Parallel Brain: The Cognitive Neuroscience of the Corpus Callosum, pp. 1–7, MIT Press.
5 Lewis, P.A. and Miall, R.C. (2003) Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia* 41, 1583–1592
6 Brown, L.N. and Sainsbury, R.S. (2002) Age- and sex-related differences in temporal judgements to visual stimuli: support for hemispheric equivalence. *Percept Psychophys.* 64, 693–702
7 Myers, J.J. and Sperry, R.W. (1985) Interhemispheric communication after section of the forebrain commissures. *Cortex* 21, 249–260
8 Corballis, M.C. and Corballis, P.M. (2001) Interhemispheric visual matching in the split brain. *Neuropsychologia* 39, 1395–1400
9 Marzi, C.A. *et al.* (1999) Pathways of interhemispheric transfer in normals and in a split-brain subject: a PET study. *Exp. Brain Res.* 126, 451–458
10 Harrington, D.L. *et al.* (1998) Cortical networks underlying mechanisms of time perception. *J. Neurosci.* 18, 1085–1095
11 Bamiou, D.E. *et al.* (2003) The insula (Island of Reil) and its role in auditory processing: literature review. *Brain Res. Rev.* 42, 143–154
12 Bushara, K. *et al.* (2001) Neural correlates of auditory–visual stimulus onset asynchrony detection. *J. Neurosci.* 21, 300–304

1364-6613/$ - see front matter
doi:10.1016/j.tics.2003.10.015

Is speech perception modular or interactive?

Jeffrey S. Bowers[1] and Colin J. Davis[1,2]

[1]Department of Experimental Psychology, University of Bristol, 8 Woodland Rd, Clifton, Bristol, BS8 1TN, UK
[2]Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia

Norris *et al.* recently reported experimental evidence that listeners learn phoneme categories in response to lexical feedback. To reconcile these findings with their modular account of speech perception, the authors argue that top-down feedback can be used to support phoneme learning, but not to influence on-line phonemic processing. We suggest that these findings have broader implications than the authors assume, and we discuss potential challenges for integrating a modular theory with top-down learning.

Theories of cognition can be divided into two general kinds: 'modular' theories, in which autonomous perceptual systems process inputs independently of contextual influences [1–3], and 'interactive' theories, in which multiple sources interact during perceptual analysis [4–6]. A classic example of the debate between these two camps is found in the field of speech perception, where the question arises as to whether phonemic processing is influenced by feedback from higher levels of processing. It is well-established that phoneme identification is often influenced by lexical context [6,7]. For example, listeners who hear an input like *?ape*, where the *?* denotes an ambiguous phoneme that is somewhere between /t/ and /d/, are biased towards interpreting the ambiguous phoneme as a /t/, so that the input is consistent with a word (*tape*) [7].

Phenomena like this have often been interpreted as evidence that the perceptual analysis of phonemes is influenced by top-down feedback from lexical and semantic levels of processing, as in interactive theories of recognition. However, this notion was rejected by Norris, McQueen and Cutler [1], who argued that lexical influences on phoneme identification result from the later (downstream) integration of information from autonomous phonemic and lexical levels of processing. These authors have now qualified their position in an important new article [2], in which they incorporate top-down feedback to explain phoneme learning.

Does top-down feedback play a role in speech perception?

Norris *et al.* [1] advanced two key arguments in support of modular theories of speech perception. First, they developed a modular model of speech perception that accommodates findings often taken as evidence for top-down processing. On the basis of parsimony, they argued that this model should be preferred to models that invoke top-down feedback. Second, they appealed to adaptive constraints, claiming that on-line top-down feedback cannot benefit spoken word identification. The optimal performance of any model of speech perception (whether modular or interactive) is to activate the lexical knowledge that best matches the input provided to the system: according to Norris *et al*. [1], nothing is gained by modifying the phonemic representations that served as input to the lexical level. Indeed, feedback might impair perception by forcing the phoneme level to interpret its input according to lexical categorizations – creating hallucinations when the inputs do not match familiar words. For example, feedback might prevent a listener from noticing the mispronunciation *edephant* by overwriting the /d/ phoneme in the perceptual input with an /l/. In practice, listeners tend not to overlook mispronunciations [8].

Does top-down feedback play a role in learning?

Norris *et al.* [2] qualified their previous argument by noting that top-down feedback could be adaptive for the sake of learning. They give an example of a British English speaker adapting to an American English speaker. In contrast to British speakers, American speakers tend to pronounce the intervocalic /t/ as a flap (/D/) in words like *motor*. Lexical information can help British listeners adapt to American accents by indicating that the /D/ in *motor* is an instance of the phoneme /t/, rather than a new phoneme. Therefore lexical feedback to the phonological level assists

Corresponding author: Jeffrey S. Bowers (j.bowers@bris.ac.uk).

in the reorganization of the perceptual system, so that /D/ is categorized appropriately. This top-down learning would immediately generalize to other words, facilitating speech recognition in previously unfamiliar contexts, and explaining how listeners are able to adapt to foreign accents so rapidly [9].

Norris *et al.* carried out a series of experiments to determine whether listeners do alter their phoneme categories in response to lexical feedback [2]. The listeners first performed an auditory lexical-decision task, which served as a form of training. The critical items were words ending in the phonemes /f/ or /s/. For one group of listeners, /f/ phonemes were replaced by an ambiguous phoneme /?/ that was equally similar to the /f/ and /s/; for a second group, /s/ phonemes were replaced by the same ambiguous phoneme. In the lexical-decision task, listeners tended to judge the items with the ambiguous phonemes as real words. For example, *dre?* would be categorized as if it were the word *dress* rather than the non-word *dref*. The participants then performed a task in which they categorized phonemes as either /f/ or /s/. The group who had heard the ambiguous phoneme in the context of /s/-ending words were strongly biased towards /s/ categorizations of this phoneme, whereas the group who had heard the ambiguous phoneme in the context of /f/-ending words were strongly biased towards /f/ categorizations. Various control conditions ruled out pre-lexical accounts of this learning. Norris *et al.* concluded that lexical feedback had affected what participants learned about the new phoneme.

Given this evidence of top-down feedback, one might be tempted to reject the modular position outlined in Norris *et al.* [1]. However, Norris *et al.* [2] take an alternative approach. They continue to argue that top-down feedback has no role in on-line processing, but propose that top-down feedback can play an important role in learning. To illustrate how a system that includes bottom-up on-line processing and top-down learning might work, they describe a simple feedforward network that learns via the back-propagation algorithm, in which learning is achieved by passing error-correcting feedback down through the network (P. Werbos, PhD thesis, Harvard University, 1974; and [10]). Over time, this feedback alters the weights in the network, changing the function computed by each level. Crucially, though, this top-down feedback does not affect on-line activation.

Does back-propagation provide a reasonable illustration of how to reconcile top-down learning with modular accounts of speech perception?

The back-propagation network described by Norris *et al.* [2] can be thought of as an existence proof showing that it is possible to combine modular bottom-up processes with top-down learning. Nevertheless, the relevance of this illustration is weakened by several considerations. First, the top-down propagation of error signals in the back-propagation learning algorithm is biologically implausible, as is readily conceded by advocates of connectionist modelling [11,12]. Attempts to make back-propagation more consistent with biology have incorporated top-down on-line activation [13]. Second, back-propagation supports a form of learning that is inconsistent with the structure of knowledge included in Norris *et al.*'s [1] modular model of speech perception: back-propagation can only support the learning of distributed representations, whereas their model employs localist representations of phonemes and words [14].

Third, and perhaps most importantly, back-propagation is unsuited for explaining the rapid learning observed by Norris *et al.* [2]. They found effects of learning after exposure to only 20 of the critical words. However, back-propagation learning is subject to 'catastrophic interference' [15,16]: that is, new learning tends to overwrite old learning – in particular, fast learning leads to fast forgetting. The remapping of phonological space that modifies the perception of the ambiguous phoneme would necessarily have an impact on the ability of the system to respond to the previously learned prototypical /f/ or /s/ phonemes. It would clearly be inappropriate for a listener's longstanding knowledge of phonemes to be 'rewritten' as a result of a small number of unusual learning instances.

It should be noted that Norris *et al.* [2] do not endorse back-propagation *per se*, and Norris has specifically criticized this learning algorithm in other places [17]. However, this serves only to highlight the difficulty of reconciling modular systems with top-down learning. If there is an example of an algorithm that supports top-down learning consistent with Norris *et al.*'s model, we expect that it would have been used. The above considerations also raise questions regarding two key arguments made in support of modularity. First, Norris *et al.* [1] argued that on-line feedback serves no function. But it might be necessary to introduce on-line feedback for the sake of learning if a suitable modular learning algorithm cannot be found. Second, their argument based on parsimony seems to be weakened if information must be communicated in qualitatively different ways for the sake of top-down learning and bottom-up activation.

In making the above points we do not intend to suggest that a modular account could not possibly accommodate the new findings reported by Norris *et al.* [2], but simply that proponents of the modular approach will need to tackle a further challenge to accommodate top-down effects in learning. The identification of an appropriate learning algorithm will allow the modular approach to be more directly compared with existing interactive models of perception in which the use of on-line top-down feedback gives rise to fast and stable learning of localist representations, while avoiding hallucinations [16,18].

Conclusion

In summary, the new research reported by Norris *et al.* [2] provides strong evidence of an effect of top-down feedback on learning. By drawing a distinction between on-line processing and learning, the authors argue that their new evidence need not affect their previous conclusion that top-down feedback has no role in on-line speech perception. Interactive theorists, on the other hand, will no doubt conclude that the new data lend support to a theoretical framework in which top-down feedback plays a key role both in perception and learning [4,16,18]. Whatever one's views, these new data will inform future debates on the issue of interactive versus modular processing, thereby

providing an important contribution to an issue of fundamental theoretical significance in cognitive science.

Acknowledgements
The writing of this paper was supported by the Biotechnology and Biological Sciences Research Council, grant number 7/S17491 awarded to Jeffrey Bowers and Markus Damian. The authors would like to thank Sven Mattys for comments on the manuscript.

References

1 Norris, D. *et al.* (2000) Merging information in speech recognition: feedback is never necessary. *Behav. Brain Sci.* 23, 299–370
2 Norris, D. *et al.* (2003) Perceptual learning in speech. *Cogn. Psychol.* 47, 204–238
3 Fodor, J.A. (1983) *Modularity of Mind: An Essay on Faculty Psychology*, MIT Press
4 Grossberg, S. and Myers, C.W. (2000) The resonant dynamics of speech perception: interword integration and duration-dependent backward effects. *Psychol. Rev.* 107, 735–767
5 McClelland, J.L. (1987) The case for interactionism in language processing. *Attention Perform.* 12, 3–36
6 Samuel, A.G. (1981) Phonemic restoration: insights from a new methodology. *J. Exp. Psychol. Gen.* 110, 474–494
7 Ganong, W.F. (1980) Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125
8 Gaskell, M.G. and Marslen-Wilson, W.D. (1998) Mechanisms of phonological inference in speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 380–396
9 Clarke, C.M. (2002) Perceptual adjustment to foreign-accented English with short-term exposure. In *Proceedings of the 7th International Conference on Spoken Langauge Processing* (Vol. 1), Casual Productions, Ltd, pp. 253–256
10 Rumelhart, S. *et al.* (1986) Learning representations by back-propagating errors. *Nature* 323, 533–536
11 Crick, F.H.C. (1989) The recent excitement about neural networks. *Nature* 337, 129–132
12 Zipser, D. and Andersen, R.A. (1988) A backpropagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* 331, 679–684
13 O'Reilly, R.C. (1996) Biologically plausible error-driven learning using local activation differences: the generalized recirculation algorithm. *Neural Comput.* 8, 895–938
14 Bowers, J.S. (2002) Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand. *Cogn. Psychol.* 45, 413–445
15 McClosky, M. and Cohen, N.J. (1989) Catastrophic interference in connectionist networks: the sequential learning problem. In *The Psychology of Learning and Motivation* (Bower, G.H., ed.), pp. 109–165, Academic Press
16 Grossberg, S. (1987) Competitive learning: from interactive activation to adaptive resonance. *Cogn. Sci.* 11, 23–63
17 Norris, D. (1990) How to build a connectionist idiot (savant). *Cognition* 35, 277–291
18 Carpenter, G.A. and Grossberg, S. (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. *Comput. Vis. Graphics Image Process.* 37, 54–115

1364-6613/$ - see front matter
doi:10.1016/j.tics.2003.10.018

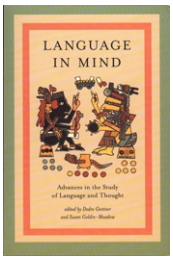
Book Review

From mind in the mouth to language in the mind

Language in Mind edited by D. Gentner and S. Goldin-Meadow, MIT Press, 2003. £22.95 (522 pages) ISBN 0 262 57163 3

Gabriella Vigliocco[1] and Luna Filipovic Kleiner[2]

[1]Department of Psychology, University College London, 26 Bedford Way, London, UK, WC1H OAP
[2]Department of Linguistics, University of Cambridge, Sidgwick Avenue, Cambridge UK, CB3 9DA



Language in Mind is a collection of papers that introduces different perspectives on the fundamental issue of the relationship between language and thought, in particular the question of whether language use can shape cognition. The book combines papers addressing the Linguistic Relativity (or Sapir–Whorf) hypothesis of whether a specific language spoken by a community affects non-linguistic cognition with papers that look at the relationship between language and thought, by discussing whether using language (regardless of which language) contributes to higher cognitive functioning (i.e. does using language make us smart?). This is not a 'position' book marshalling a single perspective; the authors do a very good job of outlining their diverse positions and the reasons for holding them. In other words, the book is not meant to provide an answer to the question of whether the language one speaks shapes cognition (which one can argue is not, in fact, a scientifically tractable question, if framed in these terms) but to provide a framework for asking questions concerning the interface between language and cognition and a number of ways in which these questions can be empirically addressed.

Despite the diversity of theoretical perspectives and empirical approaches, a fundamental commonality among the chapters in the book is a clear statement of the implications of cross-linguistic variability to any theory of cognitive functioning.

A commonly held implicit assumption in cognitive theories is what Bock [1] has labelled the 'mind in the mouth' assumption. According to this assumption, language provides us with an open window into cognition. Therefore, we can investigate cognitive functions using verbal tasks without worrying about the types of representations and processes that are required by the linguistic task. In other words, the real job is to categorize things and events, whereas assigning verbal labels to them is something that straightforwardly follows. Bock discusses the fallacy of such an assumption in terms of a lack of consideration of the contribution of language

Corresponding author: Gabriella Vigliocco (g.Vigliocco@ucl.ac.uk).