Psychological Review
1982, Vol. 89, No. 5, 483–506

0033-295X/82/8905-0483$00.75

The Problems of Flexibility, Fluency, and Speed–Accuracy Trade-Off in Skilled Behavior

Donald G. MacKay
University of California, Los Angeles

With practice, behavior sequences become more fluent (faster, less prone to error). The present article reviews existing theories of practice and proposes a new theory that better accounts for how people become more fluent in high-proficiency skills such as speech production. Under the theory, execution of behavior involves the activation of a hierarchy of nodes in proper serial order within an output system. Activating a node at any level in the system primes or partially activates its connected nodes, and practice or repeated activation increases the rate of priming per unit time, thereby allowing a faster rate of output at the lowest, muscle movement level. Relevance of the theory for several related issues is discussed: why behavior becomes more flexible with practice, transferring readily from one response mechanism to another; why there is almost perfect transfer from one hand to the other for simple skills such as Morse key tapping or moving chess pieces, but less than perfect transfer for complex skills such as handwriting with the unaccustomed hand; why skills at higher, "semantic" levels transfer to new behavioral sequences, as when bilinguals produce a word-for-word translation of a practiced sentence in their other language. The theory also provides a new way of looking at motor equivalence, automaticity, speed–accuracy trade-off, subordinate autonomy, and the motor program.

The present paper develops a theory to explain the relation between two seemingly disparate problems. One is the problem of fluency: Why do behavior sequences become more fluent (faster, less prone to error) as a function of practice? This long-standing and as yet unanswered question is the core issue in a network of interrelated and unsolved issues. Why are higher degrees of fluency associated with greater automaticity (reduced effort and conscious awareness)? What accounts for the trade-off relation between the two main components of fluency (speed and accuracy) so that we can generate behavior more rapidly only at the expense of a higher probability of error, whether low-level variability in muscle movements or interchanges of high-level components, as in speech errors (MacKay, 1971).

The other seemingly unrelated issue is the problem of flexibility, which itself consists of two components. One is the problem of transfer: When and why does skill acquired in practicing one performance transfer to other performances? The same basic question arises from our ability to generate different actions that are functionally equivalent at some level (variously labeled motor equivalence, response generalization, and action constancy). For example, movements of approximately the same form and function are generated using completely different muscle systems when we sign our name with the left rather than right hand or on a blackboard (with the muscles of the shoulder) rather than at a desk (with the muscles of the forearm) or with a pen clasped in the toes or teeth rather than the fingers (Merton, 1972).

Transfer of skill is a practical issue for education and a fundamental issue for psychology, but despite the large number of empirical studies on the question, little is known about why transfer either occurs or fails to occur (see Ellis, 1965). This gap in our un-

This research was supported in part by National Institute of Mental Health Grant 19964-08 and University of California, Los Angeles Grant 2428 to D. G. MacKay, who thanks Robert Bowman for assistance in running the experiments and Steve Keele for helpful comments on earlier versions of the manuscript.

Requests for reprints should be sent to Donald G. MacKay, Department of Psychology, University of California, Los Angeles, California 90024.

derstanding has had several unfortunate consequences. One is a lack of consensus on general conceptions of skill such as the notion of a motor program. By way of illustration, consider the following common assumptions concerning the nature of motor programs.

The muscle-specific command assumption: Under this assumption, motor programs consist entirely of muscle-specific motor commands or instructions to the muscle movement system. This assumption has difficulty with the fact that skill often transfers to novel muscle groups or movement situations, and this difficulty has been taken to indicate that motor learning does not involve the establishment of a specific set of motor commands (Schmidt, 1975).

The single-level assumption: Under the single-level assumption, motor programs involve only a single level of representation. This single-level assumption has difficulty explaining either complete or partial transfer (discussed below).

The abstract-program assumption: It is often assumed that motor programs are independent of particular muscles and are, therefore, more abstract than the muscles they guide (Keele, 1981). This assumption is consistent with the widely held view that response representation in long-term memory codes environmental goals rather than anatomic muscular activities required to accomplish these goals.

However, the abstract-program assumption has difficulty with data indicating that the motor memory code is both environmental and anatomic (see Klapp, Greim, Mendicino, & Koenig, 1979), and with the fact that transfer from one response mechanism to another is sometimes less than perfect. Consider bilateral transfer, for example. Simple motor skills, such as key tapping, transfer with little decrement from one hand to the other, but the decrement is greater for complex skills such as handwriting. If an action sequence such as signing one's name is represented abstractly rather than as a set of muscle-specific commands, what prevents perfect transfer in writing with the unaccustomed hand? And why is transfer sometimes atrocious, as in reversing the roles of the hands in guitar playing (Keele, 1981)?

The feedback-free assumption: Under the feedback-free assumption, motor programs are run off free from the influence of sensory input or feedback. This assumption has difficulty with our ability to adapt flexibly our preplanned behavior to changing circumstances. In walking, for example, if the leg touches an obstacle during its swing, the movement is automatically altered while still in progress (see Gallistel, 1980). This stumble-preventing reflex rules out a motor program for walking under the feedback-free assumption; but if walking is not controlled by program, then what is?

Adaptability is the second component of the problem of flexibility and concerns our ability to adapt or substitute components of an ongoing behavior sequence during its course of execution. The stumble-preventing reflex is one example. Another is the ability of a concert pianist to adjust a preplanned and rapidly executed sequence of notes so as to compensate for an error in the form or timing of an earlier note (from Shaffer, 1980). Even after walking and playing the piano have become fluent skills, they remain flexible and responsive to changing sensory input.

The relation between fluency and this second aspect of flexibility is a major problem for current theories. For example, according to Bindra (1978), flexibility corresponds to the segmentation of a behavior sequence into small, discrete components, which remain interruptible and changeable in the course of execution; whereas fluency corresponds to the unification of the sequence into one large underlying component, or motor program, triggered ballistically without the use of sensory feedback. Because fluency or unification can be achieved only at the expense of flexibility or segmentation, this theory predicts a negative correlation between fluency and flexibility. But, in fact, fluency and flexibility are positively correlated: One and the same action sequence becomes more fluent and more flexible as a function of practice (see, e.g., Shaffer, 1980).

The present article examines effects of practice on fluency in speech production and develops a new theory for explaining how fluency increases as a function of practice and why speed trades off with accuracy. The article then examines the problem of flexibility and extends the theory of fluency to explain both the positive correlation between fluency and flexibility and the phenomenon of high-level transfer of skill in speech production. The article concludes by applying the theory more generally to other skills and other aspects of flexibility in skilled behavior and to general conceptions of skill such as automaticity and the nature of motor programs.

The Problem of Fluency in Speech Production

Interest in speech production as a high-proficiency skill has increased dramatically over the past few years. The reason is that research into skills has shifted during the past decade toward studies of sequentially organized response components (see Holding, 1981), and speech production is a sequentially organized output system par excellence, requiring sequential organization of many different types of response components: phrases, words, syllables, and phonemes as well as muscle movements.

Until quite recently, however, research on speech production has proceeded in virtual isolation from other studies of skills, and it is instructive to understand why. One reason is that many researchers consider speech production an atypical skill and exclude it from consideration at the outset (see, e.g., Holding, 1981). This approach is counterproductive for general issues such as the problems of flexibility and fluency that apply to all human activities, including the supposedly "atypical" ones such as speech production as well as "typical" ones such as throwing the discus (after Holding, 1981).

A second reason is the belief, contradicted in the research described below, that speech production involves quite different research issues, explanatory concepts, and types of methodology from other skills. Besides being false, this belief seems detrimental to solving the problems of fluency and flexibility because speech production is the most extensively practiced, proficient, and flexible of human skills (see MacKay, 1981).

The third reason reflects a deeply rooted reaction to the scope of skills research. The domain of skills is essentially unbounded and might reasonably be said to include all human thought and behavior (see Bartlett, 1958). To help with the division of labor in a field of this scope, researchers have adopted what might be called the dichotomization strategy. Commonsense dichotomies having intuitive or practical appeal rather than demonstrable theoretical significance are used to segregate the field and create subfields with more manageable research literatures. As a consequence, the topic of skills has become splintered into large numbers of more or less separate literatures, speech production representing just one example (see Holding, 1981, for others).

Dichotomization has had unfortunate effects on the field. One is the proliferation of special theories for each class of skill, with little connection between them. Another is a concentration on the surface or muscle movement characteristics of different "types" of skills, at the expense of the underlying mental processes that are involved in the control of muscle movements. As a consequence, important generalizations applying to all skills have been missed because, as will be shown, the major effects of practice on the flexibility and fluency of skilled behavior are usually taking place at this mental level rather than at the muscle movement level.

Observations on Fluency in Speech Production

A. Newell and Rosenbloom (1981) argued for an empirical law of practice, which applies to all forms of behavior, including problem-solving skills as well as perceptual and perceptual–motor skills. Under this ubiquitous law, plotting the logarithm of the time to perform a task against the logarithm of trial number always yields a straight line (more or less) up to some asymptote.

To determine whether this empirical generalization holds for speech production, we reanalyzed some data from MacKay and Bowman (1969) and MacKay (1981). The subjects in the 1969 study were German–English bilinguals who were presented with sentences one at a time, and they simply produced each sentence as rapidly as possible. An example sentence is, "I have rearranged his bed fourteen times in morning." Half the

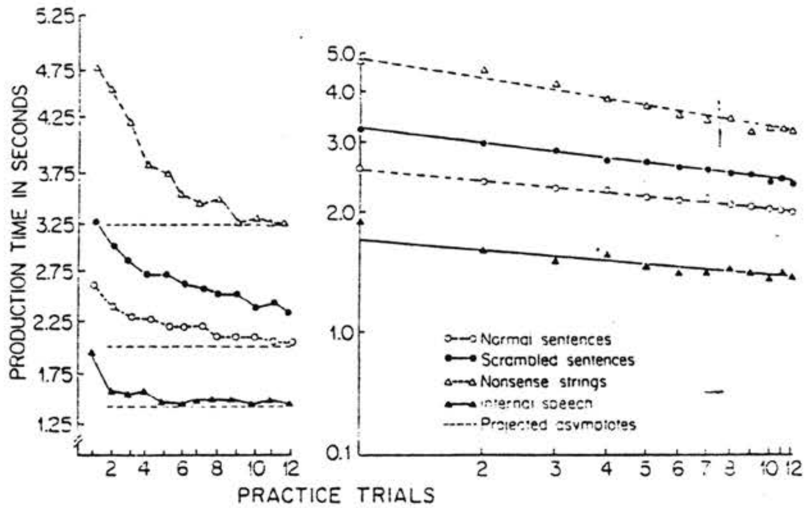

Figure 1. Practice effects for overt speech with three types of materials: normal sentences, scrambled sentences, and nonsense strings, and for internal speech with normal sentences (left panel). (The same data are plotted on log–log coordinates in the right panel.)

sentences were in German and the other half in English. Following a 20-sec pause, the procedure was repeated for a total of 12 repetitions of the same sentence. The dependent variable was the time to produce the sentence, and the independent variables were practice or trial number and types of materials. There were three different types of materials: normal sentences, such as the one above; scrambled sentences, for example, "Morning in fourteen his I bed times rearranged one have"; and nonsense strings, for example, "Morbing ni tourfeen hos i bep tiges reattanged ane hove." The scrambled sentences were derived from the normal sentences by rearranging the words, with the aim of eliminating syntactically based meaning for higher level constituents, for example, noun phrases such as "his bed." Nonsense strings were derived, in turn, from scrambled sentences by substituting or inverting letters in the words to make pronounceable nonwords such as *tourfeen.* The nonsense strings were to be produced as if they were normal sentences in English or, as was sometimes the case, German.

The average production time for 12 subjects is plotted on a log–log graph in Figure 1, and as can be seen there, the practice functions for all three conditions are linear or near linear. This implies that the function relating production time (T) to amount of practice (N) is a power function of the form $T = A\ N^X$.

The basic phenomena requiring explanation in Figure 1 are why the maximal rate of speech increased in this way and why the initial values and slopes of the power functions differed for the different types of materials. Even though the subjects were always attempting to speak at maximal rate, the average initial rate for the normal sentences (2.57 sec per sentence) was 25% faster than for the scrambled sentences, 85% faster than for the nonsense strings, and about 42% faster than normal or conversational rates for the same sentences. However, the rate of improvement as a function of practice was greatest for nonsense strings, less for the scrambled sentences, and least for the normal sentences.

Figure 1 also shows the data from MacKay (1981) for mental practice with normal sentences. The sentences and procedures for this experiment were identical to those above except that the subjects said the normal sentences silently to themselves without moving their lips and timed themselves by pressing one key as they began internal speech, and another as they finished.

As can be seen in Figure 1, there was a curious divergence from a power function on the initial trial of mental practice. Internal speech was also faster than overt speech (by about 35% on the initial trial), and this finding is consistent with the data of R. Anderson (in press) who reported a faster rate for internal speech when subjects recited the alphabet, days of the week, and months of the year as rapidly as possible, either silently or aloud.

The Theory

The theory described below was developed originally to explain how complex behaviors are sequenced and timed (see MacKay, Note 1) but has been extended here to show its relevance to the observations in Figure 1 as well as to other aspects of flexibility and fluency in skilled behavior.

According to the theory, the basic components for organizing complex actions are nodes, which are organized into at least two independently controllable systems: a mental and a muscle movement system. Nodes within the muscle movement system represent muscle-specific patterns of movement involving, in the case of speech production, articulatory organs such as the tongue, larynx, velum, and lips. In contrast, nodes within the mental system represent not specific muscles but rather cognitive units for controlling the movements making up an action sequence. For example, in producing a sentence, mental nodes represent units such as phrases, words, syllables, and phonemes, and these mental nodes and the connections between them (discussed below) constitute the mental representation of a sentence.

Hierarchical Organization of the Nodes

Each mental node represents a class of actions and is part of a syntactic domain or set of nodes serving the same syntactic function (discussed in detail below). For notational purposes the class of actions a mental node represents will be in italics, followed by its syntactic domain in brackets. For example, the mental node *practice* [noun] represents the concept practice and becomes activated whenever the noun *practice* is produced, whether whispered or voiced, unstressed or stressed (i.e., emphasized within the context of a sentence), or spoken with high or low pitch. These and other ways of expressing the concept "practice" constitute the class of actions the node *practice* [noun] represents.

This notational system can be used to describe how the nodes are organized. By way of illustration, consider the arbitrarily selected sentence "Frequent practice is helpful" and some of its constituent nodes shown in Figure 2. The highest level, [proposition] node represents the entire thought or idea underlying the sentence. This node, represented as *frequent practice be helpful* [proposition], is associated with two other nodes representing conceptual compounds: *frequent practice* [noun phrase] and *be helpful* [verb phrase]. These are connected with lexical concept nodes such as *frequent* [adjective] and *practice* [noun], which are connected with syllable nodes such as *prac* [stressed syllable] and *tice* [unstressed syllable]. Syllable nodes are connected with phonological compound nodes such as *pr* [initial consonant group] and *ac* [vowel group], which are connected with phoneme nodes such as *p* [initial consonant]. Phoneme nodes are connected with a set of phonetic feature nodes such as the one representing the frontal place of articulation of *p*. Finally, feature nodes are connected with a set of muscle movements nodes, including one for contracting the obicularis oral muscles of the lips for producing *p*.

Figure 2 also illustrates the concept of systems that, together with the concept of syntactic domain (discussed below), serves to distinguish the present theory from other hierarchical theories of speech production. The set of nodes making up a system is governed by the same timing mechanism and network of syntactic rules and can be activated independently from other node systems. According to Figure 2, the proposition nodes, conceptual compound nodes, and lexical concept nodes constitute a single system governed by timing and serial-order mechanisms that differ from those for the syllables within words and the phonemes and features within syllables. Syllable nodes, phonological compound nodes, phoneme nodes, and phonetic feature nodes constitute another independently controllable system having unique

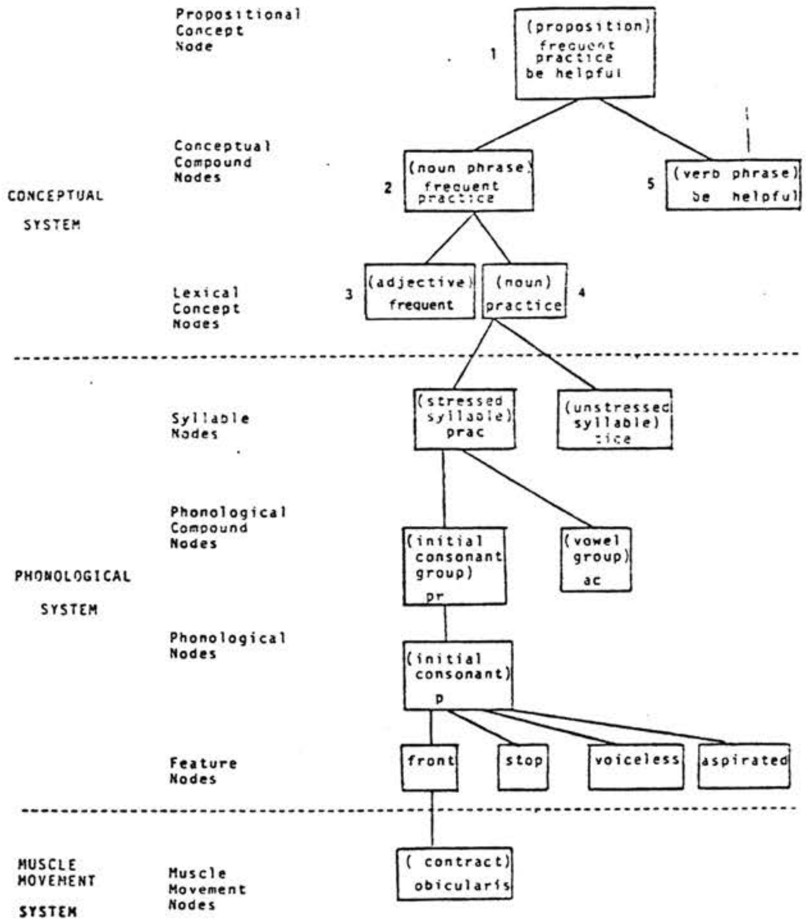

Figure 2. An action hierarchy consisting of nodes in the conceptual, phonological, and muscle movement systems for producing aspects of the sentence "Frequent practice is helpful." (The numbers indicate the order of activation of the nodes shown. Syntactic domain appears in parentheses.)

and independent timing properties and rules of sequential organization (see Figure 2). The muscle movement nodes constitute a third system that is likewise independently controllable and governed by yet another set of timing and serial-order mechanisms.

Dynamic Properties of the Nodes

Nodes have three dynamic properties that are relevant to the flexibility and fluency of behavior: activation, priming, and linkage strength.

Activation. Behavior occurs if and only if the nodes at the lowest muscle movement level in an action hierarchy become activated. Activation of nodes is serial in nature: For the components of a behavior sequence to be executed in proper serial order, the nodes at every level in its action hierarchy must be activated in proper sequence. As discussed below, a special triggering mechanism is needed to determine whether and in what order the nodes become activated.

Activation of a node is all or none and continues for a specifiable period of time, independent of the state of the source that led originally to activation. During its period of self-sustained activation, a node simultaneously primes all nodes connected directly to it.

Priming. Priming refers to transmission across a connection that produces increased strength or subthreshold activation of the connected node. Unlike activation, the state of priming, or subthreshold strength of a node, varies in degree: It summates across all simultaneously activated connections, and it increases over time during the period that any given connection remains activated. However, priming from a single connection only summates to some subthreshold, asymptotic level (see Figure 3) and cannot directly cause activation of a connected node. Moreover, priming is not self-sustaining: If a certain node, A, has been activated, it starts priming a connected node, B, but as soon as the activation of A is terminated, the priming of B begins to decay to its original or spontaneous level, which is assumed to be the same on the average for every node in a domain. Unlike activation, priming is order free or parallel in nature and requires no special triggering mechanism to determine whether and when it occurs. Also unlike activation, priming never results in behavior: No movement occurs when the lowest muscle movement nodes in an action hierarchy become primed (see MacKay, 1981).

The way that priming summates carries several important theoretical implications. One is a faster potential rate of output as a speaker progresses through a preplanned sequence, such as a common word. By way of illustration, consider the numbered nodes in Figure 2, which represent the concept "frequent practice is helpful" and must be activated in the order shown so as to achieve the proper word order in the final output. Activating Node 1 simultaneously primes Nodes 2 and 5, but because 5 cannot be activated until 2, 3, and 4 have been activated, the priming of 5 represents "anticipatory priming," which summates during the interval that Nodes 2, 3, and 4 are being activated. This accumulating anticipatory priming facilitates the activation of right-branching nodes at every level in the system, thereby

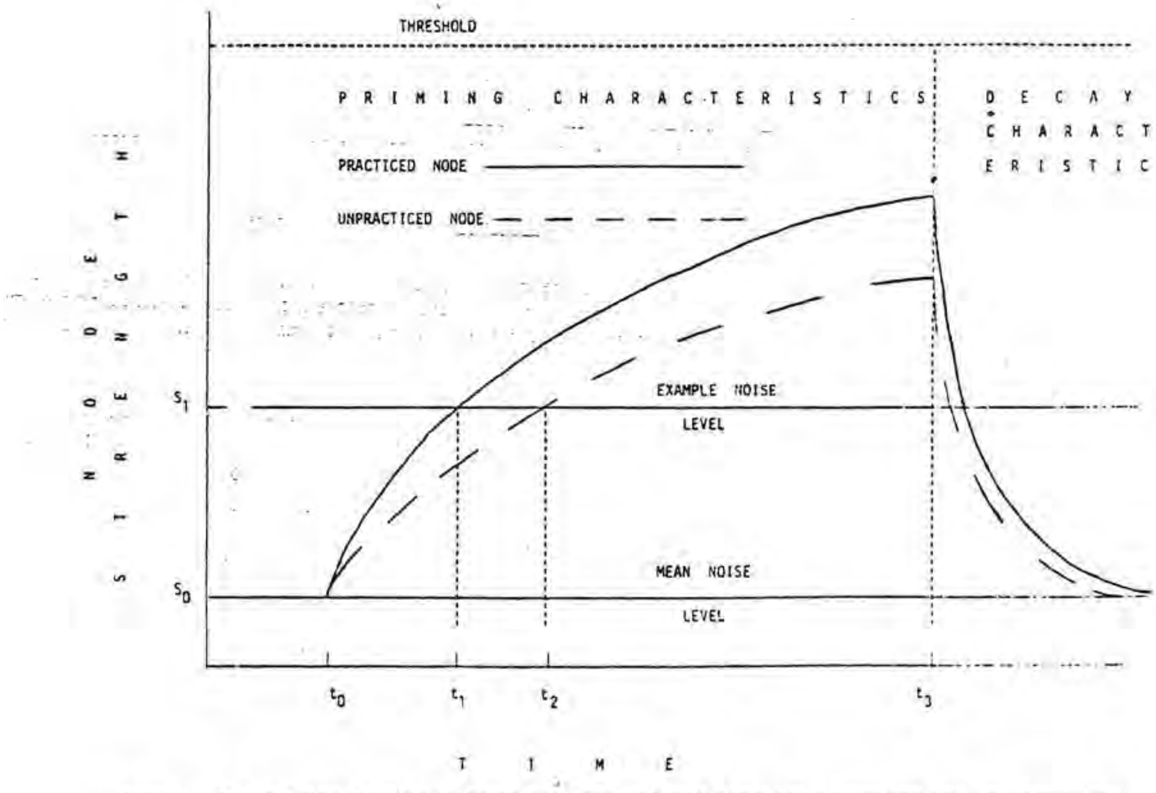

Figure 3. The priming function (relating node strength and time) for a practiced and an unpracticed node within a hypothetical domain of nodes with spontaneous or mean noise level S_0. (Both nodes are receiving priming from a superordinate node beginning at time t_0 and ending at time t_3; see text for explanation.)

speeding up the potential rate of output. Anticipatory priming may therefore explain why (all other factors being equal) components in a longer word are produced faster than otherwise identical components in a shorter word (Lehiste, 1970). By way of example, the *bi* in *responsibility* is produced faster than the *bi* in *bill* when subjects produce both words at maximal rates.

Linkage strength and effects of practice. When a node becomes activated, priming of its connected nodes increases over time up to some asymptotic level. But, as can be seen in Figure 3, both the asymptotic level and the rate of accrual per unit time (represented by the slope of the priming function) vary with practice: the frequency with which a node has been activated via a particular connection in the past. Thus, repeated activation increases linkage strength, which results in a higher asymptotic level of priming and a faster accrual of priming per unit time across one particular connection. As a long-term characteristic of a connection, linkage strength must be contrasted with node strength, which refers to the extent to which priming via any number of connections has summated on a particular node at any given point in time. Node strength is therefore variable and temporary in nature, so the distinction between node strength and linkage strength parallels that between short-term potentiation across a synapse and increased efficiency of neural transmission across synapses as a result of repeated postsynaptic activation (Eccles, 1972, p. 57).

The Sequential Triggering Mechanism: Syntax Nodes

Syntax nodes (distinguished by capital letters in the examples to follow) are the triggering mechanism that determines whether, when, and in what order the nodes become activated. Each syntax node is connected to an entire domain of nodes. For example, the syntax node NOUN is connected to all nodes representing nouns; the syntax node VERB is connected to all nodes representing verbs. An activated syntax node therefore simultaneously primes the entire domain of nodes connected with it. This priming summates quickly over time and serves to activate the node with the greatest degree of strength in its domain, normally the one that has just been primed from above via its connection to a superordinate node (see Figure 2). Following activation, the strength of syntax nodes quickly decays to resting level.

The connections between syntax nodes represent the serial-order rules of a language or any other action system and ensure that the nodes are activated in proper sequence. For example, the syntax nodes NOUN and ADJECTIVE are connected in such a way as to represent the fact that adjectives precede nouns in English. Inhibitory connections are one possible mechanism for determining these precedence relations (after Estes, 1972). Under this proposal, ADJECTIVE inhibits NOUN and dominates in strength when ADJECTIVE and NOUN are simultaneously primed. However, once ADJECTIVE has been activated and returns to resting level, NOUN will dominate in strength, thereby determining the sequence "adjective + noun" for English noun phrases. How the syntax nodes code more complex serial-order rules, including the ordering of conjoined forms such as "high and low," "near and far"; repeated forms as in "red, red rose"; and entire sentences (including those involving syntactic complexities such as center embedding) is left for a later publication (see MacKay, Note 1).

Syntax nodes code three general classes of rules for producing sentences: grammatical rules for determining the serial order of words in phrases and phrases in sentences, phonological rules for determining the serial order of phonemes in syllables and syllables in words, and muscle sequence rules for determining the serial order of muscle movements in producing speech sounds. Syntax nodes coding these three sets of rules organize the nodes into the three systems shown in Figure 2: the conceptual system, the phonological system, and the muscle movement system. Syntax nodes in one system can be activated independently of those in other systems, and rehearsal, or internal speech, represents an example where higher level systems are activated without activation of the lowest muscle movement system. During internal speech the syntax nodes representing grammatical and phonological rules become activated, resulting in activation of the phonological constituents of a sentence in proper serial order. However, the syntax nodes representing muscle sequence rules are not activated during internal speech so that no actual movement of the speech musculature ensues, even though the appropriate muscle movement nodes become primed (see MacKay, 1981).

The Timing Mechanism: Timing nodes

Timing nodes serve to organize the syntax nodes into systems. Specifically, the syntax nodes for the conceptual system are connected with a concept timing node; the syntax nodes for the phonological system are connected with a phonological timing node; and syntax nodes for the muscle movement system are connected with a muscle timing node. Timing nodes become activated at regular intervals, periodically priming the syntax nodes connected to them and activating the strongest one, following the "strongest-node-wins" principle discussed above. By determining when the syntax nodes become activated, the timing nodes therefore determine the temporal organization of the output. The concept timing node has the slowest average rate of activation (because phonemes and muscle movements are produced faster than words), and the muscle timing node has the fastest average rate of activation (because muscle movements are produced faster than phonemes and words).

Like the syntax nodes, timing nodes emit strong pulses with rapid priming and decay characteristics, and given that no syntax node has been primed, these pulses can be repeatedly applied without activating any nodes. This means that the motivational system (discussed in MacKay, Note 1) calling for, say, overt speech at some overall rate and onset time can simultaneously trigger the conceptual, phonological, and muscle timing nodes so that they begin emitting together. In producing sentences, no additional mechanism is needed to start (or stop) the timing nodes in cascade, beginning with the concept timing node and followed by the phonological and muscle timing nodes in succession.

An Example

How the timing and syntax nodes interact to determine whether, when, and in what order the nodes are activated is similar for every node in the system. I will give an example from within the conceptual system: the activation of *frequent* [adjective] and *practice* [noun] in producing the noun phrase "frequent practice" in answer to a question such as "What is the surest way to improve skill?" A more detailed account and hand simulation of the sequence of execution processes discussed below appears in the Appendix.

The concept nodes in question (rectangles) along with the connections between them are shown in Figure 4. Unbroken connections are excitatory and the dotted connection between syntax nodes (in circles) is inhibitory. It is assumed that these nodes and the connections between them are formed prior to executing this phrase in everyday speech or following the first reading of such a phrase in the Figure 1 data. This node-formation process is discussed in detail in MacKay (Note 1), and I concentrate here on the sequence of execution processes following node formation. However, it should not be assumed that concept nodes are necessary for producing the phrase "frequent practice." Foreign speakers can read these words without grasping their meaning in the same way that native speakers in MacKay and Bowman (1969) produced nonsense strings, using phonological nodes alone.

The node representing the concept *frequent practice* [noun phase] in Figure 4 is activated first, which simultaneously primes four nodes: two content nodes—*frequent* [adjective] and *practice* [noun]—and two syntax nodes—ADJECTIVE and NOUN. The inhibitory link temporarily reduces the strength of NOUN relative to ADJECTIVE, which therefore becomes activated via the first pulse from the concept timing node and primes every node in the [adjective] domain. One of these, *frequent* [adjective], having just been primed, has greatest strength, reaches threshold soonest, and becomes activated under the strongest-node-wins principle.

Following activation, ADJECTIVE returns to its spontaneous strength level, which releases the inhibition on NOUN, which now dominates in strength and becomes activated with the next pulse from the concept timing node. NOUN therefore primes the entire domain of [noun] nodes, but having just been primed, *practice* [noun] has greatest strength

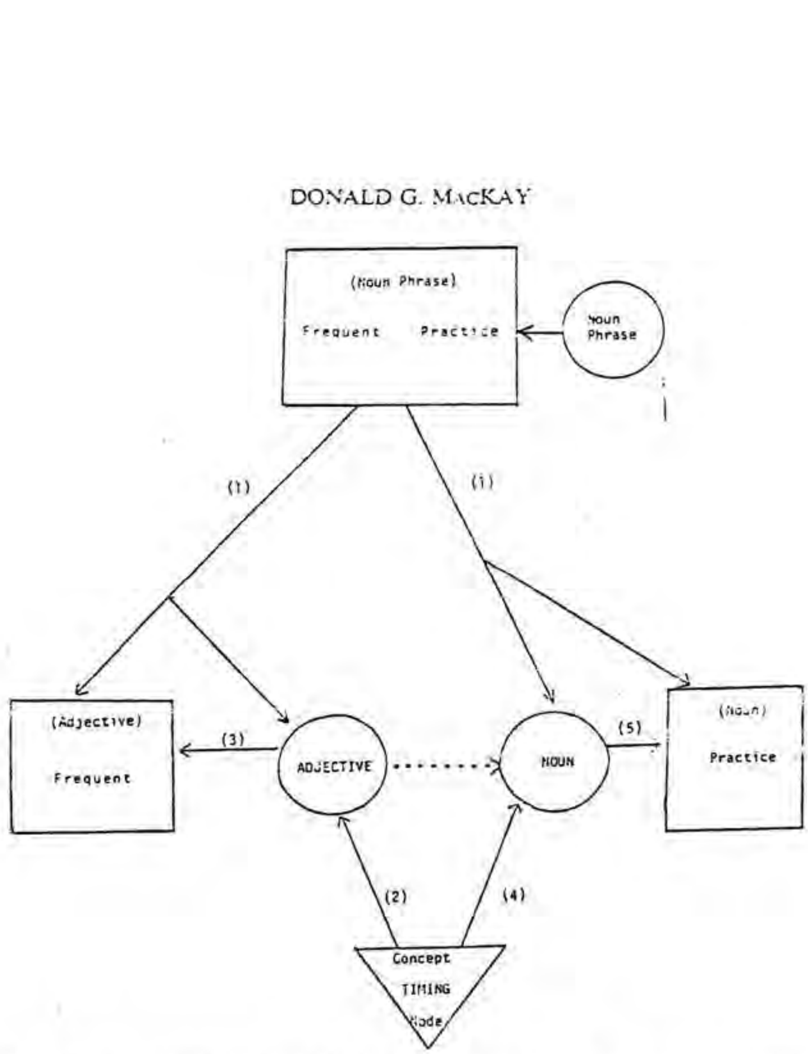

Figure 4. Processes underlying activation of three nodes (in rectangles), their corresponding syntax nodes (in circles), and timing node (in triangle). (Numbers indicate order of activation. Syntactic domain appears in parentheses.)

in the noun domain, reaches threshold soonest, and becomes activated under the strongest-node-wins principle.

Explanation of the Speech-Rate Data

The present theory provides a plausible explanation for the effects of practice on speech rate in Figure 1. A key factor in this explanation is a confound between prior practice and level in an action hierarchy: Higher level nodes have had less extensive prior practice than lower level nodes for producing sentences containing familiar words. By way of illustration, consider the word *he*. As a phonological entity, *he* is experienced over 20 million times in the course of a lifetime (see MacKay, 1980), so its phonological nodes are unlikely to benefit from further practice. Further practice is even less likely to help at the muscle movement level, because the movement components for the phonemes of *he* occur in many other frequently produced words and therefore receive even more extensive practice. However, the highest level conceptual node for a never previously encountered sentence containing *he* has received no practice whatsoever.

Log–log linearity. Under the present theory the rate of priming for a connected node increases as an exponential function of the frequency with which the connection has been activated in the past. Rate of priming in turn determines how quickly a node at any level in an output hierarchy can be activated, which in turn determines the maximal rate of output.

Speedup with practice therefore reflects the combined speedup of any and every node in the hierarchy. However, the hundreds of nodes for producing a sentence have a distribution of rates of improvement that depends on prior practice. For any given node, improvement with practice is exponential so that completely unpracticed nodes show rapid improvement; moderately practiced nodes show slow improvement; and extensively practiced nodes, which have already acquired as much linkage strength as they ever will, show no improvement whatsoever. The overall learning rate is therefore fast at first because rapidly improving, unpracticed nodes dominate the rate of change for the hierarchy during the early stages of practice, but soon make little contribution because of the effectiveness of exponential improvement. Eventually, during the later stages of practice, only slowly improving, practiced nodes remain so that the overall learning rate slows down. The overall relation between speedup and practice is therefore a log–log linear function resembling those in Figure 1.

Deviations from linearity. The theory also provides an interesting explanation for the deviations from log–log linearity that appear in Figure 1 and many other studies of speedup with practice (see Newell & Rosenbloom, 1981). Downward deviations during the initial portion of a log–log function reflect a relative scarcity of quickly improving unpracticed nodes, whereas upward deviations reflect a relative excess of unpracticed nodes. Consider the upward deviation for the mental practice function in Figure 1. As discussed above, the muscle movement system for a common action includes only slowly improving, highly practiced nodes, whereas the mental system above the muscle movement level includes proportionately more quickly improving, unpracticed nodes. Mental practice therefore produces an upward deviation or faster initial rate of improvement than physical practice, which involves activation of slowly improving muscle movement nodes in addition to quickly improving mental nodes. Moreover, there are fewer nodes to activate in the case of mental practice, which explains why mental practice is faster than physical practice at all levels of practice (see Figure 1 and Anderson, in press).

Types of material. The present theory readily explains the differences between the three types of material in Figure 1. Consider first the fact that normal sentences were produced faster than scrambled sentences and nonsense strings. Normal sentences have more higher level nodes than scrambled sentences and nonsense strings, thereby allowing more extensive anticipatory priming and a faster rate of activation.

Anticipatory priming likewise contributes to the asymptotic levels, beyond which practice has little effect. Consider the (projected) asymptotes for the three types of material in Figure 1. Normal sentences had a lower asymptote than scrambled versions because scrambling the words eliminates the possibility of anticipatory priming and practice effects at conceptual levels above the word. Nonsense strings had an even higher asymptote because of the impossibility of anticipatory priming and higher level practice effects above the level of the syllable.

Other Theories

Many other theories have been developed to explain speedup with practice, but most seem incomplete or inadequate for explaining practice effects in speech production and other skills involving high degree of proficiency. Three of these theories are discussed below.

Thorndike's Theory

The present theory synthesizes and extends several well-established theoretical ideas with long-standing historical antecedents. One of these historical antecedents is Thorndike's (1898) law of exercise: Use of a connection increases its strength, whereas disuse (not practicing a connection) leads to its weakening or forgetting. Interestingly, however, Thorndike later modified his 1898 law because of a subtle inconsistency with his law of reinforcement, which states that positive feedback (rewards or successes) automatically strengthens a rewarded connection or behavior, whereas negative feedback (punishment or failures) automatically weakens a connection and reduces the tendency to repeat the behavior (after Bower & Hilgard, 1981). Thorndike accordingly reinterpreted his law of exercise as follows: *Rewarded* or correct use of a connection between a stimulus and a response increases the strength of the connection.

The present theory differs in several respects from this modified law of effect. Its activation–speedup principle is both more specific and more general, applying to all nodes, including sensory analysis nodes, mental nodes, and muscle movement nodes, and not just to connections between stimuli and responses. Indeed, the present theory postulates no direct connection whatsoever between stimuli and responses, because, according to the theory, at least one mental node always intervenes between sensory and motor nodes.

The feedback–reinforcement issue illustrates another important difference between the two theories. Under the present theory, feedback is needed during the initial stages of acquiring a skill to determine what nodes to activate or what connections to form. However, once the proper connections have been formed in a high-proficiency skill, feedback becomes unnecessary: The potential rate of activating a node speeds up automatically as a result of repeated activation. This hypothesis fits the findings in Figure 1, where subjects improved with physical practice without receiving feedback concerning their rate of speech and where mental practice caused speedup without the occurrence of either overt action or concomitant sensory feedback.

Reinforcement nevertheless has an important role in the present theory. Subjects can control the activation of their own mental nodes, as in rehearsal or mental practice (see MacKay, 1981), and are likely to rehearse mentally those connections that lead to a desired effect more often than those that do not. As a result, naturally occurring mental practice is likely to strengthen connections, resulting in behavior that leads to reward rather than nonreward. This being the case, reinforcement is sufficient but not necessary to cause strengthening of underlying connections, and Thorndike's law of reinforcement is subordinate to his law of exercise, rather than vice versa as Thorndike himself supposed.

Crossman's Theory

Crossman (1959) provided the first attempt to explain the log–log practice function. Under Crossman's theory, a given task can be accomplished by many slightly different methods, each with a fixed execution time. Improvement occurs with practice because each method is selected stochastically according to a probability, and these probabilities are adjusted on the basis of time saved relative to previously tried methods.

Crossman's theory was for many years the accepted explanation of speedup with practice and constituted the rationale underlying industrial time-and-motion studies. Undoubtedly, there are many situations in which improvement in speed results from the omission of unecessary operations either in the final output or in the underlying programming of the output. Selecting fewer or faster movements seems especially likely during the early stages of acquiring an unfamiliar skill. For example, students take progressively fewer steps as a function of practice in solving unfamiliar mathematical problems (Neves & J. Anderson, 1981).

However, some residual component of speedup cannot be explained in this way, even for unfamiliar tasks. For example, Neves and Anderson (1981) also found that the time per step decreased with practice independent of the number of steps, a residual effect of practice that seems likely to play an even greater role in skills involving high degrees of proficiency, such as rolling a cigar following several million trials of practice (see Crossman, 1959), where the learner is already familiar with the most efficient steps. Likewise for the practice function in Figure 1, it is difficult to imagine any new methods being selected that would account for the data.

Newell and Rosenbloom's Theory

Newell and Rosenbloom (1981) outlined a class of models that explain practice effects as due to successful attempts to adjust or improve the procedure for performing a task. Under these models, learners use a single method in executing a task, and learning consists of finding and incorporating improvements in the current method. Power functions result from a law of diminishing returns in incorporating improvements: With increasing practice, either improvements become harder to find, or less time is available for finding them (because performance is speeded up), or the remaining improvements are less effective or more highly specialized and therefore less likely to be applicable.

Newell and Rosenbloom (1981) elaborated on the last of these alternatives to develop a full-fledged and general theory called the chunking theory of learning. They assumed that the performance program for a task becomes coded in terms of hierarchically organized chunks, with the time to process a chunk being less than the time to process its constituent chunks. The higher level chunks are learned at a constant rate, on the average, from the relevant patterns of stimuli and responses that occur in the specific environments experienced. However, higher level, large-sized chunks are built out of lower level chunks and are therefore learned later than lower level chunks. Thus, performance time decreases as a power function of practice (up to some asymptote) because of the increasing specialization of higher level chunks.

Newell and Rosenbloom (1981) illustrated this model for a 1,023-choice reaction time task involving patterns of finger presses and attributed the generality of log–log learning to the pervasiveness of the phenomenon of chunking. However, I had considerable difficulty applying the principles of the model to the sentence-production task discussed above. The assumption that the performance program for a sentence is coded in terms of chunks or nodes is of course widely accepted, but the assumption that higher level, large-sized chunks for, say, noun phrases are specialized and rare and are formed after lower level chunks such as nouns seems irrelevant. Once we comprehend a noun phrase, such as "the three little pigs," all of its nodes at every level are available, however rare their frequency of occurrence.

The assumption that the time to process a high-level chunk is less than the time to process its constituent chunks also seems irrelevant. All of the nodes down to the lowest muscle movement level must be activated each time a sentence is executed so that assuming shorter times for higher level nodes cannot explain effects of practice in sentence production.

Finally, I had difficulty envisaging how speedup in producing a sentence could occur as a result of adjustments in method, except perhaps at the lowest muscle movement level. However, data for the transfer condition (discussed below) indicated conclusively that the speedup is not occurring at that level.

Speed–Accuracy Trade-Off

Because speed and accuracy are the two major components of fluency, any theory of fluency must explain speed–accuracy trade-off, one of the most reliable and pervasive phenomena in the study of skilled behavior. Numerous studies using a variety of tasks and experimental conditions have demonstrated that increased speed leads to increased errors at every level in an output hierarchy, from the lowest level errors in muscle movement to the highest level lexical and phonological errors (see MacKay, 1971).

Current theories have difficulty explaining speed–accuracy trade-off. Most have concentrated on the muscle movement system, attributing errors to mechanical factors, but these are ipso facto incapable of explaining speed–accuracy trade-off for higher level phonological and conceptual errors. Theories attributing speed–accuracy trade-off to haste in the processing of proprioceptive feedback likewise have difficulty with higher level errors and with the fact that restricting proprioceptive feedback has little effect on errors (Schmidt, Zelaznik, & Frank, 1978). Finally, all current theories have difficulty with the recent results of K. Newell (1980). Newell's subjects practiced making arm movements of specific durations and, in a subsequent test, reproduced the practiced movement durations from memory. Contrary to speed–accuracy trade-off, Newell (1980) found that errors in the intended duration of the arm movements sometimes decreased as the speed of movement increased.

The explanation of speed–accuracy trade-off in the present theory overcomes the difficulties confronting other theories. As discussed below, an integral component of that explanation is how errors are caused.

The Fundamental Cause of Errors

Correct output occurs, according to the theory, whenever an "intended-to-be-activated" node has greater strength than any other node in its domain when the triggering mechanism is applied. The intended-to-be-activated node is that node that is being primed by a superordinate node controlling the action sequence, that is, the directly connected node immediately higher in the hierarchy. This priming summates systematically over time (see Figure 3) and soon exceeds the strength of all other nodes in the domain, say by Time t_1. If applied after this point in time, the triggering mechanism therefore activates the intended-to-be-activated node under the strongest-node-wins principle, and the output is error free.

Errors occur whenever an extraneous node has greater strength than the intended-to-be-activated node when the triggering mechanism is applied. Extraneous nodes include all but the intended-to-be activated node in a given domain. As illustrated below, extraneous nodes receive inputs from sources other than the immediately superordinate node in an action hierarchy, and priming from these other sources can sometimes exceed the systematically increasing strength of the intended-to-be-activated node when the triggering mechanism (syntax node) is activated. As a consequence, the extraneous node becomes activated under the strongest-node-wins principle, and an error occurs.

Errors at every level in an action hierarchy are explained in the same way, and because previous theories have ignored higher level errors, we will illustrate the explanation by means of one such error: the Freudian substitution of *bottle scarred* for *battle scarred*, spoken of a general who is convertly believed to be incompetent as a result of "hitting the bottle." Under the theory, this convert belief independently primes the node for *bottle* so that this extraneous node has greater strength than the intended-to-be-activated node for *battle* when the triggering mechanism is applied to the domain of noun concept nodes. The wrong node is therefore activated, and the word substitution occurs because the triggering mechanism automatically activates the strongest node in a domain.

The theory therefore predicts that erroneously substituted forms will invariably belong to the same domain or syntactic class as the intended form, and the available data for thousands of speech errors involving either words or speech sounds strongly support this prediction. For example, nouns nearly always substitute for nouns (as in the above example), verbs for verbs, and adjectives for adjectives; virtually never does an adjective substitute for a noun or verb (see Fromkin, 1980).

Effects of Rate

Rate of output in the theory depends on when the triggering device is activated following onset of priming: If applied soon after onset of priming for every node in the system, the overall rate of output will be fast, but if activated long after onset of priming, the overall rate will be slow. However, the sooner the triggering mechanism is activated, the greater the probability of error. By way of example, let the practiced node in Figure 3 represent the concept *battle* in the example above and let the strength of the extraneous node for *bottle* correspond to S_1. This particular "example noise level" (S_1) exceeds the spontaneous or mean resting strength of other extraneous nodes in the domain by say three standard deviations, due to priming from other sources (here, the nodes coding the belief that the general is incompetent as a result of "hitting the bottle").

The intended-to-be-activated node for *battle* receives priming from its superordinate node beginning at Time t_0, and the triggering device can be applied to the domain at any point in time after that. However, t_1 represents a critical time for activating the triggering mechanism. If applied before t_1, *bottle* will substitute the intended *battle* (see Figure 3), but if applied after t_1, the intended-to-be-activated node for *battle* will have accrued greater strength than all other extraneous nodes so that no error can occur.

In general, the extraneous nodes can be considered to assume a Gaussian distribution of strengths with mean S_0 (the spontaneous or resting level) at any given point in time. The strength of the strongest extraneous node therefore varies randomly over time, whereas the strength of the intended-to-be-activated node increases systematically over time. As a consequence, the probability of error decreases the later the application of the triggering mechanism following onset of priming. In other words, errors trade off with speed in the theory. Furthermore, the theory predicts that empirically observed relations between speed and accuracy will resemble in shape the priming functions in Figure 3, which as McClelland (1979) points out has been invariably the case for observations to date.

The Limits of Speed–Accuracy Trade-Off

Speed trades off with accuracy in the theory only during the period that priming from a superordinate node continues to summate. Once activation of the superordinate node stops (Time t_3 in Figure 3), priming stops summating and decays over time. As a consequence, the normal speed–accuracy relation will reverse itself with unnaturally slow rates, and this reversal may account for Newell's (1980) increase in errors with extremely slow rates of movement. Because of interactions between practice, potential rate, and errors (discussed below), speed–accuracy reversals resembling those of Newell (1980) can also be expected whenever fast movements that are relatively practiced are compared with slow movements that are relatively unpracticed.

Practice and the Time–Error Criterion

Rate and errors are inextricably linked in the present theory: Choosing a particular rate is equivalent to choosing a particular probability of error. Of course, speakers normally choose a conservative time–error criterion—a rate of output resulting in some acceptably low probability of error. For the example in Figure 3, a time criterion of t_1 following onset of priming yields a probability of error of .001 because, as discussed above, the strength of the intended-to-be-activated node is three standard deviations above the mean noise level for the extraneous nodes by that time.

A number of cognitive, motivational, and situational factors play a role in selecting any given time–error criterion. However, practice causes a favorable shift in the absolute values that any time–error criterion can assume. By way of illustration, compare the practiced and unpracticed nodes in Figure 3. As can be seen there, priming summates as a nonlinear function of time, and the slope and asymptote of the priming function vary with practice so that at any point in time following onset of priming, the node with more extensive practice has greater strength and therefore lower probability of error than the one with less extensive practice (see Figure 3). This means that for the same probability of error, the time of activation can be shorter for more practiced nodes. In short, practice allows a lower time–error criterion by increasing the potential rate of output (for a given probability or error) and decreasing the probability of error (for a given rate).

The distinction between actual and potential rates carries important empirical implications. By way of illustration, assume that the probability of error were negligibly small and experimentally indistinguishable for all activation times longer than t_2 for both the practiced and unpracticed node in Figure 3. This means that for this range of actual rates, practiced and unpracticed behaviors would be empirically indistinguishable. However, for all times shorter than t_2, the probability of error is experimentally determinable and will differ for practiced versus unpracticed nodes. Empirically, the potential or maximal rate for a given probability of error is faster for practiced than unpracticed behaviors only over this more rapid range of rates.

The Problem of Flexibility in Speech Production

The issue of flexibility as applied to the data in Figure 1 is this: Does skill acquired in practicing one sentence transfer to other sentences? And if so, at what level does the transfer occur? Under the present theory, practice increases linkage strength for hierarchically organized nodes controlling an action, thereby resulting in faster accrual of priming across connections and a faster potential rate of activation. However, some connections in any given action hierarchy may have had as much strengthening already as they ever will so that further strengthening is confined to particular levels in the hierarchy, thereby resulting in particular kinds of transfer or flexibility in behavior.

Specifically, the conceptual nodes for producing normal sentences have had less prior practice and therefore contribute more to the practice effects in Figure 1 than the phonological and muscle movement nodes, which have had more prior practice and are therefore closer to asymptote. As a consequence, conceptual nodes will contribute more to transfer of practice than phonological or muscle movement nodes.

To test this prediction MacKay and Bowman (1969) included a transfer-of-training condition following the 12 practice trials shown in Figure 1. Recall that the subjects were German–English bilinguals and that they produced a practice sentence 12 times at maximal rate. They then produced a transfer sentence in their other language, likewise at maximal rate. In the critical transfer condition, the transfer sentence was a word-for-word translation of the sentence they had practiced 12 times in their other language. Under the present theory, translations will be produced as fast as the fastest practice trial because components at the conceptual level are responsible for the original practice effects, and the transfer translation sentences involve these same, already practiced components.

Alternate predictions can be derived from the muscle movement hypothesis, which is based on the muscle-specific command and single-level assumptions outlined in the introduction. Under this hypothesis, speedup is completely attributable to the pattern of muscle movement, which becomes more efficient as a function of practice. This muscle movement hypothesis readily explains some of the data in Figure 1. Consider the faster rate for words than nonsense strings such as *morbing*. The specific pattern of muscle movements for familiar words such as *morning* receives extensive practice in the course of everyday speech and is therefore more efficient than for unpracticed strings such as *morbing*. The muscle movement hypothesis likewise explains the faster rate for normal versus scrambled sentences: Although the entire pattern of muscle movements for a sentence in all likelihood has never been practiced before (because sentences are rarely repeated, even over the course of a lifetime), parts of normal sentences such as *in the* or *his bed* have received more practice at the muscle movement level than otherwise similar nonsense strings such as *the in* or *bed his*. However, the muscle movement hypothesis predicts no facilitation whatsoever for the translations in the transfer condition, because the translations involve completely different muscle movements from the practiced sentences.

The Transfer Data

The language of each transfer sentence always differed from its practice sentence and was announced prior to presentation during the 20 sec between the last practice trial and the first transfer trial. The transfer sentences were repeated four times, likewise with 20 sec between repetitions.

Both the practice and the transfer sentences were obtained from published literature for which professional translations were available, and half the transfer sentences were literal translations of the practice sentences, whereas the other half were nontranslations unrelated to the practice sentence. Several controls were introduced to ensure that translations and nontranslations were comparable. All of the sentences were initially equated for minimum reading time in a pilot study, and within the experiment itself, the transfer sentences were counterbalanced across subjects so that exactly the same sentence occurred as either a translation or a nontranslation, depending on what sentence had been practiced. Finally, scoring of the data was blind, the scorers not knowing whether they were timing a translation or a nontranslation. Differences in the time to produce translations versus nontranslations could not therefore be attributed to either sentence difficulty or scoring bias.

The transfer data appear in Table 1. Production time was faster for translations (2.03 sec) than nontranslations (2.44 sec). This 17% difference was statistically reliable and reflects an effect of practice at the conceptual level, thereby supporting the present theory rather than the muscle movement hypothesis.

Even stronger support for the present theory and against the muscle movement hypothesis comes from the fact that the transfer sentences behaved as if they had received 12 trials of practice even though they had re-

Table 1
Production Time (in Seconds Per Sentence) During Practice, and Transfer Trials for Normal Sentences, Scrambled Sentences, and Nonsense Strings

	Practice trials		Transfer trials		
Material	First four	Last four	Nontranslation	Translation	Facilitation (%)
Normal sentences	2.33	2.03	2.44	2.03	17
Scrambled sentences	2.90	2.38	2.79	2.83	−1
Nonsense strings	4.33	3.20	4.33	4.33	0

ceived no practice whatsoever at the muscle movement level. Sentence durations for the last practice trial and the first transfer trial were equivalent, as were durations for the last four practice trials and the four transfer trials (see Table 1). To my knowledge this is the first study that has exhibited transfer with no measurable time decrement when the motor requirements were changed. This finding confirms the prediction that practice effects for fluent speakers producing normal sentences are completely attributable to components at the conceptual level and that practice with familiar words produces no facilitation that can be attributed to components at the muscle movement level.

Consider now the scrambled sentences. Facilitation occurred during the practice phase for scrambled sentences (see Figure 1) but did not transfer across languages: Word-for-word translations of scrambled sentences were produced no faster than were nontranslations (see Table 1). This finding is expected under the present theory because scrambled sentences cannot be meaningfully parsed and therefore have no conceptual structure that can transfer across languages. This finding also indicates that producing words in one language does not automatically activate their lexical translations in the other language and thereby rules out a low-level alternative explanation for the conceptual facilitation effect. Still, we might ask why practice with word order per se fails to transfer across languages for the scrambled sentences. This result may be an artifact of multitranslatability: Unlike words in sentences, words in isolation or scrambled sequence allow many possible interpretations, each with a unique translation into the other language. For example, the word right, taken in isolation can be translated into German with at least eleven different forms (e.g., *recht*, *richtig*, *ordnen*), each with a distinctly different meaning. As a consequence, the subjects may have comprehended the scrambled words in the English practice condition in one way but could only interpret their translations in the German transfer condition in another way. This being the case, the practice phase engaged different conceptual nodes from the transfer phase, thereby eliminating the possibility of transfer at the conceptual level for the scrambled sentences.

MacKay and Bowman (1969) introduced another condition to examine the effect of conceptual practice on errors in speech production. The bilinguals practiced sentences in one language as before, but then produced the transfer sentences in their other language while their auditory feedback (amplified over earphones they were wearing) was delayed by .20 sec, a procedure that greatly increases normal susceptibility to error. The issue was whether the probability of error would decrease for subjects producing translations of sentences they had practiced.

The data indicated that conceptual practice decreased the probability of error. Errors were 47% less frequent for the translations than nontranslations in the transfer condition, a difference significant at the .01 level. Under the present theory, this reduced probability of error reflects the increased linkage strength or rate of priming for the practiced nodes: At a given rate a practiced node has greater strength than an unpracticed node so that the probability of error (i.e., of an extraneous node having greater strength) decreases with practice.

General Discussion of Transfer

The findings discussed above illustrate an instance of transfer of skill that cannot be explained at the muscle movement level. However, the data are readily explained within the present theory. The highest level nodes for coding the meaning of a sentence and its word-for-word translation into German are identical in the case of German-English bilinguals. As a consequence, fluency resulting from practicing one sentence can transfer via these nodes to its translation. More generally, because behavior results from the sequential activation of a hierarchy of nodes, different behaviors at the muscle movement level can share some of the same nodes at higher levels in their hierarchies. And these shared nodes provide the basis for transfer of skill from one behavior to another.

Consider bilateral transfer, for example. It is relatively simple to show that bilateral transfer cannot occur within the system of muscle movement nodes. Actions performed with one limb involve different muscles and patterns of movements from actions performed with the other. For example, throwing a baseball with the left arm requires a stepping movement for the right leg and a supporting stance for the left, whereas throwing with the right arm involves stepping with the left leg and supporting with the right. The two actions share no common elements that can support transfer within the muscle movement system. Rather, concepts such as "step with the leg opposite the throwing arm" provides the basis for bilateral transfer of this sort and are of necessity coded within the system of mental nodes.

Writing with one's unaccustomed hand is another example of bilateral transfer that requires an explanation at the mental level. When we learn to write, we form and strengthen the connections between mental nodes for organizing the sequence of movements involved in producing letters, and the same mental nodes are involved whether we write with our left or right hand. As a consequence, higher level aspects of writing skill will transfer from one hand to the other. For example, personal style in writing is attributable to the mental nodes controlling letter formation, which explains why writing style remains the same whatever low-level muscle movement system is used to express the skill (see Merton, 1972).

Degrees of Transfer

The varying degrees of transfer discussed in the introduction are readily explained under the theory as follows.

Perfect positive transfer. Conceptual aspects of a skill are coded via mental nodes but can only be expressed in behavior by means of low-level muscle movement nodes. This means that perfect transfer between two tasks can only occur when the low-level, unshared (divergent) nodes for performing the transfer task are already formed and extensively practiced. Such is the case for the phonological and muscle movement nodes representing familiar words in the two languages of fluent bilinguals, likewise for the low-level nodes for moving small objects such as chess or Go pieces with either hand. Perfect bilateral transfer can therefore be expected for high-level skills such as playing chess or Go because the muscle movement nodes for expressing these skills with either hand have been formed and practiced since early childhood. The theory likewise predicts perfect high-level transfer for playing a familiar piece on the flute or violin, despite the totally different muscle movements involved in the expression of these skills. Specifically, a bi-instrumental musician, who is competent at playing both the violin and the flute, should show perfect transfer from practicing a piece on one instrument to playing the same piece on the other.

Partial transfer. The above prediction does not apply to bi-instrumental musicians who are not highly practiced or have not overlearned their skills at the muscle movement level. Only partial, imperfect transfer can be expected when the divergent or unshared nodes for the transfer activity are relatively unpracticed. For example, in writing with the unaccustomed hand, there is invariably a decrement in maximal speed and dexterity of form because speed and errors or variability in activation of muscle movement nodes depend on degree of practice. The muscle movement nodes for forming letters with the preferred hand have received much more practice than those for the nonpreferred hand and are therefore activated less quickly and with greater variability or probability of error.

The degree of positive transfer between two performances therefore depends on the existence of shared nodes and the current level of practice of the remaining, unshared (divergent) nodes for realizing the transfer performance. This general principle explains why there is little decrement in bilateral transfer for simple skills such as Morse key tapping: The unshared nodes for tapping with the nonpreferred hand have received extensive practice since early childhood, unlike those for complex skills such as handwriting, which have not.

Negative transfer. Negative transfer or interference occurs when the higher level nodes controlling an action sequence strongly prime one or more extraneous nodes within an intended-to-be-activated domain. This extraneous priming is the result of prior connections, either learned or built in. An example of the latter was found in a study by MacKay and Soderberg (1971), who examined rapidly generated patterns of finger movement carried out simultaneously by the two hands. An analysis of the finger movement errors in this task led MacKay and Soderberg (1971) to conclude that voluntary activation of a finger movement for one hand primes or partially activates the anatomically homologous finger movement for the opposite hand.

This bilateral-priming principle provides a possible explanation for the negative transfer sometimes seen in simultaneous switching of tasks from one hand to the other. If activating a finger movement for one hand primes the anatomically homologous finger movement for the opposite hand, as in MacKay and Soderberg (1971), such interactions would greatly interfere with simultaneous transfer from one hand to the other, as when the roles of the hands are reversed in playing the guitar.

The Relation Between Transfer, Adaptability, and Fluency

As discussed so far, transfer reflects the many-to-one connections between higher and lower level nodes in an action hierarchy. In bilingual transfer, for example, one lexical concept node is connected to more than one set of phonological nodes for expressing the concept in different languages, and these one-to-many connections play a critical role in transfer effects.

An additional context-dependent mechanism is needed to explain adaptability—our ability to adapt components of an ongoing behavior sequence, as in the stumble-preventing reflex and error-compensating mechanisms discussed in the introduction. But, as illustrated below, instances of transfer also require a context-dependent mechanism for determining which of the one-to-many connections to activate in any given situation.

Context-Dependent Effects in Motor Equivalence

Motor equivalence (response generalization or action constancy) refers to our ability to generate different actions that are functionally equivalent at some level. True motor equivalence is synonymous with perfect transfer in the present theory, but instances of partial transfer, where the divergent, low-level nodes controlling the different actions have some nonzero but unknown and uncontrolled degree of prior practice, are often included under this label as well.

Wickens (1938) provided an example of perfect or near-perfect motor equivalence. He first conditioned subjects to lift a finger from an electrode to avoid electric shock. The hand was palm down during training so that escape from shock required activation of extensor muscles to lift the finger. But when tested with the back of the finger on the electrode, only one subject made the extensor response and this sole error occurred on the first transfer trial. All other subjects showed perfect response generalization, transferring their withdrawal reaction to the opposite (flexor) movement even on the first transfer trial.

This example of motor equivalence requires an explanation along the following lines. Mental nodes representing concepts such as "escape shock," "lift right index finger," "palm up," and "palm down" control the observed withdrawal reactions. However, a concept such as "lift finger" is context-dependent, requiring flexion if the hand is palm up and extension if the hand is palm down. This context-dependent effect is readily achieved in the present theory. The lift-finger node primes both the flexion and the extension node; the palm-up node primes the flexion node when the hand is palm up, whereas the palm-down node primes the extension node when the hand is palm down. The appropriate context-dependent action is therefore triggered because whichever muscle movement node receives the greatest degree of conjoint priming will become activated under the strongest-node-wins principle.

Under the present theory, this and all other examples of transfer involve both one-to-many connections (representing the set of possible transfer alternatives) and many-to-one connections (which determine which of these transfer alternatives become activated in any given context), and similar mechanisms are needed to explain the second component of flexibility in behavior: context-dependent adaptability. By explaining both components in the same way, the theory therefore provides a unified account of behavioral flexibility.

Subordinate Autonomy

The context-dependent mechanism illustrated above provides a means whereby subordinate nodes can determine aspects of an action that are not directly represented by nodes higher up in the action hierarchy. Subordinate autonomy of this sort is quite general under the theory: The same context-dependent mechanism plays a role in simplifying higher level specifications at every level in the system, thereby leaving the higher level structures free to deal with more general goals. By way of illustration, consider the conceptual system for producing noun phrases containing the definite versus indefinite determiner (*the* vs. *a* as in, say, *the theory* vs. *a theory*). Without subordinate autonomy, two noun-phrase nodes are needed for coding these expressions within the conceptual system, but given subordinate autonomy, a single noun-phrase node can code both. The mechanism is as follows: A nonspecific noun-phrase node representing *the/a theory* is connected with both determiner nodes (one-to-many connections), and each of these determiner nodes receives a connection from another source (a many-to-one connection). The other source for *a* [determiner] is a node representing the concept "new or never previously mentioned," whereas the other source for *the* [determiner] is a node representing the concept "old or previously mentioned." Thus, when the triggering mechanism is applied to the domain of determiner nodes, whichever one has the greatest degree of priming from whatever source becomes activated, resulting in the appropriate context-dependent output.

The Relation Between Fluency and Flexibility

As noted above, fluency depends on linkage strength (which varies with practice—the frequency with which a node has been activated via a particular connection in the past). Flexibility likewise depends on the linkage strength of the one-to-many and many-to-one connections between nodes. Both flexibility and fluency therefore depend on the same factor under the theory, which explains the positive correlation between flexibility and fluency discussed in the introduction.

The Node-Structure Approach to Skills

The principles underlying the present theory are assumed to hold for all skills, but detailed extensions of the present theory to other skills require a new approach. Under this "node structure" approach, the goal of studies of skill is to determine which nodes control a behavior sequence and how they are organized or connected with one another. This node-structure approach contrasts sharply with the dichotomization approach outlined in the introduction. Many factors considered important in the surface characterization of skills developed under the dichotomization approach are unimportant in the node-structure approach. Consider, for example, the surface dichotomy between perceptual–motor skills (e.g., tennis) and cognitive or mental skills (e.g., chess). Under the present approach these skills display a fundamental similarity: Both involve a muscle movement component above the muscle movement level, but it is only the muscle movement nodes for transporting small objects such as chess pieces that have been formed and extensively practiced since early childhood, whereas those for playing tennis have not.

On the other hand, factors overlooked in surface characterizations of skill are important in the present approach. Consider, for example, the skills of speech, typing, and playing the piano, which are viewed in surface descriptions as highly similar, a coherent family in the kingdom of skills. Under the present approach, differences among these skills are as striking and numerous as the similarities. Speech typically proceeds at a variable and self-determined pace, unlike either playing the piano, with its largely composer-specified pace, or typing, with its characteristically steady, maximally rapid rate. The nodes for organizing speech also receive more practice than those for typing or playing the piano. The mental nodes for speech become activated not only during reading, listening, or speaking for many hours a day but also during everyday thinking incorporating either propositional thought (involving the conceptual nodes) or internal speech (involving phonological as well as conceptual nodes). By way of contrast, the nodes for typing or playing the piano virtually never become engaged during everyday thought and cannot benefit from this special source of mental practice.

The node-structure approach also suggests some insights into aspects of skill that have proven theoretically problematic in the past. Two such aspects are discussed below.

Automaticity

Given enough practice, aspects of an action hierarchy can become automatic, that is, rapid, error free, effortless, and unconscious in execution. For all multilevel skills, automaticity varies with the level under consideration. For example, try to produce a complex sentence on an unfamiliar topic such as the priming function: Choice of words and meanings to convey will be slow, conscious, effortful, and replete with errors or false starts, whereas choice of phonemes will be automatic, unconscious, effortless, and relatively error free. The question is why.

The answer under the node-structure approach is that concept nodes in general receive less prior practice than phonological nodes. New conceptual-compound and proposition nodes must be formed de novo in producing never previously encountered sentences, such as "The priming function varies with linkage strength," and forming these nodes requires time, effort, and awareness. For example, a new node must be formed within the conceptual system in order to generate the never previously encountered expression "varies with linkage strength." However, *phoneme* nodes for this expression are performed in the case of adults, primed from above, and so extensively practiced and quickly activated that conscious awareness and effort seem out of the question. As Sokolov (1960) pointed out, consciousness is reserved for what is new (unhabitual and unhabituated) and what is new in sentences is not phonemes but higher level concepts.

The Nature of Motor Programs

Under the node-structure approach, a motor program is an action hierarchy—the hierarchically organized set of nodes for controlling a preplanned sequence of actions. Under this definition, programs are performable at indefinitely many speeds because the same set of interconnected nodes is involved when a preplanned sequence of actions is executed at different rates. Programs also govern classes of actions, and the same program can result in different behaviors because the nodes representing context or ongoing sensory events contributing to subordinate autonomy are not part of a program. However, different motor programs are involved when identical actions are performed in reverse order, because a different order implies a different set of interconnections between the syntax nodes controlling the behavior.

This view overcomes the problems with the motor program assumptions discussed in the introduction. Contrary to the muscle-specific command assumption, motor programs do not consist entirely of muscle-specific commands. Mental nodes representing classes of actions are involved in the organization of even the simplest muscle movements such as the classical conditioning of finger withdrawal. Likewise, contrary to the abstract-program assumption, the lowest level nodes for movement control are concrete and specific, because nodes at the muscle movement level represent muscle-specific commands.

Contrary to the single-level assumption, programs are hierarchical under the node-structure approach, and this hierarchical organization is the basis for flexibility in behavior. If lower level nodes in an action hierarchy are termed subprograms, then flexibility reflects the one-to-many and many-to-one connections that exist between subprograms and higher level nodes within the system. For example, a noun-phrase node constitutes a subprogram that can be connected with indefinitely many proposition nodes for generating sentences (see Chomsky, 1957). The same noun-phrase node can likewise be connected with indefinitely many phonological subprograms, each representing an expression of the concept in a different language that a hypothetical speaker knows or can learn.

Finally, contrary to the feedback-free assumption, even highly fluent programs can be interrupted or altered on the basis of sensory input arising during execution, as in the stumble-preventing reflex. This is not to say that sensory input or feedback is necessary in executing a motor program. Interruptibility and adaptability are themselves flexible. For example, in playing a highly practiced tune on the piano, one can decide in advance to stop or compensate for errors signaled by sensory feedback, or to ignore them, continuing as if nothing had happened.

Given that motor programs are interruptible, it follows that the time characteristics of a preplanned behavior sequence are irrelevant to motor programming, contrary to the .5-sec limitation imposed by Bindra (1978). Besides being theoretically necessary (see introduction), liberating the concept of motor programming from such temporal constraints seems intuitively appealing. As Grossberg (1978) points out, it seems plausible to say that a motor program is called up when I decide to take my customary 5-min walk to lunch, even though I interrupt the program to talk with a friend or smell the flowers en route.

Reference Note

1. MacKay, D. G. *A general theory of serial order in behavior.* Paper presented at the Center for Theoretical Psychology, Edmonton, Alberta, Canada, September 1980.

References

Anderson, R. A. Speech imagery is not always faster than visual imagery. *Memory & Cognition*, in press.

Bartlett, F. C. *Thinking: An experimental and social study*. London: Allen & Unwin, 1958.

Bindra, D. How adaptive behavior is produced: A perceptual-motivational alternative to response reinforcement. *Behavioral and Brain Sciences*, 1978, *1*, 41–92.

Bower, G. H., & Hilgard, E. R. *Theories of learning*. Englewood Cliffs, N.J.: Prentice-Hall, 1981.

Chomsky, N. *Syntactic structures*. The Hague, Netherlands: Mouton, 1957.

Crossman, E. R. F. W. A theory of the acquisition of speed-skill. *Ergonomics*, 1959, *2*, 153–166.

Eccles, J. C. Possible synaptic mechanisms subserving learning. In A. G. Karczmar & J. C. Eccles (Eds.), *Brain and human behavior*. New York: Springer-Verlag, 1972, pp. 39–62.

Estes, W. K. An associatuve basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: V. H. Winston, 1972.

Ellis, H. C. *The transfer of learning*. New York: Macmillan, 1965.

Fromkin, V. A. Introduction. In V. A. Fromkin (Ed.), *Errors in linguistic performance. Slips of the tongue, ear, pen and hand*. New York: Academic Press, 1980.

Gallistel, C. R. *The organization of action: A new synthesis*. New York: Wiley, 1980.

Grossberg, S. A theory of human memory: Self-organization and performance of sensory–motor codes, maps and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5). New York: Academic Press, 1978.

Holding, D. H. Skills research. In D. H. Holding (Ed.), *Human skills*. New York: Wiley, 1981.

Keele, S. Behavioral analysis of movement. In V. Brooks (Ed.), *Handbook of Physiology: Motor Control*. Washington, D.C.: American Physiological Society, 1981.

Klapp, S. T., Greim, D. M., Mendicino, C. M. & Koenig, R. S. Anatomic and environmental dimensions of stimulus-response compatibility: Implication for theories of memory coding. *Acta Psychologica*, 1979, *43*, 367–379.

Lehiste, I. *Suprasegmentals*. Cambridge, Mass.: MIT Press, 1970.

MacKay, D. G. Stress pre-entry in motor systems. *American Journal of Psychology*, 1971, *84*, 35–51.

MacKay, D. G. On the goals, principles and procedures for prescriptive grammar. *Language in Society*, 1980, *9*, 349–367.

MacKay, D. G. The problem of rehearsal or mental practice. *Journal of Motor Behavior*, 1981, *13*, 274–285.

MacKay, D. G., & Bowman, R. W. On producing the meaning in sentences. *American Journal of Psychology*, 1969, *32*, 23–39.

MacKay, D. G., & Soderberg, G. Homologous intrusions: An analogue of linguistic blends. *Perceptual and Motor Skills*, 1971, *32*, 645–646.

McClelland, J. L. On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 1979, *86*, 287–330.

Merton, P. A. How we control the contraction of our muscles. *Scientific American*, 1972, *226*, 30–37.

Neves, D. M., & Anderson, J. R. Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, N.J.: Erlbaum, 1981, 57–84.

Newell, A., & Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.). *Cognitive skills and their acquisition*. Hillsdale, N.J.: Erlbaum, 1981, 1–56.

Newell, K. M. The speed–accuracy paradox in movement control: Errors in time and space. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. New York: North Holland, 1980.

Schmidt, R. A. A schema theory of discrete motor skill learning. *Psychological Review*, 1975, *82*, 225–260.

Schmidt, R. A., Zelaznik, H. N., & Frank, J. S. Sources of inaccuracy in rapid movement. In G. E. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

Shaffer, L. H. Analyzing piano performance: A study of concert pianists. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. New York: North Holland, 1980.

Sokolov, E. N. Neuronal models and the orienting reflex. In M. A. Brazier (Ed.), *The central nervous system and behavior*. New York: Josiah Macy Foundation, 1960.

Thorndike, E. L. Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs*, 1898, 2(4, Whole No. 8).

Wickens, D. D. The transference of conditioned excitation and conditioned inhibition from one muscle group to the antagonistic muscle group. *Journal of Experimental Psychology*, 1938, *22*, 101–123.

Appendix

The Processes Underlying Activation of Timing, Syntax, and Concept Nodes

Table A1 illustrates a hand simulation of the processes underlying activation of the seven nodes in Figure 4. The columns in Table A1 represent the simulated degrees of activation for the concept timing node, syntax nodes, and conceptual nodes as a function of time, whereas the rows represent the "instantaneous strength" of these nodes at each of 12 instants in time ($t_1, t_2, \ldots, t_{12}$), sepa-

Table A1
A Hand Simulation of the Processes Underlying Activation of the Timing, Syntax, and Concept Nodes in Figure 4

		Syntax nodes			Concept nodes		
Instants in time	Concept timing node	NOUN PHRASE node	ADJECTIVE node	NOUN node	*Frequent practice* [noun phrase]	*Frequent* [adjective]	*Practice* [noun]
t_1	0	20	0	0	20	0	0
t_2	100	100	50	50	100	20	20
t_3	0	0	20	0	100	25	25
t_4	100	50	100	50	100	100	30
t_5	0	0	0	20	100	100	35
t_6	100	50	50	100	100	100	100
t_7	0	0	0	0	0	100	100
t_8	100	50	50	50	0	100	100
t_9	0	0	0	0	0	0	100
t_{10}	100	50	50	50	0	0	100
t_{11}	0	0	0	0	0	0	0
t_{12}	100	50	50	50	0	0	0

Note. The numbers represent node strengths (see text for explanation).

rated by equal but arbitrary intervals. These strength values vary from 0 (resting or spontaneous level) to 100 (full activation), with intervening values representing priming. Time t_1 represents the hypothetical state of the nodes when NOUN PHRASE and *frequent practice* [noun phrase] have been primed or readied for activation.

Timing, syntax, and concept nodes share three characteristics: their strength level during activation is 100, following activation it is 0, and priming from a single source summates to a maximum (asymptotic) value of 50. All other characteristics of concept versus timing and syntax nodes differ. At the time of activation, a concept node adds 20 units of priming to its connected nodes, whereas syntax and timing nodes add 80 units of priming to their connected nodes (within limits discussed below). For the sake of simplicity, these initial cross-connection increases in node strength are counterfactually assumed to occur instantaneously . For example, the concept timing node has become activated at t_2 and activates NOUN PHRASE, which simultaneously passes its 80 units priming to *frequent practice* [noun phrase].

Activation is self-sustained over five time intervals in the case of concept nodes (unlike syntax and timing nodes) and during that time adds five priming units per interval to connected concept nodes. The effect of this anticipatory priming (see text for discussion) is especially evident for the concept node *practice* [noun] over the period from t_2 to t_5. (Nonlinearities and differences in rise times and asymptotes as a function of prior practice [see Figure 3] are omitted here, again for the sake of simplicity.) A relatively slow decay rate for unmaintained priming is assumed for concept nodes, but a virtually instantaneous decay rate is assumed for timing and syntax nodes. The inhibitory interaction between ADJECTIVE and NOUN leaves ADJECTIVE with full priming and the NOUN with 0 priming and is likewise represented as occurring instantaneously. The inhibited syntax node (NOUN) only assumes its full priming value (20) after its inhibiting syntax node (ADJECTIVE) is no longer activated (t_5 in Table A1).

Received February 28, 1982
Revision received May 4, 1982 ■