

12 Modularity, Working Memory, and Reading Disability

Ignatius G. Mattingly
*Haskins Laboratories, and
University of Connecticut*

One of the many fine things about the research of Isabelle Liberman and her colleagues is that their experimental comparisons of good and poor readers have not been restricted to reading performance. They have also investigated their subjects' speech perception, verbal short-term memory, and linguistic awareness, not only demonstrating clear differences between the two reading groups but also providing evidence that poor readers' problems may be specifically phonological. Brady's current research extends this approach, comparing good and poor readers' performance not only in perceptual and memorial tasks but also in various speech production tasks (Rapala & Brady, 1990).

A GENERALIZED PHONOLOGICAL DEFICIENCY?

In her contribution to this volume, Brady argues that poor readers perform badly in short-term recall tasks not because they fail to employ sophisticated mnemonic strategies, or because they are using a visual or a semantic rather than a phonological coding, but because the phonological processes they employ in "working memory" operate inefficiently. She then considers a more general question: Are these working-memory phonological processes to be equated with those required for linguistic tasks carried out by other cognitive functions? If so, then it is possible that the problems of poor readers are due to a generalized phonological deficiency. Perhaps, she suggests, "the difficulty observed in encoding phonological information is not restricted to memory tasks, but occurs at a more abstract level, whenever it is necessary to create and maintain a phonological representation."

Brady reviews a considerable body of evidence bearing on this question and seems to think it very likely that working-memory phonological processes may overlap with at least some of the phonological processes associated with other functions. Consistent with this possibility, she points out, poor readers do less well than good readers in speech perception, speech articulation, and naming. Thus, she says, "a link appears to be justified among the more basic language tasks associated with reading disability." However, she finds that the evidence is more equivocal with respect to the possibility of overlap between the phonological processes required for segmental awareness tasks and those required for tasks that stress other functions. Some studies find correlations; others do not. Moreover, naive adult illiterates and readers of nonalphabetic orthographies are lacking in segmental awareness, although their primary linguistic behavior is normal and their phonological processes presumably operate efficiently. In the end, Brady decides that, given the evidence available, it would be premature to come to any definite architectural conclusions.

I would like to point out a difficulty with Brady's proposal, and offer a friendly amendment. I agree with her basic insight that the phonological processes employed in different modes of linguistic behavior can hardly be independent of one another. But the concept of "working memory," apparently accepted unquestioningly by Brady, as by many other psychologists of language, is fundamentally inconsistent with this proposal.

"WORKING MEMORY" AND ITS PROBLEMS

According to Baddeley and Hitch (1974), its original proponents, working memory exists to support such high-level cognitive activities as reasoning, language comprehension, and learning. It consists of a central processing system, or executive component (whose precise role, however, has never been clearly defined), and a peripheral slave system for phonemic rehearsal, said to be (Baddeley, 1986) "specifically adapted to retaining speech-based material" (p. 75) and later called the "articulatory loop" (Baddeley, 1979, 1986). At least one other slave system in working memory, the "visuospatial scratchpad," is proposed (Baddeley, 1986).

The details of the representation in the articulatory loop have never been made clear. It is said to be "phonemic" (Baddeley & Hitch, 1974), a term that suggests, at least to linguists, a more abstract level of representation than does "articulatory loop," unless, indeed, the phonemic units are taken to be articulatory gestures (cf. Browman & Goldstein, 1986; Liberman & Mattingly, 1985). Moreover, if this representation is to be equated with "inner speech," as Baddeley (1986) suggests at one point, it might be expected to include morphological and syntactic information also (yet the inclusion of the latter, at least, would be inconsistent with the idea that working memory merely

supports higher level language-comprehension processes). On the other hand, there is some recent evidence that the representation includes subphonemic information (Xu, in press). However, for convenience, I continue to speak of the "phonological representation."

A related question is just how the rehearsal loop is supposed to work. If the phonological representation is subject to decay, how does rehearsal help? Why does the representation not just become less correct with each rehearsal? This seems possible to explain only if there are multiple, redundant levels of linguistic representation and some process that exploits this redundancy to correct the decaying representation.

To see why working memory is inconsistent with a common set of phonological processes, recall Fodor's (1983) distinction between "horizontal" and "vertical" approaches to psychology. For the horizontalist, the primary divisions of cognition are such broad functions as action, perception, understanding, and memory, and the task of psychology is to work out, for each of these functions, the general principles that must hold across differences of modality and domain. Thus, tones, colors, shapes, and speech sounds are all supposed to be perceived according to the same basic laws. For the verticalist, on the other hand, specific domains or faculties are primary. From this point of view, the psychologist's job is to find the common principles underlying action, perception, understanding, and memory for, say, music; these principles are naturally specific to music and are not assumed to have any particular kinship with those underlying action, etc., for other domains. Indeed, such functional terms will have quite different meanings for different domains and little general explanatory value.

Horizontal approaches have dominated psychological research in the twentieth century, and Baddeley's working-memory proposal is squarely in this tradition. Working memory is of course part of memory in general. One central executive component controls all forms of working-memory activity, and whatever is unavoidably domain-specific is relegated to an appropriate low-level slave system. Note that Baddeley's proposal has the effect of fragmenting linguistic processing, and in two ways: Memory for language is clearly separated from the production, perception, and understanding of language; and sentence-parsing and semantic analysis are separated from phonological processes, the former taken to be central, the latter, to be peripheral.

VERTICALITY OF BRADY'S PROPOSAL

Though Brady adopts the notion of working memory, thus proclaiming herself a horizontalist, her proposal for a generalized set of phonological processes is in fact a move toward verticality. If it is correct, the articulatory loop cannot be just the slave of working memory, as in Baddeley's account.

On the other hand, if the notion of working memory is correct, there must be some other explanation of the evidence Brady discusses.

There is, in fact, a possible escape from this dilemma that Brady might have considered. This is to attribute the apparent correlations between phonological activity in production, perception, and memory to a common store of phonological *knowledge*, part of the grammar, accessed by motoric, perceptual, and memorial processes that may be themselves distinct and diverse. If the grammar is in a poor state, then all these diverse phonological processes will seem to function less effectively. From this standpoint, the performance of poor readers might be due simply to inadequate phonological knowledge. This is essentially the "phonological lag" hypothesis, considered and rejected by Shankweiler and Crain (1986). This hypothesis seems particularly appealing as a way of accounting for poor performance on memory tasks. The phonological generalizations that are part of the knowledge of mature speakers make much low-level phonetic structure predictable. If the poor readers lack these generalizations, then their memorial representations of particular utterances will be necessarily more detailed and will use up more working-memory capacity.

But Brady herself has provided the evidence that blocks this way out. She and her colleagues found that, in order to demonstrate a difference between good and poor readers in speech perception, it was necessary to add noise to the stimuli, thereby increasing the processing load (Brady, Shankweiler, & Mann, 1983). From this result it can be inferred that good readers do have the correct phonological generalizations and can demonstrate this under conditions of minimal stress. But under more difficult conditions, their performance is poorer than that of good readers, suggesting that they are having more serious processing difficulties.

SHANKWEILER AND CRAIN'S MODULARITY

Brady is not alone in mingling horizontal and vertical concepts. Shankweiler and Crain (1986), arguing for "a unitary phonological deficit" and considering much of the same evidence, offer an account of reading performance in which the contradiction is even clearer. Following Fodor (1983) and Liberman and Mattingly (1985), they adopt the thoroughly vertical idea of a language module. To accept this view is to reject the horizontal idea that linguistic processing is accomplished by a temporary alliance of general motor, perceptual, and memorial processes, and to view it rather as the business of a special system responsible for all primary linguistic processing and nothing else. Shankweiler and Crain's language module, reasonably enough, includes as one of its parts a phonological component. This is the obvious locus for the common phonological processes proposed by Brady, and its disorders are responsible, according to Shankweiler and Crain, for the "unitary phonologi-

cal deficit." But also included in their language module is horizontal working memory, its executive component now controlling the transfer of information from one linguistic subcomponent to another. This arrangement, a drastic demotion of working memory, is obviously inconsistent with Baddeley's proposal.

CONFLICTS BETWEEN WORKING MEMORY AND MODULARITY

In fairness, it should be mentioned that Shankweiler and Crain are aware of impediments to this marriage. However, they seem to think it is modularity that is causing the problem:

On Fodor's [1983] view, the language module as a whole is an "input system": Its operations are fast; they are mandatory; they are largely sealed off from conscious inspection; they are also insulated from cognitive inferencing mechanisms external to all language. Working memory, as we understand it, does not conform to all of these criteria. Some of its operations consume appreciable time, and some are open to conscious inspection, as in the rehearsal and reanalysis of linguistic material. Nevertheless, it seems to us that working memory belongs in the language module by reason of its intimate association with the parsers which assign phonological, syntactic, and semantic structure to linguistic input. (p. 151)

It may seem that I am merely making a fuss about terminology, and that the contradictions I complain of in Brady's and in Shankweiler and Crain's proposals could be resolved simply by calling their versions of working memory by other names, or by explicitly distinguishing their brand of working memory from Baddeley's, as Shankweiler and Crain do in a later paper.¹ But there is a substantive issue here. The reason these investigators (and many others) invoke working memory is that they take it as self-evident that the linguistic representation of which one is aware as "inner speech," and which is tapped in "short-term memory" experiments, is the product of an intrinsically memorial process. It is because this process is memorial that the representation endures after it has been used to support higher-level processes, as a kind of by-product.

If one makes this assumption, it inevitably shapes one's view of the

¹... we conceive of verbal working memory a domain-specific system that subserves the language apparatus... This distinguishes our conception of working memory from proposals by Baddeley (1986) and Carpenter and Just (1988). These researchers see working memory as a general purpose device, which plays a central role not only in language, but also in reasoning, problem solving and in other forms of complex thinking" (Crain, Shankweiler, Macaruso, & Bar-Shalom, in press).

architecture of linguistic processing: A self-refreshing memory store has to fit into the block diagram somehow. Thus, some suppose that the store provides a way of retaining information needed for parsing at a later point in the sentence. Shankweiler and Crain (1986) see it as the first of several stages in the analysis of an utterance, permitting phonetic information to be retained until it can be transformed into lexical, syntactic, and semantic information. Perfetti and Lesgold (1977) see it as a tool for decoding in the case of young readers. What all these suggestions have in common is the implication that the phonological representation is a survival from some intermediate memorial stage of linguistic processing.

This is not an unreasonable notion as long as language processing is assumed to be carried out by general-purpose cognitive functions. On this assumption, what is to be regarded an intermediate and what as a final representation may depend on the cognizer's intentions. But the notion of a by-product is inconsistent with modularity. A module is, by definition, a special-purpose device that provides *specific* output representations to central cognition (Fodor (1983): "Input analyzers have 'shallow' outputs" [p. 86]). In the course of its computations, the module may form intermediate representations, but these representations, not being outputs, are not available to cognition. To suggest that cognition, which could get no information at all without the support of the modules, could have some means of looking into a module and obtaining access to intermediate representations, would render incoherent the whole notion of modularity. Thus, one of the earmarks of a module, according to Fodor, is "limited central access . . . only such representations as constitute the final consequences of input processing are fully and freely available to . . . cognitive processes" (pp. 55-56).

It follows that any representation that *is* cognitively available is a true output, and we cannot plausibly treat it as a by-product of an intermediate stage. It may well be that the first thing the language module does when it receives a speech signal is to form some kind of phonological representation that is temporarily retained, as Shankweiler and Crain propose. But we are not entitled to equate this hypothetical intermediate representation with the *final output* phonological representation. Yet it is just this latter representation that has been the object of short-term memory research.

A MODULAR ACCOUNT OF READING DISABILITY

Let us see if it is possible to dispense with "working memory" altogether, and to offer a strictly modular account of reading and of the problems of poor readers. Of course, we are also obliged to explain how other mental processes usually said to be supported by working memory—reasoning, learning, and remembering telephone numbers—are to be accomplished without it.

The fact that reading is possible is itself a problem for those who take a modular view of language. A module is supposed to have its own specific domain, and for the language module, in its perceptual mode, this domain consists of acoustic speech signals. How is it that the module also is able to cope with arbitrary optical signals?

I have suggested elsewhere (Mattingly, 1991) how this problem might be dealt with. It is proposed there that the language module has several modes of operation, in all of which the processes are essentially similar, though the inputs and outputs vary. Thus, in its perceptual mode, as in Fodor's (1983) original proposal, the input is the speech signal; the outputs are a semantic representation (perhaps Logical Form) and a linguistic representation. In its production mode (Lieberman & Mattingly, 1985; Mattingly & Lieberman, 1985), the input is a semantic representation, the outputs, a sequence of motor commands and a linguistic representation. In its rehearsal mode, which is the mode specially relevant to reading, a decaying linguistic representation—itsself the output, originally, of perception or production—becomes the input, and the outputs are a semantic representation and a refreshed linguistic representation. It is, of course, the redundancy of the linguistic representation that makes this refreshment possible. (See Mattingly, 1990, for speculation on the biological basis for the rehearsal mode and for the linguistic representation.)

By virtue of its rehearsal capability, the language module can support high-level cognitive processing in much the way that working memory is said to do. It can also facilitate the retention of telephone numbers and performance on short-term recall tasks. But there are some important differences. The first is that the parsing of sentences takes place in the language module, and the representation the module provides to cognition is linguistic, including syntactic, lexical, morphological, and phonetic, as well as phonological information. The second is that the module is not a memory system: Though the language user is free to exploit the module for mnemonic purposes, its mnemonic utility is incidental to its linguistic functions. On this account, some of the points that bother Shankweiler and Crain about modularity can be disposed of. If rehearsal is a mode of using the language module, rather than a process purely internal to the module, then its output is naturally "open to conscious inspection." A similar point can be made for reanalysis. When one finds that one has been led up the garden path by a bizarre sentence, one may reanalyze, adjusting one's picture of the world and then rehearsing the sentence, but this is not the same as having access to the language module's internal representations.

In reading, the module's rehearsal mode is exploited in a different way. From the orthographic form of an utterance, the reader forms what I have termed a "synthetic" linguistic representation. This is what is usually called "decoding." Decoding is a cognitive task and requires both a knowledge of

the orthographic system and an awareness of those features of linguistic representations to which the orthography appeals: phonemic segments, in the case of alphabetic writing; syllables, in the case of morphosyllabic writing. The synthetic linguistic representation is very incomplete; no practical orthography gives anything like a complete representation of the utterance it transcribes. But such a complete representation would be unnecessary. Because of the redundancy of linguistic representations, the module can accept the incomplete, synthetic representation as an input, and, by rehearsing, produce a semantic representation and a complete linguistic representation as output.

On this view, there are two potential sources of reading difficulty that can be characterized as "phonological." One of them, as we have seen, is the phonological component of the module, and some weakness here must underlie the problems of those poor readers who also have difficulty in primary linguistic processing: speech production, speech perception, and short-term recall. The other source of reading difficulty is not modular but cognitive; the need for the reader to be aware of the units of phonological structure in linguistic representations that the orthography appeals to, in order to decode. Lack of such cognitive phonological awareness is quite a different matter from a weakness internal to the module. The two cannot be entirely unrelated, obviously, for in order for awareness to develop, the language module must presumably have been producing the right natural linguistic representations. But it is not surprising that a strong correlation between performance on cognitive segmental awareness tasks and performance on tasks that stress precognitive phonological capacities should prove difficult to demonstrate.

ACKNOWLEDGMENTS

Support from NIH grant HD-01994 to Haskins Laboratories is gratefully acknowledged.

REFERENCES

- Baddeley, A. D. (1979). Working memory and reading. In P. A. Kollers, M. E. Wrolstad, & H. Bouma (Eds.), *The proceedings of the conference on the processing of visible language* (I, 355-370). New York: Plenum Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation*, V, VIII (pp. 47-90). New York: Academic Press.
- Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology*, 35, 345-367.

- Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.
- Carpenter, P. A., & Just, M. A. (1988). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crain, S., Shankweiler, D., Macaruso, P., & Bar-Shalom, E. (in press). Working memory and comprehension of spoken sentences: Investigations of children with reading disorder. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge University Press.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-37.
- Mattingly, I. G. (1991). Reading and the biological function of linguistic representations. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 339-346). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mattingly, I. G., & A. M. Liberman (1985). Verticality unparallelled. *Behavioral and Brain Sciences*, 8, 24-26.
- Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141-183). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rapala, M., & Brady, S. (1990). Reading ability and short-term memory: The role of phonetic processing. *Reading and Writing*, 2, 1-25.
- Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder A modular approach. *Cognition*, 24, 139-168.
- Xu, Yi (in press). The depth of phonological encoding in short-term memory. *Memory & Cognition*.