COMPETITION AS A MECHANISM FOR PRODUCING SENSITIVE PERIODS IN CONNECTIONIST MODELS OF DEVELOPMENT

MICHAEL S. C. THOMAS

Developmental Neurocognition Lab, School of Psychology
Birkbeck College, University of London, Malet St., London, UK

Marchman's [1] framework for simulating sensitive periods in development was extended to investigate whether *competition* is a mechanism that might contribute to reductions in functional plasticity with age. Under this view, the ability to learn new behaviors is reduced when old established behaviors are unwilling to give up shared representational resources. The simulations supported this hypothesis, but indicated that a range of factors modulated competition effects: the similarity between old learning and new learning, the level of representational resources, the prevailing plasticity conditions within the system, the timing of introduction of new learning, and the complexity of the problem domain.

1. Introduction

1.1. *Age-related changes in plasticity*

Cognitive development frequently exhibits non-linear profiles of change. One of the most studied is the phenomenon of the critical period in development, in which functional plasticity appears to reduce with age. Research on plasticity is typically informed by three sources of empirical evidence: (i) the rate and upper limit of behavioral change at a given age; (ii) the effects of early deprivation on subsequent development; and (iii) recovery from damage at different ages. Research in this area has drawn several conclusions [2]. First, because plasticity is rarely ever eliminated in older individuals, the term *sensitive period* is viewed as more appropriate for age-related changes. Second, there are multiple varieties of sensitive period ranging from sensory processing to high-level cognitive abilities, often exhibiting different profiles. Third, multiple neurocomputational mechanisms may underlie changes in plasticity, including processes of entrenchment, pruning and assimilation [3]. Fourth, many sensitive periods may be a consequence of the basic processes underlying postnatal brain development, including the emergence of functionally specialized structures. In this paper, we

employ connectionist modeling to explore another potential source of reductions in plasticity with age: *competition for resources*.

1.2. *Evidence for competition as a mechanism that mediates plasticity*

Competition effects on plasticity are proposed to occur under circumstances of sequential learning, where one behavior is learned after another (and the first behavior is continued). It is the later age at which the second behavior is acquired that would give the appearance of a sensitive period in development, all other things being equal. Strabismus (squint) in young children provides evidence for the importance of competition effects. Where the input to one eye is disrupted by a squint, infants often develop dominant fixation in the other eye. In this case, the squinting eye can lose vision, termed *amblyopia*. The loss of function is viewed as a consequence of competition for synaptic sites, where the dominant eye invades the areas of representation of the squinting eye, thereby disrupting the development of the normal geniculo-cortical connections [4]. Amblyopia can be reversed by patching the dominant eye and forcing the child to use the squinting eye. That all other things are not equal is illustrated by the fact that such reversal is only possible up until 6-7 years of age [5], implying that other conditions of plasticity are altering within the visual system.

Modulation of competition has been argued to affect recovery from acquired brain damage. For example, in adult humans suffering hand paralysis following stroke, recovery of function in the paralyzed hand is aided by restraining the other functionally intact hand [6]. This is consistent with the view that recovery of the motor behavior in the paralyzed hand exploits the representational resources of the intact hand, but cannot do so while the intact hand is utilizing these resources. Similarly, it has been argued that following insult to the left cerebral hemisphere, language can be developed in the right hemisphere only before the age at which the right hemisphere has developed its normal functional circuits [7]. That is, once the right hemisphere resources are being used for their typical function, they are less plastic to accommodate other functions.

Competition for resources implies that the resources are themselves limited. Evidence regarding exact levels of resource allocation in cognitive development is hard to come by. However, *crowding effects* following focal brain damage in children are suggestive that resources are not effectively infinite in cognitive development. Following focal damage, the long-term consequences are a general lowering of abilities across the child's cognitive profile, rather than specific behavioral deficits [4]. The implication is that the residual resources are no longer sufficient to develop cognition to the expected level.

If competition reduces plasticity, then when the conditions of the competition alter, new functional plasticity should be revealed. However, such plasticity may not always be adaptive. It has been argued that phantom limb pain in amputees may in part be due to reorganization of primary somatosensory cortex, following an attenuation of the normal input from the amputated limb and invasion of adjacent areas into the limb's representational zone. For example, following the loss of an arm, the representation of the cheek was observed to take over the arm and hand representations in somatosensory cortex [8].

Second language acquisition provides a strong test of the sequential learning paradigm. Here, it has been observed that when the second language *replaces* the first language before the age of 6, the second language can overwrite the first language, removing evidence of it both at behavioral and brain level [9]. Again, the fact that this is a phenomenon of early childhood suggests that there are other plasticity conditions changing across development. By contrast, when the second language is *added to the first*, brain-imaging evidence suggests that the higher the proficiency of the second language, the more its use is associated with activation of areas used by the first language, as if the second language must compete for the optimal functional circuits established by the first language [10]. Moreover, bilingualism is not a static condition but subject to continuous slow shifts in dominance depending on relative usage of the two languages.

1.3. *Computational modeling of changes in functional plasticity*

Connectionist models have been used to explore how different mechanisms may produce a reduction in functional plasticity, providing a necessary level of analysis intermediate between behavior and brain [11]. The following four models have generated influential findings. O'Reilly and Johnson [12] constructed a model of filial imprinting in the chick brain using a self-organizing network. After a certain level of exposure to an input stimulus, no amount of training on a subsequent stimulus could cause the network to shift its preference to the second stimulus. McClelland *et al.* [13] also used a self-organizing network to simulate the ending of a sensitive period in non-native language phoneme discrimination. After a certain amount of training on a given phonemic class, the system could no longer learn a subtle distinction between two new similar-sounding phonemes. Marchman [1] used an associative connectionist network acquiring the English past tense to explore sensitive period effects in recovery from brain damage, arguing that connectionist networks exhibit slower recovery from damage the later the damage occurs in development. Lambon Ralph and Ehsan [14] used similar associative networks to explore age-of-

acquisition effects – the idea that patterns appearing earlier in training show stronger learning and are subsequently less vulnerable to damage.

Connectionist networks typically contain a learning rate parameter in their learning algorithms. This parameter modulates the size of connection weight changes induced by each learning experience. A reduction of plasticity with age could be achieved simply by reducing the value of this parameter across training. It is of note that none of the four preceding models used this method, holding the learning rate constant. Effective reductions in functional plasticity thereby revealed alternative mechanisms for generating sensitive periods, characterized respectively as self-terminating sensitive periods, assimilation and entrenchment.

The simulations presented here extend the framework employed by Marchman [1] to consider the role of *competition* in reducing functional plasticity. The following **key questions** were addressed:

1. In a sequential learning paradigm, does the functional plasticity of the learning system reduce when a later training set must compete for representational resources with an earlier (and continuously refreshed) training set?
2. Are competition effects modulated by the level of resources, in this case the number of hidden units in a three-layer backpropagation network?
3. Are competition effects modulated by the similarity between later and earlier training sets?
4. Are competition effects modulated by on-going changes in the intrinsic plasticity conditions within the system, for instance if resource levels are reducing (equivalent to pruning) or the learning rate is reducing (equivalent to a reduction in the availability of functionally unspecified synapses [4])?
5. If intrinsic plasticity conditions do change across development, does the timing of the introduction of the later training set matter? In other words, are competition effects worse at older 'ages' of the network?
6. Are competition effects modulated by the encoding of the problem domain, which alters the complexity of the mappings that the network must learn?
7. How do all of the above factors interact?

2. Simulation Methods

2.1. *Basic model*

For the base learning system, Marchman's [1] model of acquisition, loss, and recovery in connectionist networks was extended to consider competition. The simulations employed the well-understood domain of English past tense formation, which has often served as a test bed to illustrate the importance of the frequency and consistency of associative mappings in shaping performance.

These features are argued to influence many aspects of cognitive development (e.g., [15] for wider arguments in language development). The English past tense is of note because it is characterized by a predominant rule (e.g., *talk-talked*, *drop-dropped*, etc.) that extends to novel stems (e.g., *wug-wugged*), but also contains irregular verbs of different types (*go-went*, *hit-hit*, *sing-sang*). The past tense serves as a useful test domain because it allows us to examine the effects of problem type on competition, and in particular the modulating effects of consistency, type frequency, and token frequency on functional plasticity.

The regular past tense has the highest type frequency and forms a consistent set of mappings (reproduce the stem, add an inflection). The different classes of exception verb fall on a continuum of inconsistency, with no-change past tenses (*hit-hit*) least inconsistent (reproduce the stem but don't add an inflection), vowel change past tenses (*sing-sang*) at an intermediate level (partly reproduce the stem, no inflection), and arbitrary past tenses (*go-went*) most inconsistent (no relation between stem and past tense). Arbitrary past tenses have the lowest type frequency but require the highest token frequency in order to be acquired.

The following simulations had a modest objective. Marchman's model was relatively simple, both in its topology (a 3-layered feed forward network) and in its learning (the backpropagation rule). It was a cognitive model rather than a model of a language-related brain area, and some of its assumptions have restricted biological plausibility. Nevertheless, it provides a transparent framework in which to consider the direct consequences of five potentially interacting factors on competition effects: resource level, pattern consistency, intrinsic plasticity conditions, timing, and mapping complexity.

2.2. *Simulation details*

Architecture: The architecture comprised a three-layer feedforward network with 90 input units and 100 output units. Resources levels were manipulated by varying the number of hidden units, from 30 (low resources: just sufficient for successful learning) to 50 (medium resources) to 100 units (high resources).

Intrinsic plasticity: Models were trained with the backpropagation algorithm using the cross-entropy error measure. Three conditions of intrinsic plasticity were considered, shown in Figure 1. Under the *constant* condition, the learning rate was fixed at 0.1 (momentum = 0). Under the reducing learning rate or *redux* condition, the learning rate started higher and stepped downwards across training. The network's 'lifetime' was 400 epochs. The learning rate changed as follows: epoch 1=0.3; 2=0.2; 12=0.1; 200=0.05; 300=0.01. Under the *pruning* condition, after a certain point in training, unused connections were

probabilistically lesioned. Beginning at 100 epochs, connections were permanently lesioned with a probability of 0.1 if their magnitude was less than 0.5. Figure 2 demonstrates the result of pruning on the number of connections in an equivalent weight layer in a medium or a high resource level network. Solid lines indicate the number of weights when the network underwent training, while dashed lines indicate the number of weights when the network was untrained across the same period. This plot shows that many fewer connections are pruned when the network undergoes training (because the weights become larger than the pruning threshold), and proportionately fewer connections are lost in networks that start with lower resource levels.

Simulation parameter values – hidden units, learning rate (constant or reducing), pruning onset, threshold, and rate – were calibrated to ensure successful learning of the full training set within the 400 epochs allotted to each network's 'lifetime', under conditions of minimal resources.

Overall training and testing sets: The training set was an artificial language comprising tri-phonemic verbs stems represented using an articulatory feature-based phonological code (30 bits per phoneme). There were 410 regular verbs (adding *–ed* to form past tense); irregular verbs were of three types: 68 vowel change verbs, 20 no-change verbs, and 10 arbitrary verbs. The generalization set comprised 410 novel verbs that rhymed (shared two phonemes) with the regular verbs in the training set. The output layer included 10 units to encode the inflectional morpheme. To explore the role of increased mapping complexity, an alternative coding of the training set was employed. Phonemes were represented over a much more compact code, using 6 bits per phoneme [16]. This latter architecture employed 18 input units and 20 output units.

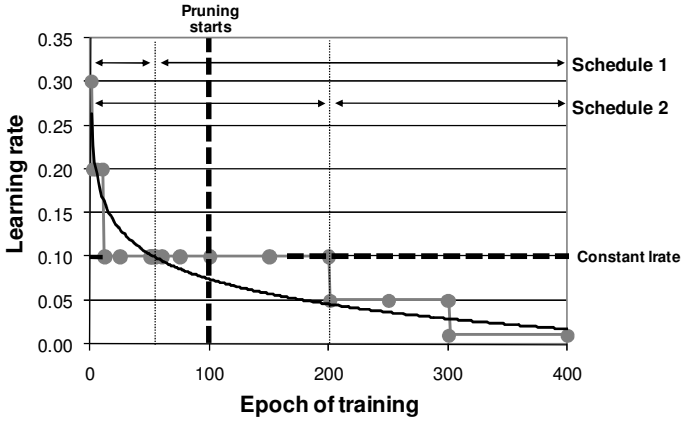

Figure 1. The three *Intrinsic plasticity* conditions

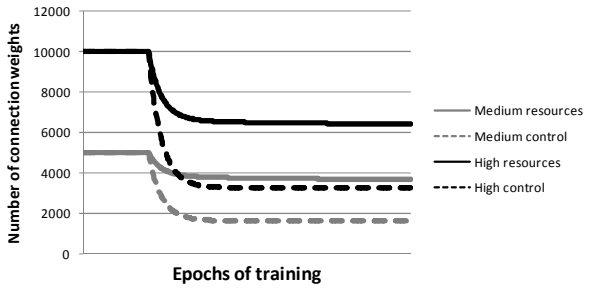

Figure 2. The effects of pruning on the number of connection weights in equivalent layers of a medium resource and a high resource network. Solid lines indicate networks undergoing training. Dashed lines indicated networks that do not undergo training.

Early and late training sets: To capture the sequential learning paradigm, the overall training set was split in two ways, either to *maximize* the inconsistency between early and late training sets, or to *minimize* the inconsistency. To maximize the inconsistency, the overall set was split into regular vs. irregular past tenses. To minimize the inconsistency, the training set was split into two equal halves, A and B.

Training schedules: For simplicity, we refer to the early training set as L1 and the later training set as L2. In the maximum inconsistency condition, we considered both the situations where L1 comprised regular verbs and L2 irregular verbs, and the situation where L1 comprised irregular verbs and L2 comprised regular verbs. In the minimum inconsistency condition, L1 and L2 comprised the two equal halves of the training set, which were counterbalanced.

To explore *timing effects*, competition either occurred when the network was relatively 'young' or when it was somewhat 'older'. *Young:* L1 was trained for 50 epochs; L2 was then either added to or replaced L1 and training continued to 400 epochs. *Older:* L1 was trained for 200 epochs; L2 was then either added to or replaced L1 and training continued to 400 epochs.

2.3. *Design*

The simulations were run in two parts. In the first part, a combinatorial design was used with four factors: Resources (low, medium, high), Intrinsic plasticity (constant, redux, pruning), Similarity (maximum inconsistency, minimum inconsistency), and Timing (Young, Older). There were six replications of each condition with different random seeds. This yielded 3x3x2x2x6=216 individual simulations. In the second part, this design was repeated but omitting the Resource factor (medium resources were used throughout) and adding a Domain Complexity factor (normal encoding, atypical encoding), yielding a further 2x3x2x2x6=144 individual simulations.

3. Results

Our central question was whether competition reduced functional plasticity. The key contrast was the rate and final level of acquisition of the later training set when it was added to, and therefore had to compete with, the earlier training set (L1+L2) compared to when it replaced the earlier training set (L2). The *rate* was assessed by deriving the mean disparity between the developmental trajectories of L2 in the two conditions; the *final level* reflected performance on L2 at the end of training (400 epochs).

Network performance was assessed via percent correct on each verb class in the training set (**Regular**s and the three exception types), as well as the network's ability to extend the past tense rule (add *–ed*) to novel verb stems, referred to here as **Rule** patterns. Network responses were derived by using a nearest neighbor method to convert output activations to phoneme strings, which were then compared to the target output. In the following, performance on exception patterns is referred to via the abbreviations **EP1** (*hit-hit*), **EP2** (*sing-sang*), and **EP3f** (*go-went*). The **EP** number emphasizes the degree of inconsistency of the verb class with the regular past tense, while the **f** indexes the greater token frequency of the arbitrary verb class. Mean results are reported averaged over the six replications, along with the standard error of the mean.

Figures 3 and 4 depict the delay and final performance levels for the Similarity condition. These two figures illustrate several points. First, competition indeed reduced functional plasticity, both in the rate of learning and final level of proficiency: there are positive differences in Figure 3 for all pattern types in the training set. Second, competition effects were much smaller in the *minimum inconsistency* condition. In the *maximum inconsistency* condition, reductions in plasticity were more marked for generalization; they increased with greater inconsistency (from EP1 to EP2) but were attenuated by high token frequency (EP3f). Lastly, delays in learning were sometimes (irregular verbs) but not always (regular verbs) associated with reductions in the final level of proficiency. High type frequency was therefore a protective factor for final levels of performance in the face of competition.

Figure 5 demonstrates that the competition effects were significantly worse when there were fewer representational resources available. With high resource levels, competition effects still reduced plasticity in terms of rate of learning (albeit to a lesser extent) but final performance levels were now unaffected. Figure 6 indicates that reductions in plasticity due to competition interacted with the intrinsic plasticity conditions prevailing within the learning system, and were particularly exaggerated when processes of pruning reduced the level of

resources available. By contrast, Figure 7 indicates that as a main effect, the timing of the introduction of the second training set did not modulate reductions in plasticity due to competition.

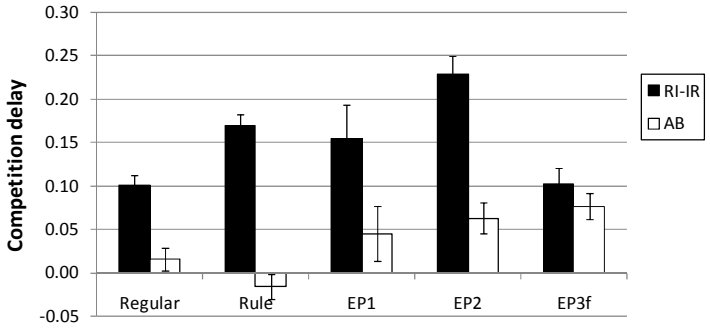

Figure 3. The delay in acquiring each pattern in its role as L2. RI-IR shows the *maximum inconsistency* condition: For Regular and Rule, irregular verbs were the corresponding L1. For EP1, EP2, and EP3f, regular verbs were the corresponding L1. AB shows the *minimum inconsistency* condition where for all pattern types, the corresponding L1 was the other half of the training set.

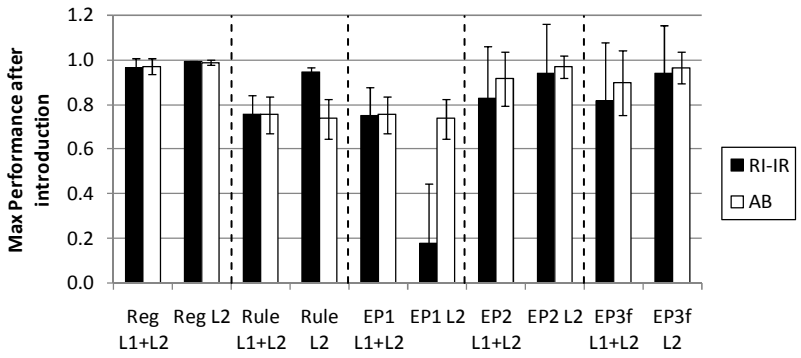

Figure 4. Final performance level for each pattern type in its role as L2, either when competing with the earlier training set (L1+L2) or replacing the earlier training set (L2). Results are shown separately for maximum (RI-IR) and minimum (AB) inconsistency conditions.

However, Figure 8 suggests that timing did interact with other factors. This figure depicts a 4-way interaction between timing, similarity, resource level, and intrinsic plasticity, and reveals complex interactions between these factors. Most notably, competition effects were exaggerated late in training when there was high inconsistency and intrinsic plasticity had itself reduced. Lastly, Figure 9 depicts the situation of an increase in the complexity of the mappings that the

system was required to learn, implemented via a change in the encoding of the problem domain. This factor was also found to increase competition effects.

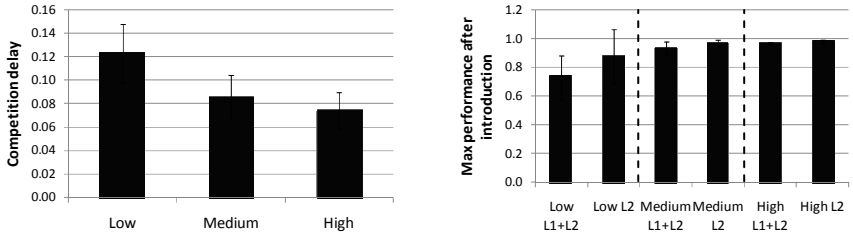

Figure 5. Delay and final performance level depending on the amount of representational resources.

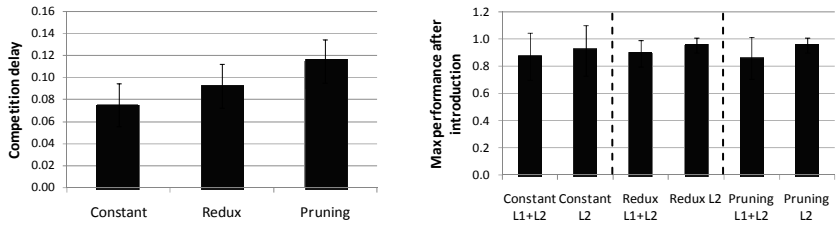

Figure 6. Delay and final performance level depending on intrinsic plasticity conditions.

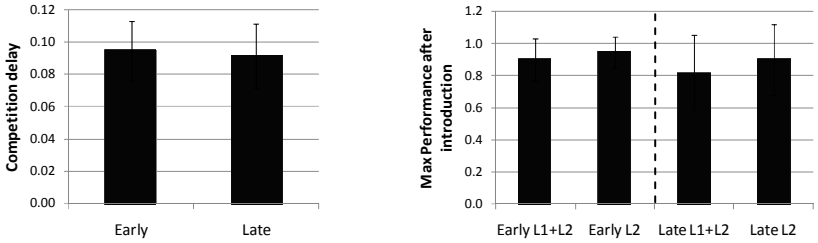

Figure 7. Delay and final performance level depending on the timing of introduction of L2.

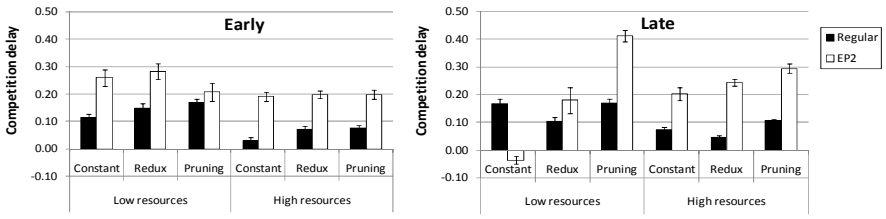

Figure 8. 4-way interaction: Timing x Pattern similarity x Intrinsic plasticity conditions x Resources.

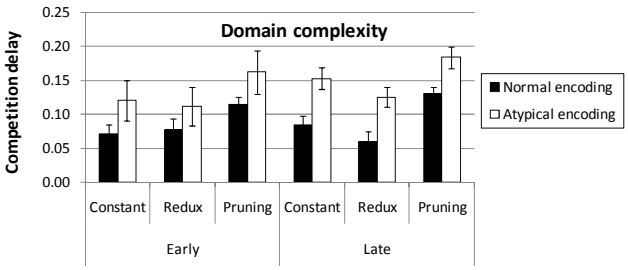

Figure 9. An increase in the complexity of the problem domain also exaggerated reductions in functional plasticity via competition for resources.

4. Discussion

The Marchman [1] simulation framework was extended to explore whether competition for representational resources may be a contributory mechanism to reducing functional plasticity across development. Under this view, the ability to learn new behaviors is reduced when old established behaviors are unwilling to give up shared representational resources. The simulation results offered support for the hypothesis.

Mechanistically, the reduction in functional plasticity due to competition can be viewed as related to effects of entrenchment. Both competition and entrenchment are consequences of previous learning. Entrenchment puts the system in a state where it can less readily respond to new learning. Competition additionally interferes with the attempts of the new learning to reconfigure the representational resources. Competition effects may have contributed to changes in plasticity reported in previous computational models of development [3].

The simulation results were notable for the number of interactions observed. Competition effects depended on consistency between old and new learning. They were mitigated by high type and token frequency in the later training set; but they were exaggerated by inconsistency, since this increased the extent to which the later training set had to fight to reconfigure the representational resources. Competition effects were made worse by other intrinsic reductions in plasticity conditions and when this occurred, the timing of introduction of the later training set became important. Competition operates over a restricted amount of resources: competition effects were attenuated when those resources were greater. They were exaggerated when the complexity of the mappings was increased, because greater complexity requires greater resources. Lastly, reductions in rate and final performance level did not always co-occur. There was therefore a dissociation between the two behavioral markers of functional plasticity.

These are, of course, preliminary findings, which need to be extended to other training sets and other learning architectures to evaluate their generality. The motivation for modeling work of this type is ultimately to establish an inventory of the mechanisms that may serve to reduce functional plasticity with age, and to identify the behavioral hallmarks of each mechanism. The aim is to prescribe appropriate interventions where people are struggling to achieve a desired behavioral change. What are the hallmarks of reductions in plasticity due to competition? Timing effects, consistency effects, and similarity effects were observed but are unlikely to be unique to competition. Competition is implicated if there are any indications of low resource levels (e.g., low IQ, developmental disorders, previous damage to the system). However, the most salient evidence that competition is responsible for reducing functional plasticity is that greater plasticity is revealed when the competing behavior is removed.

Acknowledgments

This work was supported by MRC grant G0300188.

References

1. V. Marchman, *J. Cog. Neurosci*., **5**, 215-234 (1993).
2. M. Thomas & M. Johnson, *Curr. Dir. Psych. Sci*., **17**, 1-5, (2008).
3. M. Thomas & M. Johnson, *Dev. Psychobiol*., **48**, 337-344 (2006).
4. P. Huttenlocher, *Neural plasticity*, Harvard Uni. Press: Cam. Mass. (2001).
5. A. Assaf, *Br. J. Opthalmol*., **66**, 64-70 (1982).
6. E. Taub et al., *Arch. Phys. Med. Rehab*., **74**, 347-354 (1993).
7. S. Knecht, *Brain*. **127**, 1217 (2004).
8. H. Flor, L. Nikolajsen and T. Jensen, *Nat. Neuro. Rev*., **7**, 873-881 (2006).
9. C. Pallier et al., *Cerebral Cortex*, **13**, 155-161 (2003).
10. J. Abutalebi, S. Cappa and D. Perani, in J. Kroll and A. de Groot (eds.) *Handbook of Bilingualism*, 497-515, OUP: Oxford (2005).
11. M. Seidenberg and J. Zevin, in Y. Munakata and M. Johnson (eds.) *Processes of Change in Brain and Cognitive Development*, OUP: Oxford (2006).
12. R. O'Reilly and M. Johnson, *Neural Computation*, **6**, 357-390 (1994).
13. J. McClelland *et al*., in J. Reggia, E. Ruppin and D. Glanzman (eds.) *Disorders of Brain, Behavior and Cognition*, 75-80, Elsevier: Oxford (1999).
14. M. Lambon Ralph and S. Ehsan, *Visual Cognition*, **13**, 928-948 (2006).
15. E. Bates & B. MacWhinney, in B. MacWhinney (ed.) *Mechanisms of Language Acquisition*, 157-193, LEA: Hillsdale, NJ (1987).
16. K. Plunkett and V. Marchman, *Cognition*, **38**, 1-60 (1991).