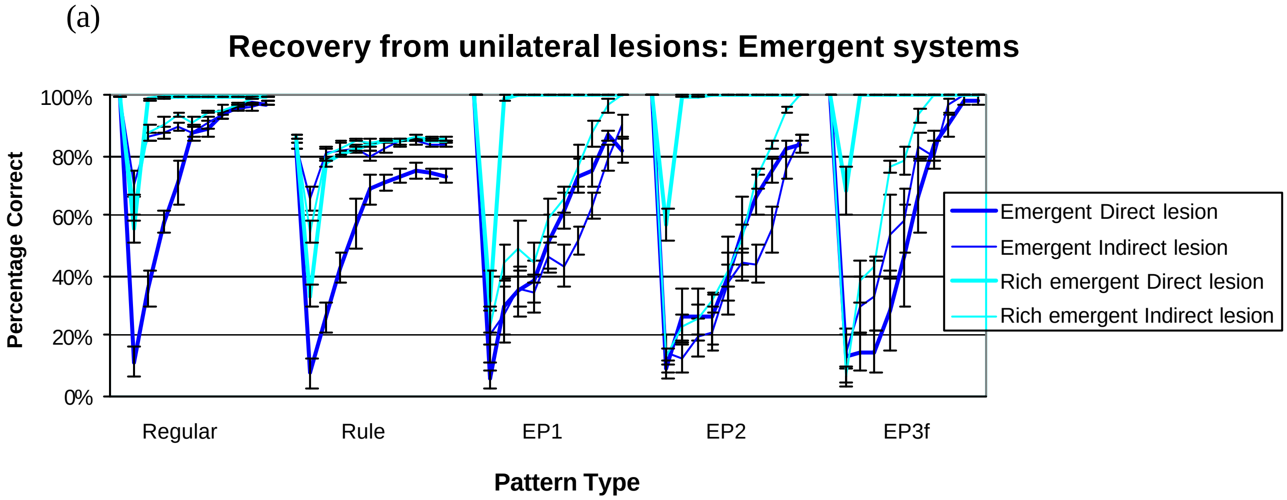
The redundant system was robust to damage but did show some unilateral vulnerability. This stemmed from the different domain relevancies of the two processing routes in the network. The Direct route was less able to learn both parts of the domain than the Indirect route, therefore loss of the Indirect route was more serious. Unless the two routes are equipotential, even redundant systems will show unilateral vulnerability under some circumstances.

The system with emergent specialisation and rich computational resources demonstrated the strongest recovery. It illustrated the importance of having abundant computational resources available to relearn the problem domain, but also the ability to exploit remaining resources in a co-operative fashion. Most notably, this system reproduced the pattern found in the recovery of language after early unilateral damage (Bates & Roe, 2001). Recovery patterns were conditioned by side of damage but final performance levels were independent of side of damage.

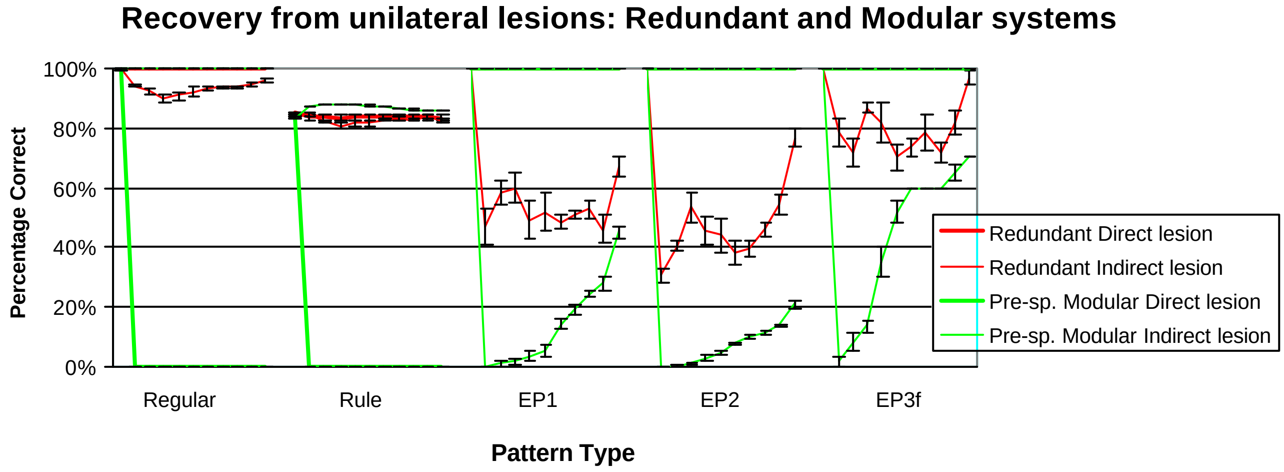
The emergentist position is therefore an account that may be able to explain both uniformity of outcome (driven by initial domain relevancies) and recovery from early injury (because structure-function correspondences are relevant rather than specific). By contrast, pre-specified modularity appears only able to explain uniformity of outcome. However, its strong side-of-damage effects and limited recovery is characteristic of adult aphasia. If the emergent system best explains developmental damage and the pre-specified modular system best explains the adult state, this suggests some qualitative change of the system with age, presumably in its effective plasticity or available resources.

Figure 6. Recovery profiles for unilateral lesions. (a) Systems with emergent specialisation. Rich = use of 100 hidden units in the Indirect route. (b) Redundant and Pre-specified modular systems. (c) The route that drives behaviour during recovery for the Redundant and Modular systems.

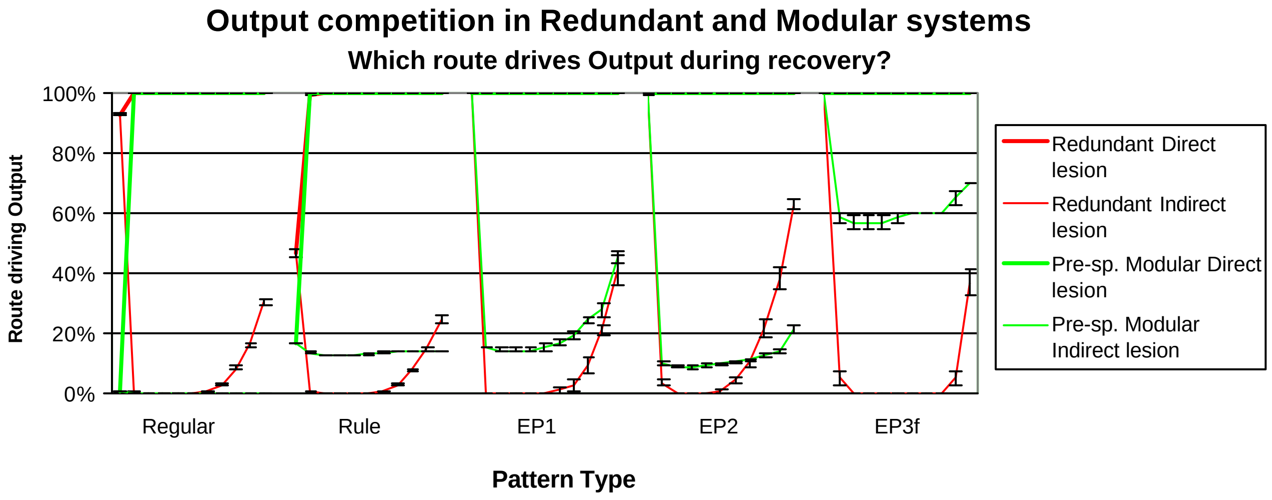
(a)



(b)



(c)



Conclusion

We first summarise the main findings with regard to the specific aims of the simulation work. We then relate these to our wider theoretical questions. The modelling work indicated the following: (1) For self-organising systems, changes in competition disrupted specialisation, with changes to intrinsic competition in the output layer more powerful than changes to the input (for the parameter variations we considered). Reductions in resources led to the same functional structure but poorer categorical discrimination. (2) For error-driven systems, we evaluated a version of the mixture-of-experts model. This revealed that differential properties of areas of heterogeneous computational substrate (i.e., the experts) could mediate patterns of functional specialisation. Differential settings of plasticity and levels of resources between areas were particularly powerful modifiers of emergent specialisation. (3) In error-driven systems, once again, competition was a powerful factor. Indeed the choice of competition settings was sufficient to mediate between outcomes with pre-specified modularity, emergent specialisation, and redundancy in a system with otherwise identical domain-relevant computational properties. (4) Consideration of recovery after damage indicated that systems with emergent specialisation (defined by Update competition) were best positioned for recovery so long as they were sufficiently resourced, because this form of emergent specialisation arises from conditions encouraging co-operative processing. The system with emergent specialisation demonstrated a recovery profile found in language development in children experiencing early unilateral brain damage. This supports the idea that the emergentist position can reconcile normal uniformity of outcome with flexibility after early damage, by specifying suitable constraints on learning. Pre-specified modularity could account for uniformity of outcome but (as implemented here) it was not a robust solution for successful acquisition. Moreover, it showed greater vulnerability and poorer recovery after damage than either emergent or redundant systems, a pattern more in common with deficits after adult focal brain damage.

We now return to our broader theoretical questions regarding the origins of specialised functional structures. First, of the neurocomputational factors that drive specialisation, can separate parameters affect specialisation versus functioning within the subsequent individual components? The answer is a qualified ‘yes’. In self-organising systems, resources could affect within-component processing but leave emergent structure untouched. However, this did not lead to uneven performance across components. In error-driven systems, competition was a parameter that altered specialisation without changing the functioning of components. However, changes to the computational properties of individual components led to an alteration in specialisation. In short, we did not find conditions were uneven profiles of performance arose at the same time as normal functional structure.

Second, we asked when does full specialisation occur, if at all? Self-organisation appeared to drive fairly complete specialisation. In the error-driven system, when the competition parameter guided the system to emergent modular structure or redundancy, specialisation was only partial. For error driven systems, complete specialisation required pre-specification of functional modules.

Third, we asked if specialisation becomes fixed. We did not consider alterations to plasticity across the ‘age’ of the system. However, the recovery data suggested that emergent modularity captured recovery after focal damage in childhood, while pre-specified modularity was more reminiscent of more limited adult recovery. These two findings would be reconciled if there were a change of state with age in the parameters affecting functional specialisation.

Fourth, we asked whether recovery (such as in the case of language and unilateral brain damage in children) might be better explained by bilateral redundancy rather than reorganisation. The results indicated that a resource-rich emergent model captured the data better than a redundant system. Notably, however, in systems with Output competition, focal damage could immediately cause a different mechanism to drive behaviour. Depending on whether the alternate mechanism could support normal behaviour, this could represent robustness or an immediate source of errors. In both cases, the damaged component could take over driving performance if it later recovered. This finding demonstrates the difficulty of drawing inferences from deficits, since behaviour after damage may reflect the functioning of a different component.

Fifth, we asked whether specialisation necessary conveys a behavioural advantage. The results indicated that pre-specified modularity was not a robust solution for acquisition where both components were required to drive the same output, since their competition needed to be calibrated. Both the emergent modular system and the redundant system were flexible after damage, the emergent system more so. The emergent modular system had the additional advantage of being resource efficient. However, pre-specified modularity may be advantageous where the modules must drive separate outputs - see Calabretta, Di Ferdinando, Wagner and Parisi (2003), for a model where pre-specified modules are a superior solution for learning to output the identity versus the location of an object on an input retina.

Evidence of more bilateral brain activation patterns in some developmental disorders (for example, in face processing in Williams syndrome; see Karmiloff-Smith, 1998) might be taken as evidence for *reduced* specialisation. This could be explained as less competition operating in the cognitive system, either an absence of Input competition (via initial over-connectivity or insufficient pruning) or an absence of Update competition. However, reduced specialisation may also be a consequence of reduced processing resources. When we carried out severe startstate lesions to both routes of the base model (Thomas & Karmiloff-Smith, 2002b), the result was poorer acquisition and an absence of specialisation. In effect, the system exhibited an all-hands-to-the-pump approach, because both mechanisms were necessary to produce any kind of correct performance. Emergent specialisation and redundancy are a luxury born of sufficient resources.

Finally, what do the simulation results tell us about the possible causes of uneven cognitive profiles in developmental disorders? They point to two possible sources: (1) Focal changes in the domain-relevant computational properties of different areas of the initial substrate (i.e., changes restricted to a subset of future specialised components) or focal changes in connectivity modifying the input to a restricted subset of future components. For either of these initial differences not to affect the emergence of other modular structures, the atypical area of substrate must either be self-organising or not interact / compete with unaffected regions during normal development. (2) The uneven profile is caused by an atypical overall structure, where shifts in the conditions of competition or the domain-relevant properties cause different specialisation to emerge. Importantly, the simulations suggested that variations in functional architecture might only modulate the external behavioural trajectories of development in subtle ways. Sensitive empirical measures may be necessary to discriminate between different possible functional architectures underlying variations in development.

In conclusion, the indication is that the emergence of normal modularity despite focal problems in a subset of functionally specialised components could only occur under circumscribed conditions. This implies that *innate modularity* and *neuroconstructivism* represent diverging explanations of uneven cognitive profiles in developmental disorders. The key question remains the extent to which the functional structure varies in atypical development, either for disorders with wide-ranging cognitive deficits such as autism and Williams syndrome, or for disorders in which the deficits are apparently narrower, such as Specific Language Impairment and dyslexia. The answer to this question will tell us much about the origins of modular structure in the human cognitive system.

References

Baron-Cohen, S. (1999). Does the study of autism justify minimalist innate modularity? Learning and Individual Differences, 10, 179-191

Bates, E., & Roe, K. (2001). Language development in children with unilateral brain injury. In C. A. Nelson & M. Luciana (Eds.), Handbook of Developmental Cognitive Neuroscience (p. 281-307). Cambridge, Mass: MIT Press.

Blank, S. C., Bird, H., Turkheimer, F., & Wise, R. J. S. (2003). Speech production after stroke: The role of the right pars opercularis. Ann. Neurol., 54, 310-320.

Calabretta, R., Di Ferdinando, A., Wagner, G. P., & Parisi, D. (2003). What does it take to evolve behaviorally complex organisms? Biosystems, 69, 245-262.

Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recogition in modular neural network models. Neural Networks, 12, 1053-1073.

De Haan, M. (2001). The neuropsychology of face processing during infancy and childhood. In C. A. Nelson & M. Luciana (Eds.), The handbook of developmental cognitive neuroscience, (p. 381-398). Cambridge, Mass.: MIT Press.

Demuth, H., & Beale, M. (2002). Neural network toolbox for use with Matlab: User's guide, Version 4. The Mathworks.

Dudek, S. M. (2001). Multidimensional gene expression in cortical space. Developmental Science, 4(2), 145-146.

Elman, J. L, Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking innateness: A connectionist perspective on development. Cambridge, Massachusetts: MIT Press.

Fodor, J. A. (1998). In critical condition: Polemic essays on cognitive science and the philosophy of mind. Cambridge, Mass.: MIT Press.

Fodor, J. A. (2000). The mind doesn't work that way: The scope and limits of computational psychology. Cambridge, Massachusetts: MIT Press.

Huffman, K. J., Molnar, Z., Van Dellen, A., Kahn, D. M., Blakemore, C., & Krubitzer, L. (1999). Formation of cortical fields on a reduced cortical sheet, Journal of Neuroscience, 19(22), 9939-9952.

Huttenlocher, P. R. (2002). Neural plasticity. Cambridge, Mass.: Harvard University Press.

Jacobs, R. A. & Jordan, M. I. (1992). Hierarchical mixtures of experts and the EM algorithm. Neural Computation, 6, 181-214.

Jacobs, R. A. (1997). Nature, nurture, and the development of functional specializations: a computational approach. Psychonomic Bulletin Review, 4, 229-309.

Jacobs, R. A. (1999). Computational studies of the development of functionally specialized neural modules. Trends in Cognitive Science, 3, 31-38.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural Computation, 3, 79-87.

Johnson, M. H. & Karmiloff-Smith, A. (1992). Can neural selectionism be applied to cognitive development and its disorders? New Ideas in Psychology, 10(1), 35-46.

Kahn, D. M. & Krubitzer, L. (2002). Massive cross-model cortical plasticity and the emergence of a new cortical area in developmentally blind mammals. Proceedings of the National Academy of Science (USA), 99(17), 11429-11434.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. Trends in Cognitive Sciences, 2, 389-398.

Karmiloff-Smith, A., Plunkett, K., Johnson, M. H., Elman, J. L., & Bates, E. A. (1998). What does it mean to claim that something is 'innate'? Response to Clark, Harris, Lightfoot and Samuels. Mind & Language, 13, 558-597

Kennedy, D. N., O'Craven, K. M., Ticho, B. S., Goldstein, A. M., Makris, N., & Henson, J. W. (1999). Structural and functional brain imaging brain asymmetries in human situs inversus totalis. Neurology, 53(6), 1260-1265.

Kimura, D. (1992). Sex differences in the brain. Scientific American, 267, 118-125.

Knecht, S., Drager, B., Floel, A., Lohmann, H., Breitenstein, C. Deppe, M., Henningsen, H., & Ringelstein, E. B. (2001). Behavioural relevance of atypical language lateralisation in healthy subjects. Brain, 124, 1657-1665.

Kohonen, T. (1995). Self-organizing maps. New York: Springer.

Mills, D. L., Coffey-Corina, S., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. Developmental Neuropsychology, 13(3), 397-445.

Monaghan, P. & Shillcock, R.C. (2004). Hemispheric asymmetries in cognitive modelling: Connectionist modelling of unilateral visual neglect. Psychological Review, 111, 283-308.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. Trends in Cognitive Sciences, 2(11), 455-462.

Oliver, A., Johnson, M. H., Karmiloff-Smith, A., & Pennington, B. (2000). Deviations in the emergence of representations: A neuroconstructivist framework for analysing developmental disorders. Developmental Science, 3, 1-23.

Pinker, S. (1994). The Language Instinct. Penguin books.

Pinker, S. (1999). Words and rules. London: Weidenfeld & Nicolson

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E. Herscovtich, P, & Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature, 427, 448-451.

Price, C. J. (2003). An overview of the functional anatomy of speech comprehension and production. In R. S. J. Frackowiak (Ed.), Human brain function. Elsevier Science, USA.

Reggia, J. A. & Schulz, R. (2002). The role of computational modelling in understanding hemispheric interactions and specializations. Cognitive Systems Research, 3, 87-94.

Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models (pp. 216-271). Cambridge, MA: MIT Press.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Fletcher, J. M., Shankweiler, D. P., Katz, L., & Gore, J. S. (1995). Sex differences in the functional organization of the brain for language. Nature, 373, 607-609.

Shrager, J. & Johnson, M. H. (1996). Dynamic plasticity influences the emergence of function in a simple cortical array. Neural Networks, 9, 1119-1129.

Small, S. L., Hart J., Nguyen, T., & Gordon, B. (1996). Distributed representations of semantic knowledge in the brain: Computational experiments using feature based codes. In J. Reggia, E. Ruppin, & R. S. Berndt (Eds.), Neural modelling of brain and cognitive disorders. World Scientific.

Strauss, E., Wada, J., & Hunter, M. (1992). Sex-related differences in the cognitive consequences of early left-hemisphere lesions. Journal of Clinical and Experimental Neuropsychology, 14, 738-748.

Szaflarski, J. P., Holland, S. K., Schmithorst, V. J., & Weber-Byars, A. (2004). An fMRI study of cerebral language lateralisation in 121 children and adults. Paper presented to the American Academy of Neurology 56th Annual Meeting, April 24 2004, San Francisco.

Taatgen, N. A. & Anderson, J. R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. Cognition, 86(2), 123-155.

Tager-Flusberg, H. (2000). Differences between neurodevelopmental disorders and acquired lesions. Developmental Science, 3, 33-34.

Tager-Flusberg, H., Plesa-Skwerer, D., Faja, S., & Joseph, R., M. (2003). People with Williams syndrome process faces holistically. Cognition, 89, 11–24.

Temple, C., & Clahsen, H. (2002). How connectionist simulations fail to account for developmental disorders in children. Behavioral and Brain Sciences, 25, 769-770.

Thomas, M. S. C. & Karmiloff-Smith, A. (2002a). Modelling typical and atypical cognitive development. In U. Goswami (Ed.), Handbook of Childhood Development (pp. 575-599). Oxford, Blackwells Publishers.

Thomas, M. S. C. & Karmiloff-Smith, A. (2002b). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. Behavioural and Brain Sciences, 25(6), 727-780.

Thomas, M. S. C. & Karmiloff-Smith, A. (2003). Modelling language acquisition in atypical phenotypes. Psychological Review, 110(4), 647-682.

Walker, S. F. (2003). Misleading asymmetries of brain structure. Behavioural and Brain Sciences, 26, 240-241.

Acknowledgements

This work was supported by UK Medical Research Council Grant G0300188 to Michael Thomas. We thank to two anonymous reviewers for their helpful comments.