Reading

John G. Holden
Psychology Department
California State University, Northridge

and

Guy C. Van Orden
Cognitive Systems Group
Arizona State University

Running Head: Recurrent Networks and Word Recognition

John G. Holden
Psychology Department
California State University, Northridge
18111 Nordhoff Street
Northridge, CA 91330-8255
jay.holden@csun.edu

To appear in The Handbook of Brain Theory and Neural Networks, Second edition, (M.A. Arbib, Ed.), Cambridge, MA: The MIT Press, 2002.
http://mitpress.mit.edu

Introduction

A skilled reader can recognize many thousands of printed words, each in a fraction of a second, with no noticeable effort. A child who is a developmental dyslexic does not easily acquire this skill. For a dyslexic child, recognizing a printed word, as a particular word, can be effortful to the point of frustration. Dyslexia may plague an otherwise bright and articulate child, and the fact of dyslexia illustrates how recognition of printed words, as words, is the crux of reading. Reading is not exclusively the recognition of printed words, but it is word recognition that sets reading apart from natural language. In effect, to become a reader is to master this special skill. A vast empirical literature concerns word recognition, and most "neural" networks of reading are models of word recognition.

One tool with which to diagnose dyslexia is a naming task, that presents individual pseudoword spellings, such as glurp, to be read aloud. A dyslexic child may have great difficulty with this task, and success in this nonword task is generally correlated with word reading skill . Trouble decoding the pronunciations of letter strings is a symptom of the most common form of dyslexia. In this form, dyslexia is a specific problem in translating words' spellings into their phonology (roughly, their sounds or pronunciations). And the key to word recognition would seem to lie in the derivation of phonology from words' printed forms.

All written languages have systematic relations between words' printed and spoken forms, so perhaps word recognition includes an analytic letter-by-letter process, that translates spelling into phonology. This possibility has pre-occupied reading scientists for over 100 years. Nevertheless, word recognition is not simply analytic. Evidence that supports an analytic hypothesis has always existed side-by-side with evidence that word recognition is synthetic (or holistic). The consequent synthetic-analytic debate defined reading theory throughout the 20th century, and it provides the organizing theme for this article.

Early Reading Research

Nineteenth century studies introduced almost all topics of current reading research. A key early finding, from eye-movement studies, was that readers' eyes make a series of jumps, in moving across a line of text, pausing about a quarter of a second when fixated. Discrete eye movements implied the fixation (recognition) of individual printed words in natural reading. Another key development was the invention of the t-scope (or tachistoscope). A t-scope can flash individual words, a few milliseconds at a time, well within the range of fixation times observed in eye movements.

More recent t-scope studies, in the second half of the 20th century, perpetuated the synthetic-analytic debate. Ulric Neisser's seminal book Cognitive Psychology includes a review of this debate concerning word recognition. For example, pseudoword spellings such as glurp, that obey the letter-to-sound pronounceability patterns of English, are more easily recognized and recalled than random strings of letters. The advantage for prounceable letter-strings suggests an analytic process. A more contemporary finding, however, indicated that word recognition could be synthetic. Words flashed for a fraction of a second (and then replaced by a "pattern mask" of letter features) are more accurately reported than their component letters, presented individually under the same extreme conditions, a phenomenon known as the word superiority effect.

Dual Route Theories

Dual route theories emerged in the 1970's with the advent of cognitive psychology. As the name implies, traditional dual route theories were an ad hoc resolution of the synthetic-analytic debate. Both options were included as separate processing modules (mechanisms). Early dual route theories were important, in this regard, because they moved past a contentious theoretical debate that could not be resolved empirically.

The two modules of dual route theories accomplish word recognition in two different ways. Skilled readers, reading frequently encountered words, rely on a fast, synthetic, lexical module to translate a visual representation into an entry in the lexicon, the mind's dictionary. The speed of access to dictionary entries, in the lexical module, is determined by the relative frequency with which a word appears in print. Word frequency is estimated by counting the occurrence of each word (per million words) in large samples of text. Higher frequency words are more readily available in the lexicon. (Different dual route theories proposed different frequency sensitive mechanisms.)

By contrast, novice readers, and skilled readers who encounter a novel word, recognize words via a slow, analytic, sublexical module. At the heart of the sublexical module are rules to translate minimum units of spelling (graphemes) into minimum units of phonology (phonemes). A combination of phonemes may guide assembly of pronunciation for completely novel letter-strings, such as glurp, or may achieve lexical access when the unfamiliar spelling translates into the phonology of a familiar word. Access to a lexical entry allows access, in turn, to lexical representations of words' pronunciations and conventional meanings, as one would find in a dictionary. (Different dual route theories proposed different representations and translation procedures.)

Word Naming

Word naming experiments measure the time required to pronounce individually presented words. At one time, dual process theories provided a sufficient account of results from naming experiments, with skilled readers as participants. Low-frequency regular words, such as mint, that obey sublexical grapheme-phoneme-correspondence rules, are named faster than low-frequency exception words, such as pint that entail exceptions to the rules. No regularity effect was found to high-frequency words. Hence naming of low-frequency words is accomplished by the sublexical module, but naming of all other words is accomplished by the lexical module. Eventually, however, this categorical regular-exception distinction was contradicted. New studies found graded effects of semi-regular relations between spelling and phonology, not simply the regular versus exception distinction predicted by traditional dual route theories. For instance, although both wave and wade obey the grapheme-phoneme rules, the existence of have, an exception “neighbor” to wave, induces slower naming times to wave itself.

Lexical Decision

Word recognition is also studied using the lexical decision task. Lexical decision experiments measure the time required to indicate that an individually presented word is a word (with catch trials that present pseudoword spellings). Previously, lexical decision time did not appear to be affected by regularity, only by relative frequency. High-frequency words are recognized faster than low-frequency words. Regularity effects arise in the sublexical module, and frequency effects arise in the lexical module. Hence recognition, for lexical decisions, appeared to rely on the lexical module, exclusively. New studies contradicted this hypothesis, however. Key findings were subcategories of exception words, such as weird or choir, called strange words, that produced slow and error prone performance, reliably worse performance than to garden variety exception words, such as pint. These findings, like the graded-regularity effects in naming experiments, contradict the categorical regularity distinction of dual route theories.

Patient Studies

At one time, dual process theories also provided a reasonable account of neuropsychological findings. Lexical and sublexical modules were corroborated in the patterns of naming errors, described in case studies of brain damaged individuals. For example, surface dyslexics incorrectly regularize exception words (pint pronounced to rhyme with mint) but correctly name regular words. Regularized pronunciations of exception words dissociate the sublexical module (the source of regularization errors) from the damaged, or absent, lexical module (the source of correct pronunciations). Alternatively, deep dyslexics produce visual errors (bush pronounced as brush) and semantic errors (bush pronounced as tree). These errors were attributed to a dissociated, but damaged, lexical module.

Evidence from patient studies proved to be problematic, however. No general agreement emerged concerning which behavioral deficits actually count as dissociated components of word recognition. All of the patient case studies that concerned reading were challenged by competing theorists, who claimed they did not actually dissociate synthetic versus analytic components, or that they simply did not pertain to reading. More recent brain imaging studies appear to have arrived at a similar impasse. No general agreement has emerged in the brain imaging literature that uniformly implicates specific brain regions in a large sample of reading tasks. Indeed, small differences in reading tasks, and experimental methods, appear to implicate different brain regions in what appear, intuitively, as very similar reading acts.

Additive Factors Method

As noted, dual process theories emerged in the 1970's, when cognitive psychology was predominantly concerned with information processing. Within that framework, the mind was conceived as an information processing device, that could be described much like a flowchart of information processing in a computer program. Simon

(1973) described how complex systems, such as cognitive systems, could be partly decomposed, if the system's components interact approximately linearly. If interactions (exchanges) between components are linear in their effect, then the components can be identified, even if their internal dynamics are nonlinear. Cognitive components, thus described, work as a chain of single causes—a metaphorical extension of domino causality. Push the first domino in a chain of standing dominoes and each will fall in its turn.

The previous theoretical rationale coincided with a methodological tool to individuate cognitive components: the additive factors method, proposed by Sternberg (1969). Factorial experiments allow simultaneous manipulations of candidate variables that may influence distinct hypothetical components. The main effects of two or more manipulations are additive if they add up to the total behavioral effect. In this idealization, separate experimental manipulations selectively influence (e.g., slow) the falling times of individual "dominoes," but do not change the falling times of dominoes that precede or follow an affected domino. Alternatively, when non-additive interaction effects are observed, manipulations do not satisfy the assumption of selective influence, and influence (at least) one common component. Thus, to Sternberg's lasting credit, his method included an empirical failure point: ubiquitous non-additive interaction effects.

Additive main effects are rarely observed in reading experiments. It is not possible to manipulate all factors simultaneously, in one experiment, but it is possible to trace chains of non-additive interactions across published experiments that preclude the assignment of any factors to distinct components. For example, factorial manipulations of word frequency and regularity yield non-additive interaction effects. This raises the question of whether the respective effects actually arise from separate synthetic and analytic processes.

Parallel Distributed Processing (PDP) Models

PDP models allowed a new position in the debate; they did not require distinct synthetic and analytic processes. PDP models are connectionist models in which constraints (connection weights) determine the activation values of nodes. Nodes represent words' spellings, pronunciations, or meanings and patterns of response times are simulated in a model's "error term" (the difference between a model's pronunciation, for example, and an ideal correct pattern of pronunciation-node activation). A learning algorithm shapes a matrix of connection weights to reflect statistical relations among node representations. This is referred to as statistical or covariant learning. PDP models introduced covariant learning algorithms to a broad audience of cognitive scientists. PDP models were equally important as existence proofs, which advanced the synthetic-analytic question. Covariant learning may reflect, in a single process, both subword ("analytic") and whole-word ("synthetic") covariation, as we illustrate next.

Distinctions in the relations among English spellings and pronunciations, at a variety of scales, may all be construed as statistical relations. Covariant learning tracks all scales of covariation, simultaneously, in the connection weights of a PDP model. For example, consonant spellings and pronunciations are more reliably correlated than vowel spellings and pronunciations, and in both cases there are statistically dominant and subordinate relations. “Regular” words, comprised of dominant relations, are named more quickly than “exception” words that include subordinate relations. Likewise, spelling bodies (e.g., the spelling pattern _uck, in duck) and pronunciation rimes (e.g., pronunciation /_ uk/, in /duk/) are invariantly correlated; but some other body-rime relations are dominant though less strongly correlated (_int pronounced as in mint); and still other body-rimes are subordinate and only weakly correlated (_int pronounced as in pint). This rank order is corroborated in naming times, words like duck are named faster than words like mint, and words like mint are named faster than words like pint (all other things equal). Finally, a word’s relative frequency estimates the strength of the correlation between the word’s (whole-word) spelling and pronunciation, and high-frequency words are named faster than low-frequency words.

As the examples illustrate, the outcome of covariant learning will be determined by the pattern of statistical relations in a training set—the sample of words used to train the model. Each training set entails a sample of constraints (relations between spelling, phonology, and meaning) from the body of constraints in a literate culture, and covariant learning attunes the network to the sampled constraints. Thus, implicit in the training set is a description of the cultural artifact—a particular language’s pattern of relations—which is crucial for cognitive theory. Jared (1997) used this implication of PDP models to derive a nonintuitive empirical test. Careful attention to statistical relations among the spellings and pronunciations of high-frequency words, predicted a statistical advantage for high-frequency words with invariant body-rime relations. Jared subsequently corroborated this prediction—a previously unobserved “regularity” effect in naming for high-frequency words. However, no such effect was observed in a lexical decision experiment.

Hybrid, partially recurrent, connectionist models moved the PDP approach further in the direction of fully recurrent, iterative “neural,” network models. In a hybrid model, the output of nonrecurrent (strictly feedforward) portions of a PDP network sets the initial conditions in a recurrent subnetwork, that includes feedback connections. The recurrent portion behaves as an attractor network, tuned to fixed points that correspond to word pronunciations (for example) , and pronunciation times are simulated in the number of iterations before the network reaches a “stable” attractor.

Partly “damaged” hybrid networks simulated the bizarre semantic and visual errors of deep dyslexic patients as well as the regularization errors of surface dyslexics. Simulated lesions have been implemented in several ways , including (a) random cutting of some connections between nodes, (b) random changes in connection weights, and (c) random selection of nodes whose values are fixed at zero. The various patterns of dyslexic patient’s naming errors have been mimicked using one, or a combination, of these simulated lesions.

Iterative Network Models

Hybrid feed forward PDP models were actually proposed as a first step toward fully recurrent, iterative networks. Iterative networks are attractor networks simulated as nonlinear iterative maps. A nonlinear iterative map may approximate solutions of a system of nonlinear differential equations. Thus, iterative network models, as dynamical systems, invoke the most sophisticated mathematical framework available to scientists. An iterative map takes its output at one time step as input on the next time step, until a stable pattern of node activity emerges—an attractor state. A stable attractor state corresponds to an iterative model's naming response.

Iterative network models may include covariant learning algorithms that reflect relations that map from patterns of spelling to pronunciations, and from pronunciations to spelling patterns (and meanings). Invariant, bi-directional relations correspond to stable attractors in the network, which extends the previous view of statistical relations among spellings and pronunciations. Some consonants have invariant bi-directional relations with their pronunciations. For example, the grapheme B, at the beginning of a word, is always pronounced /b/, and the /b/ pronunciation is always spelled B. Likewise, some spelling-body pronunciation-rime relations are invariant (e.g., _uck and /_uk/, as in duck, always co-occur). And most words have a bi-directional invariant relation between their particular whole-word spelling and their particular pronunciation. As noted, invariant bi-directional relations correspond to stable attractor states in an iterative network. By contrast, ambiguous spelling-pronunciation relations correspond to multiple, mutually inconsistent multistable, or more precisely metastable, attractor states in an iterative network.

Empirical studies of nonlinear systems typically focus on their less stable behavioral regimes, because very stable regimes reveal less of system dynamics. Pioneering studies have examined how ambiguous relations between spelling and pronunciation affect empirical patterns in naming performance. By definition, the relation between spelling and phonology is ambiguous if more than one reliable pronunciation is elicited by the same spelling. For example, a homograph, such as wind, has two legitimate pronunciations, and is thus an ambiguous spelling. In the case of a homograph, such as wind, its "regular" pronunciation (to rhyme with pinned) is produced faster than its "exception" pronunciation (to rhyme with find). Thus, the dynamics of word naming must unfold in way that respects this ordering.

Kawamoto and Zemblidge (1992) simulated homograph naming using an iterative network. Relations (connections) among letter, phoneme, and semantic node families, were recurrent (including both feedforward and feedback connections) and excitatory, but within each node family, recurrent connections were (predominantly) inhibitory. The multistable unfolding of homograph pronunciations was simulated as a transcritical bifurcation. In a transcritical bifurcation, for example, wind's two possible "pronunciations" exchange stability at a bifurcation point. That is, initial dynamics unfold in favor of one potential solution, but over successive iterations additional constraints (that may unfold on a slower time scale) begin to favor an alternative solution. In the case of wind, the faster "regular" pronunciation reflected a strong local attractor between letter and phoneme nodes. However, coherent interactive-activation among phoneme and semantic nodes slowly emerged to favor the "exception" pronunciation. The bifurcation point occurred when the balance of constraints switched to favor the "exception" pronunciation. At the bifurcation point, the "regular" pronunciation (attractor) exchanged stability with the "exception" pronunciation.

Discussion

The previous simulation of ambiguous homograph pronunciations, as bifurcation phenomena, illustrates how nonlinear dynamical systems theory has been applied to reading performance. However, empirical methods appropriate to nonlinear analysis are not widely applied. In large part, connectionism has inherited the empirical methods of information processing psychology. However, statistical analyses that assume the general linear model, and theories implemented as strongly nonlinear dynamical systems, are incompatible at their root. They entail contradictory notions of cause and effect.

Circular Causality

Previously we discussed how the factorial logic of additive factors method assumes a one-way, domino notion of causality. In contrast, bifurcation phenomena entail circular causality. Circular causality requires a strategic (not morphological) reduction of system behavior, due to emergent properties. In a strategic reduction, generic emergent properties may be found at multiple levels of systems, but emergent properties at "higher" levels do not reduce to causal properties of "lower" levels.

Plausible nonlinear models allow that it may not be productive, for scientific purposes, to view word recognition (or reading) as a component process, but linear methods are directed at the discovery of component processes. Moreover, the results of a vast linear analysis actually raise the question of whether a distinct process of word recognition may be distinguished. All reading tasks would seem to include word recognition, but they do not yield any converging pattern of word recognition effects.

Consider the word frequency effect in the lexical decision task, for example. The same set of words that produce a large word frequency effect in the lexical decision task may produce a reduced, or statistically unreliable, word frequency effect in naming (or other tasks). Within the lexical decision task, itself, it is possible to modulate the word frequency effect by making the nonwords more or less word-like (and, in turn, modulate nonadditive interaction effects among word frequency, regularity, and other variables). Across languages, Hebrew produces a larger word frequency (familiarity) effect than English, and English produces a larger effect than Serbo-Croatian (which tracks the analytic transparency of their print-to-sound relations—less transparent equals larger frequency effect) .

All empirical phenomena of word recognition appear to be conditioned by task, task demands, and reference language. These non-additive interactions allow the question of whether "word recognition effects" may be attributed to a distinct process of word recognition. Consider the previous examples together, within the guidelines of additive factors logic. Word recognition factors cannot be individuated from each other, and they cannot be individuated from the context of their occurrence (task, task demands, and language). Because additivity of task effects, or language effects, is never observed, we lack evidence that may individuate word recognition.

Despite these problems, most theorists, including connectionist theorists, share the intuition that a distinct separable component of word recognition may yet exist. We speculate, however, that the intuition persists because most cognitive theorists were trained exclusively in linear methods. If we are correct, then rigorous tests of iterative network models await a reliable logic of nonlinear analysis, consistent with nonlinear dynamical systems theory, and appropriate to reading performance. Thus the historical question of analytic versus synthetic processes with which we began is replaced by the question of linear versus nonlinear dynamics—a question motivated, in part, by the success of nonlinear iterative network models.

References

Farmer, J. D. 1990. A Rosetta Stone for connectionism, Phys D, 42:153-187.

*Frost, R., and L. Katz, eds. 1992. Orthography, Phonology, Morphology, and Meaning. Amsterdam: North Holland.

Jared, D. 1997. Spelling-sound consistency affects the naming of high-frequency words, J Mem Lang, 36:505-529.

Kawamoto, A. H., and J. H. Zemblidge, 1992. Pronunciation of homographs, J Mem Lang, 31:394-374.

Mattingly, I. G. 1992. Linguistic awareness and orthographic form. In R. Frost and L. Katz (Eds.), Orthography, phonology, morphology, and meaning (pp. 11-26). Amsterdam: North-Holland.

*Neisser, U. 1967. Cognitive Psychology. Englewood Cliffs, NJ: Prentice-Hall.

*Perfetti, C. A. 1985. Reading Ability. New York: Oxford University Press.

*Pennington, B. F. 1991. Diagnosing Learning Disorders: A Neuropsychological Framework. New York: Guilford Press.

Plaut, D. C., J. L. McClelland, M. S. Seidenberg, and K. Patterson, 1996. Understanding normal and impaired word reading: Computational principles in quasi-regular domains, Psychol R, 103:56-115.

*Lukatela, G., and M. T. Turvey, 1998. Reading in two alphabets, Am Psychol, 53:1057-1072.

*Rayner, K., and A. Pollatsek, 1989. The Psychology of Reading. Englewood Cliffs, NJ: Prentice Hall.

Seidenberg, M. S. (1995). Visual word recognition: An overview. In J. L. Miller & P. D. Eimas (Eds.) Speech, Language, and Communication (pp. 137-179). New York: Academic Press.

Simon, H. A. 1973. The organization of complex systems, in The Challenge of Complex Systems (H. H. Pattee, ed.), New York: George Braziller, pp. 1-27.

Sternberg, S. 1969. The discovery of processing stages: Extensions of Donders' method, Acta Psychol, 30:276-315.

Van Orden, G. C., B. F. Pennington, and G. O. Stone, 2001. What do double dissociations prove?, C Sci, 25:111-172.