



HAL
open science

MROM-p: An Interactive Activation, Multiple Readout Model of Orthographic and Phonological Processes in Visual Word Recognition

Arthur M. Jacobs, Arnaud Rey, Johannes C Ziegler, Jonathan Grainger

► **To cite this version:**

Arthur M. Jacobs, Arnaud Rey, Johannes C Ziegler, Jonathan Grainger. MROM-p: An Interactive Activation, Multiple Readout Model of Orthographic and Phonological Processes in Visual Word Recognition. Jonathan Grainger; Arthur Jacobs. *Localist Connectionist Approaches To Human Cognition*, Lawrence Erlbaum Associates, pp.147-188, 1998, Scientific Psychology Series, 9780203774533. hal-04450865

HAL Id: hal-04450865

<https://hal.science/hal-04450865v1>

Submitted on 13 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MROM-p: An Interactive Activation, Multiple Readout Model of Orthographic and Phonological Processes in Visual Word Recognition

Arthur M. Jacobs
Philipps University of Marburg

Arnaud Rey
Laboratoire de Neurosciences Cognitives - CNRS

Johannes C. Ziegler
Jonathan Grainger
CREPCO-CNRS & Université de Provence

The world is worded before it is sentenced.

—Variation on a theme by Humboldt

READING, WORD RECOGNITION, AND THE LEXICAL DECISION TASK

The world of words is just as wondrous as the world of syntax, or even more so. For not only are people as infinitely creative with words as they are with phrases and sentences, but memorizing individual words demands its own special virtuosity.

—S. Pinker, *The Language Instinct*

The subject of this chapter is reading, one of the finest achievements of human civilization and one of the most complex activities of the human mind. Explaining the whys and hows of reading skill represents an outstanding intellectual challenge for cognitive scientists. Word recognition is the fundamental process underlying reading skill; it provides a favorable focus for experimental reading research. At the level of word representations, all lower and higher level processes involved in reading seem to meet. Word repre-

sentations are the central building blocks of language learning and processing (Miller, 1993). Here, sensory, orthographic, phonological, morphological, semantic, and syntactic operations converge and diverge. The problem is to find how these central representations are organized and how they interact with both lower and higher level processes that depend on task or language contexts. Solving this problem necessarily involves a multistep approach. Before tackling, for example, word or sentence production problems that reflect being "infinitely creative with words," we try to solve the problems of the perceptual and mnemonic organization of lexical memory, that are the basis of the "special virtuosity." Before tackling problems of morphosemantic and -syntactic processing of words, we try to answer simpler questions about orthographic and phonological processing.¹

Psychologists have studied word recognition and reading by using experimental techniques that still require methodological and theoretical unification (Jacobs & Grainger, 1994; see also discussion section). The most widely used modern experimental method for investigating visual and auditory word recognition is the lexical decision task (LDT). Like any other experimental technique, the LDT provides only indirect and incomplete information about the processes underlying word recognition and therefore requires cognitive modeling as a complement to experimental analyses. This chapter focuses on such a model of word recognition performance as assessed by the LDT.

ORTHOGRAPHIC AND PHONOLOGICAL PROCESSING: A MODEL-GUIDED, MULTILINGUISTIC, MULTITASK PERSPECTIVE MOTIVATED BY THREE SKEPTICISMS

From an evolutionary perspective, writing and reading evolved because humans needed a convenient means of coding and decoding oral language for purposes of storage, transmission, and tradition. As a means of externalization of thoughts, writing and reading presumably had already withstood a test on the scale of survival values. In alphabetic writing systems, the individual elements of the alphabet correspond to the elementary sounds of the spoken language. The degree of this correspondence (its consistency) is variable and

¹Orthographic processing refers to the use of orthographic information (i.e., knowledge of the spellings of words). In alphabetic languages such as English, French, or German, we assume that such knowledge is letter based. Knowledge of how to spell a word is thought to be stored as a set of abstract representations that code both the identity and position of a word's component letters. Phonological processing refers to the use of phonological information (i.e., knowledge of the sounds of language) in processing written and oral language. The question of which functional units code this knowledge is more complex than for orthographic processing, as discussed in the text.

the object of much linguistic and psycholinguistic research. In some places many-to-many (homomorphic) mappings evolved (e.g., English and French), whether "naturally," following invasions, or via spelling reforms. In other places something closer to one-to-one (isomorphic) mappings between script and sound evolved (e.g., Serbo-Croatian or Spanish). This considerable variation across languages in the degree of correlation or the consistency of the spelling-to-sound and sound-to-spelling mappings ("deep" vs. "shallow" orthographies) has provided a rich playground for cross-linguistic experimental studies of reading skills (Frost, Katz, & Bentin, 1987). Such studies have their costs, but for understanding the reading process, they are more interesting than are monolingual studies. Moreover, the considerable cross-linguistic variations in consistency also provide a challenge for researchers who aim at building computational multilingual, multitask models of reading (Carreiras, Perea, & Grainger, 1997; Jacobs, 1995; Ziegler, 1996; see also chap. 6). Our research program, of which the present modeling efforts are an integral part, is multilingual, because we are skeptical about the view that the reading process can be understood by studying a single language. Instead, as many examples have shown, cross-linguistic research can avoid the dangers of premature or false conclusions drawn from the results of monolingual work (Hagège, 1986; Lass, 1995; MacWhinney, Bates, & Kliegl, 1984; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Van Orden & Goldinger, 1994). Currently, our research program includes three languages (French, German, and English), and indirectly encompasses three others (Dutch, Spanish, and Italian) via scientific collaborations.

Orthographic-phonological processing is well suited to illustrate the benefits of both multilingual research (Frost et al., 1987) and model-guided multitask research (Grainger & Jacobs, 1996). Well-planned combinations of multilingual and multitask research could quickly advance our understanding of the constraints involved in reading (Jacobs, 1995; Ziegler, 1996). Under the optimality assumption, would we expect users of English orthography to develop the same reading strategies (and underlying representations and processes) as do users of French?² The authors of some recent descriptive statistical studies on spelling-to-sound and sound-to-spelling consistency for English and French (Stone, Vanhoy, & Van Orden, 1997; Ziegler, Jacobs, & Stone, 1996; Ziegler, Stone, & Jacobs, 1997) estimated that about 72% of all English monosyllabic words are feedback inconsistent (i.e., their phonologic bodies can be spelled in multiple ways) whereas about 31% are feedforward

²Regardless of the reasons for the variability in the spelling-to-sound and sound-to-spelling mappings, evolutionary perspectives of cognitive psychology (e.g., Shepard, 1994) must start with the premise that users of variable linguistic environments optimally adapt(ed) to their corresponding orthography-to-phonology and phonology-to-orthography mappings. Such an optimality assumption facilitates formal analyses (Massaro & Friedman, 1990), which are applied here to the domain of orthographic and phonological processing in different tasks and languages.

inconsistent (i.e., their spelling body has more than one pronunciation). In comparison, about 79% of all monosyllabic French words are feedback inconsistent, whereas only 12% are feedforward inconsistent. In view of this data, could we expect users of English and French to have problems in tasks that require a solid spelling knowledge (e.g., proofreading or LDT), but users of French to have fewer problems in naming tasks than do users of English? What about users of "shallow" orthographies, such as German? Do they fare well regardless of task demands? If so, they could serve as a control group for comparison with the performance of the two other populations and thus could make the estimation of language-dependent effect sizes possible (Jacobs, 1995; Ziegler, 1996).

We are also skeptical about the view that the reading process can be understood by using a single experimental paradigm. Different reading tasks (e.g., LDT, naming task, perceptual identification task) capture both identical and different aspects of the reading process, but there is no model-free way to determine which of those aspects is relevant to an understanding of the reading process and which is purely task specific. Pursuing our stratagem of modeling functional overlap illustrated in Figure 5.1 (Grainger & Jacobs, 1996; Jacobs, 1994; Jacobs & Grainger, 1994), we attempted to gain an understanding of phonological processes that might be common to silent reading and reading aloud, as assessed by the LDT, perceptual identification, and naming task.

Finally, we are skeptical about the view that reading can be fully understood by viewing it as a one-way process, which exclusively proceeds from

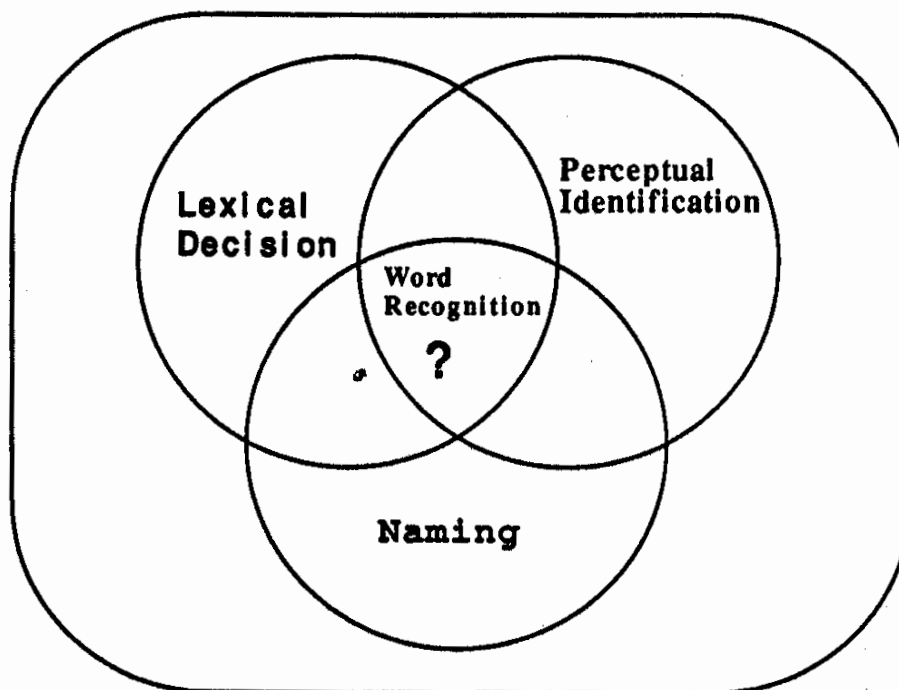


FIG. 5.1. Venn diagram illustrating the concept of functional overlap (for details, see Grainger & Jacobs, 1996; Jacobs, 1994; Jacobs & Grainger, 1994).

print to sound. In practice, this classical view has led to a separation of models, methods, factors (variables), and effects stressing either orthographic or phonological aspects. An example is the separation of experimental psychologists in "orthographic" and "phonological" camps. For example, the orthographic camp prefers the LDT, orthographic variables (e.g., measures of orthographic neighborhood), and models that focus on the explanation of orthographic effects. In contrast, the phonological camp favors the naming task, phonological variables (e.g., consistency measures), and models that focus on the explanation of phonological effects (see recent special section of the *Journal of Experimental Psychology: Human Perception and Performance*, 1994, on modeling visual word recognition). If the reading system is interpreted as an interactive, dynamic system (Grossberg & Stone, 1986; Korte, 1923; Rumelhart, 1977; Stone et al., 1997; Van Orden & Goldinger, 1994), models, methods, and measures must be developed to adequately reflect and help analyze the functioning of such a complex nonlinear system.³ According to our view, reading is a two-way system: Phonological information and phonological skills influence orthographic processing, and orthographic information and orthographic skills also influence phonological processing (Dijkstra, Roelofs, & Fieuws, 1995; Jakimik, Cole, & Rudnicky, 1985; Wagner & Torgesen, 1987; Ziegler & Ferrand, 1997; Ziegler, Montant, & Jacobs, 1997; Ziegler, Van Orden, & Jacobs, 1997). From this perspective, single-task approaches to reading, measures of orthographic neighborhood (Coltheart, 1978), one-way metrics of spelling-to-sound consistency (Rosson, 1985; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995; Venezky & Massaro, 1987), or monotask models of performance must necessarily remain incomplete approximations. The study of bidirectional consistency effects (Stone et al., 1997; Ziegler et al., 1996; Ziegler, Montant, & Jacobs, 1997), which we discuss here, and our present attempt to model them represent a step beyond this one-way approach to reading.

A PRINCIPLED APPROACH TO COGNITIVE MODELING

Our approach to understanding the reading process by help of formal cognitive models follows a set of pragmatic stratagems and principles that are outlined in several recent works (Grainger & Jacobs, 1996; Jacobs, 1994; Jacobs & Grainger, 1994) and further discussed by Grainger and Jacobs (chap. 1). The most relevant stratagem for the present chapter is *nested* modeling: the idea that a new model should either include the old one as

³A view now shared by extant former representants of traditional modular, noncomputational, feedforward models of the reading process (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart & Rastle, 1994; for a different view, see Massaro & Cohen, 1994).

a special case by providing formal demonstrations of the inclusion or dismissal after falsification of the core assumptions of the old model. The development of our *Multiple ReadOut Model* (MROM) gives a detailed example of nested modeling in the domain of orthographic processing in lexical decision and perceptual identification tasks (Grainger & Jacobs, 1996). Here we have further pursued our efforts at nested modeling of visual word recognition by including elementary phonological processes in the MROM that thus becomes the MROM-p.

SUBJECT AREA

In this section, we use two empirical phenomena that are considered as evidence for bidirectional orthographic-phonological influences on visual word recognition, the *pseudohomophone effect* and the *bidirectional consistency effect*, as empirical touchstones to test the MROM-p.

The Pseudohomophone Effect

The first phenomenon is the classical pseudohomophone effect. As regards the LDT, it refers to the observation that nonwords sounding like words when read aloud (e.g., BRANE) are more difficult to reject than are nonpseudohomophonic control stimuli (e.g., FRANE; Rubenstein, Lewis, & Rubenstein, 1971). Since the precursor of our present model, the MROM (briefly discussed in the next section), does not include any phonological processes, it should not be able to simulate the pseudohomophone effect in the LDT if the effect is genuinely phonological. Thus, the first critical test for the MROM-p was to evaluate its ability to capture the pseudohomophone effect. As a testing ground, we chose the classical set of data by Coltheart, Davelaar, Jonasson, and Besner (1977) and a set of data from a later replication study by Seidenberg, Petersen, MacDonald, and Plaut (1996).

The Feedforward and Feedback Consistency Effects

The second phenomenon, only recently discovered, provides perhaps even stronger experimental evidence for an interaction between orthographic and phonological processes in visual word recognition. It combines two effects: the traditional feedforward consistency effect and the newly discovered feedback consistency effect (Stone et al., 1997; Ziegler & Ferrand, 1997; Ziegler, Montant, & Jacobs, 1997).

Effects of spelling-to-sound (feedforward) consistency have been studied extensively in the naming task, a task requiring overt pronunciation. The standard result is that naming latencies are longer, pronunciation errors more

frequent, or both, for *inconsistent* words that have multiple spelling-to-sound mappings than for *consistent* words whose spelling bodies are always pronounced the same. Thus, a feedforward inconsistent word whose body has several possible pronunciations like -OUGH in COUGH, DOUGH, THROUGH, BOUGH, TOUGH is harder to pronounce than a consistent word like DUCK that has a unique spelling body (-UCK). The known conditions favorable to obtaining this effect (see Ziegler, Montant, & Jacobs, 1997) include inconsistent words of low frequency (Andrews, 1982; but see Jared, 1997) and words with a consistency ratio smaller than .5 (i.e., the ratio given by the summed frequency of *friends*—words with the same spelling pattern and the same pronunciation—and the summed frequency of *enemies*—words with the same spelling pattern but a different pronunciation; Jared, 1997; Jared, McRae, & Seidenberg, 1990; Treiman et al., 1995).

In contrast, in the LDT, feedforward inconsistency effects are much less clear. To the extent that the LDT does not require an overt pronunciation, it is also less likely to be sensitive to feedforward consistency (Jared et al., 1990). Two studies (Brown, 1987; Jared et al., 1990) that used more carefully controlled stimuli than did older studies failed to find an effect. In contrast, Stone et al. (1997) provided one of the first experimental demonstrations of a feedforward consistency effect in the LDT using English-speaking participants (see Pugh, Rexer, & Katz, 1994, for an earlier demonstration). Stone et al. (1997) found that mean LDT latency to feedforward inconsistent words was 48 milliseconds longer than for feedforward consistent words when all words were feedback consistent. For words that were feedback inconsistent (i.e., whose phonological body maps onto more than one spelling, such as /-ip/ in DEEP and HEAP), the effect decreased to 8 milliseconds. This finding suggests that previous studies might have failed to detect the effect because they did not control for feedback consistency. Ziegler et al. (in press-a) replicated this effect in French. They obtained effects similar to Stone et al.: A 53-millisecond feedforward consistency effect when words were feedback consistent decreased to 12 milliseconds when feedback inconsistent words were used. To the extent that this effect can be successfully replicated and because the LDT requires no overt pronunciation, the feedforward consistency effect in the LDT provides stronger evidence for bidirectional influences of orthographic-phonological processes in visual word recognition than do the results from the naming task reported earlier.

Effects of sound-to-spelling (feedback) consistency are a recent discovery in psycholinguistics. Feedback consistency effects have been reported in the visual and auditory LDT and the naming task (Stone et al., 1997; Ziegler & Ferrand, 1997; Ziegler, Montant, & Jacobs, 1997). For the present chapter, we concentrate on the effect reported by Stone and collaborators. In two lexical decision experiments, Stone et al. found a reliable feedback consistency effect. In their Experiment 2, they used a factorial design that included four types of

words: bidirectionally consistent words such as DUCK, in which the spelling body (-UCK) could be pronounced only one way and the pronunciation body (/ -uk/) could be spelled only one way; feedforward inconsistent words such as MOTH, in which the spelling body could be pronounced more than one way (e.g., BOTH), but the pronunciation body (/ -oth/) could be spelled only one way; feedback inconsistent words such as HURL, in which the spelling body could be pronounced only one way, but the pronunciation body could be spelled more than one way (e.g., GIRL); and bidirectionally inconsistent words such as WORM, in which the spelling body could be pronounced more than one way (e.g., DORM) and the pronunciation body could be spelled more than one way (e.g., FIRM). Stone et al. found that lexical decision performance was equally affected (longer reaction times [RTs] and more errors) for feedforward inconsistent words, feedback inconsistent words, and bidirectionally inconsistent words. Only words that were both feedforward and feedback consistent produced better performance. Bidirectionally inconsistent words did not affect performance beyond what was obtained for words that were inconsistent in only one way.

Ziegler, Montant, and Jacobs (1997) replicated Stone et al.'s results in English in more carefully controlled conditions in French. They excluded the possibility that the feedback consistency effect obtained in English resulted from a failure to match feedback consistent and inconsistent items on orthographic neighborhood variables. This replication is of particular interest. Because statistical analyses showed that the structure of French and English with respect to feedback inconsistency is highly similar for these two languages (Ziegler, Jacobs, & Stone, 1996, 1997), similar feedback consistency effects were predicted for English and French. Their results, like those of Stone et al., suggested that visual word perception is affected by both feedforward and feedback consistency.

An important aspect of this effect is that feedback inconsistency might explain small, unreliable consistency effects in previous studies. Ziegler et al. (1996) analyzed all French words that would traditionally have been classified as consistent on the basis of spelling to phonology correspondences (87.6% of all monosyllabic words). In traditional experiments on consistency effects, these consistent items have served as control items against which the processing cost of inconsistent items has been tested. Ziegler, Stone, and Jacobs (1997) calculated that 77.4% of these presumably consistent items were, however, feedback inconsistent. Thus, small, unreliable consistency effects in previous studies might have resulted from the possibility that the major part of the presumably consistent control items was feedback inconsistent. Another interesting aspect of feedback inconsistency is that it should be an important variable for cross-linguistic research on spelling. If multiple possibilities of mapping phonology onto spelling affect spelling performance, it should be harder in a feedback inconsistent language (e.g., French) than in a relatively feedback consistent language (e.g., Spanish).

In view of these arguments, it seems clear that psycholinguistic experiments should be controlled for feedback consistency, and further research is needed to specify the influence of this effect. The present attempt to give a formal account of this effect, if successful, provides us not only with a tool for making predictions, but also with a formal means for stimulus selection and control. For example, simulations by MROM-p could be used in the planning phase of an experiment—together with statistical analyses—to ensure that the stimuli are well matched on the feedback consistency variable.

To our knowledge, no computational model has yet explained either the pseudohomophone effect or the feedforward and feedback consistency effects in the LDT. There are preformal (verbal or boxological) accounts of both effects in the literature, but we do not consider them here (for arguments about the strengths and weaknesses of, and the complementarity between, different model formats, see Jacobs & Grainger, 1994). Any model that does not assume automatic activation of—and feedback from—phonological representations in the visual LDT would not have predicted feedforward and feedback consistency effects to start with. In contrast, models of the resonance–interactive activation (IA) family (for a classification of models, see Jacobs & Grainger, 1994) suggest such an effect naturally, without going through the trouble of adding auxiliary assumptions (Stone et al., 1997; Ziegler, Montant, & Jacobs, 1997). The present simulation studies will tell us whether this intuition matches the computational evidence.

MODEL PRESENTATION

Model History

The scientific adventure of IA models of cognitive processing has a rich history. The conceptual ingredients that characterize this family of models can be traced back to many authors in fields such as biological cybernetics, artificial intelligence, and psychology (e.g., Arbib & Kaplan, 1979; Erman & Lesser, 1975; Grossberg, 1976, 1980; Levin, 1976; Marslen-Wilson & Welsh, 1978; Morton, 1969; Rumelhart, 1977). As far as word recognition is concerned, the adventure started for us with the publication of the two papers by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982), in which all the different conceptual ingredients were synthesized in an original and formal way that allowed direct applications to psycholinguistic studies.

Why did we choose this model format and type? Before the interactive activation model (IAM), basically two model formats were used in the word recognition literature: verbal (V-type) models (any model that is expressed verbally or graphically without making use of closed form or algorithmic formulations) and mathematical (M-type) models (models that use closed form expressions to represent the modeled section of reality). The IAM

introduced algorithmic (A-type) models (models that are implemented in form of a simulation program, including production systems and neural nets of the localist or distributed families) to the field.

Apart from well-known innovative aspects that distinguished it from its precursors (Jacobs & Grainger, 1992), the IAM offered three possibilities that neither V-type nor M-type models could provide as a whole. First, it possessed dynamics and thus offered the two possibilities of time-dependent predictions and interval-scaled modeling of RT as a dependent variable.⁴ McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) had exploited only the first of these possibilities. We were interested in the second as RT is the major dependent variable in psycholinguistic research (mainly because of the popularity of the LDT and naming task). Second, the IAM possessed a (toy) lexicon that made item-specific (fine-grained) predictions possible. This fact seemed a logical necessity in a field that has to evaluate its empirical effects with respect to subject- and item-specific data. In addition, for those who believe in the virtues of strong scientific inference (see chap. 1), fine-grained analyses are a necessity: Falsificationism and strong inference are not the only research strategies, but they seem to be the best ones whenever theorists deal with specific assumptions that can be tested at the level of fine-grained analyses (Jacobs & Grainger, 1994; Massaro & Cowan, 1993). Third, contrary to other models, the IAM, rather than being definitive, possessed rich structural potential and appeared to include the promise of interesting further developments. For a reasonable application of the strategies of nested and canonical modeling, structural potential is a necessary (but not a sufficient) condition (Grainger & Jacobs, 1996). The IAM, as the prototype of a canonical resonance model (Stone & Van Orden, 1994), allows the testing of system and design principles, to which we can attribute explanatory credit and blame independently of other aspects of the model (chap. 1).

In sum, by its original combination of formal preciseness, structural-computational richness, and computational transparency (a feature that distinguishes it from most parallel distributed processing [PDP] models), the IAM intuitively seemed the right model at the right time to allow falsifiable quantitative predictions and the discovery of new phenomena via simulations. An example of the latter, the so-called neighborhood frequency effect, is discussed next.

Predicting a New Phenomenon: The Discovery of the Neighborhood Frequency Effect

Only theories tell us what can be observed.

—Variation on a theme by Einstein

⁴For a critique of interval-scaled RT models, see Van Orden and Goldinger (1994). For a reply, see Jacobs and Grainger (1994).

The ability to predict a new phenomenon (and the conditions under which it must or must not appear) is one of the higher criteria for model evaluation (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Jacobs & Grainger, 1994). One feature of connectionist models in general, and the IAM in particular, is that they are rich enough to allow emergence of effects that have not yet been observed.

Playing with a variant of the original IAM, one of the present authors discovered that the neighborhood frequency effect in visual word recognition is possible in the model system. Looking at the activation function for the word *blur*, he observed an attenuation of the rise of this function during the early phase (a crossover between the functions for *blur* and *blue*), because *blur*, a low-frequency word, shares all but one letter with *blue*, a high-frequency word.

The activation functions for *blur*-type words reach a criterion level of activation (arbitrarily defined for response generation in the model) more slowly than do low-frequency words that have no high-frequency orthographic neighbors (e.g., *idle*; see Figure 5.2). Further simulations with the IAM showed that a selection of low-frequency words with many high-frequency neighbors (e.g., *beal*) did not differ from low-frequency words with a single high-frequency neighbor in terms of the number of cycles required to reach criterion activation levels. This simulation result was important with respect to our application of strong scientific inference in model testing. In contrast to the IAM, serial search-verification models of visual word recog-

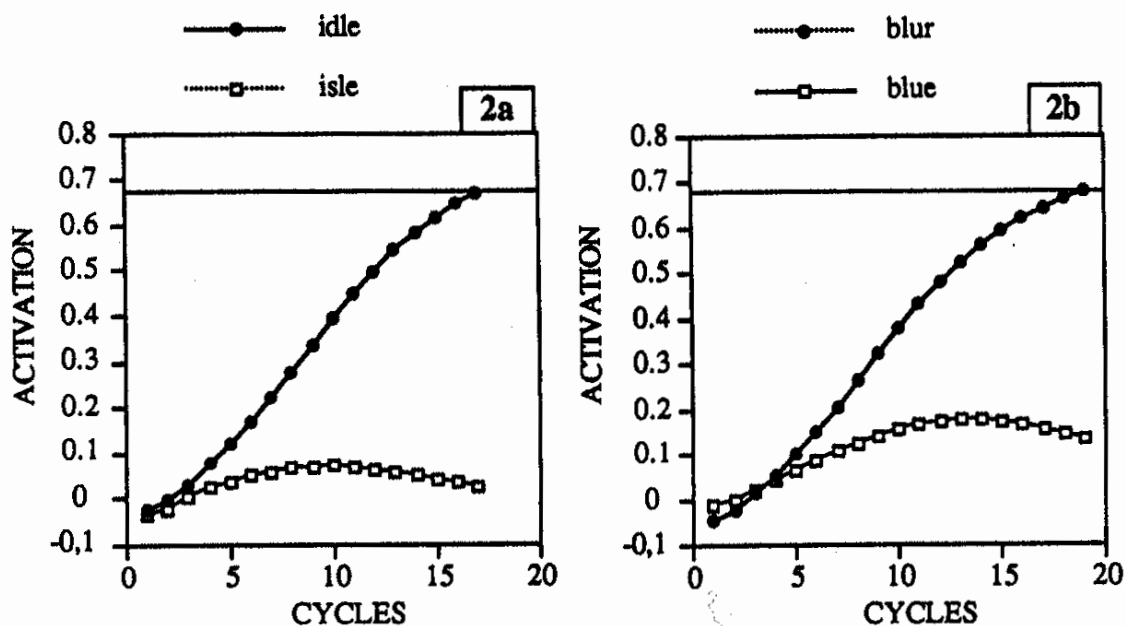


FIG. 5.2. Panels 2a and 2b show two simulations obtained with the target words IDLE and BLUR, respectively. Both these words have orthographic neighbors (IDLY, ISLE, and BLUE, respectively) and the same frequency (15 per million), but only BLUR has a higher frequency neighbor. Consequently, IDLE reaches the response threshold two cycles earlier than does BLUR (i.e., 17 vs. 19 cycles).

dition (Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982) predicted a further decrement in performance to such stimuli. The pattern predicted by the IAM was observed by Grainger, O'Regan, Jacobs, and Segui (1989, 1992) and Grainger (1990). Although subsequent research has complicated the neighborhood frequency story (e.g., Sears, Hino, & Lupker, 1995), the important point is that the IAM simulations, using the same stimuli as in the human experiments, accurately predicted the observed pattern for that particular stimulus set.

Thus the IAM, which, like all connectionist models, has structural and processing features that were built-in specifically to create known empirical phenomena (e.g., the resting level parameter that creates the frequency effect; cf. Dell, 1988), predicted an unknown effect that has now been observed under a variety of conditions (Grainger & Jacobs, 1996). This finding provides an encouraging example for solving the recurring epistemological issue of a theory-centered approach as seemingly opposed to a result-centered approach (Greenwald & Pratkanis, 1988) by showing that a model can specify the conditions under which previously unobtainable results occur. This demonstration is clearly theory centered for one of the two complementary result-centered approaches (i.e., the design approach), advocated by Greenwald and Pratkanis in their attack on theory-centered approaches to psychology.

Model Structure

The MROM

It is useful to give a short description of the MROM here (see Grainger & Jacobs, 1996, for more details). The MROM is an extension of the IAM incorporating the design principle of multiple readout, which states that a response in a given experimental task is generated (read out) when at least one of the codes appropriate for responding in that task reaches a critical activation level. This principle is particularly relevant to our explanation of performance in the LDT. With respect to this task, we hypothesized that unique word identification is not the only process that can lead to a correct Yes decision in the LDT and that an extralexical process controls the production of No responses. In the functional context of LDT, word-nonword discrimination requires that participants use a reliable source of information that allows them to make rapid and accurate judgments about the "word-likeness" of stimuli (e.g., their familiarity; Balota & Chumbley, 1984). In the MROM, we postulated three processes underlying a speeded binary lexical decision response. Two of the processes use intralexical information to generate a Yes response, and the third uses extralexical information to generate a No response. The two intralexical sources of information are the overall (global) activity in the orthographic lexicon, operationalized in the simulation model as the sum of the activation levels of all word units (hereafter referred

to as σ) and the (local) activity of functional units in the lexicon, operationalized as the activation level of individual word units (μ). The extralexical source of information is time (t) from stimulus onset. In the MROM, a criterion value set on each of the three information dimensions determines the type (yes/no) and speed of a response. The criterion on the (local) μ dimension is referred to as M , the criterion on the (global) σ dimension as Σ , and the temporal deadline as T . Figure 5.3 illustrates how these three response criteria combine to determine the type and speed of a response in the LDT.

If either the local M or the global Σ response criterion is reached before the T criterion, the response is positive; otherwise, the response is negative. Errors to word stimuli (false negatives) therefore arise when the T criterion is set too low, both the M and Σ response criteria are set too high, or both. Errors to nonword stimuli (false positives) arise in exactly the opposite circumstances (high T criterion, low M criterion or low Σ criterion, or both). In the example in Figure 5.3, both the M and the Σ response criteria are reached before the T criterion, a result giving rise to a positive lexical decision response. The speed of this response is determined by the earliest moment in time that either the M criterion is reached (i.e., a specific word has been identified) or the Σ criterion is reached (i.e., a fast guess has occurred). RT for a negative response is given simply by the value of the T criterion.

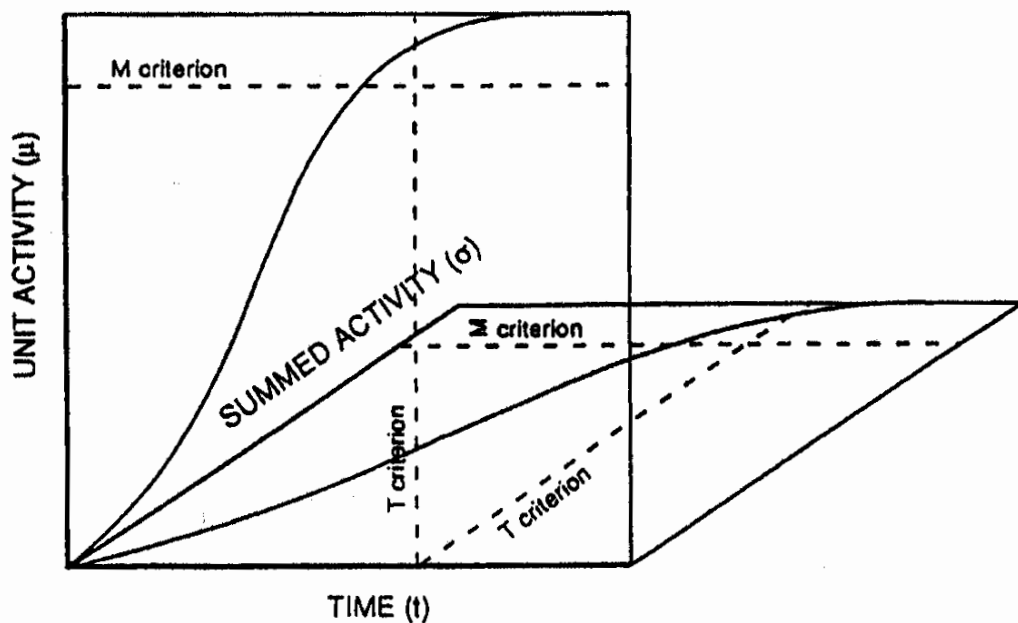


FIG. 5.3. Application of the multiple readout model to the lexical decision task. Three response criteria (M , Σ , T) are set on three information dimensions: unit activity in the mental lexicon (μ), summed lexical activity (σ), and time (t). Increases in μ and σ over time follow the sigmoid function of an interactive activation network (McClelland & Rumelhart, 1981). In general, word recognition is said to occur when the M criterion is reached, whereas a positive lexical decision response can be triggered when either the M or the Σ criterion is reached before the T criterion. A negative lexical decision response is given in the converse situation.

The MROM-p

The starting point for the MROM-p's coding scheme is the V-type (boxological) model of orthographic-phonological processing by Ferrand and Grainger (1994, see also 1996). This model (see Figure 5.4) was empirically motivated by results from a series of masked priming studies (Ferrand & Grainger, 1992, 1993, 1994) and represents the simplest possible (global) phonological coding scheme in an IA-type architecture that includes sublexical phonological structure. Typical of V-type models, Ferrand and Grainger did not specify the nature of the phonological processing units.

According to the principles of canonical and nested modeling (chap. 1), we started the construction of the MROM-p with the original structure, processing assumptions, and parameters of the MROM. These elements had already been kept constant in our previous "English" and "French" extensions of the IAM, the semistochastic IAM or SIAM (Jacobs & Grainger, 1992), the letter-frequency model (Grainger & Jacobs, 1993), the dual readout model or DROM (Grainger & Jacobs, 1994), the semistochastic IAM for the fragmentation task or SIAM-FRAG (Ziegler, Rey, & Jacobs, in press), and the MROM (Grainger & Jacobs, 1996).

Multilength Lexicon and the Coding of Letter-in-Word Position.

In our previous IA models, we used the simplification of a length-specific lexicon representing a single word length (either four or five letters). In view of the absence of an isomorphism between the size of orthographic and phonological representations (i.e., grapheme and phoneme units), the present "English" MROM-p is equipped with a much richer lexicon (albeit still a very simplified one) than the MROM, including all three- to five-letter monosyllabic English words extracted from the CELEX database (Baayen,

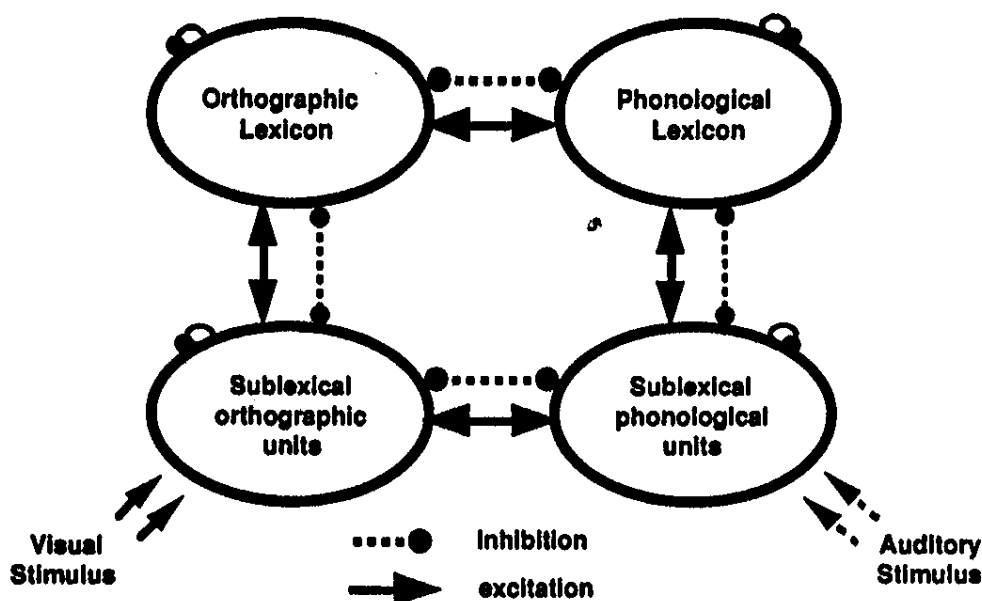


FIG. 5.4. Ferrand and Grainger's model (1994).

Piepenbrock, & van Rijn, 1993).⁵ This process led to an orthographic lexicon of 2,494 words (for a detailed description of the cleaning procedure applied on this database to extract the selected lexicon, see the Appendix).

This multilength lexicon raises the issue of how to code letter-in-word position and the relation between word units of different lengths. Some recent empirical studies have suggested that the cognitive system uses relative rather than absolute position coding (Grainger & Jacobs, 1993; Peressotti & Grainger, 1995; see also chap. 1). Thus, letters in a word are not supposed to be represented in terms of their absolute position in the word (i.e., U at the third position in BLUE), but in terms of their relative position, which is calculated from the word boundaries (e.g., U is one position before the final letter, and L is one position after the initial letter in BLUE; i.e., $U = \text{final} - 1$, and $L = \text{initial} + 1$, respectively). This coding scheme allows us to establish more plausible similarity relations between words of different lengths than the original position coding scheme of the IAM allowed.⁶ As an example, the stimulus WORD is encoded as W in initial position, O in initial plus one, D in final position, and R in final minus one. This coding scheme was applied to the orthographic lexicon and to the coding of the input stimuli presented to the model. Table 5.1 gives a general description of our coding scheme for three- to five-letter words.

Whole-Word Phonological Lexicon. In MROM-p we assumed that each orthographic representation of a word has a corresponding whole-word phonological representation. The phonological lexicon containing these representations is smaller than is the orthographic one because of homophones that have distinct orthographic entries but share the same phonological code. There are 2,323 whole-word phonological representations in contrast to 2,494 orthographic word units. Connections were established between each whole-word orthographic unit and its corresponding whole-word phonologic unit, and specific values of resting level activation (i.e., coding frequency of occurrence) were given to each unit using the CELEX frequency count (see Appendix for a description of the procedure).

Phonological Units. We assumed that a reader's cognitive system possesses some elementary phonological representations or coding units, and that these basic phonological units, or phonemes, are grouped according to

⁵Focusing on monosyllabic words is a simplification that has been adopted in most experimental and modeling studies in the field. In future work, this practice must be revised.

⁶The evaluation of the plausibility of this coding scheme requires further investigation and, probably, further refinements. The use of a more structured coding scheme, introducing a grapheme level composed of functional pronunciation units as proposed by Berndt, D'Autrechy, and Reggia (1994; Berndt, Reggia, & Mitchum, 1987), is a possible candidate for such refinements, but more constraining results are needed to specify the nature and boundaries of such units.

TABLE 5.1
 General Description of Relative Position
 Coding Scheme for Three- to Five-Letter Words

<i>Word Length</i>	<i>L Initial</i>	<i>L Initial + 1</i>	<i>L Initial + 2</i>	<i>L Final - 1</i>	<i>L Final</i>
5	L ₁	L ₂	L ₃	L ₄	L ₅
4	L ₁	L ₂	—	L ₃	L ₄
3	L ₁	L ₂	—	—	L ₃

a consonant, vowel, consonant (CVC) or, more precisely, an onset, nucleus, coda (ONC) scheme. This scheme was motivated both by current linguistic theory (e.g., Wiese, 1996) and by empirical data (Dell, 1988). We chose this subsyllabic organization as a pragmatic, parsimonious compromise. Higher sublexical phonological structures, like rime units, for example, are less flexible and less general. Lower structures, like ungrouped phoneme strings, for example, complicate the connectivity between letter and phoneme units.

The phonemic representation level thus contains onset, nucleus, and coda positions. These units can code single phonemes or phoneme clusters. Furthermore, a "silent phoneme" is added at the onset and coda positions to represent monosyllabic words that have no pronounced consonant at their beginning or end. As an example, the stimulus GOOD is phonologically coded as /gUd/ and decomposed as Onset equals /g/; Nucleus equals /U/; Coda equals /d/. Similarly, BLUE /blu/ is decomposed as Onset equals /bl/; Nucleus equals /u/; Coda equals /*/ (the asterisk represents the silent phoneme in the coda position). Table 5.2 shows the MROM-p's three sets of phonological units, which include all possible phonemes or combinations of phonemes contained in the phonological lexicon of the model at each position.

Connections Between Letters and Phonological Units. Figure 5.5 illustrates our connection scheme for linking orthographic and phonological units, a tentative solution to the problem of discovering an optimal graph-

TABLE 5.2
 Description of Three Sets of Phonological Units

ONSET 51 units	*, b, bl, br, d, D, dr, dw, f, fl, fr, g, gl, gr, h, J, k, kl, kr, kw, l, m, n, p, pl, pr, r, s, S, sk, skr, skw, sl, sm, sn, sp, spl, spr, Sr, st, str, sw, t, T, Tr, tr, tw, v, w, z, _
NUCLEUS 34 units	1, 2, 3, 4, 5, 6, 7, 8, 9, #, \$, @, E, i, I, j#, j\$, j1, j3, j5, j7, j8, j9, jE, ji, jQ, ju, jV, jl, Q, u, U, V, l
CODA 78 units	*, b, d, D, dst, dT, Dz, dz, f, fT, ft, g, gz, h, J, k, ks, ksT, kst, m, md, mf, mp, mps, mt, mz, kt, l, lb, ld, lf, lk, lm, ln, lp, lS, ls, lt, lT, lv, lz, l_, n, N, nd, ndz, nj, Nk, Nks, nS, ns, nt, nT, nts, nz, nZ, Nz, n_, p, ps, pt, pT, R, s, S, sk, s

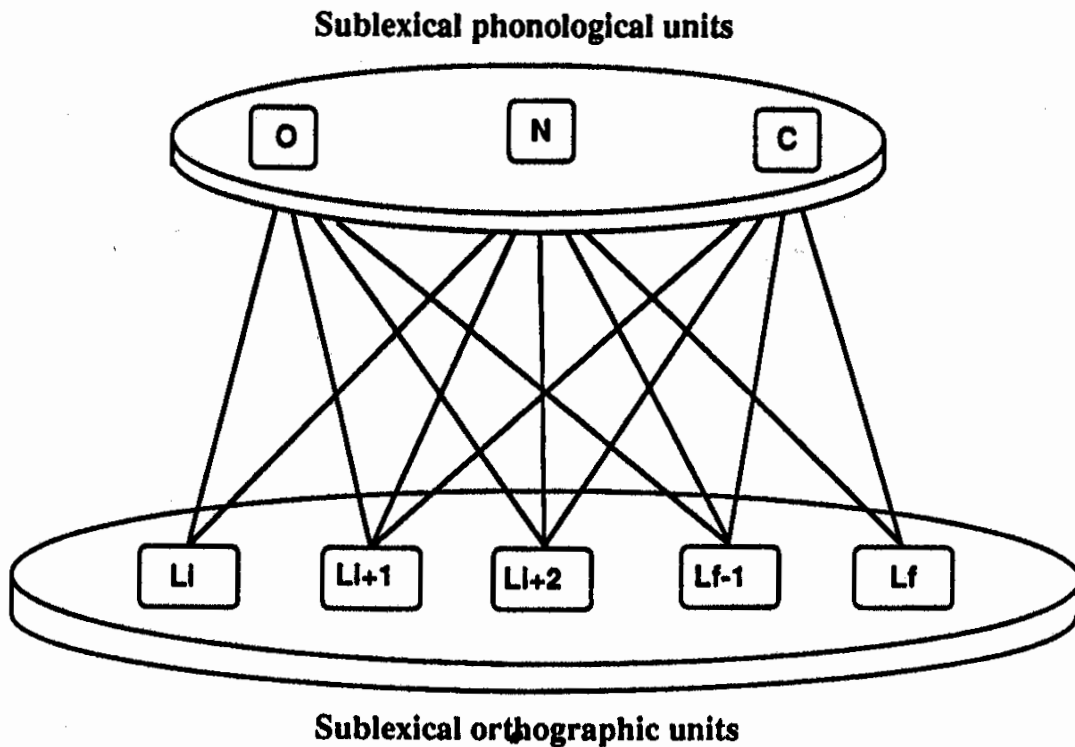


FIG. 5.5. Connectivity between letter and phonological unit levels.

eme-to-phoneme correspondence (GPC) scheme for IA-type models. To be consistent with the (system) principle of spreading activation of IA models, we carried out an exhaustive analysis of the present lexicon and recorded all links between each letter-grapheme at each relative position and between each phonological unit at the ONC positions. The results were stored in large matrices that code the GPCs, such as the fact that B in initial position (as in BLUE) is connected to the phonological unit /b/ at the onset position. Similarly, for the same word, L at the initial plus one position is connected to /b/ in onset, U at the final minus one is connected to /u/ in nucleus, and so on. Thus, in terms of the spreading activation mechanism of the model, when a letter like B is activated in initial position, it sends excitation to all corresponding phoneme units in onset position, for example, /b/, /bl/, and /bR/. Figure 5.6 gives an illustration of what we might call the phonemic space in the MROM-p. It shows activation functions of the phonological units at the ONC positions obtained for the target word BLUE.

Parameter Tuning. In our parameter-tuning approach, we also followed the strategy of nested modeling. Because this strategy demands that MROM is an integral part of MROM-p, we faced the problem of having to find new parameters for the phonological substructures of the model while keeping the original parameters of MROM as constant as possible. Because adding new interactive substructures to the original IAM leads to a new global dynamic system, the original parameter set had to be adjusted (the

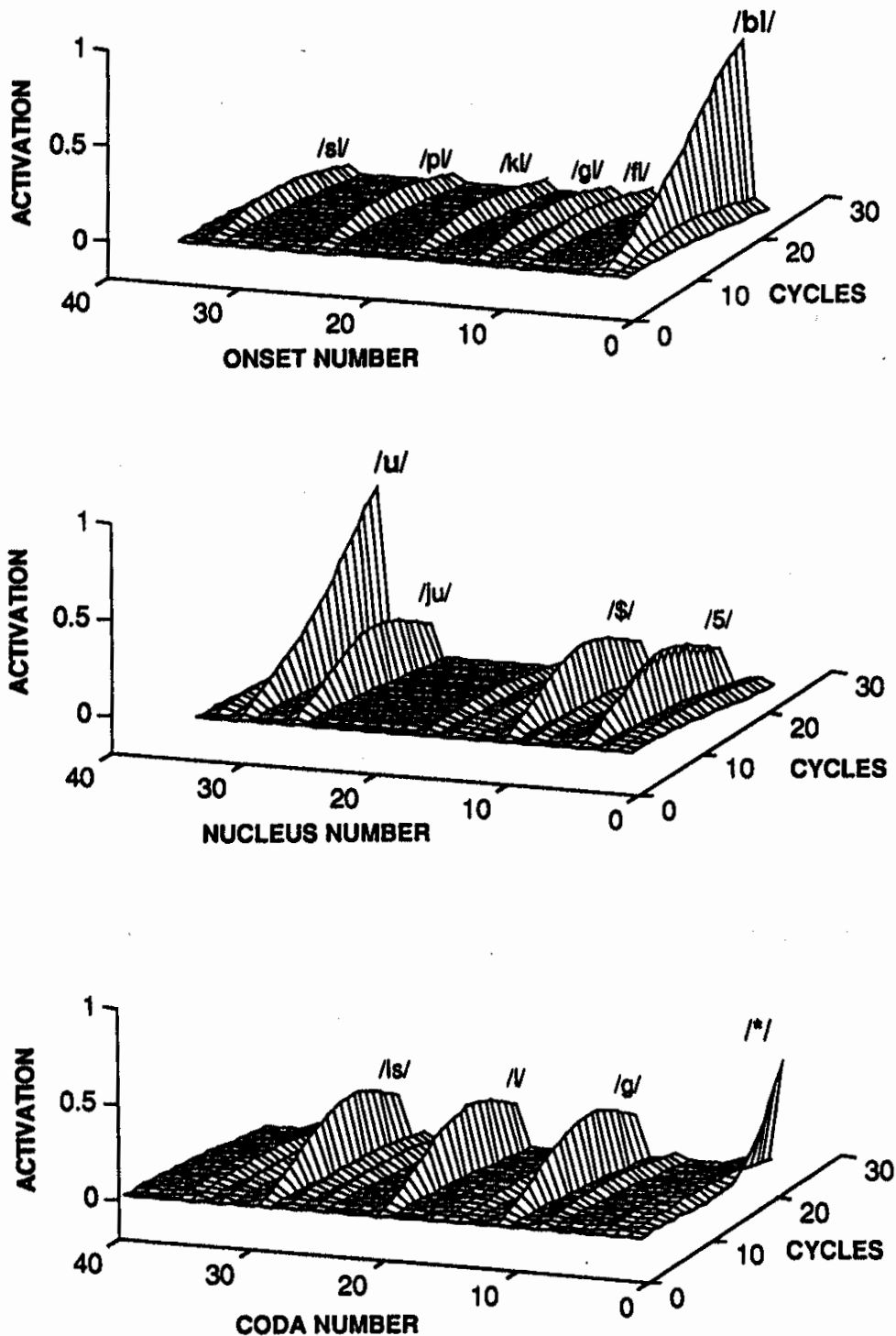


FIG. 5.6. Activation functions of phonological units at the onset, nucleus, and coda positions for the target word BLUE.

old one led to catastrophic model behavior). Table 5.3 gives the parameter set fixed for the present simulation studies on the basis of the parameter-tuning studies discussed later. We acknowledge that more work has to be done to precisely determine the role of each parameter for the model dynamics. Furthermore, although connectionist modelers in general seem to ignore the issue, we acknowledge that a central aim for future work in the field must be to solve the nontrivial problem of the identifiability of complex A-type models in general and models of the IA family in particular.

TABLE 5.3
Parameter Set Used in Simulation Studies

<i>Connections</i>	<i>Alpha</i>	<i>Gamma</i>
Feature unit—Letter unit	.005	.15
Letter unit—Feature unit	0	0
Feature unit—Feature unit	0	0
Letter unit—Letter unit	0	0
Letter unit—Lexical orth. unit	.035	.02
Lexical orth. unit—Letter unit	0	0
Lexical orth. unit—Lexical orth. unit	0	.2
Letter unit—Sublex. phono. unit	.1	.004
Sublex. phono. unit—Letter unit	0	0
Sublex. phono. unit—Sublex. phono. unit	0	.05
Sublex. phono. unit—Lexical phono. unit	.1	.02
Lexical phono. unit—Sublex. phono. unit	.3	.2
Lexical phono. unit—Lexical phono. unit	0	.2
Lexical orth. unit—Lexical phono. unit	.1	0
Lexical phono. unit—Lexical orth. unit	.2	0

Note. *Alpha* refers to the weight of excitatory connections; *gamma* refers to the weight of inhibitory connections.

Simulation Method. In the following studies, we simulated effects obtained in the LDT. Because the task-specific mechanisms of the LDT have been well specified in the MROM (Grainger & Jacobs, 1996), the same design principles (i.e., principles that determine the behavior of a class of models and the observed dependent variables) were adopted here. At this stage, we did not carry out full-blown simulations. The present version of MROM-p is still prototypical, a fact implying that the present study provides no criteria-oriented falsification test of the MROM-p but rather a test of its appropriateness as a working model of phonological coding (see discussion section). Full-blown simulations would have necessitated, for example, use of stochastic response thresholds and of many simulated subjects equal to the number of subjects in the real experiments (Grainger & Jacobs, 1996). Consequently, rather than response (time) distributions, we simply obtained activation functions for each item used in the simulated experiments. For each experimental condition, the mean of these activation functions was calculated across the different items. The resulting activation functions (and derived mean RT bar charts) are presented as illustrations of how the model can quantitatively account for the experimental data.

For the task-specific readout procedure, the MROM-p follows the design principle of the MROM in that No responses are produced by monitoring the global orthographic activity generated by the stimuli and Yes responses are generated by looking at the orthographic unit activity (for simplicity, we did not consider the possible role of global orthographic activity on Yes

trials here; compare Figure 5.3). The rationale for this decision is given by Ferrand and Grainger (1996), who discussed the qualitative predictions of MROM-p for a masked priming LDT. In accord with the assumptions of Grainger and Jacobs (1996), the results of Ferrand and Grainger suggested that in an LDT using pseudohomophones, participants use readout from the orthographic lexicon, because readout from the phonologic lexicon would lead to too many false positive errors.

Model Predictions and Tests

Testing Strategy

The literature provides no generally accepted testing policy for complex A-type models. The situation is anarchic; for example, consider PDP or artificial neural network (ANN) learning models. Whereas classical mathematical learning theory provided a wealth of testing principles for M-type modelers (Tack, 1976; see also Myung & Pitt, chap. 10, this volume), ANN or A-type modelers today seem little concerned with this issue (Prechelt, 1996; Simon & Kaplan, 1989). Surely, the "fit-or-die" strategy of classical learning theory has both statistical, inferential pitfalls (Collyer, 1985, 1986) and epistemological drawbacks (Greenwald & Pratkanis, 1988; Lakatos & Musgrave, 1970). On the other hand, the question arises whether the current anarchic or laissez-faire testing strategy for computational A-type models will eventually give the positive results anticipated by laissez-faire anarcho-epistemologists like Feyerabend (1975).

Without any claims that the testing strategy adopted here is the right or optimal one, we nevertheless prefer a critical-rational, (mildly) Popperian approach. Such an explicit approach can be constructively criticized and thus potentially advances our enterprise. Our testing approach is made up of the following five steps, inspired by testing procedures in psychometrics and mathematical psychology.

1. **Parameter tuning studies:** During the initial phase of model construction, these tests check the global appropriateness of the architectural and parametric model assumptions in a simple way to see whether the model is not fundamentally flawed (e.g., does not include parameter configurations that produce catastrophic model behavior). The meaning of "in a simple way" depends very much on each model builder's implicit assumptions and preferences. We have found no systematic, explicit approaches for parameter tuning in complex A-type models in the literature (cf. McClelland & Rumelhart, 1988). We can provide only motivated examples for how we proceeded.

2. **Estimator set studies:** In analogy with classical procedures of cross-validation in psychometrics, an estimator study provides the data set from which model parameters are estimated (cf. Collyer, 1986). For the present

purposes, the difficult questions of how parameters of M-type and A-type models are best estimated and how such models' identifiability can best be determined must be put aside (see chaps. 1 & 10). Once a new model has stood the test of Step 1, it is economical to run it against already available data from the literature in Step 2 before carrying out time-intensive new experiments. We adopt this method here.

3. Criterion set studies: In analogy with the procedures of psychometrics, a criterion set study provides data with which the model predictions are compared once the parameters have been fixed after the estimator set study. With a criterion of descriptive or behavioral accuracy (for a discussion of this criterion, see Jacobs & Grainger, 1994), the criterion set study provides the first serious cross-validation test of the model. We provide two such tests here: The first uses data about the same effect as that used in the estimator set study, but from a different empirical study. The second test uses data about a different effect from a different study.

4. Strong inference studies: In chapter 1, we discuss the stratagem of strong scientific inference in detail. This relatively costly but worthwhile testing phase involves formal, criterion-guided comparisons of alternative models against the same data sets. We cannot provide such testing here (for a typical approach, see Massaro & Friedman, 1990).

5. Model refinement or replacement: As theoretically firm believers but practically mild (nondogmatic, non-naive) users of a theory-building approach adhering to Popper's (1935) and Platt's (1964) principles (see chap. 1), we would continue with a process of model refinement (after which we return to Step 1) as long as the model is only mildly discredited and no better alternative is available. In view of the current state of the art in modeling visual word recognition (Jacobs & Grainger, 1994), in a pluralistic perspective of canonical modeling (Stone & Van Orden, 1993, 1994; see also chap. 1), such an adoption of a hybrid between "falsificationism" and "refinementism" seems in order. We are nevertheless aware of the dangers of confirmation bias (Greenwald & Pratkanis, 1988) and believe that eventually IA-type models—including the present one—will be no longer refined but replaced by better models. At present, we have reasons to believe that IA-type models have a lot to offer (Jacobs & Grainger, 1994). The present volume is perhaps the nicest expression of and justification for this belief.

The MROM as Null Model

In the following model tests, we used the MROM as a null model of phonological effects. That is, as the MROM has no explicit phonological processing units, it should not predict any difference between stimuli having phonological properties, such as pseudohomophones, and supposedly control stimuli that lack these properties. The MROM could very well produce

“pseudophonological” effects when the pseudohomophones differ on dimensions other than phonological ones from the controls, for instance, when they were badly matched for orthographic neighborhood properties. To the extent that the MROM provides a successful model of orthographic neighborhood effects in word recognition (Grainger & Jacobs, 1996), it can also be used as a tool for precisely selecting stimuli, for example, for avoiding pseudophonological effects. In contrast, if we included adequate phonological processing units in the MROM-p, it should predict clear differences between control stimuli and pseudohomophones, for example. In the present MROM simulations, the parameters governing the phonological parts of MROM-p were simply set to zero.

Step 1: Parameter-Tuning Studies: The Pseudohomophone Test.

Once the initial parameter-tuning procedure gave satisfactory results, a first simple test of the ability of MROM-p to account for phonological effects consisted in presenting the model with “watertight” pseudohomophone stimuli, whose correct pronunciation is empirically confirmed (Van Orden, Johnston, & Hale, 1988). As an example, we presented both MROM and MROM-p with stimulus triples, such as FEEL (base word), FEAL (pseudohomophone), and FEED (control). Figure 5.7 shows activation functions for both MROM and MROM-p at the level of orthographic word units, which we take to be the critical level for assessing interactive phonological effects in the LDT (Ferrand & Grainger, 1996). The simulation results are clear-cut. Whereas stimuli like FEEL generate sufficient lexical activity in both MROM and MROM-p to be correctly recognized, FEAL and FEED generate the same lexical activity in MROM but not in MROM-p. Here, pseudohomophones like FEAL generate activity intermediate between real words like FEEL and control pseudowords like FEED. Thus, in stochastic simulations under data-limited conditions (i.e., brief, backward-masked stimulus exposure), MROM-p occasionally (i.e., depending on the noise level) identifies FEAL as FEEL (Ziegler & Jacobs, 1995; Ziegler, Van Orden, & Jacobs, 1997). We took this result as suggesting that MROM-p’s architectural-parametric assumptions were adequate and fixed the parameters to the values yielding this result (see Table 5.3).

Step 2: Estimator Set Study: Coltheart et al. (1977) Test. Although the previous study hinted at the appropriateness of MROM-p’s structural-parametrical assumptions, it was no serious estimator set study. For this, we chose the stimuli and data of the classical study of Coltheart et al. (1977, Experiment 1, table 1, and app. A), which provided one of the first falsifications of serial search models of word recognition (Forster, 1976). This study had already given good service in this respect during the construction phase of SIAM (Jacobs & Grainger, 1992).

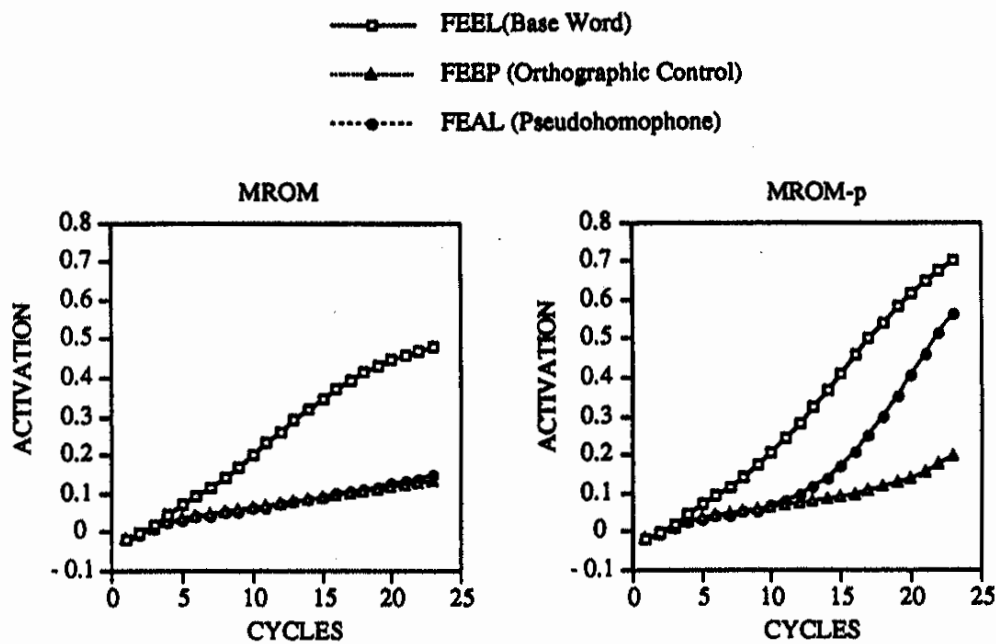


FIG. 5.7. Activation functions for the most activated word unit (local activity) with the MROM and the MROM-p for a pseudohomophone (*feal*), an orthographic control (*feep*), and their base word (*feel*).

The crucial result of Coltheart et al. for the present purposes concerns the longer mean latencies for correct No responses to pseudohomophones than to control pseudowords in the LDT. Coltheart et al. observed a 62-millisecond difference in the subject analysis and a 35-millisecond difference in the item analysis. Instead of using Coltheart's data for a full-blown parameter-fitting study, as we could have done with an M-type model, here we simply checked whether the MROM-p, as structurally and parametrically defined during the previous test phase, could simulate the data from Coltheart et al. If not, we would have gone through another phase of parameter tuning or model restructuring.

Figure 5.8a gives the mean overall lexical (orthographic) activity over time, according to MROM, the critical information dimension determining No responses (Grainger & Jacobs, 1996), generated by the 30 base words, 30 pseudohomophones, and 30 control pseudowords of the Coltheart et al. study contained in our lexicon (18 of the stimuli had to be excluded because their base word was bisyllabic or absent in our model lexicon). In MROM, both pseudohomophones and control pseudowords generate the same activity, which yields identical mean No RTs. In contrast, in MROM-p, overall lexical activity nicely reflects the differences between pseudohomophones and controls. Because pseudohomophones generate more activity than do controls, No responses to them take longer to reject. This finding is shown in the bar charts of Figure 5.8b, which captures the pattern of results obtained by Coltheart and collaborators.

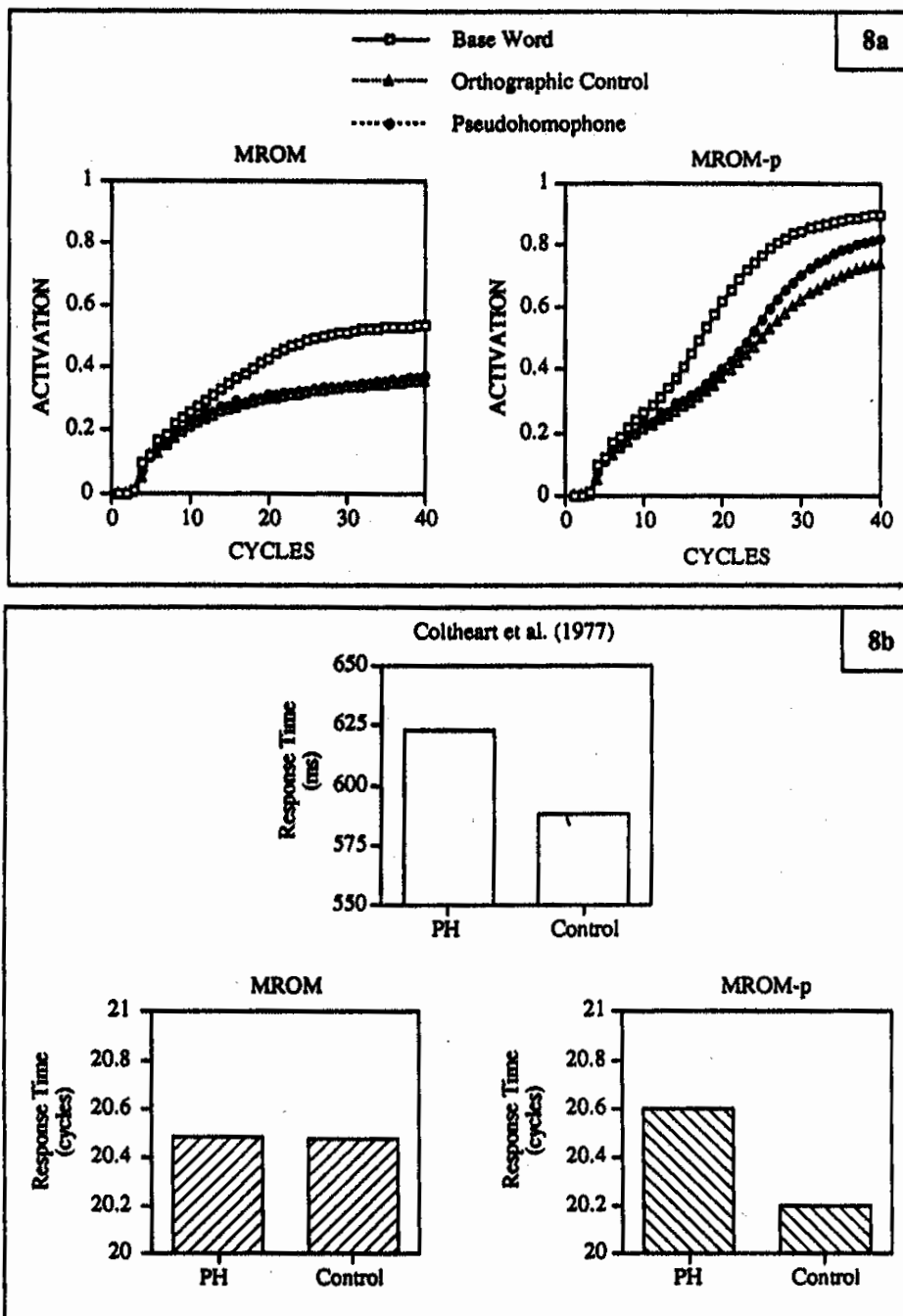


FIG. 5.8. Simulations run with Coltheart et al.'s (1977) stimuli. Figure 5.8a shows the mean overall orthographic activity over time for pseudohomophones, orthographic controls, and base words, obtained with the MROM and MROM-p. Figure 5.8b gives obtained and predicted effects on RTs to pseudohomophones and orthographic controls.

Step 3: Criterion Set Studies: Study 1—Seidenberg et al. (1996)

Test. As a first criterion set test of MROM-p's ability to capture the pseudohomophone effect in the LDT, we ran simulations using the stimuli of a study comparing pseudohomophone effects in the LDT and the naming task (Seidenberg et al., 1996). These authors observed a 31-millisecond inhibitory pseudohomophone effect on mean No RTs in the LDT. This criterion set study used the parameter set fixed during the previous two steps and provided a cross-validation test of MROM-p for the same experimental effect

as used in the estimator set study, but obtained in an independent study with different stimuli and subjects.

Figure 5.9 summarizes our simulation results. The activation curves show the same trends as those for the Coltheart et al. study. In MROM, both pseudohomophones and control pseudowords generated the same activity, whereas in MROM-p, overall lexical activity reflected the experimentally observed differences between pseudohomophones and controls. Thus, contrary to the MROM, MROM-p captures the well-replicated pseudohomophone effect in the LDT using the parameters from the estimator set study. This is first encouraging evidence for the descriptive accuracy, (cross) validity, and generality of the model. A stronger test involves confronting the model not only with data different from the estimator set study, such as in the previous test, but also with a different effect, that is, data from the manipulation of experimental factors different from the estimator set study. Thus, in the second criterion set study, the model is not tested with respect to the pseudohomophone effect (i.e., a difference in LDT latencies to nonword stimuli) but with respect to two effects of LDT latencies to words: the feedforward and feedback consistency effects described earlier.

Criterion Set Study 2—Stone et al. (1997) Test. The critical feature of this second criterion set study is that it tests the model with data exhibiting a different effect coming from an independent study. If the MROM-p successfully simulates the feedforward and feedback consistency effects for word stimuli found by Stone and collaborators with the parameter set that

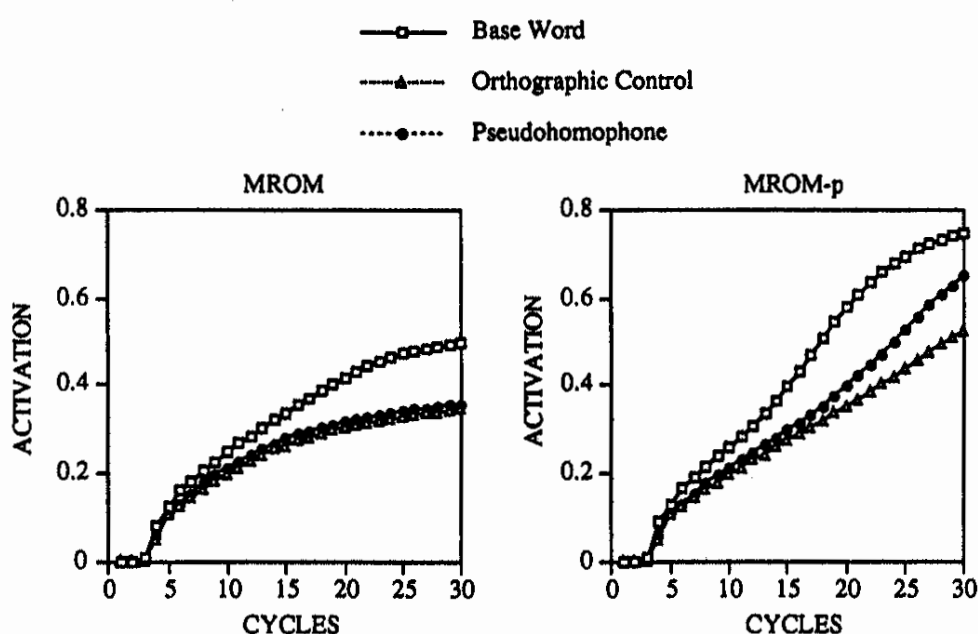


FIG. 5.9. Mean overall orthographic activation functions for Seidenberg et al.'s (1996) stimuli obtained with the MROM and MROM-p for pseudohomophones, orthographic controls, and base words.

simulated the pseudohomophone effect for nonword stimuli, we can hypothesize that both effects are different manifestations of the same underlying interactive mechanisms.

Figure 5.10 summarizes the results of our simulations using the word stimuli of Stone et al. (1997). The figure gives mean individual lexical (orthographic) activity over time for the four groups of word stimuli used in

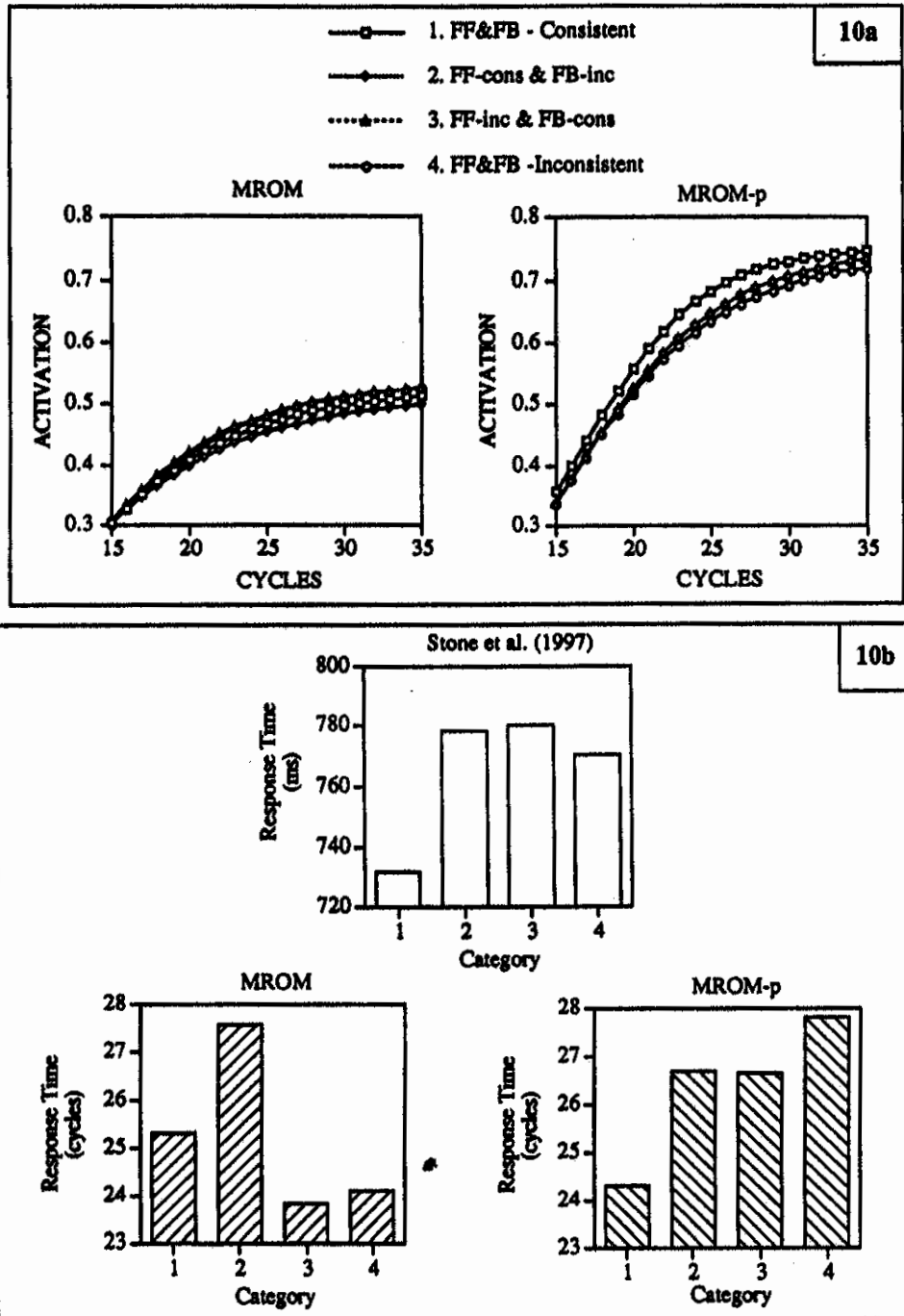


FIG. 5.10. Simulations with Stone et al.'s (1997) stimuli. Panel 5.10a shows mean orthographic unit activity over time for the four categories of words, for both the MROM and MROM-p. Panel 5.10b gives response time obtained (ms) and predicted (cycles) to the four word groups: 1 = Feedforward and Feedback consistent; 2 = Feedforward consistent and Feedback inconsistent; 3 = Feedforward inconsistent and Feedback consistent; 4 = Feedforward and Feedback inconsistent.

this study. The correct mean RTs and rounded error rates obtained by Stone et al. are indicated in the following: feedforward and feedback consistent (e.g., COIN, STAB: 732 ms, 2%), feedforward consistent and feedback inconsistent (HEAP, MOAN: 778 ms, 9%), feedback consistent and feedforward inconsistent (PINT, COUCH: 780 ms, 9%), and doubly inconsistent (NEAT, SWARM: 770 ms, 15%). The data of Stone et al.'s Experiment 2 indicate an overall 20-millisecond feedforward consistency effect (775 – 755 ms) and an 18-millisecond feedback consistency effect (774 – 756 ms).

Figure 5.10 shows that the MROM fails to capture these bidirectional consistency effects: It wrongly predicted that doubly inconsistent words (and feedforward inconsistent–feedback consistent words) are processed faster than are fully consistent words. According to MROM, feedforward consistent and feedback inconsistent words yielded the slowest processing. In contrast, MROM-p captured Stone et al.'s data pattern much better, although not perfectly. In MROM-p, fully consistent words were processed fastest and doubly inconsistent words were slowest, with the two monodirectionally inconsistent words being intermediate. Thus, although in Stone et al.'s data, the difference between the three inconsistent word groups was not important, MROM-p predicted a difference between monodirectionally and bidirectionally inconsistent words. Thus the model predicts a pattern actually closer to the one observed by Ziegler, Montant, and Jacobs (1997) in French: In contrast to the finding of Stone et al., these authors observed that doubly inconsistent words produced slightly longer RTs than did one-way inconsistent words. In light of this incompatibility in the empirical results and because Ziegler et al.'s study used more carefully controlled stimulus materials than did Stone et al.'s, it would be premature to conclude that the MROM-p failed this second criterion set test. Tests using the French stimuli of Ziegler et al. (1997a) must wait until a French version of the MROM-p is available.

As an additional guard against a possible confirmation bias, we ran another test of the model using the nonword stimuli of Stone et al. (1997). The interesting result for RTs to nonwords was the absence of a significant difference between feedback consistent and feedback inconsistent nonwords. If MROM-p captures this null effect, we can be more confident that its failure to simulate the null difference in RTs to doubly inconsistent versus one-way inconsistent words does not represent a fundamental problem with the model. The data in Figure 5.11 show that this is the case.

Thus, MROM-p successfully stood the second criterion set test. This result allowed us to hypothesize that the pseudohomophone and bidirectional consistency effects are different stimulus-specific manifestations of interactive processes operating between orthographic and phonological representations and that the present MROM-p provides a viable “model-system” in which such effects can be understood at sufficient levels of clarity, transparency, formality, and precision.

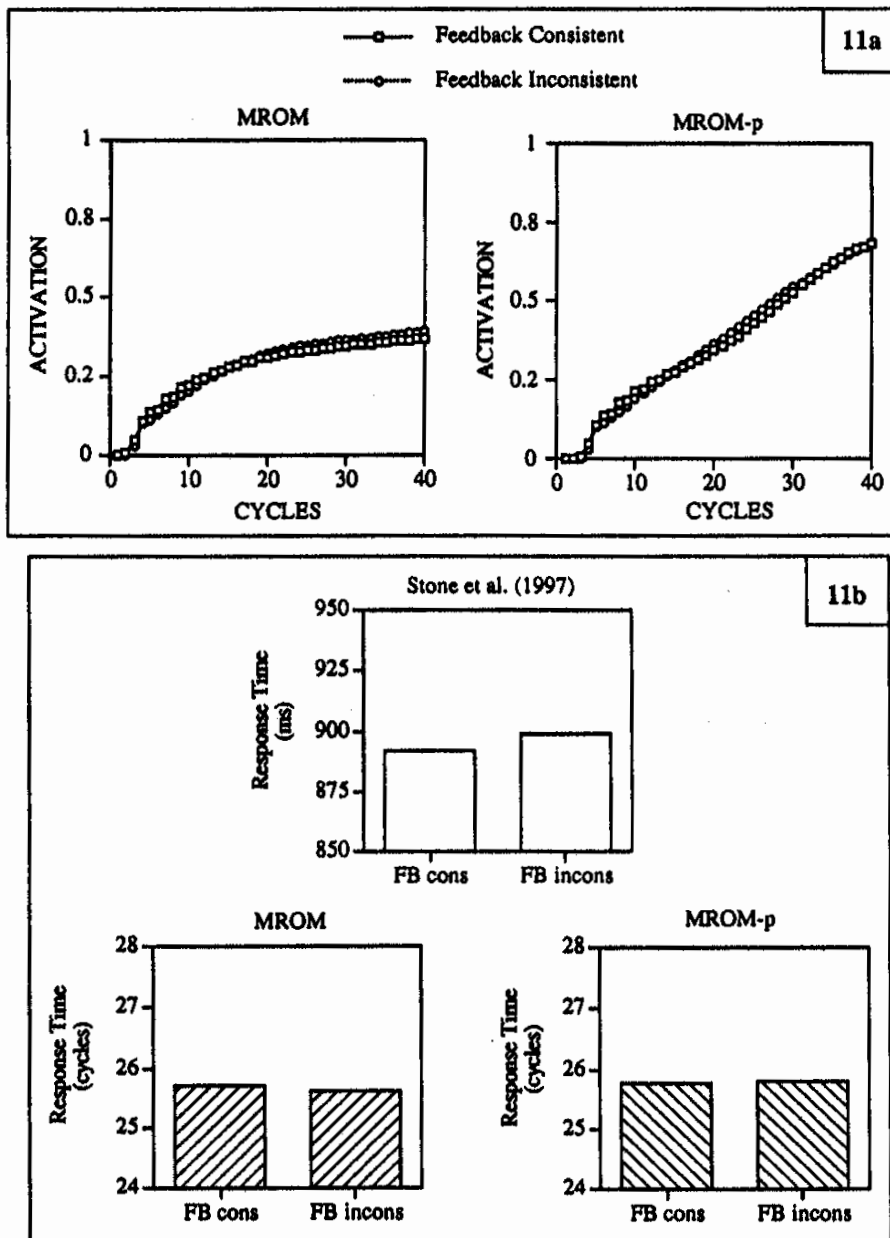


FIG. 5.11. Simulation of Stone et al.'s (1997) nonword data.

Model Evaluation

In this section, we follow the tentative criteria for model comparison and evaluation proposed by Jacobs and Grainger (1994), in particular, potential and actual descriptive accuracy, horizontal and vertical generality, and simplicity and falsifiability.

Potential and Actual Descriptive Accuracy

The first aspect of this accuracy criterion is potential descriptive accuracy. Does the model allow predictions at the level of scale at which the dependent variables of interest are actually measured? For example, when the dependent variable that reflects the effect is interval scaled (e.g., a frequency effect measured in ms), a model has potential descriptive accuracy if it allows predictions on the scale of milliseconds. Any current psychological A-type

model can achieve this only indirectly, for example, by transforming cycle times into RTs via regression analyses.

Despite complex considerations about measurement problems, we think that A-type modelers should take this issue seriously and provide a means by which model users can formally compare the values of the dependent variable(s) they want to study with those of the model's output. This process is necessary to develop standards for evaluating the relative goodness-of-fit for competing models. Why should "eyeballing" or "hand waving" be accepted as alternatives to model-to-data fitting for complex A-type models when such methods are not accepted for any other formal model format? Massaro and Friedman (1990) provided an encouraging example for evaluating the descriptive accuracy of a set of comparable M- and A-type models when percentage correct is the dependent variable. As concerns the more complex issue of predicting RT means and distributions with A-type models, first steps toward progress in this direction have been made in some recent studies (Grainger & Jacobs, 1996; Jacobs & Grainger, 1992).

Like its precursor, the MROM, the present MROM-p has potential descriptive accuracy for a variety of dependent variables, including RT means and distributions for both correct and incorrect responses, as well as hit or false alarm rate for the LDT and percentage correct for perceptual identification tasks (see Grainger & Jacobs, 1996).

As concerns the actual descriptive accuracy of the MROM-p, the tests are encouraging but not conclusive. To be conclusive, we would need to test the model in both a deeper and broader fashion, similar to our extensive tests of the MROM. Deeper tests would imply providing graphs showing linear regression between predicted and observed mean RTs for item or participant analyses, as well as distributional and error analyses (Grainger & Jacobs, 1996). Broader tests would imply running simulations of tasks other than the LDT (e.g., perceptual identification tasks).

In accord with Step 4 of the testing strategy previously proposed, the MROM-p should be tested in competition with comparable, alternative models on the same broad range of tasks and dependent variables before any conclusions about its actual descriptive accuracy can be made. The time is not now ripe for this, but we hope that easily comparable, broadly testable variants of the MROM-p and, for example, the dual-route cascaded model (Coltheart & Rastle, 1994) are soon available.

Even if quantitative, strong inference comparisons between models of phonological processing in visual word recognition become possible in the near future, a problem remains: Finding that one model fits the data better than competing models does not establish the best fitting model as the probable source of the data (Collyer, 1985, 1986). Developing methods to overcome Collyer's almost totally neglected problem represents one of the interesting challenges for A-type (and M-type) model builders in the future.

Horizontal and Vertical Generality

Jacobs and Grainger (1994) distinguished between horizontal and vertical generality. Horizontal generality refers to a model's ability to generalize across different stimulus sets, configurations (stimulus generality), or both, different tasks (task generality), or response types and measures (response generality). Vertical generality refers to a model's ability to generalize across different scales of the modeled process, such as (macrostructural) static-asymptotic behavior versus microstructural dynamics, or different types or sizes of a processing structure, such as the number of entries in the lexicon of a simulation model. Vertical generality has received little attention in comparison with horizontal generality, but it might become an important issue in a field that provides more and more complex algorithmic models, some of which might have severe limitations for scaling up (e.g., distributed connectionist models; Feldman-Stewart & Mewhort, 1994; Jacobs & Grainger, 1994).

We have discussed the vertical generality of the SIAM elsewhere (Jacobs & Grainger, 1994). By virtue of nested modeling (i.e., SIAM is an integral part of MROM, which is an integral part of MROM-p) and because MROM-p includes a richer lexicon than does MROM, MROM-p has higher vertical generality. As for horizontal generality, again thanks to our application of the nested modeling strategy, we can say the following: Because MROM has stood an extensive series of tests in different tasks and languages (and thus has reasonable horizontal generality) and to the extent that we can show that the MROM-p behaves at least qualitatively like the MROM (e.g., as for the simulations of the nonword data in Figure 5.11), the MROM-p has a higher degree of horizontal generality than does the MROM because it allows adequate simulation of stimuli processing that the MROM cannot account for (i.e., pseudohomophones and inconsistent words). Apart from this verified higher stimulus generality, it remains to be seen to what extent the MROM-p also includes the promise of higher task generality, the capacity to simulate data from tasks other than the LDT or the perceptual identification task (the naming task).

Simplicity and Falsifiability •

This criterion is one of the trickiest in model evaluation (Jacobs & Grainger, 1994). The current state of affairs and our adoption of a moderate "Popperianism" make evaluation relatively easy: It is premature to make any sensible statement about the simplicity and falsifiability of the MROM-p. Popper (1935) linked the criterion for simplicity to that of falsifiability (i.e., a model's ability to generate predictions that can be falsified), in proposing that, with two models in the same domain with equal success, we should prefer the simpler. He defined simplicity as a property that places the greatest restrictions on the world, that is, on how the empirical data can turn out to

be.⁷ Thus, we should prefer the model that is more easily falsified (cf. Estes, 1975; Massaro & Cowan, 1993), but only when we already have two viable models (or model variants) for which equal success has been established in the same domain. As concerns the present subject area, this is not the case, and we can only postpone evaluation with respect to this criterion.

Notwithstanding the repeatedly appearing critique of connectionist models as being too powerful (Massaro, 1988) and therefore not falsifiable in any easy way, a nontrivial question about falsifiability can be answered here: Is the MROM-p falsifiable at all? Following other theoreticians, we have proposed that any A-type model should offer a clear answer to the question: What cannot be or happen, if the model is correct? In other words, which effects or phenomena does the model exclude? An example is given by Grainger and Jacobs (1996), who showed that if the MROM is correct, a facilitatory effect of orthographic neighborhood density (as measured by Coltheart's *N*) is possible in both the Yes–No and Go–No–Go variants of the LDT, but *not* in the perceptual identification task. Thus, replicable experimental demonstrations of a facilitatory *N* effect in the perceptual identification tasks would falsify the MROM.

Another example was given earlier (see Figure 5.8). According to the MROM as included in the present MROM-p, a phonological pseudohomophone effect in the LDT is impossible, because the MROM possesses no phonological processing units whatsoever. Thus, although the MROM was explicitly designed to deal with orthographic processing in the LDT and other reading tasks that do not include pseudohomophonic stimuli, the simulation data in Figure 5.8, for example, present a falsification of the (nonphonological) MROM. Evidently, simply falsifying a model by using conditions outside its explicitly stated validity space (domain of application) is not necessarily useful. As we have demonstrated in this chapter, using the MROM as a null model against which to test models of phonological coding is a more useful variant of falsification studies.

In a field that lacks universal laws, we cannot expect models to have universal validity (cf. Newell, 1990). On the other hand, we can hardly want to continue with models that accurately account only for a single effect, as measured by a single variable in a single task, but whose validity stops there (cf. Jacobs & Grainger, 1994; Newell, 1990; Roberts & Sternberg, 1993).

The MROM-p is also falsifiable in several nontrivial respects. Like the MROM, it allows making qualitative predictions that can be tested in a straightforward way. An example was discussed by Ferrand and Grainger (1996). They used a prequantitative version of MROM-p (called a bimodal

⁷For Popper's simplicity–falsifiability criterion to work, certain rules have to be respected in this game of non-naive, undogmatic falsificationism and strong inference (e.g., constraints on the use of auxiliary assumptions; see Popper, 1935, 1966, 1972). Otherwise, the strength of Popper's approach can easily be turned into weakness (cf. Feyerabend, 1975).

extension of the MROM) to make qualitative predictions about the existence and direction of priming effects in a masked priming LDT manipulating prime type (homophones, pseudohomophones, or unrelated controls) and list composition (pseudohomophones, legal pseudowords, or illegal nonwords). The strongest qualitative prediction of MROM-p, the one most easily falsifiable, was that it predicts a null effect with homophone primes in the presence of illegal nonwords.⁸ The rationale for this finding is that the presence of illegal nonwords encourages participants to use the Σ criterion, because such nonwords can easily be discriminated from words on the basis of summed lexical-orthographic activity; homophone primes generate high levels of orthographic inhibition when readout is from the orthographic M criterion. The facilitatory effects from increased use of the Σ criterion (i.e., the fast guess mechanism producing decreases in RT) is canceled by the inhibitory effects of homophone primes. A null effect is the predicted result. In Ferrand and Grainger's (1996) experiment, this result was the case.

DISCUSSION AND OUTLOOK

Although a complete, criteria-oriented evaluation of the MROM-p is impossible at present, the results of our partial evaluation stand the general test criterion that we had fixed as our objective, that is, whether the present MROM-p is an appropriate prototype for developing a general model of phonological coding in visual word recognition. The model presented here is definitely a prototype, not in the sense of representing an ideal, but in the sense of being a working model. If we accept the principles of model development we adhere to, the model has some virtues. In the constraint of nested modeling, it represents what we think to be the simplest possible localist connectionist network that allows an account of two critical empirical effects indicating the influence of phonological processes in what is still the most widely used reading task in experimental psychology and psycholinguistics—the LDT.

⁸A word on null effects and their significance for theory building is in order, because many psychologists are firm believers in the virtues of null hypothesis testing (but see Gigerenzer & Murray, 1987; Rouanet, 1996; Van Orden, Alchison, & Podgornik, 1996). With models permitting quantitative predictions with reasonable precision, a null effect is a strong prediction to make. Perhaps the most famous example is the prediction of the null effect of the speed of light in Michelson and Morley's experiments by Einstein's special theory of relativity (Spielberg & Anderson, 1985). No known psychological A-type model can be compared with Einstein's theory: We simply want to make clear that the existence of theoretical tools allowing quantitative predictions of empirical effects frees us from the use of null hypothesis testing as exclusive inferential method. Thus, contrary to standard practice, accepting the null hypothesis can become a valid inference whenever there is sufficient faith in the validity and precision of a formal model or theory.

Moreover, the MROM-p, like our other work involving A-type modeling, is essentially a heuristic device in the sense discussed in chapter 1. It provides a heuristic, algorithmic description of phonological coding, but—needless to say—it falls short of presenting a computational theory in Marr's sense (1982). Few theoreticians in the field of cognitive science have achieved, or come close to, a computational theory (Marr, 1982; see also Jacobs, 1994; Pylyshyn, 1989), and the fields of memory or reading research are still not fully prepared for such an enterprise (Humphreys, Wiles, & Dennis, 1994; Jacobs, 1994). Nevertheless, these areas are open to some theoretical unification.

Neither the recent special section of the *Journal of Experimental Psychology: Human Perception and Performance* (Jacobs & Grainger, 1994) on modeling visual word recognition nor any other literature we have come to know since provides a computational model that can formally account for the pseudohomophone and bidirectional consistency effects in the LDT we have simulated here. At least one other model, however, has the potential to provide such an account, the dual route cascaded (DRC) model (Coltheart & Rastle, 1994). Because both the present model and the DRC belong to the family of localist connectionist models (and therefore are easily comparable) but differ with respect to one crucial structural feature—the presence of a GPC rule mechanism in the DRC—there might be some exciting strong inference studies ahead, but for such studies to be efficient, certain methodological issues have to be solved.

Future work on the MROM-p will involve adding phonological representations to multilingual models such as the bilingual IAM presented by Dijkstra and Van Heuven in chapter 6. How do multilingual speaker-readers deal with the different sets of spelling-to-sound correspondences in each language? If we postulate, as here, that the phonological coding component of the MROM-p is automatic and strategically nonmodifiable, how do bilingual readers deal with the potential interference caused by automatically generating all correspondences in both languages? The notion of language node introduced by Grainger and Dijkstra (1992; see also chap. 6) provides one solution to this problem. Top-down inhibition from the unattended language node to the corresponding word units (i.e., all words in the unattended language) would block resonance between irrelevant phoneme units and word units in the unattended language. Clearly, much exciting theoretical and empirical work is yet to be done in the multilingual domain.

Going Beyond MROM-p: A Challenge for Cognitive Modelers in “Word-Nerd” World

*Science, even more profoundly than politics, is the art of the possible.
It does only what can be done next.*

—A. Newell, *Unified Theories of Cognition*

The 1994 special section of the JEP:HPP entitled "Modeling Visual Word Recognition" gives an impression of the empirical and theoretical richness of this classical field of experimental cognitive psychology. Is it time for some serious effort of theoretical unification? Whoever thinks that this question is idle might look at table 1 in our editorial of that special section (Jacobs & Grainger, 1994), which gives a selective overview of 15 models of visual word recognition starting with Morton's (1969) logogen model. Originally, our taxonomic work started with the ambition to give as complete as possible a synopsis of models of word recognition to be used for the tasks of theoretical unification and development of standards for model comparison and evaluation. That aim was clearly too ambitious! Even the published version is far too complex to have a fair chance of being used in the way we wanted it to (previous versions of the table, including more than 40 different models of the past 30 years, were worse!).

Nevertheless, we continue to think that the answer to our question is a clear Yes. Perhaps pressure for theoretical unification in the word recognition literature is not as high as in the general field of cognitive science. As Newell (1990, p. 25) stated: "In my view, it is time to get going on producing unified theories of cognition—before the data base doubles again and the number of visible clashes [i.e., between theory and data] increases by the square or cube." On the other hand, if Newell (1990) could convincingly present four harbingers of unified theories of cognition, why should we not be able to agree on a limited number of harbinger models that have the potential to become unified models of word recognition? After all, we are dealing with only a small part of the cake of cognition (although perhaps with one of the most complex parts). Models of the steadily growing IA family (like the present volume) are clearly among the harbinger candidates, and because we have already argued for this goal (Jacobs & Grainger, 1994), we do not repeat ourselves (see chap. 1). Suffice it to say that the present results showing the generalizability of the original IAM to conditions involving phonological processing are clearly in favor of our view.

To facilitate theoretical integration in the "word-nerd" field, the following challenge has to be met: Agree on a minimal set of standards for model comparison and evaluation. This challenge has two facets. First, it involves agreement on a minimal set of standard effects and tasks, which any model of word recognition that competes for "harbingership" should be able to predict in a way that can enter into strong inference competition. In Jacobs and Grainger (1994, table 1), we made a minimalistic proposal of four such effects. Whether they were the right ones or whether they have to be augmented by other effects is a question that can be solved only by ongoing published scientific debate.

The second and more problematic facet of this challenge is to agree on a minimal set of criteria for model evaluation and on a standard way of

applying the criteria. Among other things, this problem implies the tricky question of complex A-type models being made comparable, in at least several critical features or dimensions (e.g., the "currency" problem discussed by Massaro & Friedman, 1990). This issue is nontrivial (cf. Estes, 1975), as model builders who have problems of keeping different variants of their own models comparable over the years of development admit (for some examples, see Jacobs & Grainger, 1994). Our stratagem of nested modeling might appear to lead us close to what Newell (1990) called the "Popperian damnation," the risk of killing theories via falsificationism before we can know their true potential. Apart from the fact that Newell's interpretation of Popper's ideas is naive and uninformed, nested modeling is a valuable method to tackle this basic problem: By facilitating unification of a single model or family of models, nested modeling should also facilitate unification of competing models. Even Newell (1990)—not to be suspected of being a Popperian—urged cognitive theoreticians to start by unifying their own theories before attacking a grand unified theory of cognition, which according to our view, can result only from strong inference competition in a pluralistic, pragmatic, but principle-oriented modeling perspective. Nested modeling can considerably contribute to that unification.

We think that the establishment of a set of standard criteria for model comparison and evaluation serves this aim of theoretical integration. Standard criteria need not at all exclude parallel anarcho-creative modeling enterprises à la Feyerabend (1975). Why should we not give the venture a try?⁹

APPENDIX

Cleaning procedure for all three-, four-, and five-letter English words taken from the Celex database (Baayen, Piepenbrock, & van Rijn, 1993):

All three-, four-, and five-letter English words were extracted from the Celex database.

All words with a written (Fwrit) *and* a spoken (Fspok) frequency smaller than or equal to one per million were excluded.

Abbreviations, proper names, and so on were excluded.

Homograph entries were reduced according to the following procedure: Homophonic homographs (e.g., verb/noun: *to beat, the beat*) were

⁹It is indeed a nontrivial problem to lay out an epistemological position that tries to benefit from the creative tension between the charming, anarchic, and droll island occupied by Feyerabend and the rigorous, critical-rational, but also liberal-Socratic, castle constructed by Popper.

reduced to a single entry in the orthographic and phonologic lexicons. F_{writ} and F_{spok} were summed across the multiple homograph entries. Nonhomophonic homographs (e.g., *lead, lead*) were kept as two separate entries in the phonological lexicon and one entry in the orthographic lexicon. Their written frequencies (F_{writ}) were summed; their spoken frequencies (F_{spok}) were kept separate (not summed). Nonhomophonic homograph alternatives with identical frequencies but slight variations in their phonology depending on regional or contextual constraints were reduced to one standard entry in both lexicons according to Harrap's *Dictionary*. (Those multiple entries are alternative pronunciations; they must be separated from real nonhomophonic homographs because they have identical frequencies.)

For all homophones (e.g., *sea, see*; $n = 444$), the spoken frequencies (F_{spok}) were summed.

All words with two or more syllables were excluded.

All words with a grapheme decomposition of two were excluded (e.g. *mayor*, with /m8R/ as phonological code, with *may-or* as graphemic decomposition).

This cleaning procedure resulted in an orthographic lexicon consisting of 2,494 entries (409 three-letter words, 1,151 four-letter words, and 933 five-letter words).

The readout procedure used in these simulation studies follows the design principles described by Grainger and Jacobs (1996). For the Coltheart et al. (1977) and Seidenberg et al. (1996) tests as well as the second Stone et al. (1997) test, the critical dependent variable was correct RT to nonwords. In the MROM, a No response is generated when neither the activation of a lexical unit (μ value) nor the global lexical activity (s value) has reached the response criteria (M and Σ , respectively) before the temporal deadline (T criterion). Thus, RT for a No response is given simply by the value of the T criterion. This value is a function of the global lexical activity (σ) generated by the target stimulus. A high global lexical activity is interpreted as a high probability that the target stimulus is a word. Practically, we assume that during the early phases of stimulus processing, the computed σ value indexes the likelihood that the stimulus is a word. A high σ value encourages participants to set a longer deadline, that is, a higher T criterion (Coltheart et al., 1977; Jacobs & Grainger, 1992). For the present simulations, we monitored the σ value at Processing Cycle 15 for each stimulus. In the MROM-p simulations, if $\sigma(15) > 0.33$, then $T = 27$ cycles, or else $T = 25$ cycles. For the MROM simulations, if $\sigma(15) > 0.3$, then $T = 27$ cycles, or else $T = 25$ cycles.

For the first Stone et al. (1997) test, the critical dependent variable was correct RT words. In the MROM, a Yes response is generated when either the activation of a lexical unit (μ value) or the global lexical activity (σ

value) has reached the response criteria (M and Σ , respectively) before the temporal deadline (T criterion). For simplicity, in the present simulations we did not consider the possible role of global orthographic activity on Yes responses. Thus, the RT was computed by determining when the activation of lexical units reached the Σ criterion. This criterion was 0.67 for the MROM-p and 0.47 for the MROM. These values correspond to 90% of the asymptotic activation values for lexical units in each model (Jacobs & Grainger, 1992).

ACKNOWLEDGMENTS

This research was supported by a grant from the Deutsche Forschungsgemeinschaft to A. Jacobs (Teilprojekt 7 der Forschergruppe "Dynamik kognitiver Repräsentationen"), a grant from the French Ministry of Education and Research (No. 95124, Programme sciences cognitives) to A. Rey, and a German Academic Exchange Service Grant (DAAD-Doktorandenstipendium aus Mitteln des zweiten Hochschulsonderprogramms) to J. Ziegler.

REFERENCES

- Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory and Cognition*, 10, 565-575.
- Arbib, M. A., & Kaplan, D. (1979). Neurolinguistics must be computational. *Behavioral and Brain Sciences*, 2, 449-483.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340-357.
- Berndt, R. S., D'Autrechy, C. L., & Reggia, J. S. (1994). Functional pronunciation units in English words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 977-991.
- Berndt, R. S., Reggia, J. S., & Mitchum, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments, and Computers*, 19, 1-9.
- Brown, G. D. A. (1987). Resolving inconsistency: A computational model of word naming. *Journal of Memory and Language*, 26, 1-23.
- Carreiras, M., Perea, M., & Grainger, J. (1997). Orthographic neighborhood effects on visual word recognition in Spanish: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23.
- Collyer, C. E. (1985). Comparing strong and weak models by fitting them to computer-generated data. *Perception and Psychophysics*, 38, 476-481.
- Collyer, C. E. (1986). Goodness-of-fit patterns in a computer cross-validation procedure comparing a linear and a threshold model. *Behavior Research Methods, Instruments, and Computers*, 18, 618-622.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). London: Academic Press.

- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route approaches and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.
- Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1197–1211.
- Dell, G. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124–142.
- Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology*, *49*, 264–271.
- Erman, L. D., & Lesser, U. R. (1980). The Hearsay-II speech understanding system: A tutorial. In W. A. Lea (Ed.), *Trends in speech recognition* (pp. 361–381). Englewood Cliffs, NJ: Prentice Hall.
- Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology*, *12*, 263–282.
- Feldman-Stewart, D., & Mewhort, D. J. K. (1994). Learning in small connectionist networks does not generalize to large networks. *Psychological Research*, *56*, 99–103.
- Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked nonword priming. *Quarterly Journal of Experimental Psychology*, *33A*, 325–350.
- Ferrand, L., & Grainger, J. (1993). The time-course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society*, *31*, 119–122.
- Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *Quarterly Journal of Experimental Psychology*, *47A*, 365–382.
- Ferrand, L., & Grainger, J. (1996). List context effects on masked phonological priming in the lexical decision task. *Psychonomic Bulletin and Review*, *3*, 515–519.
- Feyerabend, P. (1975). *Against method*. London: Verso.
- Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. W. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.
- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 104–115.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *4*, 506–528.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*, 228–244.
- Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In D. Harris (Ed.), *Cognitive processing in bilinguals*. Amsterdam: North Holland.
- Grainger, J., & Jacobs, A. M. (1993). Masked partial-word priming in visual word recognition: Effects of positional letter frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 951–964.
- Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1158–1176.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.
- Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception and Psychophysics*, *45*, 189–195.

- Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception and Psychophysics*, *51*, 49–56.
- Greenwald, A. G., & Pratkanis, A. R. (1988). On the use of "theory" and the usefulness of theory. *Psychological Review*, *95*, 575–579.
- Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback expectation, olfaction, and illusions. *Biological Cybernetics*, *23*, 187–202.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, *87*, 1–51.
- Grossberg, S., & Stone, G. O. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, *93*, 46–74.
- Hagège, C. (1986). *L'Homme de paroles* [Speaking man]. Paris: Arthème-Fayard.
- Humphreys, M. S., Wiles, J., & Dennis, S. (1994). Toward a theory of human memory: Data structures and access processes. *Behavioral and Brain Sciences*, *17*, 655–692.
- Jacobs, A. M. (1994). On computational theories and multilevel, multitask models of cognition: The case of word recognition. *Behavioral and Brain Sciences*, *17*, 670–672.
- Jacobs, A. M. (1995). *Zur Rolle phonologischer Prozesse beim Lesen* [On the role of phonological processes in reading] (Antrag zum DFG Projekt No. 7 der Forschergruppe "Dynamik kognitiver Repräsentationen"). Marburg, Germany: Philipps-Universität.
- Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1174–1188.
- Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1311–1334.
- Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language*, *24*, 165–178.
- Jared, D. (1997). Spelling–sound consistency affects naming of high-frequency words. *Journal of Memory and Language*, *36*, 505–529.
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687–715.
- Korte, W. (1923). Über die Gestaltauffassung im indirekten Sehen [On Gestalt-perception in peripheral vision]. *Zeitschrift für Psychologie*, *93*, 17–82.
- Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge*. New York: Cambridge University Press.
- Lass, U. (1995). *Einfluß von Sprache auf Verarbeitungsprozesse im Kurzzeitgedächtnis* [Influence of language on processing in short-term memory]. Unpublished *Habilitationsschrift*, Universität Göttingen, Germany.
- Levin, J. A. (1976). *Proteus: An activation framework for cognitive process models* (Tech. Rep. No. ISI/WP 2). Marina del Rey: University of Southern California, Information Sciences Institute.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.
- Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189–256.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.
- Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language*, *27*, 213–234.
- Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1107–1128.

- Massaro, D. W., & Cowan, N. (1993). Information processing models: Microscopes of the mind. *Annual Review of Psychology*, 44, 383–425.
- Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, 97, 225–252.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part I: An account of basic findings. *Psychological Review*, 88, 375–407.
- McClelland, J. L., & Rumelhart, D. E. (1988). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 3). Cambridge, MA: Bradford Books.
- Miller, G. A. (1993). *Wörter: Streifzüge durch die Psycholinguistik* [The science of words]. Frankfurt, Germany: Zweitausendeins.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Norris, D. G. (1994). A quantitative model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1212–1232.
- Paap, K., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review*, 89, 573–594.
- Peressotti, F., & Grainger, J. (1995). Letter-position coding in random consonant arrays. *Perception and Psychophysics*, 57, 875–890.
- Pinker, S. (1994). *The language instinct*. Harmondsworth, England: Allen Lane.
- Popper, K. R. (1935). *Logik der Forschung* [The logic of scientific discovery]. Vienna: Springer.
- Popper, K. R. (1966). *Conjectures and refutations*. New York: Harper & Row.
- Popper, K. R. (1972). *Objective knowledge*. Oxford, England: Clarendon Press.
- Popper, K. R. (1984). *Auf der Suche nach einer besseren Welt* [Searching for a better world]. Munich, Germany: Piper.
- Prechelt, L. (1996). A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice. *Neural Networks*, 9, 457–462.
- Pugh, K. R., Rexer, K., & Katz, L. (1994). Evidence for flexible coding in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 807–825.
- Pylyshyn, Z. W. (1989). Computing in cognitive science. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 49–91). Cambridge, MA: MIT Press.
- Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention & performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611–653). Cambridge, MA: MIT Press.
- Rosson, M. B. (1985). The interaction of pronunciation rules and lexical representations in reading aloud. *Memory and Cognition*, 13, 90–99.
- Rouanet, H. (1996). Bayesian methods for assessing importance of effects. *Psychological Bulletin*, 119, 149–158.
- Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 10, 647–657.
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI* (pp. 573–603). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception. Part II: The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60–94.
- Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 876–900.

- Seidenberg, M. S., Petersen, A. S., McDonald, K., & Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 48–62.
- Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review*, *1*, 2–28.
- Simon, H. A., & Kaplan, C. A. (1989). Foundations of cognitive science. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 1–47). Cambridge, MA: MIT Press.
- Spielberg, N., & Anderson, B. D. (1985). *Seven ideas that shook the universe*. New York: Wiley.
- Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, *36*, 337–359.
- Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 744–774.
- Stone, G. O., & Van Orden, G. C. (1994). Building a resonance framework for word recognition using design and system principles. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1248–1268.
- Tack, W. H. (1976). *Stochastische Lernmodelle* [Stochastic models of learning]. Berlin, Germany: Kohlhammer.
- Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.
- Van Orden, G. C., Aitchison, C. S., & Podgornik, M. N. (1996). *When a ROWS is not a ROSE: Null effects in the absence of cognitive structures*. Manuscript submitted for publication.
- Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1269–1291.
- Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 371–385.
- Venezky, R. L., & Massaro, D. W. (1987). Orthographic structure and spelling–sound regularity in reading English words. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Shared mechanism in listening, speaking, reading and writing* (pp. 158–179). London: Academic Press.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skill. *Psychological Bulletin*, *101*, 192–212.
- Wiese, R. (1996). *The phonology of German*. New York: Oxford University Press.
- Ziegler, J. C. (1996). *La Lecture en tant que système de résonance: Une Approche expérimentale et computationnelle de l'activation et de la rétroaction phonologique* [Reading as a system of resonance: An experimental computational approach of phonological activation and retroaction]. Unpublished doctoral dissertation, Université de Provence, Marseilles, France.
- Ziegler, J. C., & Ferrand, L. (1997). *Orthography shapes the perception of speech*. Manuscript submitted for publication.
- Ziegler, J. C., & Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language*, *34*, 567–593.
- Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bi-directional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, and Computers*, *28*, 504–515.
- Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language*, *37*, 533–554.

- Ziegler, J. C., Rey, A., & Jacobs, A. M. (in press). Simulating individual word identification thresholds and errors in the fragmentation task. *Memory and Cognition*.
- Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for _OUGH and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, and Computers*.
- Ziegler, J. C., Van Orden, G. C., & Jacobs, A. M. (1997). Phonology can help or hurt the perception of print. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 845-860.