Psychological Review

VOLUME 86 NUMBER 4 JULY 1979

On the Time Relations of Mental Processes: An Examination of Systems of Processes in Cascade

James L. McClelland
University of California, San Diego

This article examines the possibility that the components of an information-processing system all operate continuously, passing information from one to the next as it becomes available. A model called the *cascade model* is presented, and it is shown to be compatible with the general form of the relation between time and accuracy in speed–accuracy trade-off experiments. In the model, experimental manipulations may have either or both of two effects on a processing level: They may alter the rate of response or the asymptotic quality of the output. The effects of such manipulations on the output of a system of processes are described. The model is then used to reexamine the subtraction and additive factors methods for analyzing the composition of systems of processes. The examination of the additive factors method yields particularly interesting results. Among them is the finding that factors that affect the rates of two different processes would be expected to have additive effects on reaction times under the cascade model, whereas factors that both affect the rate of the same process would tend to interact, just as in the case in which the manipulations affect the durations of discrete stages. On the other hand, factors that affect asymptotic output tend to interact whether they affect the same or different processes. In light of this observation, the conclusions drawn from several studies about the locus of perceptual and attentional effects on processing are reexamined. Finally, an outline is presented of a new method for analyzing processes in cascade. The method extends the additive factors method to an analysis of the parameters of the function relating response time and accuracy.

When we analyze performance in an information-processing task, we often proceed by assuming that performance may be decomposed into a set of separate subprocesses. Sternberg (1969a), following Donders (1868–1869), has noted that we can attempt to study the supposed component processes themselves using reaction-time data if we make some additional assumptions about their temporal relations. In Sternberg's formulation, the important assumptions are (a) that only one component process may be active at any one time, and (b) that the amount of time taken up by one component process does not influence the time required for another. In practice, these assumptions have usually been embodied in a model of performance in which the subprocesses are identified as successive temporal *stages*, each of which occupies a separate interval of time. I call this model the *discrete stage model*.

Explicitly or implicitly, the discrete stage model is the cornerstone of a huge experimental literature addressing itself to the nature and organization of mental processes. The logic underlying this literature is direct and compelling. If the discrete stage model is correct, and if we can find two tasks that differ in that a particular component process is used in performing one and not the other, then the difference in the time it takes to perform the two tasks will indicate the duration of the process that differs between them (Donders, 1868–1869). By the same token, if we can find one manipulation that affects the duration of one stage and a second manipulation that affects the duration of another, then the effects of these two manipulations should combine additively and there should be no interaction (Sternberg, 1969a).

This logic has been used as the rationale for some very important theoretical conclusions. I will mention two examples. (a) Hunt has studied the difference in reaction times between two tasks, one of which requires accessing information in memory and the other of which does not (Hunt, 1978; Hunt, Frost, & Lunneborg, 1973; Hunt, Lunneborg, & Lewis, 1975). The difference in reaction times between these two tasks is smaller for individuals who score high on verbal abilities tests. According to discrete stage logic, this difference indicates that the speed of memory access is a source of individual differences in verbal ability. (b) Meyer, Schvaneveldt, and Ruddy (1975) have studied the time it takes subjects to recognize visually presented words. They have found that semantic context facilitates word recognition and that the magnitude of this effect is larger when the visual quality of the word is degraded than when it is normal; that is, the effects of visual quality and semantic context interact. According to discrete stage logic, this interaction means "that semantic context and stimulus quality influence a common stage" (Meyer et al., 1975, p. 107).

In this article I consider what happens if we take an alternative approach. I introduce a model called the *cascade model* to represent systems of processes that operate continuously to let their outputs reflect the situation represented by their inputs, and I show that specific versions of the general model are compatible with the literature on the relation between response time and accuracy. Then I reexamine the logic of the subtraction method and the additive factors method. It turns out that many of the conclusions based on the discrete stage model, including those mentioned above, are open to question. The difference in reaction time between conditions differing by the inclusion or exclusion of a single process does not necessarily provide an unambiguous indication of the rate or efficiency of the manipulated process. Similarly, the manipulation of the parameter of two different processes sometimes produces additive effects and sometimes produces interactions. Finally, I suggest a method for analyzing systems of processes in cascade, extending the additive factors method of Sternberg (1969a) to the parameters of the curve relating time and response accuracy.

Do Component Processes Take Place in Strict Succession?

The basic assumption of the discrete stage model is that the components of an information-processing act take place in strict succession. Is this assumption correct? Consider as an example, the lexical decision task used by Meyer et al. (1975) and many others (Meyer & Schvaneveldt, 1971; Rubinstein, Garfield, & Millikan, 1971). The task is simply to determine whether a visually presented letter string is a word or a nonword and to press one response key if the string is a word and another if it is a nonword. We might postulate several subprocesses in the performance of this task. For example, we might assume that there is an initial light analysis process to determine the pattern of light and dark; a feature analysis process; letter identification, word identification, and decision processes; and response

Preparation of this article was supported by National Science Foundation Grant BNS 76-14830 to the author and by Grant PHS NH 15828 to the Center for Human Information Processing, University of California, San Diego.

I would like to thank Jon Baron, Albert Corbett, Tom Erickson, Heidi Feldman, George Furnas, Robert Glushko, Jonathan Grudin, James Hollan, Mark Jackson, Jeff Miller, Donald Norman, Roger Ratcliff, Wayne Wickelgren, and especially David Rumelhart for their useful comments and suggestions. I would also like to thank Mark Wallen for help with the programs that generated the figures and Marguerite Moreno for assistance with the production of the manuscript.

Requests for reprints should be sent to James L. McClelland, Department of Psychology, C-009, University of California, San Diego, La Jolla, California 92093.

selection and execution processes. Having postulated this or any other specific set of processes we can ask: Is there any reason to adopt a discrete stage model and assume that none of the processes can begin until the preceding process is completed? Is it even necessary to imagine that *any* of the processes must wait until the preceding process is completed?

It is, of course, widely accepted that some processing operations may take place in parallel. For example, it has often been suggested that subjects can analyze several different dimensions or attributes of a visual input in parallel. (See Egeth, 1966, for a discussion of these models.) But when processes are logically contingent upon each other, as many of them would appear to be in the lexical decision task, the assumption of successive processing seems more compelling. We naturally assume that subjects would not be able to identify the letters in a stimulus without the results of feature analysis. Thus, the assumption of a strict succession of at least some of these processes seems to follow from the logical requirements of the task.

Recently, however, a number of theorists have questioned the view that one component of processing must be completed before a second can start, even when the second process depends on the output of the first. For example, Norman and Bobrow (1975) suggest that the output of a process may be continually available to other processes. In their formulation, the output of each process could be a set of quantities, each one indicating the degree of confidence that one of several possible conclusions about the input is correct. For example, at some instant in time, the output of a feature analysis process might indicate a 20% chance that there is a vertical line on the left of the input pattern and a 5% chance that there is a horizontal line across the middle. A bit later, the same outputs might indicate values of 35% and 60%. Given that the outputs are always available, there is no reason why the letter identification process could not be using them as they are changing. In general, there is no logical reason why any number of contingent processes should not in fact be operating at the same time so long as the outputs of each process are always available to the others.

One very influential model that postulates this type of contingent relationship between processes has been proposed by Turvey (1973). In his model, a peripheral visual process passes information about crude features of the input to a central information-processing mechanism as the information is extracted from the visual input. The central mechanism monitors the output of the peripheral process to determine the identity of the contents of the display. The rate at which information is passed on to the central mechanism is affected by the contrast, brightness, and other physical parameters of the stimulus. The rate of processing at the central level depends on the availability of input from the peripheral level. Turvey uses the term *parallel–contingent* to describe the relation between the central and peripheral processes; the processing at the central level is contingent on the results of processing at the peripheral level, even though it actually takes place in parallel with it.

Turvey's assumption of parallel–contingent processing is implicit in Selfridge's (1959) *pandemonium* model of pattern recognition. Pandemonium postulates several different levels of processing, and each level makes use of the results of processing at the immediately preceding level. Each level actually consists of a set of *demons*, or detectors. The demons work continuously to evaluate evidence provided by the activations of demons at the preceding level and to update their own activations. Each demon's activation is a continuous variable whose value represents the extent to which the input to the demon is consistent with the attribute of the input the demon stands for. Several investigators have now applied models like Selfridge's to account for the process of word recognition (Henderson, 1977; LaBerge & Samuels, 1974; McClelland, 1976).

Most of the applications of parallel–contingent processing have been in models of perceptual processes, but it can easily be extended to memory processes as well. It has frequently been suggested that memory consists of a network of associative links joining memory units called *nodes*. Retrieval of information from such a network is accom-

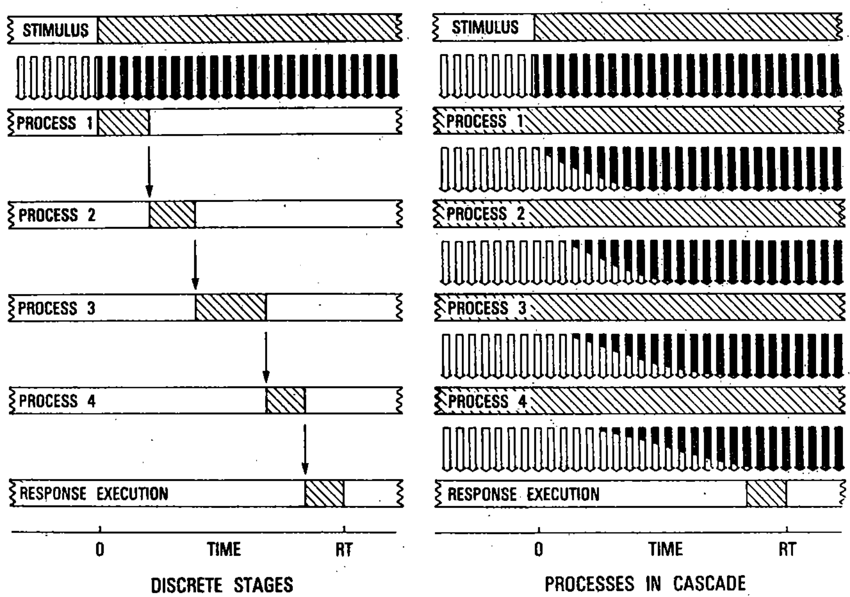

Figure 1. The events that occur between the presentation of a stimulus and the execution of a response, according to the discrete stage model and the cascade model. (Arrows represent the transfer of information from one process to the next, and shading is used to indicate when a process is at work. The blackening of the arrows indicates the degree to which the signals represented by the arrows reflect the stimulus input. Note that in the cascade model, information is transferred between processes all of the time. This aspect of the model is only imperfectly represented by the finite number of arrows.)

plished by the spread of activation from node to node via the associative links (Collins & Loftus, 1975; Collins & Quillian, 1969; Quillian, 1968). There is no reason at all to presume that the activation of a node in the network needs to be all or nothing. In fact, Wickelgren (1976) has suggested that the level of activation of a particular node to which activation is spreading may increase with time to some asymptotic value. Suppose the activation of a node depended on the strength of the input to it—that is, upon the degree of activation of its predecessor. Then if several such activation processes were concatenated in an associative chain, we would have a perfect example of parallel-contingent processing.

General Assumptions of the Cascade Model

In spite of the plausibility of parallel-contingent processing systems in several areas of perception and cognition, we have no general way of representing them. The model presented here is an attempt to bridge the gap. The model is designed to capture the properties of systems of processes conforming to the following postulates:

1. The system is composed of several subprocesses or processing levels.
2. Each subprocess is continuously active, working to let its outputs reflect the best conclusions that can be reached on the basis of its inputs.
3. The output of each process is a set of continuous quantities that are always available for processing at the next level.
4. Processing at each level is based on the results of processing at the preceding level only. Outputs are passed in only one direction through the system of processes, with no skipping or bypassing of subprocesses.

When a set of processes conforms to these assumptions, the processes are said to be organized in cascade. The model presented embodies these assumptions, so it is called the *cascade model.*

To relate the cascade model to behavior, it is necessary to add some way of translating the outputs of internal processes into responses. It is therefore necessary to add some assumptions like the following:

5. The output of a final continuously active process called the *response activation process* indicates which of the possible alternative responses should be executed.

6. Actual response execution is assumed to be a discrete event that adds the duration of a single discrete stage to the time between the presentation of the stimulus and the registration of the overt response.

Figure 1 contrasts a cascade model and a discrete stage model of the events taking place between the presentation of a stimulus and the registration of a response in a schematic reaction-time experiment. Both models postulate a number of component processes, and both assume that information flows in only one direction with no skipping of levels or feedback. The cascade model also shares with the discrete stage model the assumption that the execution of a response is a discrete event. However, in the discrete stage model, only one process is at work at a time, whereas in the cascade model, all processes except response execution are at work all of the time. In the discrete stage model, information is transferred at the termination of one process and the start of the next. In the cascade model, the transfer of information between processes is taking place all of the time.

The Cascade Model

The presentation of the details of the cascade model begins with some necessary background assumptions about the structure and function of individual processing levels. This is followed by a discussion of the central assumptions of the model concerning the temporal dynamics of processing. These ideas are drawn from the theory of linear differential equations, which has many applications in physics. In psychology it has been used in formulating models of flicker and luminance increment detection (Sperling, 1964; Sperling & Sondhi, 1968) and has been extended to account for some simple masking phenomena (Ganz, 1975).

Processing Levels and Processing Units

Following Postulate 1, the model assumes that the system underlying performance in a task is composed of a number of processes or processing levels. A simplified example of a processing system for performing the lexical decision task is illustrated in Figure 2. Each processing level consists of a number of processing units. At perceptual levels, the units are like demons or detectors for the properties of stimuli. At higher levels not necessarily directly involved in the lexical decision task, units might correspond to representations of semantic features or nodes within a semantic network. There are also units for making comparisons and decisions and units for indicating which response should be performed.

Information as activation. Units accumulate information in the form of activation, as in pandemonium. For example, perceptual units that become positively activated as a result of the presentation of a stimulus signal the presence of the unit they represent in the stimulus. Perceptual units that become nega-

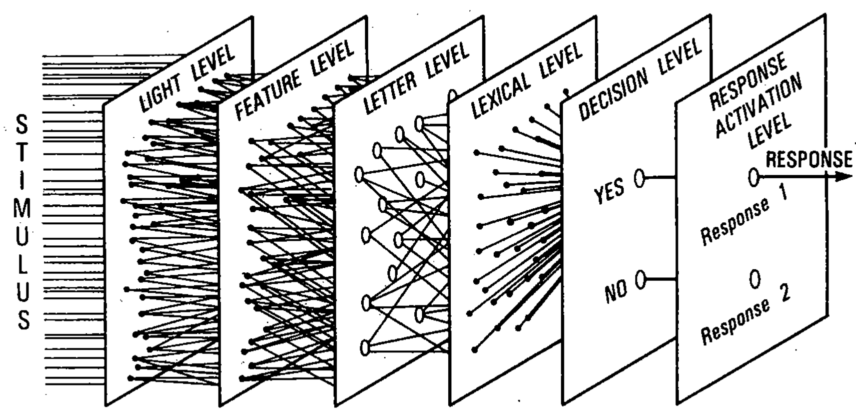

Figure 2. Possible connections between units at different levels of processing in a simplified hypothetical system for determining whether a string of letters is a word or a nonword.

tively activated signal the absence of the unit they represent. The same general characterization applies to units at other levels as well. The level of activation of a response unit signals the appropriateness of the corresponding response, and so on.

Linear integrators. The derivation of the cascade equation, which represents the time course of activation of units at any arbitrary processing level, is based on the assumption that all the units at each processing level are linear integrators. Linear integrators are very simple general-purpose processing units. They simply take a weighted sum of a subset of the outputs of the units at the preceding level (Anderson, 1977; Anderson, Silverstein, Ritz, & Jones, 1977). They can be used to perform a variety of perceptual and cognitive functions. Cortical line detector cells are linear integrators, at least over some range of operation: They combine excitatory inputs and inhibitory inputs to give their maximal response to a particular line orientation in a particular position. Linear integrators can also be used as detectors for higher order units of all types, including letters, words, propositions, and others. Of course, more sophisticated types of units may be required to carry out various types of processing operations, but in many cases their dynamic properties, which are the focus of interest here, will be quite similar to those of linear integrators. The "maximum" units discussed in the Appendix are one example; for another see the discussion of "comparitor" units in McClelland (Note 1).

Strength and weight constants. The strength of the effect exerted by one unit on another is represented in terms of the weight constants in the weighted sum taken by the linear integrator. Consider a unit at some level n and index the unit by j. The asymptotic effect of the units at level $n-1$ on unit nj (represented by a_{nj}) is given by

$$a_{nj} = \sum_{j'} w_{j'j} a_{(n-1)j'}, \qquad (1)$$

where $w_{j'j}$ is the weight of the effect of the activation of unit j' at level $n-1$ on unit j at level n. So, for example, there will be weight constants for the effects of feature detectors on letter detectors, for the effects of letter detectors on word detectors, and for the effects of word detectors on decision units. When a stimulus is presented to someone in a task like the lexical decision task, the weight constants, in conjunction with the characteristics of the stimulus itself, will determine how strongly the stimulus will ultimately activate any particular unit. Consider, for example, the effects of various stimuli on the detector for the word *cat.* If a bright, high-contrast, clearly printed version of the word cat is shown, the detector for cat will presumably reach a high level of activation. Visually degrading the stimulus will have the effect of lowering the effectiveness of the stimulus in activating all of the relevant feature detectors. Similarly, distorting the letters so that they are nearly unrecognizable versions of C, A, and T will have the effect of lowering the effectiveness of the stimulus in activating the letter detectors, even if it is not degraded in the usual sense. Misspellings of the word will lower the effectiveness of the stimulus in activating the relevant word detector, and if an entirely different word is shown, the activation of the detector for cat may even become negative. All of these manipulations of the stimulus then will presumably lower the asymptotic activation of the detector for the word cat compared to the first case. In addition, of course, if the subject's knowledge of the relevant transformations is impoverished at any level, the ultimate activation level will likewise be reduced. For example, if the type font were unfamiliar to the subject, the features would not activate the relevant letter detectors, and therefore the ultimate activation of the detector for the word cat would probably be quite low. As this example suggests, the weight constants are thought to be influenced by learning (at least at some levels) and they can be taken as equivalent to associative strengths.

Thus far we have seen how the cascade model represents the ultimate effect of activations of units at one level on the activations of units at the next level in terms of weight constants. As shown in the Appendix, these weight constants can be put together for all of the units at all of the levels of processing in a system to give a representation of the effects of a simulus on units at any arbitrary level (Anderson et al., 1977). But so far we have been considering only the *ultimate* or *asymptotic* effect of the outputs of one level on the outputs of the next. I now consider how the activations of units reach these eventual levels as a function of time.

The Dynamics of Processing

The inputs to a unit can be thought of as a driving force pushing the activation of the unit toward the level dictated by the inputs. The central assumption of the cascade model is that the rate of activation of a unit depends on the difference between the level its inputs are driving it to and the level of activation the unit has already reached. When the discrepancy is large, the change will be rapid, and when the discrepancy is small, the change will be slow. This assumption characterizes the accumulation of charge in a capacitor and a variety of other simple physical systems. For a *step* input (sudden onset of a stimulus), it results in an exponential approach to asymptote. The assumption can be represented by a very simple differential equation. Consider unit j at level n in a system of processes. Let $i_{nj}(t)$ represent the activation level the input is driving the unit to at time t, and let $a_{nj}(t)$ represent the level of activation the unit has already achieved at t. Then the assumption is represented by the equation

$$\frac{d}{dt}(a_{nj}(t)) = k_{nj}(i_{nj}(t) - a_{nj}(t)). \quad (2)$$

The constant k_{nj} is called the rate constant; it indicates the rate of response of unit nj. The larger the value of k, the faster the activation of the unit will change in response to its inputs.

In the remainder of this article it is assumed that the rate constant associated with each unit at a given level is the same, that is, that $k_{nj} = k_n$ for all j. This very important simplifying assumption allows us to use k_n to represent the general rate of processing at level n.

The temporal form of the input. Since the activations of the units determine the timing and selection of responses, we need to derive an expression describing how these activations will vary as a function of time since the onset of a stimulus. The expression will depend on the temporal form of the stimulus input and on the *initial conditions* of activation of the units at the different levels. For present purposes, the analysis is restricted to stimuli that are suddenly presented at time $t = 0$ and left on until all events of interest have transpired. This situation would arise in most reaction-time tasks in which a stimulus is instantly illuminated at a point in time identifiable as $t = 0$ and extinguished upon registration of a response. In deriving the cascade equation, all initial activations are assumed to be 0. The effects of activations arising from other sources, such as priming and response biases, are discussed below. The activation function can be obtained by solving Equation 2 for Level 1, plugging the result into the equation for Level 2 and solving again, and so on successively up to any desired Level n. A general solution for an arbitrary n can also be obtained using Laplace transformations as discussed in the Appendix.

The cascade equation. The general solution is called the cascade equation. It gives an expression for the activation of unit j at processing level n to a stimulus S presented at time $t = 0$. The expression is

$$a_{nj/S}(t) = a_{nj/S}(1 - \sum_{i=1}^{n} K_i e^{-k_i t}), \quad (3)$$

where $a_{nj/S}$ is the asymptotic activation of the unit that would result if the stimulus were left on indefinitely, and the k_is are the rate constants of the different processes in the system, which are indexed by the subscript i. The K_is are constants attached to the various exponential terms in the sum. They are given by

$$K_i = \prod_{l \neq i}^{n} \frac{k_l}{k_l - k_i}. \quad (4)$$

The cascade equation only holds when the k_is are all different, since when two or more of the ks are identical, the denominators of some of the terms in the product represented by the K_is contain 0s. Formulas for the case in which an arbitrary set of ks are equal may be derived, but there is no practical need for these, since it is possible to achieve any desired degree of approximation to equal ks by using values that are almost, but not exactly, equal.

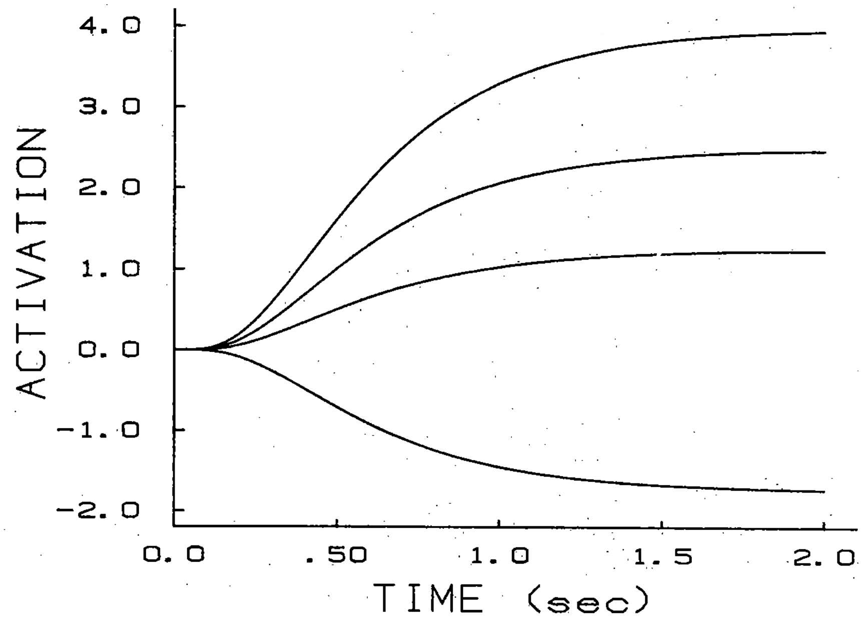

Figure 3. Activation functions for four different units at the fifth level in an arbitrary hypothetical system of processes. (All of the curves are identical up to multiplication by a constant.)

Independence of asymptotic and dynamic parameters. A very important property of the cascade equation is that the terms that vary as a function of time are identical for every unit at the same processing level, under the influence of every possible stimulus. The only part of the expression that depends on which unit is considered and which stimulus is presented is the asymptotic activation of the unit, $a_{nj/S}$. This fact leads to a very important conclusion:

All of the units of the same processing level have the same activation function for every possible stimulus, up to multiplication by a (positive or negative) constant.

This conclusion is illustrated in Figure 3, in which several hypothetical activation functions differing by a multiplicative constant are shown. Although the curves in the figure do grow at different rates because of the effect of the multiplicative constant, each curve reaches a given proportion of its asymptotic height at the same instant in time. In cases such as this, the curves are said to have *equivalent dynamics.*

The dynamic portion of the activation function. Let us now consider the time varying portion of the cascade equation. This expression is the cumulative form of what McGill (1963) has called the general gamma function. I will designate it by $\Gamma_n(t)$, where the subscript represents the processing level whose activation function it describes.

By exploring $\Gamma_n(t)$ we can find out how the activations of units at Level n vary with time since stimulus onset. The function is

$$\Gamma_n(t) = 1 - \sum_{i=1}^{n} K_i e^{-k_i t}. \tag{5}$$

Essentially, $\Gamma_n(t)$ is 1 minus the sum of n negative exponential terms, where each term is weighted by a multiplicative combination of the rate constants. The first thing to note is that as t goes to infinity, all of the exponential terms approach 0, so the expression has an asymptotic value of 1. The asymptotic value, therefore, of the cascade equation is $a_{nj/S}$, as it should be. When $t = 0$, on the other hand, the sum is equal to 1 (McGill & Gibbon, 1965), and $\Gamma_n(0) = 0$. So $\Gamma_n(t)$ always varies from 0 to 1 as t goes from 0 to infinity.

Effects of concatenating processes. The exact shape of $\Gamma_n(t)$ (and therefore the time course of activation of the units of Level n) depends on n and the k_is. There are some general facts about these activation functions, however, that turn out to have very important implications for our understanding of the results produced in experiments investigating the time course of information processing. Figure 4 shows four versions of $\Gamma_n(t)$ for $n = 1$–4. One can think of the curves as describing the activation functions of units at each of four successive levels in a hypothetical system of four processes, with the asymptotes equated so that we may concentrate on the shapes of the curves.

The units at the first level receive the stimulus input directly. Their response to that input is simply the exponential approach to an asymptote. In this case Equation 5 reduces to

$$\Gamma_1(t) = 1 - e^{-k_1 t}. \tag{6}$$

The next curve to the left shows the response of units at the second level. The rate constant for these units is almost exactly the same as the rate constant of the units at Level 1, but their activation function is different because their input depends on the output of Level 1. If the input came directly to the second level, the response would be identical in form to the response of units at the first level. However, the units of the second level get their input from first-level units, which are only slightly activated at early points after stimulus onset. As a result, the activation function for units at the second level grows slowly at first. As the output of the units making up the first process approaches asymptote, the units of the second process receive stronger inputs, so their activation functions grow more quickly. It is for this reason that the curve for $n = 2$ begins with a relatively slow rise and then accelerates into a steeply rising section before finally leveling off. As more and more processes are added, the flat portion gets longer and longer, and it begins to look as though nothing is happening initially at later levels. Actually, the activation is just growing so slowly at first that it is impossible to detect it.

The two processes that contribute to the form of the second curve have nearly identical rate constants, but the third process, which is added to determine the form of the third curve, has a smaller rate constant. This slow rate constant dominates the resulting activation function, making it grow much more slowly, and extending the initial flat portion of the curve. The fourth curve shows the effect of adding another process with a relatively large (fast) rate constant to the system represented by the third curve. Here the effect of adding the additional process is simply to shift the activation function to the right. Although the (vertical) difference in activation between the two curves becomes smaller and smaller as $t \to \infty$, the (horizontal) difference in the time it takes to reach any

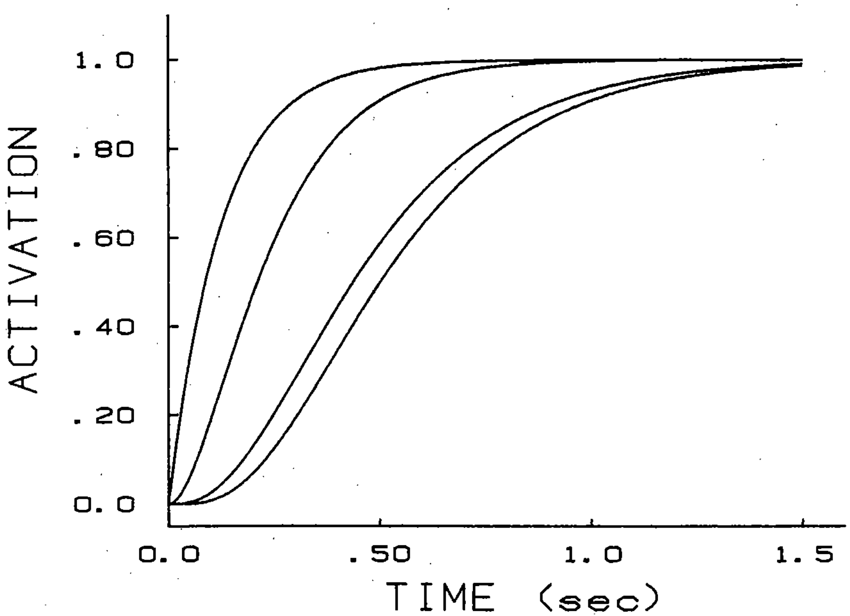

Figure 4. The response of units at the first, second, third, and fourth levels in a hypothetical system to the onset of a stimulus display at time $t = 0$. (The rate constants are 8.0 for the first level, 8.1 for the second, 4 for the third, and 16 for the fourth.)

given level of activation stays remarkably constant once the curves have passed the initial portion of the rise.

The mathematical basis for the fact that the rate of approach to asymptote is dominated by the slowest process in the system is straightforward. The slope of the activation function, that is, the derivative of $\Gamma_n(t)$ is

$$\frac{d}{dt}\Gamma_n(t) = \sum_{i=1}^{n} k_i K_i e^{-k_i t}. \tag{7}$$

As $t \to \infty$, $e^{-k_i t} \to 0$ for all i; but the rate of approach to 0 is greatest for large k_i, so in the limit, the only term left in the sum is the one for which k_i has the smallest value, corresponding to the process with the slowest rate. Thus, we find that

$$\lim_{t\to\infty}\left(\frac{d}{dt}\Gamma_n(t)\right) = k_s K_s e^{-k_s t}, \tag{8}$$

where the subscript s indexes the slowest process. This slope is equal to the slope that we would get for a single process with rate k_s (i.e., $k_s e^{-k_s t}$) shifted over in time:

$$\begin{aligned} k_s K_s e^{-k_s t} &= k_s e^{\ln[K_s]} e^{-k_s t} \\ &= k_s e^{-k_s[t-(1/k_s)\ln(K_s)]}. \end{aligned} \tag{9}$$

The constant $(1/k_s)\ln(K_s)$ determines the size of the shift. Inserting the value of K_s, we find that

$$\begin{aligned} &\frac{1}{k_s}\ln\left(\prod_{i\neq s}^{n}\frac{1}{1-k_s/k_i}\right) \\ &\qquad = \sum_{i\neq s}^{n}\left(\frac{1}{k_s}\ln\left(\frac{1}{1-k_s/k_i}\right)\right). \end{aligned} \tag{10}$$

Since $\ln[1/(1-k_s/k_i)]$ can be approximated by k_s/k_i for small values of k_s/k_i, the size of the shift is approximated by

$$\sum_{i\neq s}^{n}\frac{1}{k_i}. \tag{11}$$

Let us call this temporal shift T_s. What we have seen is that the size of T_s is determined only by the rate constants of the relatively fast processes.

Thus far we have only considered what happens as t approaches infinity. In fact, it can be shown by simplifying a proof for a related situation that arises in biochemical cascades (Sharney, Wasserman, Gevirtz, Schwartz, & Tendler, 1965) that as the rate of each of the relatively fast processes grows large relative to the rate of the rate-limiting process, the form of the activation function approaches the exponential shifted by the sum of the reciprocals of the rate constants, given by

$$1 - e^{-k_s(t-T_s)} \tag{12}$$

for all values of $t - T_s \geq 0$. (For values of t less than T_s, the value of the expression is zero.) Of course, most systems will not typically correspond closely to these limiting cases. However, as Figure 4 illustrates, the approximation is adequate as long as the slowest process has a rate constant one-fourth the size of the relatively fast processes under consideration.

Thus, we observe an important general property of systems of processes in cascade:

Relatively slow or rate-limiting processes determine the slope of the activation function, whereas relatively fast ones determine where it will begin to rise without altering its slope.

It is worth noting two additional properties of the activation functions given by $\Gamma_n(t)$. The first is that their form depends only on the values of the rate parameters and not on their arrangement. Thus, the shape of the activation function of a particular unit tells us nothing about the arrangement of the processes in the system. The rate constants of the n processes could be permuted in any order and the same activation function would result. In addition, it should be apparent that the general form of the activation function will not be very useful for determining the value of n (McGill, 1963). In the absence of converging evidence, it is generally possible to replace a process that is not rate limiting with a number of faster ones and obtain equivalent results.

Does the Model Fit the Data?

The cascade model is a plausible alternative to the discrete stage model, and so it would be of interest, merely on the basis of its plausibility, to determine what its implications would be for the interpretation of the results of reaction time experiments. However, before we undertake such an exercise, it is worth establishing whether the model is compatible with what we know about the time course of information processing. There is a small but growing literature in which the accuracy of responding is measured at several points in time after the presentation of a stimulus. These studies permit us to compare the shapes of empirically obtained time–accuracy curves to the curves that would be generated by a system of processes in cascade.

Response timing in speed–accuracy experiments. In the studies we will consider, subjects are induced in either of two ways to respond at particular points in time after the presentation of a stimulus. One method (the *response-signal* method; Reed, 1973, 1976; Shouten & Bekker, 1967) is to present a response signal at different lags after stimulus presentation and to instruct the subject to respond within a brief period of time after the presentation of the signal. The other method (the *deadline* procedure; Fitts, 1966) is to instruct the subject to respond before a designated amount of time passes, providing feedback about whether the deadline has been met. Here, presumably, the response is induced when an internal temporal criterion set in response to the instruction is met. Both of these cases amount to a decision to respond based on time alone, as illustrated by the vertical line in Figure 5.

The data generated in a particular deadline or signal lag condition are plotted as a single point. For simplicity, it is assumed that the responses represented in each point are all generated at the same time after stimulus onset and that the variability in the reaction times simply represents variability in the time it takes to execute the response. Execution itself does take some time of course; McGill (1963) suggests a mean of about .1 sec, and I have adopted this value. As an approximation, mean execution time is assumed to be invariant over stimulus and signal lag conditions. Response latencies are longer for shorter signal lags, but this may be because subjects cannot help delaying response execution in short signal lag conditions, in which their responses are based on very little information and a little more time has a chance of producing a big improvement in accuracy. An alternative possibility is that the processing of the response signal itself is delayed when it comes immediately after the target. In any case, I assume that the mean reaction time measured from stimulus onset (i.e., signal lag plus latency) reflects the state of the activations of the response units .1 sec before.

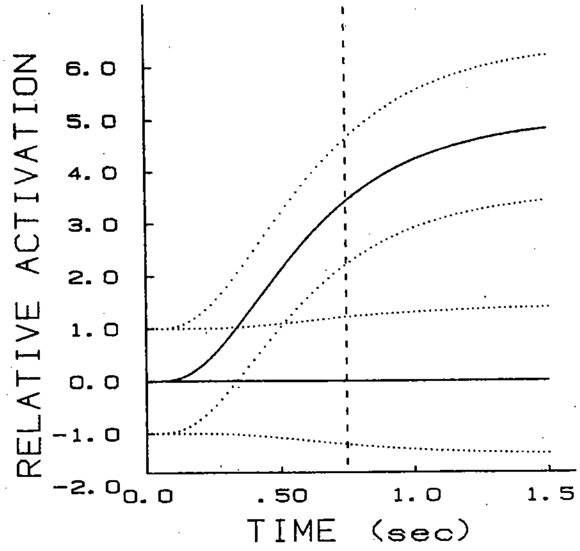

Figure 5. Hypothetical activation functions for the response unit associated with a yes response in a yes–no task such as the lexical decision task. (The solid rising curve represents mean relative activation for stimuli associated with the yes response; the solid flat curve is the mean relative activation for stimuli associated with the no response. The dotted lines indicate the variability in these mean activation functions; each pair indicates one standard deviation above and below the mean. Five processes are assumed with rate constants of 3, 6, 12, 24, and 48. The standard deviation due to stimulus variability is set to 1.0.)

Response selection and sources of variability. Response selection in the time-controlled situation is straightforward; subjects simply execute the response corresponding to the most strongly activated response unit at the instant the decision to execute the response is made. Errors in response execution are assumed to arise from the presence of variability in the activations of response units. Two sources of variability are assumed. The first arises from extraneous inputs to the response units. These extraneous inputs result in what are called *base activations* of the units. A variety of factors might give rise to base activations, including priming and context effects, expectations, and anticipations. When they are unsystematic, these influences will act as noise in the set of activations; when they are systematic, they will introduce biases in the activations of units.

The model assumes that the base activations stay constant within an individual trial, so the effects of inputs from the preceding level are simply added onto them. It is assumed for simplicity that the effect of the complete ensemble of base activations is to produce uncorrelated base activations of the response units drawn from a normal distribution with unit variance. This assignment scales the activations produced by input. When there is no systematic bias, the mean of the distribution will be set to zero.

The second source of variability arises from the stimuli themselves. We can expect that different stimulus items in the same condition (e.g., different words or nonwords in a lexical decision task, different old or new probes in a recognition task, etc.) will tend to drive the units in the system to different asymptotic activations (Ratcliff, 1978). The effect of differences between stimuli is to add random variability to the asymptotic activations produced in a given response unit in a given experimental condition. The variability is assumed to be distributed normally, with standard deviation σ. The value of σ may vary between experiments. When the effects of stimulus variability are not homogeneous across stimulus conditions, σ stands for the pooled variability.

Now consider the activation of a response unit in a typical processing system. Figure 5 indicates both the central tendency and the range of variability of hypothetical activations of the r sponse unit associated with a yes response in a yes–no experiment, such as the lexical decision task, at various points in time after the presentation of a stimulus associated with the yes response (e.g., a word; rising curve) or a no response (nonword; flat curve). I have arbitrarily set the mean asymptotic activation to 5 for yes trials and to 0 for no trials. Although there may be some general increase or decrease in the activations of the response unit on no trials, the 0 point is arbitrary, since only the difference in activation turns out to matter. The stimulus variability parameter $\sigma = 1$ so that both sources of variability make equivalent contributions.

Accuracy as a function of time. In tasks that require a yes–no decision, we might assume, following signal detection theory, (Green & Swets, 1966; Swets, Tanner, & Birdsall, 1961) that the subjects establish a criterion and execute a yes response if activation of the yes response unit exceeds the criterion and a no response if it does not. It seems reasonable to imagine that subjects adjust the criterion as time goes on, effectively asking, "Is the present activation level more indicative of the activation level I should expect at this point in time if the correct answer is yes?" This criterion adjustment is tantamount to assuming that the no response unit is activated by the mere passage of time since stimulus onset.[1] In this situation, the probability of choosing the yes or no response will depend on the temporal form of the variation in the activation of the no response unit. Even in the absence of any assumptions about the form of this variation, we can use signal detection theory to derive a formula for the accuracy of responses initiated at time t. Since responses initiated at time t actually get recorded .1 sec later, the observed values of d' at time t are based on activations of response units .1 sec earlier. Let $\hat{a}_{y/y}$ represent the mean asymptotic activation of the yes response unit given a presentation of a stimulus appropriate to that response, and let $\hat{a}_{y/n}$ represent the mean asymptotic activation of the yes response unit given a stimulus appropriate to the no response. Then the observed value of d' at time t is given by

$$d'(t) = (\hat{a}_{y/y} - \hat{a}_{y/n})\frac{\Gamma_n[t-.1]}{\sqrt{1+\sigma^2(\Gamma_n[t-.1])^2}}, \tag{13}$$

where $[t-.1] = t-.1$ for $t > .1$ and 0 otherwise. The derivation of this expression is given in the Appendix. A similar expression can

[1] In ordinary reaction-time tasks, *no* responses are influenced by the similarity of the stimulus to the items in the *yes* response category (Ratcliff, 1978). In such cases, it seems reasonable to suppose that those activations that are exciting the yes response unit might also be inhibiting the no response unit. Although such influences may be operative in the case of time-controlled responding as well, they would not materially alter the results.

be derived for the simple case in which the subject has to categorize the stimulus into one of two alternative categories.

The Shape of the Time–Accuracy Curve

The solid curve in Figure 6 illustrates the time–accuracy curve given by Equation 13 for the parameters used in generating the response activation functions depicted in Figure 5. The shape of this curve is characteristic of the curves generated by Equation 13. As with the underlying activation functions themselves, this theoretical time–accuracy curve stays close to chance for a few hundred msec and then increases rather rapidly over a short period, finally leveling off at some asymptotic level. Empirical time–accuracy curves reported in the literature all have this general form (see Pachella, 1974; Pew, 1969; and Wickelgren, 1977, for reviews). In fact, as with the underlying activation functions, the time–accuracy curves given by Equation 13 can be closely approximated by an exponential approach to asymptote following a delay, as illustrated by the dashed curve in the figure. Wickelgren (1977) has noted that empirical time–accuracy curves are quite closely approximated by such a delayed exponential function, given by

$$d'(t) = a(1 - e^{-k[t-t_d]}), \tag{14}$$

where $[t - t_d] = t - t_d$ for $t > t_d$ and 0 otherwise. The parameter a determines the asymptote of the curve, k determines its rate of approach to that asymptote, and t_d represents the duration of the delay before the curve begins to rise above chance. The only notable difference between the curve generated by Equation 13 and Wickelgren's curve is that the former has a small initial rise into the steep rise rather than an abrupt transition. An examination of empirical time–accuracy curves reveals that this subtle lip is in fact present (McClelland, Note 1). Generally, then, the shape of the time–accuracy curve seems to be very close to what we would expect from the cascade model.

Effects of Experimental Manipulations

Several experiments have examined the effects of different sorts of experimental manipulations on the form of the time–accuracy curve. The effects of these manipulations are consistent with plausible variations

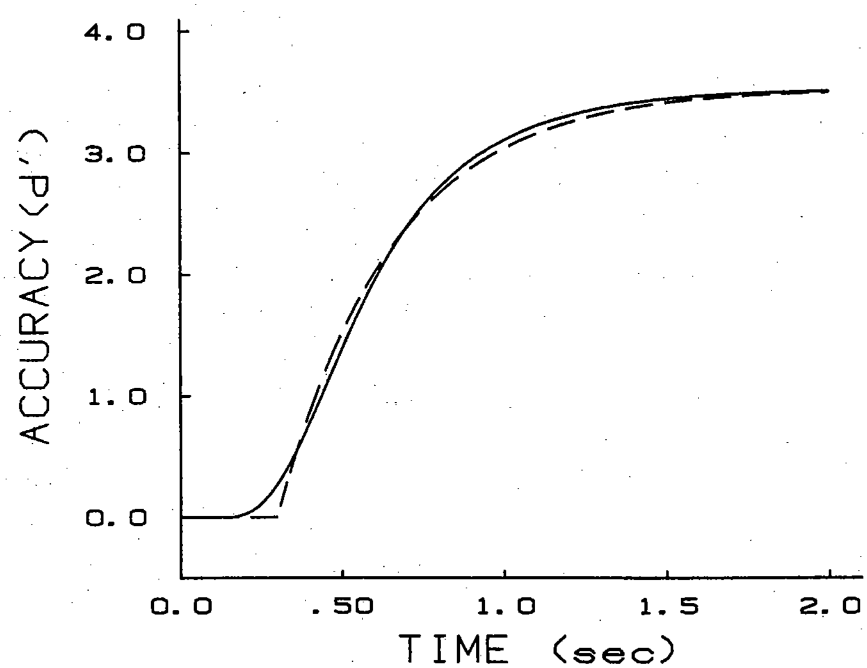

Figure 6. The solid curve shows the relationship between time and accuracy for the hypothetical system of processes that generated the activation functions represented in Figure 5; the dashed curve is fit to the solid curve, using Wickelgren's equation.

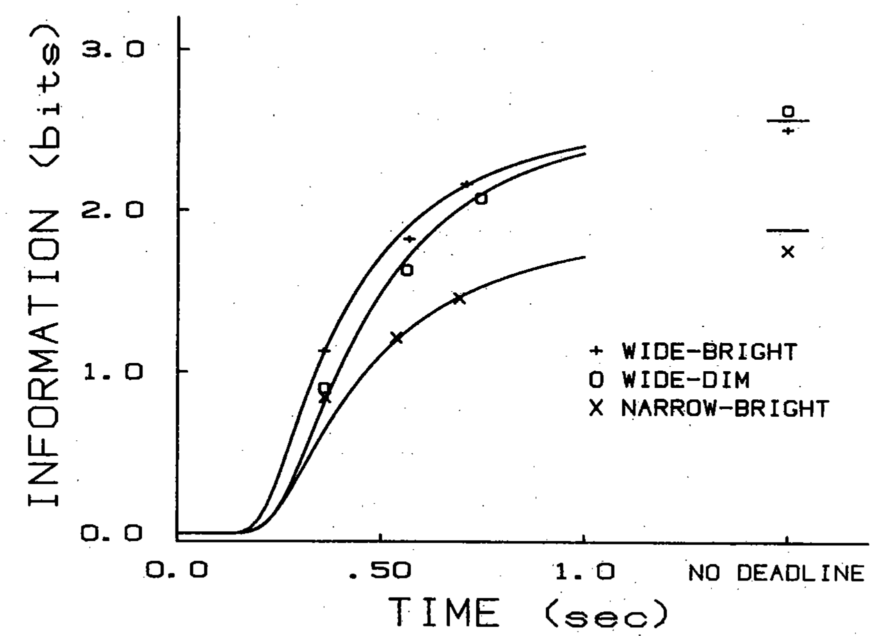

Figure 7. Data from Pachella and Fisher (1969), along with time–accuracy curves generated from the model described in the text. (The relation between activation and response accuracy is based not on Equation 13, but on Luce's [1963, pp. 168–177] method for translating the exponential transform of the activations given by Equation 15 into the bits of information measure used by Pachella and Fisher. As in the case of Equation 13, responses at time t are assumed to be based on activations of response units at time $t - .1$ sec. See McClelland [Note 1] for further details. Three processes are assumed to underlie performance, with the rate of the slowest process equal to 2.7; the rate constants of the other processes are 20 and 19.9 in the bright–wide and bright–narrow conditions, 9 and 19.9 in the dim–wide condition. The asymptote parameter is the same for both of the wide spacing conditions, but smaller for the narrow spacing condition. The fits to the points in the *no-deadline* condition, in which responses were not timed, are simply the asymptote parameters for the conditions in question.)

in the parameters of the underlying activation function entering into the time–accuracy function given by Equation 13.

Before considering specific cases, some important basic facts should be noted. For a fixed value of σ, Equation 13 is similar to the cascade equation, in that factors affecting the asymptote are separated from factors affecting the dynamics of processing. Asymptotic d' is determined by the *difference* in asymptotic activation of the yes response unit by stimuli associated with the yes and no responses. Factors that would affect the size of this difference (degradation, confusability of stimuli, strength of association, etc.) would not be expected to affect the dynamics of the time–accuracy curve, but only its asymptote. In addition, it is easy to show that the time–accuracy curve is affected in the same way as the underlying activation functions by factors that affect the dynamics of processing. Insertion of a relatively fast process, or manipulation of the rate of such a process, will simply shift the time–accuracy curve along the time axis, whereas insertion of a relatively slow process or manipulation of its rate will alter the rate of approach to asymptote.

These general facts are the basis of the interpretations of the effects of experimental manipulations described in the following sections. In each case, I describe in qualitative terms a plausible interpretation of the effects of the various experimental manipulations and show that these interpretations are consistent with the effects of the corresponding parameter manipulations on the shape of the curve generated by Equation 13. In fact, in each case the data may be closely approximated by the cascade model, as illustrated by the curves in the figures.

The effects of spacing and contrast on absolute judgment. Pachella and Fisher (1969) studied subjects' judgments of the position of a vertical bar under different stimulus and deadline conditions. On each trial, they presented a single black bar on a white background. The bar could appear in any 1 of 10 equally spaced positions along a horizontal line, and the subjects had to press 1 of 10 keys to indicate their judgment of the position.

Three different visual conditions were run. In one, the bars were spaced widely apart, and the display was brightly illuminated. In a second, the spacing of the bars was reduced. In a third, the wider spacing of the first condition was used, but luminance was reduced. There were three response deadline conditions, as well as a no-deadline condition in which subjects were instructed to respond as accurately as possible without regard to time.

The results, shown in Figure 7, indicate that the spacing manipulation lowered asymptotic accuracy without affecting the shape of the activation function, whereas the luminance manipulation shifted the accuracy function to the right without affecting the asymptote.

The effect of spacing on accuracy makes sense if we assume that the presentation of a particular stimulus partially activates the response units for neighboring stimuli as well as the correct response unit and inhibits response units for stimuli that are some distance away. Specifically, let j index the stimuli and corresponding responses, numbering from left to right. When stimulus j is presented, the activation of response unit j' is given by

$$a_{rj'/j}(t) = a_{rj/j}(t)(1 - s|j - j'|), \quad (15)$$

where s depends on the spacing of the stimuli. (This hypothesis is adapted from Luce, 1963.) Activation will be maximal for the appropriate response unit (i.e., when $j = j'$) and will fall off linearly on either side. The closer the spacing, the smaller s will be. Since accuracy of response selection is determined by relative activation levels, the effect of this assumption will be to reduce asymptotic accuracy when spacing is narrow.

Now consider the effect of the luminance manipulation. The first impulse might be to expect that this manipulation would also affect the asymptote of the curve, but the results do not support this impulse. Asymptotic performance in the dim condition was, if anything, slightly more accurate than asymptotic performance in the bright condition, though the difference was not significant. In view of this, we must rethink the matter. It is likely that a wide range of variation in luminance would be compatible with effectively optimal sensory registration. Of course, if luminance is reduced sufficiently, the stimulus will not drive the light analysis process to this maximal effective level. But when the effective ceiling is reached, the luminance manipulation will only affect the time it takes for the output of the light analysis process to reach its maximal effective level. The luminance manipulation can therefore be approximated as a manipulation of the rate parameter of the light analysis process (see Appendix). Such an effect of the luminance manipulation would produce the obtained shift in the time–accuracy curve, as illustrated in Figure 15.

Effect of imagery on retrieval of paired associates. Corbett (1977) used the response-signal method to examine the retrieval of associations studied either by rote rehearsal or by the use of mental imagery. In trials during the test phase of his experiment, pairs of words were presented, and subjects were instructed to determine whether the pair was one they had studied or not (foils were incorrect pairings of words from the learned pairs). After a variable lag, the response signal was presented, and subjects were instructed to respond within the next 200 msec.

The data from the tests for both types of items are shown in Figure 8. The results were closely fit by Wickelgren's equation, but the parameters differed between conditions. A good fit to the data was obtained by holding the rate parameter constant across rote and imagery conditions. The asymptote parameter was higher in the imagery condition, but the delay parameter was longer. Apparently, then, the imagery manipulation increased asymptotic accuracy but shifted the time–accuracy function to the right.

Corbett (1977) suggested that the results might be accounted for by assuming that imagery either slowed some process or resulted in the addition of an extra process into the processing system. A cascade model that can account for the obtained time–accuracy curves can be generated from either of these hypotheses. Insertion of a process or reduction of the

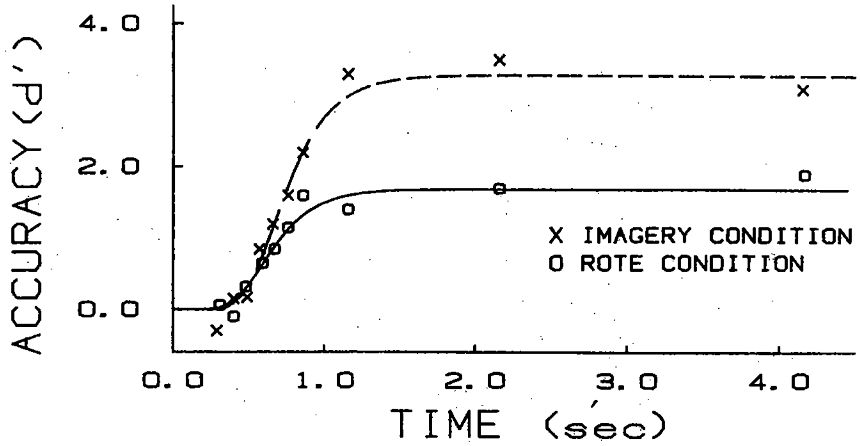

Figure 8. Combined data from early tests of three subjects in Corbett (1977), along with curves generated by the model described in the text. (Both curves share seven processes in common with rate constants of 6, 5, 15, 15.1, 15.2, 15.3, and 15.4. The curve fitting the data for the imagery condition includes an additional process with a rate constant of 9.)

rate of a process both have the effect of shifting the time–accuracy curve to the right, increasing the delay parameter of the best-fitting Wickelgren curve. The curves illustrated in Figure 8 are based on the assumption that the manipulation resulted in the insertion of a process whose rate ($k = 9$) was only slightly slower than the rate of the rate-limiting processes (the two slowest processes in the simulation had rates of 5 and 6).

Set-size effects in immediate memory. Reed (1976) investigated the time–accuracy function for recognition of members of a predesignated set of letters held in immediate memory (Sternberg, 1966, 1967). The experimental factor, as in the standard Sternberg task, was the number of items in the memory set.

One model for this task might be formulated as follows Before each trial, the memory set is encoded and a special *comparator* unit is established for each member of the memory set. When the probe is presented, it is encoded, and the output of the final encoding level (e.g., a set of activations of phonological features of the probe item) is input to each of the comparators. The comparator units compute the product of their inputs from the memory set items and the probe, so the more closely the memory inputs and the stimulus-produced inputs to a comparator correspond, the larger its asymptotic activation will be (Anderson, 1973). Under the assumption that the memory inputs to the comparators are static on a given trial, the comparators behave just like linear integrator units (McClelland, Note 1). The outputs of the comparator units provide the input to a decision unit that signals the largest of its inputs and finally provides the input to the response unit associated with the yes response. As in the previous model, the no unit would be activated (if at all) by the mere passage of time. This model is similar to one recently suggested by Ratcliff (1978); although Ratcliff's formalism incorporates only a single continuous process (analogous to the comparison process in this cascade model), the general nature of the account of the effects of experimental manipulations in the Sternberg memory search task is quite similar.

In the present model, at least two effects of the set-size manipulation are possible. An increase in set size might lower the activation levels of the inputs to the comparator units from memory. If these activations are maintained through attention and if attention is limited, it may not be possible to keep several sets of inputs at maximal levels. Lower memory inputs would result in smaller asymptotic outputs of the comparator units, thus reducing the relative asymptotic output of the response units. Alternatively, it is possible that increasing set size reduces the rate of the memory comparison process. It may take some limited processing capacity to drive several different comparator units simultaneously. If so, increasing the number of comparisons would slow the rate of the comparison process.

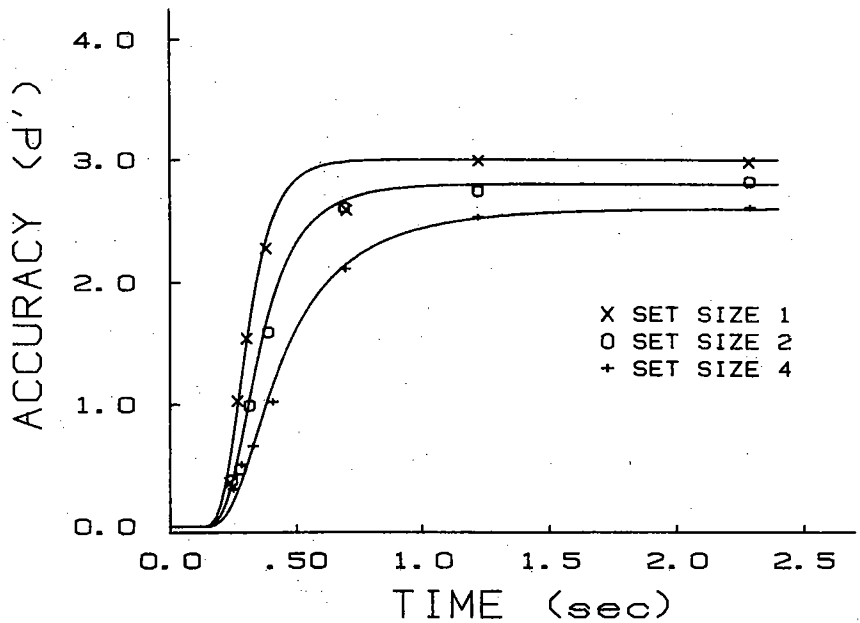

Figure 9. Group time–accuracy functions from each memory-set size and signal-lag condition of Reed (1976), along with curves generated from the cascade model of performance described in the text. (All three curves assume a total of six processes, with five relatively fast processes, $k \cong 30$, and one slow process for the comparison operation, with $k = 12/m$, where m is the number of items in the memory set. The asymptote parameters were chosen to achieve the best possible fit in each signal-lag condition.)

The results of Reed's (1976) experiment (Figure 9) suggest that neither of these interpretations is correct by itself. Clearly, the set-size manipulation does have a small asymptote effect (although it only shows up in the data of some of the individual subjects), but there is also a difference in the rate of approach to these asymptotes. It appears, then, that the set-size manipulation may have both of the effects described above.

Independence of rate and strength effects. The models of the Corbett (1977) and Reed (1976) experiments assume that the same manipulation has both a dynamic and an asymptotic effect. Are dynamic and asymptotic effects generally correlated? The answer to this question has important implications for the usefulness of the cascade model. The model assumes that the rate constants of all of the units at the same processing level are equal. It would be necessary to discard this useful simplifying assumption if factors that affected the strength of the connection between a unit and its inputs also affected the rate parameter of the activation of a unit by its inputs. Units with different strengths would then have different rate constants, violating the assumption. Fortunately, factors that would be expected to affect the strength of associative connections tend not to have noticeable dynamic effects.[2]

For example, when subjects in a semantic judgment task had to determine whether an exemplar was a member of a predesignated category, Corbett and Wickelgren (1978) found that the typicality of the exemplar to the category affected the asymptote but not the dynamics of the best-fitting Wickelgren curve. Further, Dosher (1976) found that asymptotic recognition accuracy for fragments of sentences varied as a function of the type of fragment, but type of fragment had no

[2] It may be worth pointing out that the decision as to whether a manipulation has an effect on information-processing dynamics depends on the exact nature of the model that is being used to interpret the data. Different models might take the fact that the slope of the rising portion of the time–accuracy curve increases with the placement of the asymptote as evidence that there is in fact a dynamic effect. The only claim made here is that within the context of the definition of information-processing dynamics as formulated for the cascade model, manipulations that affect placement of the asymptote do not necessarily influence the dynamic parameters of the model.

apparent effect on the dynamics of the time–accuracy curve. In addition, repeated testing of the same fragment increased the asymptote but did not affect either the rate or the delay parameter. Repeated testing of items learned by rote in Corbett (1977) also had no dynamic effect. There was a dynamic effect of repeated testing of items learned by imagery, as Wickelgren and Corbett (1977) also found. However, this finding does not necessarily mean that repeated testing affects the rate constants of the units of a particular process. Corbett suggested that subjects might simply be learning to bypass the extra processing initially required to retrieve items learned by imagery.

In short, the findings are consistent with the view that factors affecting the strength of associative connections have little or no effect on the dynamic properties of the hypothesized units. This conclusion makes it reasonable to keep the assumption that the rate constants associated with all units of the same processing level are equal, at least for the processes studied to date. If only from the standpoint of tractability, this is an extremely important result, since processing in a system with units responding at many different rates within the same level can be mathematically cumbersome to describe.

In summary, it seems fair to say that the cascade model provides reasonable accounts of the time–accuracy data obtained from a nice variety of experiments. In view of this fact, it becomes very important to consider its implications for the interpretation of the results of more standard reaction-time experiments.

Analysis of Conclusions Based on Reaction Times

On the Timing of Responses in Reaction-Time Experiments

We are almost ready to discover the implications of the cascade model for the interpretation of the results of standard reaction-time experiments. First, however, we need to understand how subjects decide that the time has come to initiate a response in these experiments. Typically, subjects in reaction-time experiments are instructed to respond as rapidly as possible consistent with a high degree of accuracy. In the context of a discrete stage model, this instruction seems reasonable enough. According to the discrete stage model, the output of the response selection process becomes available at some particular instant in time. After that time, the correct response can be executed; before it, the subject would simply have to guess. However, in terms of a cascade model (or as Pachella, 1974, and Wickelgren, 1977, have pointed out, any other model postulating the availability of intermediate activation levels), the instructions are vague. There is no identifiable instant in time before which responding would be at chance and after which it would be perfect. The activations of response units (and, therefore, the potential accuracy of responding) increases continuously, gradually leveling off at some maximal level. In such a situation, the instruction to respond as rapidly as possible consistent with a high degree of accuracy has no specific meaning.

The activation criterion hypothesis. How do subjects deal with standard reaction-time instructions? One possibility is that they set an implicit deadline consistent with a low enough error rate and respond so as to beat it (Ollman, 1977). However, this possibility does not explain the fact that reaction times are different for different experimental conditions, even when these are mixed within the same block of trials. An alternative possibility is that subjects adopt an *activation* criterion and respond when activation of a response unit reaches a level that is sufficient to ensure an acceptably low error rate (Grice, 1968).

As a first approximation, I will discuss a version of this model that assumes that subjects select a fixed criterion of relative activation for each of the two responses. Consider, for example, the activation of the response unit associated with a yes response in a yes–no task like the lexical decision task we have considered before. For this case, the criterion may be specified in terms of the difference between the observed activation of the yes response unit and the activation level it would be expected to have at the same point in time after stimulus onset if the stimulus were appropriate to a no response (i.e., a nonword). The situation is illustrated in Figure 10.

The model for no responses is simply the mirror image of the model for yes responses. The criterion for a no response is specified in terms of the difference between the observed activation of the no response unit and the activation that would be expected if the correct response were yes. It has been noted (e.g., Ratcliff, 1978) that no responses are affected by such factors as the similarity of the no response items to items in the yes response category. In the present model, this would be accounted for by assuming that the no response unit is inhibited by the same inputs that excite the yes response unit. Errors are produced when the activation of the incorrect response unit exceeds its criterion before the activation of the correct response unit exceeds its criterion. The same sources of variability apply to the activations of response units in this situation as in the case of time-controlled responding discussed above. An additional potential source of variability, indistinguishable from the variability in the base activations of the response units, is variability in the placement of the criterion from trial to trial (Grice, 1968).

The fixed-criterion hypothesis is superficially similar to the fixed random-walk criterion that has been used successfully by a variety of investigators in accounting for reaction-time and error-rate data (Link, 1975; Ratcliff, 1978; Stone, 1960), but it is not identical to it. Although it can be shown that the probability of exceeding any given count criterion approaches unity as time in the trial goes to infinity, this is not the case for the fixed criterion in the cascade model, since on any given trial the activation function for a particular response unit could end up at a point below the criterion. In the limiting case, in which asymptotic accuracy of performance approaches perfection, this problem would not arise. For the time being, then, conclusions will be restricted to these ideal conditions. When mean asymptotic activation is reasonably close to the activation criterion, it would appear to be necessary for subjects to relax their criteria as time goes on to insure that a response is eventually selected. The effects of criterion relaxation will be discussed below.

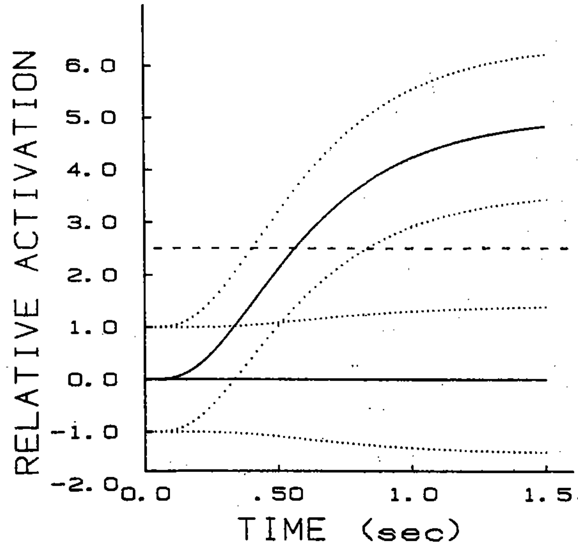

Figure 10. Illustration of the fixed activation criterion hypothesis of response execution. (The solid curves give the central tendency of the activation function for the yes response unit in a yes–no task such as the lexical decision task for the case in which the yes response is correct [rising curve] or incorrect [flat curve], as in Figure 5. An indication of the range of variability [plus or minus one standard deviation] is also given, as in Figure 5. The horizontal dashed line represents an arbitrary placement of the activation criterion.)

Determining the effects of manipulations on reaction times. Experiments that implicitly or explicitly adopt the discrete stage model use mean reaction time for each response in each experimental condition as the basis for interpreting the organization of the underlying processes. To evaluate these interpretations, it is necessary to determine how mean reaction time will vary with the manipulation of the parameters of systems of processes in cascade. An exact expression for mean correct reaction time can only be computed by determining the exact distributions of *crossing times* (time to reach criterion) for each response, taking into account the effect of trials that do not contribute to the distribution because of errors, convoluting the obtained distribution of response execution times with the assumed distribution of the time required to actually execute the response, and finally, computing the mean of this distribution. Luckily, we do not need to go through all of this to understand the general character of the effects of experimental manipulations under ideal conditions in which accuracy is high. It is only really necessary to see how experimental manipulations would effect the time it takes the function representing the mean relative activation function for the correct response unit (in Figure 10, the solid rising curve) to reach the activation criterion. The distribution of crossing times will be positively skewed for reasonably high placements of the criterion, so the crossing time for the mean relative activation function will tend to underestimate the mean of the crossing times. The extent of this underestimation is unaffected by manipulations that simply shift the activation functions, and it is increased (or decreased) by manipulations that decrease (or increase) the slope of the mean activation function at the point where it crosses the criterion. This pattern of divergence does not affect the conclusions we will reach about the behavior of the correct reaction time under the influence of the experimental manipulations we will be interested in. Furthermore, as long as response execution time is unaffected by experimental conditions (as we have assumed), the exact form of the distribution of execution times cannot affect our conclusions.

The fact that errors take out trials that would otherwise contribute to mean reaction times is more troublesome, especially since errors often occur on trials that would otherwise produce reaction times longer than the mean. The magnitude of the distortion introduced in this way depends on the relative sizes of the various sources of variability and, of course, on the error rate itself. This problem is another reason why firm conclusions must be restricted to cases in which error rates are negligible.

Our analysis, then, applies to a case that is idealized in two related respects: The model assumes that the response criterion remains flat as time in the trial goes on, and it assumes that the error rate is negligible. Analogous idealizations are, of course, implicit in the discrete stage model. After we consider the implications of the cascade model for the interpretation of reaction-time results under these ideal conditions, we shall return to a brief consideration of the effects that errors and criterion relaxation would have on our conclusions.

Simple Manipulations of Rate and Asymptote Parameters

Suppose we perform a reaction-time experiment in which we make a single manipulation, and we find that it increases reaction time by some amount. According to the discrete stage model, we would infer that the manipulation increases the duration of one of the processes in the system under investigation. In the cascade model, of course, processes do not have durations, but manipulations might affect either the rate parameter of a process or the relative asymptotic activation of the response units. In fact, a manipulation that decreases either the rate of a process or the relative asymptotic activation will increase the time it takes the activation function to reach criterion, as Figure 11 illustrates. These facts lead to the conclusion that a manipulation that simply increases reaction time does not indicate by itself whether it affects the relative asymptotic activation or the rate of one of the processes in the system. This conclusion is not restricted to the cascade model, of course, and has been pointed out previously by Wickelgren (1977; Corbett & Wickelgren, 1978).

In many reaction-time experiments, it may seem odd to suggest that a manipulation could actually affect the asymptotic activation level, since in the majority of cases, subjects could always be correct if they were not under time pressure. In these sorts of cases, it is natural to assume that asymptotic activation levels must reach the same ceiling level in all cases. However, this natural assumption is not correct. A d' of about 4 or 5 is sufficient to produce performance levels that are perfect for all practical purposes, but there is no reason why relative activation levels could not be compatible with d's of 10 or even 20. Thus, the mere fact that accuracy would reach perfection in each of two experimental conditions if subjects were given unlimited time is not sufficient evidence that the manipulations produce the same relative asymptotic activation levels.

Of course, it is possible that there is a ceiling activation level somewhere in the system of

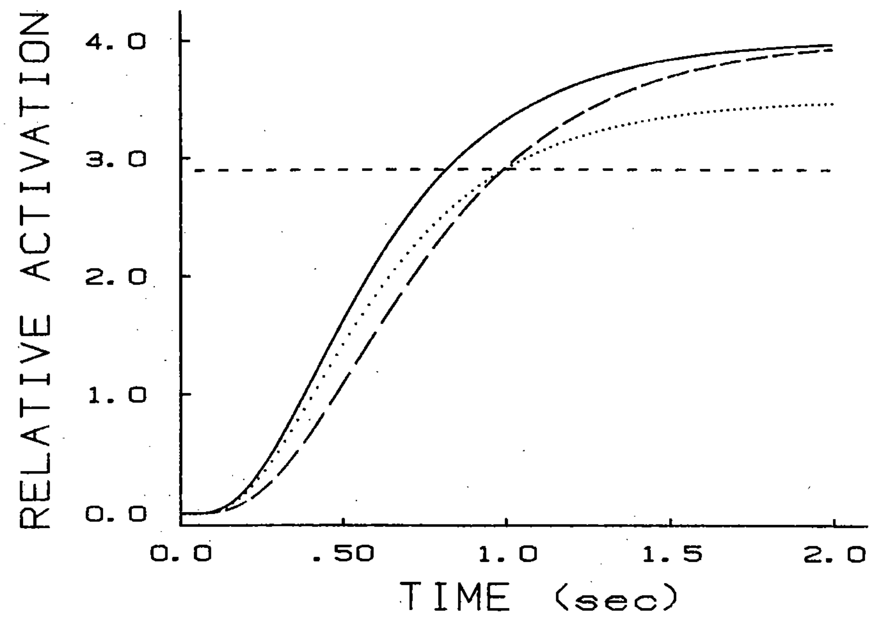

Figure 11. Ambiguous effects of rate and asymptote manipulations in reaction-time experiments. (The dashed curve differs from the solid curve in the rate of one of the processes, and the dotted curve differs from the solid curve only in the asymptote.)

processes that forces equal activation levels when accuracy levels get very high, but there is no necessary reason to suppose that there are such ceilings at interesting levels. We may reach the ceiling on our measuring stick long before we reach a ceiling on the actual internal activations themselves.

Stimulus quality and lexical decision. A large number of studies are subject to multiple interpretations on the basis of this conclusion. For example, Corbett and Wickelgren (1978) point out that the effect of typicality on reaction time in semantic verification tasks could be due to an asymptote rather than a rate effect. As another example, consider the finding of Meyer et al. (1975) that degrading the visual input increases the time it takes to make a lexical decision. This finding is compatible with the possibility that degrading slows the rate of some process or with the possibility that it reduces the size of the difference in asymptotic activation of the yes response unit on word versus nonword trials. Just such a reduction would be expected if degrading made it more difficult to tell what visual features were present in the displays, since the effects of degrading would be expected to carry through to affect asymptotic activation at the final processing level. And, according to the preceding discussion, the fact that subjects could usually read the degraded words correctly does not demonstrate conclusively that the degrading manipulation did not have an effect on relative asymptotic activation levels.

Although rate and asymptote manipulations will tend to have indistinguishable simple effects, we must be careful to distinguish between them conceptually. It turns out that manipulations that affect the rate parameter of a process behave differently than those that affect the relative asymptotic activation level in factorial combination with other manipulations.

The Effect of Insertion of a Process: The Subtraction Method

Donders's (1868–1869) subtraction method attempts to measure the duration of a component of an information-processing system by taking the difference in reaction time between a condition in which the process is present in the system and a condition in which it is not present. The method is based on two assumptions—(a) that the experimental manipulation actually does result in the insertion of a process with no alteration of the other processes, and (b) that insertion of a process will result in

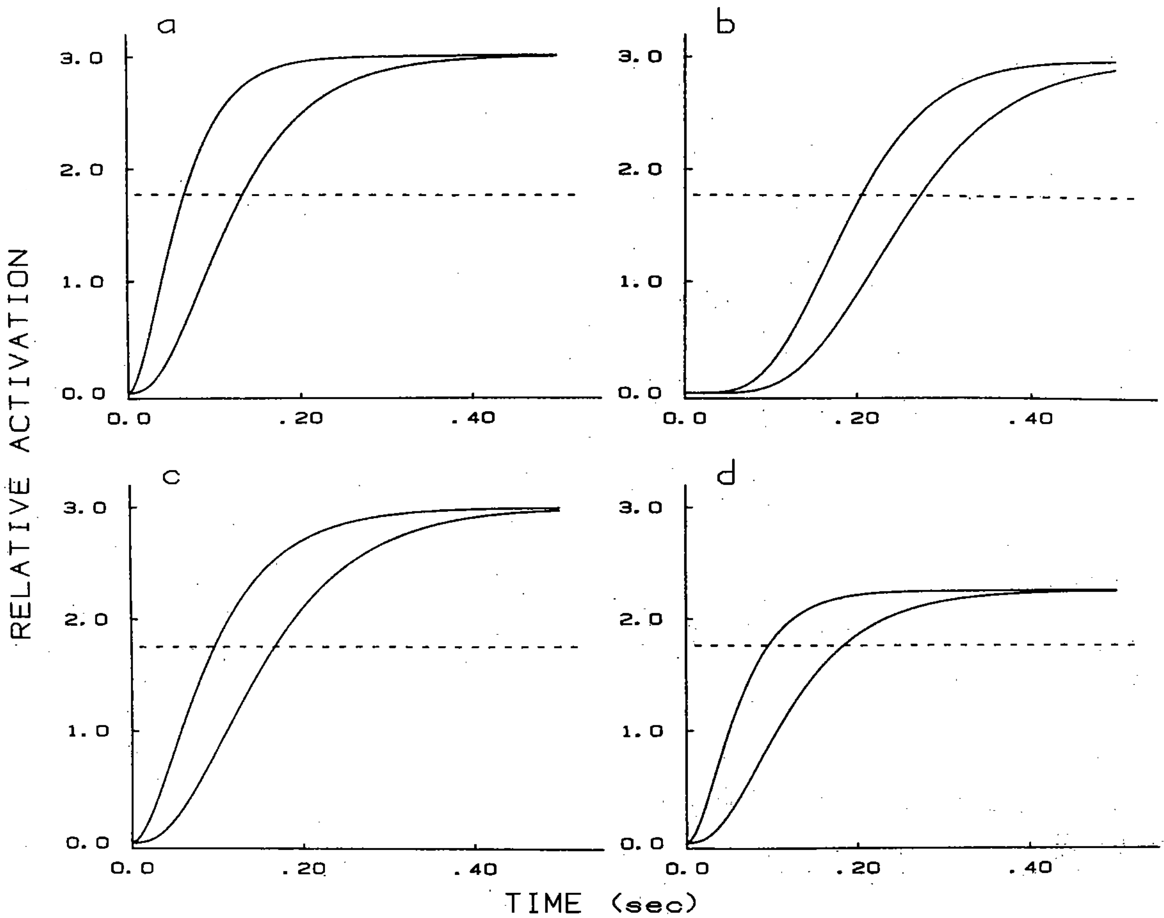

Figure 12. The effect of insertion of a process with a rate constant of 15 into base systems containing relatively fast processes. (Panel a illustrates the effect of the insertion into a system containing two processes with rate constants of 30. In b the base system contains 4 additional processes with rate constants of 30. In c one of the rate constants of the base system in a has been reduced from 30 to 15. In d the asymptote of the base system in a has been reduced 25%.)

the addition of a fixed time interval to the reaction time, independent of the other processes in the system. Sternberg (1969a, 1969b) has pointed out the difficulties with the first assumption, but it is probably true enough in some situations for a brief examination of the second to be worthwhile.

The assumption that an inserted process adds an invariant amount of time follows naturally from the discrete stage model, but would there be an invariant effect in a system of processes in cascade? It turns out that insertion will have an invariant effect as long as (a) the inserted process does not alter the asymptotic activation level of the response units, and (b) it is not a rate-limiting process. When these conditions hold, the inserted process will simply shift the activation function to the right by an amount equal to the reciprocal of its rate constant and therefore delay the time it takes the activation function to reach the criterion, as we saw earlier in this article. Of course, it is understood that one of the other processes must be substantially slower than the inserted process. Invariance is good as long as the rate of the inserted process is four times that of the slowest process in the base system.

Matters are more complicated when the inserted process is among the slowest in the system, even if it does not alter the asymptote. In this case, the size of the increase in reaction time will depend on the relative placement of the accuracy criterion with respect to the asymptote, as illustrated in Figure 12. The reason the effect is not invariant is that the insertion of a process that has a relatively slow rate constant will reduce the slope of the time–accuracy curve. In fact, if the inserted process is not the slowest but is within a factor or two of the rate of the slowest process in the system, the effect on the slope of the curve may be noticeable. As the rate of the slowest process other than the inserted process increases, the magnitude of the effect of the placement of the criterion increases. Thus, insertion of a relatively slow process will tend to increase reaction time, but the size of the effect will depend on the placement of the criterion with respect to the asymptote and on the rate parameters of other relatively slow processes in the system.

Even more complications arise if the inserted process affects the asymptote of the time–accuracy curve, as illustrated in Figure 13. If the inserted process lowers relative asymptotic activation, the magnitude of the effect of insertion will increase as the placement of

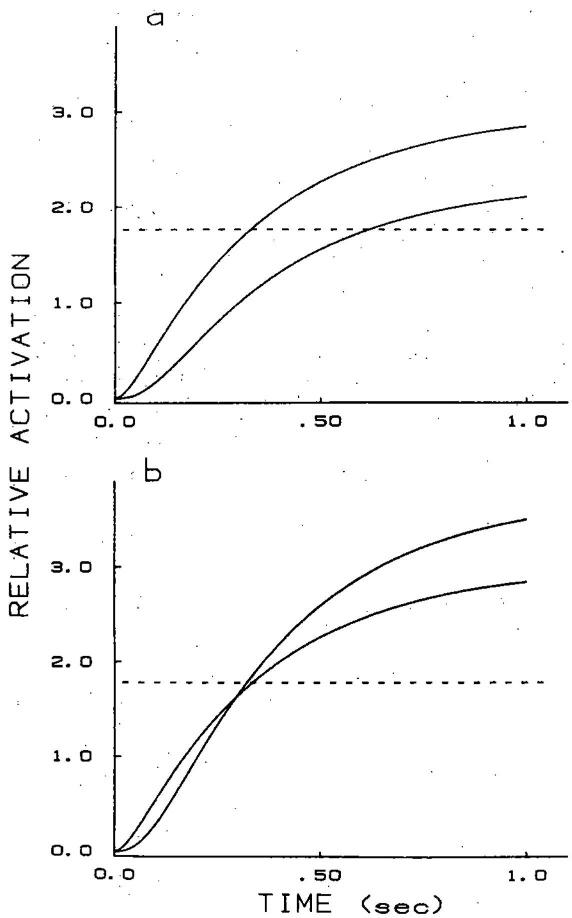

Figure 13. Effects of insertion of processes that alter asymptotic accuracy. (In panel a, the inserted process lowers the asymptote; in panel b, it raises it.)

Table 1
"Same" Reaction Time for Physically Identical and Name-Identical Stimuli

Stimulus	High verbals	Low verbals
Name identical	588	632
Physically identical	524	542
Difference	64	90

Note. From Hunt, Lunneborg, and Lewis (1975).

the criterion is moved higher on the relative activation scale. On the other hand, if the inserted process *increases* relative asymptotic activation, the effect of the insertion of a process will actually decrease as the placement of the criterion with respect to the asymptote increases, to the point that reaction times can actually be faster in the condition including the inserted process.

In summary, it is apparent that the insertion of a process into a system of processes in cascade can have an invariant effect on reaction time, but only under a very specific set of conditions. The inserted process must be relatively fast compared to the slowest process in the system, and it must not affect the asymptotic activation of response units. Otherwise, the size of the effect of the inserted process on reaction time may vary with other parameters of the activation function or even simply with the location of the response criterion in relation to the location of the asymptote.

Verbal ability and memory access. The best known application of the subtraction method in the contemporary literature can be found in analyses of individual differences in the letter-matching task developed by Posner and Mitchell (1967). In this task, subjects view pairs of letters, which are either physically identical (A A), name identical (A a), or different, under two instructions—(a) to respond same if the letters are physically identical (PI condition) or (b) to respond same if the letters are identical in name (NI condition). For the average subject, the time it takes to respond same to physically identical pairs is something less than 100 msec faster than the time it takes to respond same to name-identical pairs.

Hunt and his collaborators (Hunt, 1978; Hunt et al., 1973, 1975) have taken the

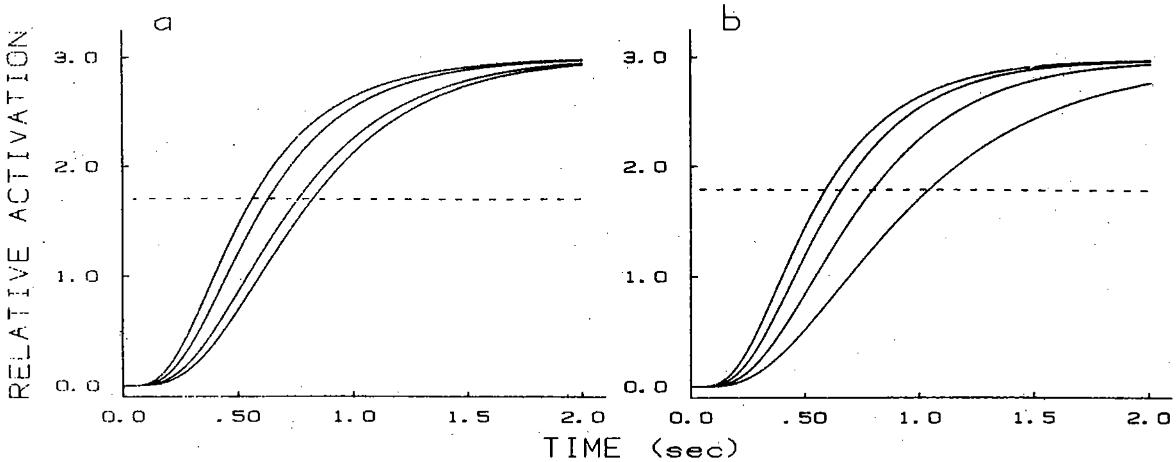

Figure 14. (a) The effects of joint manipulation of factors affecting the rate parameters of two different processes, and (b) the effects of joint manipulation of factors affecting the rate parameter of the same process.

NI–PI difference as a reflection of the extra processing required to access the names of letters. They have shown that the NI–PI difference is correlated with performance on verbal abilities tests. From this correlation they have concluded that the efficiency of activating memory codes for familiar stimuli is correlated with verbal ability. No very explicit model is presented, but I think what follows describes at least one version of the implicit model underlying this conclusion. It is assumed that performance in the physical match task requires several component processes leading up to the formation of visual codes for the two display characters, followed by a comparison process, a decision process, a response selection process, and finally, a response execution process. It is also assumed that the determination that two stimuli are identical in name requires the insertion of an additional memory retrieval process between visual encoding and comparison. The comparison operations are thought to be at least roughly equivalent in the two cases.

Typical results of the NI–PI task, as a function of verbal ability, are shown in Table 1. As in this case, there is generally a difference in reaction times in the PI task between high and low verbal subjects, as well as a difference in the size of the NI–PI difference. Within the context of the discrete stage model, the difference in the PI task presumably reflects a difference in some process other than accessing name information in memory. So long as we hold to the view that processing stages occur in strict succession, such a difference in the base reaction time need not cloud our interpretation of the NI–PI difference. But will the NI–PI difference provide a stable indication of the rate of accessing information in memory if we assume that the processes operate in cascade? The answer is that it might, but it does not necessarily. Insertion of a process into two different environments will only have an invariant effect on reaction time if the inserted process is not rate limiting, and if it does not affect the asymptotic accuracy of performance. It is by no means clear whether either of these assumptions is correct. In particular, it may very well be that there is a difference in relative asymptotic activation levels between the NI and PI tasks. Although it is probably correct to assume that all of the subjects studied know all of the letters well enough to avoid making mistakes in identifying the letters if they are given unlimited time, this is not a guarantee that relative asymptotic activation reaches the same level in both tasks. Thus, these data leave open the possibility that the larger size of the NI–PI difference for high verbal subjects is due to a difference in some process other than accessing memory codes.

In all fairness (and for the record) it must be noted that Hunt (1978) has produced considerable converging evidence in favor of the conclusion that accessing information in memory is a source of differences between high and low verbal individuals. The conclusion is not based merely on the NI–PI difference, but on evidence from a variety of other tasks as well. And in fact, Jackson (1978) has been able to eliminate the difference in the PI condition; at the same time, the NI difference remains. When the base reaction time is the same, the inference that the supposedly inserted process is responsible for the difference between groups is somewhat more compelling.

Factorial Manipulations of Parameters: Additive Factors?

Except in the study of individual differences, we are usually not so much interested in the actual parameters of a particular process as we are in its nature and function. If we adopt the discrete stage model, Sternberg (1969a) has shown how we can use experimental manipulations that are assumed to selectively influence specific stages to study what stages are affected by other manipulations. Within the discrete stage model, the assumption that one experimental manipulation influences the duration of one stage and another manipulation influences the duration of another stage leads to the conclusion that the two factors will have additive effects on reaction time. On the other hand, factors that influence the duration of the same stage will generally interact with one another. By determining whether combinations of factors interact, we can therefore determine whether they affect the duration of a common stage. This analysis is based upon the discrete stage model. If we adopt the cascade model, the logic is only partially correct.

We have seen before that rate constants of the relatively fast processes determine the size of the shifting effect these processes exert on the activation function. This means that if we manipulate the rate of a relatively fast process in factorial combination with a manipulation of the rate of another process, the magnitude of the effect of the manipulation of the fast process will be the same at both levels of the factor that affects the rate of the other process. This will be true even if, as in Figure 14a, the other process whose rate is manipulated is a relatively slow process. The result, then, will be additivity of the two effects.

How good is the additivity and over what range of variation does it apply? It appears to be extremely good. Additivity is virtually exact if the rate constant of either of the two manipulated processes is at least four times the rate of the slowest process in the system at all levels of both factors. Even outside this range, additivity is surprisingly good. I have performed several factorial simulations of the time it takes the activation of the response elements in different systems of processes to reach different criterion levels. In one of these simulations, which is representative, a total of six processes were assumed. Four of the processes were invariant *base processes* unaffected by experimental factors. One of the base processes had a relatively slow rate constant of 2.5. The other three base processes were assigned rate constants of 20. The rate constants of the remaining two processes were varied factorially from 2 to 50. The results of this simulation, for a criterion located 75% of the way to the asymptote, are shown in Figure 15. As the rate constants of the two processes are varied over the range from 2 to 50, they have nearly completely additive effects.

When manipulations of the rates of two processes are arranged in such a way that they switch roles as the rate-limiting process, the additivity does break down. Making a process rate limiting increases its effect on the upper portion of the activation function. This means that when one of the two processes is relatively fast, the other will have a relatively greater effect than it would have had if the first process had been slow. The magnitude of this effect becomes greater and greater as the asymptote is approached. This sort of situation results in an *underadditive* interaction between the two processes. However, the underadditivity is slight for reasonable criterion levels (50%–90% of asymptote) and would probably go undetected in the normal reaction-time experiment. For example, I have not been able to simulate any of the reported instances of underadditivity that I know of in the literature in this way. So as a rule of thumb, we can conclude that factors that affect the rate constants of two different processes will have additive effects on reaction times.

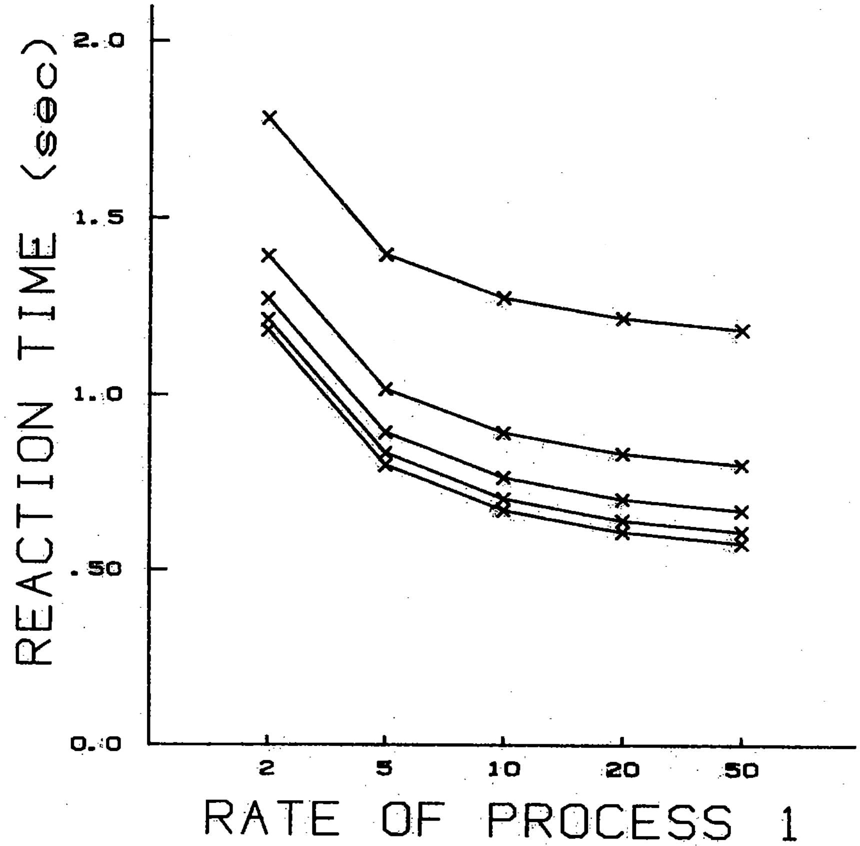

Figure 15. Simulated results indicating the additive effects of factorial manipulation of two factors, each affecting the rate constant of a different process.

What about two factors that both affect the rate of the same process? Assuming that the manipulations each reduce the rate constant by a fixed proportion, the result will be an interaction of the overadditive type: At the level of one factor that produces longer reaction times, the other factor will have a greater affect. The reason for the overadditivity is that a proportional change in the rate of a process will make a larger difference if the rate is already slow than if it is fast. This situation is illustrated in Figure 14b. We can conclude that factors that affect the rate of the same process will generally have interactive effects. Taking this conclusion together with the additivity of manipulations that affect the rate parameters of different processes, we find that additive factors logic applies to systems of processes in cascade, as long as the factors only affect the rate constants of the processes under consideration and do not alter relative asymptotic activation.

When one or both of two factors affect the asymptotic output of a process, the results are different. Consider first a manipulation that affects the relative asymptotic activation levels in conjunction with a manipulation of the rate of a relatively fast process (Figure 16a). These two manipulations will have additive effects. The reason for this is that the shifting effect of the manipulation of the rate parameter is independent of the placement of the asymptote. Note, further, that this is true regardless of the level at which the asymptote manipulation actually has its effect. We could lower the asymptote by visual degrading, by reducing the association between stimuli and designated responses, or by making the stimuli associated with different responses more similar; this manipulation would have thc same additive effect when combined with a manipulation that affects, say, the rate of activation of associations.

Now consider a manipulation of the asymptote of a process in conjunction with a manipulation of the rate of the rate-limiting process

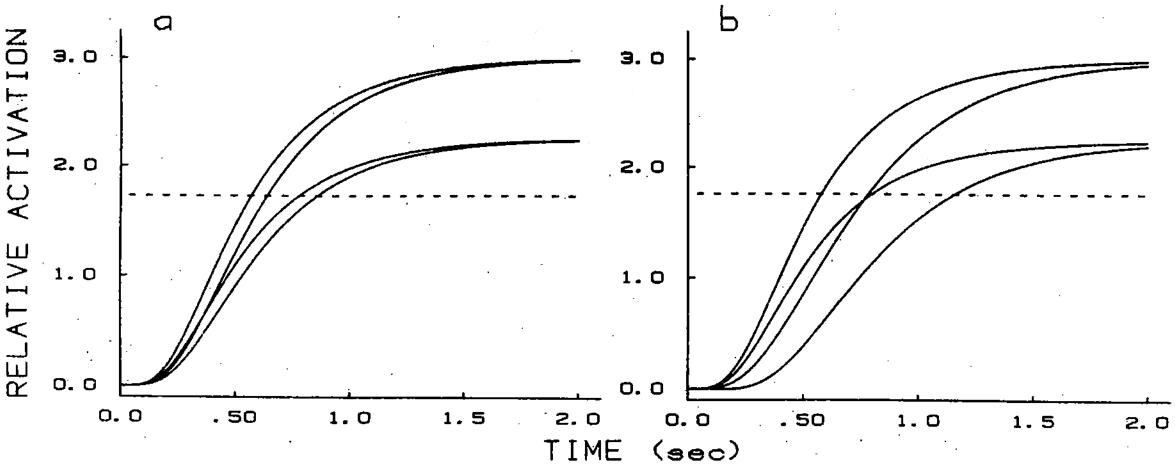

Figure 16. (a) The effects of joint manipulation of the rate parameter of a relatively fast process and the asymptote of the relative activation function, and (b) the effects of joint manipulation of the rate parameter of the rate-limiting process and the asymptote of the relative activation function.

(Figure 16b). The rate constant of the rate-limiting process determines the rate of growth of the curve to a constant proportion of its final height, and the asymptote determines what proportion of the final height is needed to reach the criterion. Their joint manipulation will therefore result in an overadditive interaction. Again, since asymptote effects are indistinguishable as to level, it does not matter where the asymptote manipulation actually has its effect. Thus, we see that rate and asymptote manipulations will produce overadditive interactions as long as the rate manipulation applies to the rate-limiting process. Of course, whether a process is rate limiting is to some extent a matter of degree, as noted earlier in discussing the effect of the insertion of a process. Manipulations of the rates of processes whose rates are close to the rate of the slowest process will produce slight interactions with asymptote manipulations.

What happens when two factors that affect the asymptote are manipulated? The result will generally be an overadditive interaction.

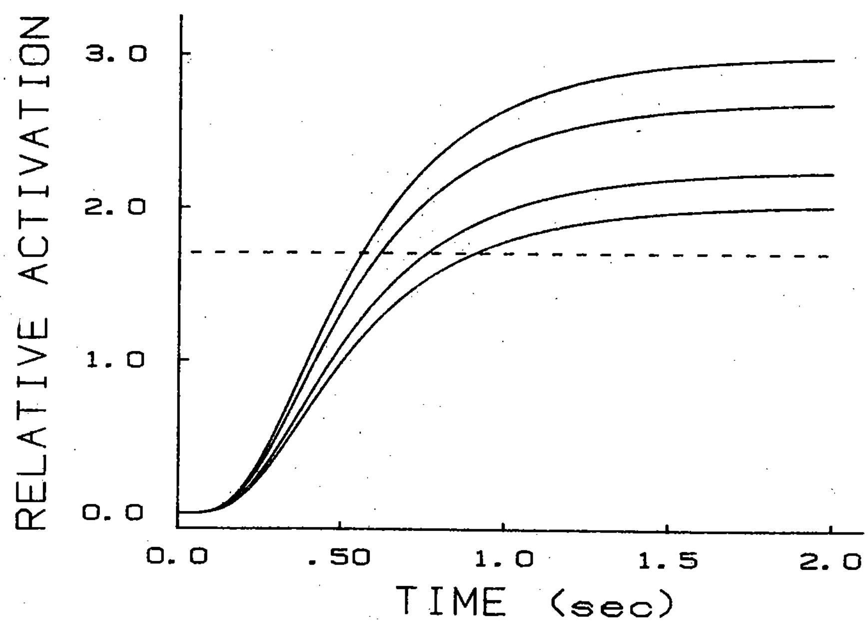

Figure 17. The effect of joint manipulation of two factors affecting the asymptote of the relative activation function. (Each manipulation reduces relative asymptotic activation by a fixed percentage of the value it would otherwise achieve.)

Table 2
Comparison of Inferences Derived From the Discrete Stage Model and the Cascade Model

Condition	Discrete stage model	Cascade model
If factors interact	They affect the duration of the same process.	They affect the rate of the same process, *or* they both affect relative asymptotic activation, *or* one affects the rate of the rate-limiting process and the other affects the relative asymptotic activation.
If effects are additive	They affect the durations of different processes.	They affect the rates of different processes, *or* one affects the rate of a fast process and the other affects the asymptote.

Note. Inferences from the cascade model assume a fixed criterion and very low error rates. The locus of an asymptote effect cannot be determined from the pattern of additivity and interaction.

The case in which two manipulations each reduce the asymptote by a fixed percentage is illustrated in Figure 17. An even larger overadditive effect is produced if the two manipulations each have additive, rather than proportional, effects on the asymptote. We will consider a manipulation that would give rise to such an effect in the next section.

The conclusions I have described are summarized in Table 2. We can see that the cascade model paints a more complicated picture than the discrete stage model. As long as the asymptote is unaffected, the two models have the same implications for the interpretation of reaction-time data, but as soon as we admit the possibility that a manipulation might affect the asymptote, several more possibilities arise. The conclusions summarized in Table 2 have implications for the interpretation of the results of several experiments in the literature.

Reexamination of Interpretations of Interactions Based on the Discrete Stage Model

In this section I reconsider some of the interpretations of interactions based on the discrete stage model with the foregoing analysis in mind. What we will see, in general, is that overadditive interactions that have been taken as evidence that two factors affect a common process need not be so taken after all.

Locus of the effect of context in word recognition. Consider again the lexical decision experiment of Meyer et al. (1975). As previously mentioned, these authors studied reaction time to decide whether a string of letters was a word or a nonword and manipulated the visual quality of the display of the target word in factorial combination with the relatedness of a preceding context word to the target. The result was an overadditive interaction of the quality and relatedness factors. Relatedness had a stronger effect for degraded than for intact words, as illustrated in Table 3.

According to the discrete stage model, these results suggest that priming and degrading the stimulus with dots have their effects on a common stage. Since it appears likely that degrading with dots affects a relatively early stage, Meyer et al. (1975) suggested that priming may affect an early stage as well, perhaps the translation of a visual representa-

Table 3
Effect of Stimulus Quality and Priming on Reaction Time

Stimulus	Intact	Degraded
Unassociated	566	728
Associated	528	657
Difference	38	71

Note. From Meyer, Schvaneveldt, and Ruddy (1975).

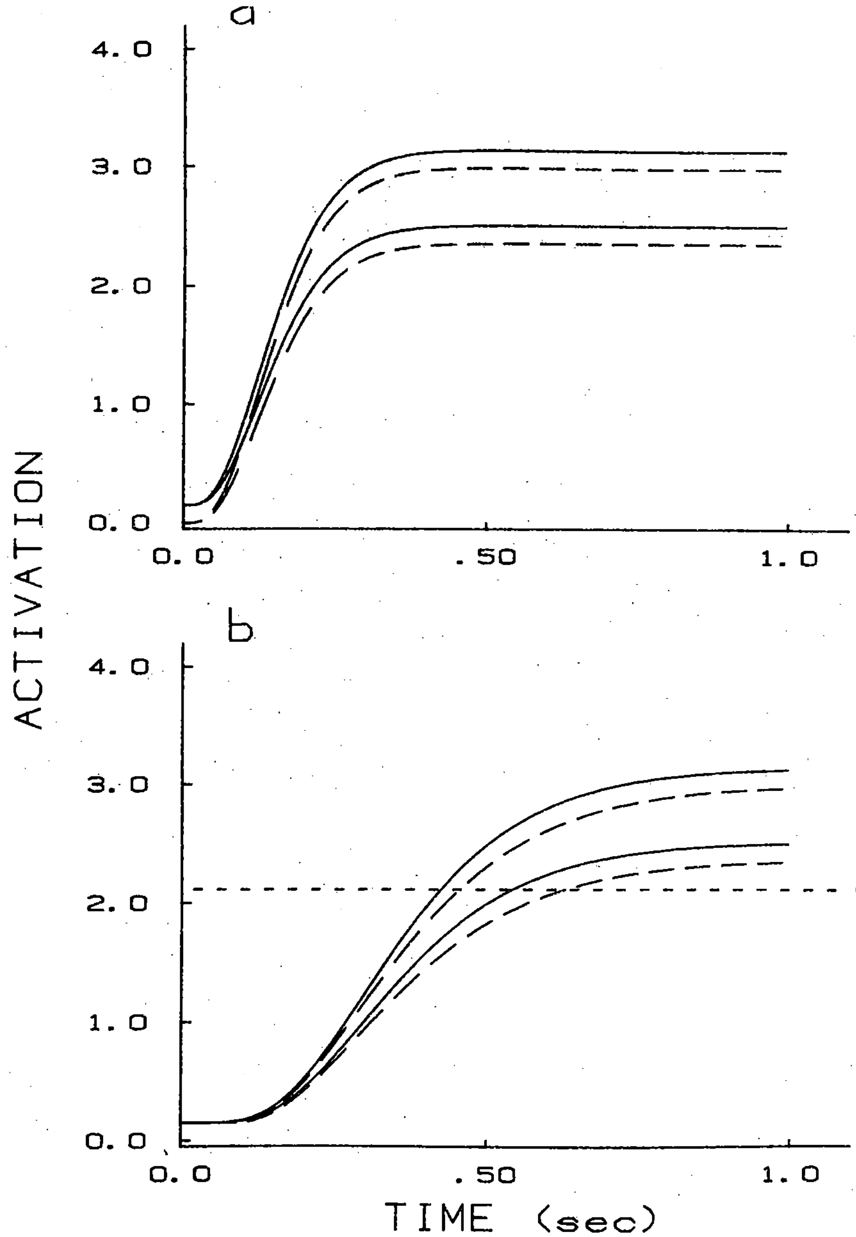

Figure 18. (a) Activation functions for the target word detector under normal (upper pair of curves) and degraded (lower pair of curves) conditions for words preceded by related (solid curves) or unrelated (dashed curves) context, according to the model described in the text. (Four processes are assumed with rate constants of approximately 25.) (b) Activation functions for the yes response unit under the same four conditions. (The two added processes have rate constants of 10 and 8. The parameters were chosen to approximate the results of Meyer et al. [1975]; reaction time is 100 msec longer than time to reach the criterion indicated by the horizontal line.)

tion into graphemic codes. However, the conclusion that the factors affect a common process does not follow if we adopt a model in which the processes operate in cascade. As Table 2 indicates, the interaction could be due to the joint effects of the two factors on the rate parameter of the same process, as Meyer et al. suggested, but it could also be due to the joint effects of two factors each affecting asymptotic activation or to the joint effects of a factor affecting the asymptote and another affecting the rate of the rate-limiting process. In other words, the two factors could well be influencing different processes.

In fact, a quite simple interpretation of the Meyer et al. results may be given in terms of the simplified cascade model of the lexical decision task that has been used for illustrative purposes throughout this article. Recall that this model (illustrated in Figure 2) postulates light, feature, letter, and word analysis processes, a decision process, and a final response activation process. All processes consist simply of linear integrators, except the decision process, in which the yes response unit is a maximum unit driven by the largest of the outputs of the word analyzers. In this model, it seems reasonable to assume that degrading affects the asymptotic output of feature analysis. The dots would be expected to produce spurious activations of detectors for features not present in the display and to reduce activations for detectors for features that are present in the display. This degradation effect would be propagated all the way through the system, reducing the relative asymptotic activations of the appropriate letter, word, decision, and finally, response units. The effect of relatedness may be placed at the word detector level. We could simply adopt Morton's (1969) assumption that related context words produce base activations of the detectors for words associated with the context.

In this model, relatedness and degradation will have additive effects on the asymptotic activation of the detector for the target word. The relatedness will determine the starting activation of the detector, and the visual quality of the target display will determine how high the activation function will rise above this starting level (Figure 18a). These effects will be carried through to the final response activation level (Figure 18b), with some minor distortion at the low end owing to the behavior of the maximum unit (see the Appendix). As the figure illustrates, the time it takes the activation functions to cross the response criterion is increased more by degradation when the word is preceded by unassociated context, thereby accounting for the interaction.

Differential effects of degrading by dots and by contrast reduction. The cascade model suggests a simple interpretation of an interesting pattern of results investigated by Miller (1976). Miller found that degrading with dots produced an overadditive interaction with the probability of stimulus occurrence, but degrading by contrast reduction had additive effects in conjunction with the probability manipulation. In terms of the discrete stage model, the interaction would indicate that degrading with dots affected the same stage as the probability manipulation, presumably a stage having to do with the visual processing of the figure, but that contrast reduction affected a different and therefore presumably earlier stage. However, the results can be interpreted in terms of a cascade model in which the effects of the probability manipulation and the degrading manipulation are placed at completely different levels. Perhaps contrast reduction influences the rate of approach of the output of an early light analysis process to its effective ceiling, whereas dot degradation lowers relative asymptotic activation by affecting the output of the feature analysis process. These possibilities are consistent with the effects of contrast reduction in Pachella and Fisher (1969) and with the interpretation of the effect of dot degrading given earlier for the results of Meyer et al. (1975). If we accept these contrasting interpretations of these two different kinds of degrading, we would expect to find Miller's results even if the probability manipulation only affected the asymptotic activations of response units. The results are also consistent with the possibility that the probability manipulation affects the rate of a late process if the process is rate limiting. In either case, the analysis suggests that Miller's pattern of results is consistent with effects of the probability manipulation on processes that have nothing directly to do with the visual analysis of the figure. Of course, I do not mean to argue against the possibility that probability influences an early processing level. The point is only that the pattern of additivity and interaction obtained by Miller is consistent with an alternative interpretation.

Attention and word recognition. Finally, consider the interesting experiment of Becker (1976) on the locus of the effect of attention in word recognition. Becker's subjects had to perform a lexical decision on a visually presented word while carrying out a secondary task requiring responses to presentations of tones. The two factors manipulated were the frequency of the word in the lexical decision task and the complexity of the secondary task (simple or choice reaction). Word frequency and secondary task complexity produced an overadditive interaction for reaction times in both the primary and the secondary task. The choice task produced generally longer reaction times, and the effect of word frequency on reaction time was greater when the secondary task required choice.

According to a discrete stage model, if word frequency affects the process of lexical access, and if performance in the choice secondary task requires more attention than performance in the simple secondary task, the interaction would suggest that lexical access requires attention. However, this conclusion only follows if we accept a discrete stage model. A simple cascade model can be formulated as follows: Assume that the rate of activation of response units depends on attention and that the asymptotic activation of the detector for the word depends on word frequency. Then words of lower frequency will produce lower asymptotic activations of response units. The lower the asymptotic activation, the more the attentional demands of the secondary task will increase the time it takes the output of the response activation level to reach the response criterion. This account explains the effect of the secondary task on the reaction times in the primary task. To account for the reaction times in the secondary task, we need only suppose that the full resources required for efficient performance in the secondary task are not freed until the response in the primary task is initiated.

Indirect effects within the discrete stage model. These reinterpretations of interactions do not depend on acceptance of all of the assumptions underlying the cascade model. As Sternberg (1969a, 1969b) has noted, interactions of this type may be produced by indirect effects: Factors that affect the output of one stage can indirectly influence the duration of a subsequent stage. It turns out that it is possible to frame interpretations of the interactions considered above in terms of models in which the two factors manipulated have their direct effects on different processing stages while maintaining the assumption that processes operate in strict temporal succession. All we have to do is assume that one of the factors affects the output of a processing stage, thereby affecting rate of processing at later stages. For example, Meyer and Schvaneveldt (1975) have suggested that degrading letter strings by dots might directly affect the output of feature analysis, and therefore indirectly affect the duration of a matching process that compares the results of feature analysis against representations of words stored in memory. Priming, then, could directly affect the matching process, and the interaction of priming and degrading would be explained. Similar sorts of interpretations might be given for the interactions of dot degrading and stimulus probability in Miller (1976) and for the interaction of word frequency and secondary task complexity in Becker (1976).

Perhaps the reason why such interpretations have not generally been suggested lies in an implicit tendency to think of the output of a processing stage as a discrete code or set of codes, rather than as a set of continuous quantities. Certainly this tendency is strong for some of the processing stages that have been postulated. For example, it seems natural to think of the output of a stage called *stimulus identification* as a code indicating what the stage has determined that the stimulus is. Given the assumption that the output of a stage is such a discrete code, the duration of the stage and even the *accuracy* of its output might be affected by stimulus degrading, but we would not expect either the duration of the early stage nor the accuracy of its output to affect the duration of later processing stages. An example of the application of this reasoning can be found in the experiment of Sternberg (1967). This experiment compared reaction time to determine whether a probe letter was a member of a predesignated memory set, using both intact and dot-degraded probes. In the first session, degrading affected the slope of the linear function relating reaction time to number of items in the memory set: The size of the set-size effect was greater for degraded than for intact probes. If we do not think that degrading can affect the quality of the output of a stimulus identification stage (i.e., if the result of stimulus identification is a discrete code), the finding that degrading increases the effect of memory set size can be taken as an indication that the comparison process operates on unidentified visual representations of the probes, as Sternberg suggested. However, in a cascade model, visual degrading could affect the asymptotic activations of units at a stimulus identification level. If visual degrading affected the quality of the output of a process at one level, it would do so at all later levels as well, unless there were an intervening effective ceiling of some sort. In fact, even if processes occurred in discrete stages, the assumption that the output of each process is a set of continuous variables rather than a discrete code or set of codes is sufficient to leave open the possibility that factors affecting the output of any processing stage might produce overadditive interactions with factors directly influencing the duration of any subsequent stage.

The possibility that the outputs of processes can be affected by experimental factors considerably weakens the additive factors method, as Sternberg (1969a, 1969b) was aware. In fact, he suggested that it might be necessary to include as part of the definition of a stage the stipulation that its output be unaffected by factors that affected its duration. Another way of saying the same thing is simply to note that additive factors logic will not generally apply if factors affect the outputs as well as the durations of stages. Perhaps one reason why this problem has not been more seriously explored lies in the many instances in the literature in which factors have produced additive effects. Until now, many such instances of additivity have seemed to be strong support for the discrete stage model and the additive factors method that rests on it. Although it has been shown that parallel, limited capacity models could imitate some predictions of serial models (Townsend, 1974), no previous treatment of parallel–contingent processes has indicated that the additivity of rate manipulations might hold up in such conditions.[3] Additive effects are even compatible with manipulations that affect the asymptotic output of a process in conjunction with a manipulation that alters the rate of a relatively fast process in the system under consideration. Since discreteness can no longer be inferred from additivity, it behooves any researcher to consider carefully whether the discrete stage model should be adopted before using it to analyze patterns of additivity and interaction in reaction-time experiments. Even adopting the discrete stage model, it may be more important than has generally been acknowledged to consider the possibility that factors might affect the outputs, and not only the durations, of hypothesized processing stages.

I am not suggesting that we abandon the additive factors method. There are many information-processing tasks in which the idea that component processes occur in successive stages seems very plausible, and the assumption that factors that affect the duration of a process would not affect the quality of its output seems reasonable in many of these cases. I leave it to further research to determine which tasks are better thought of as operating in cascade and which as operating in discrete stages. The point is simply that when we do not know which model to choose, the possibility that processes are operating in cascade or even just the possibility that their outputs are sets of continuous quantities may make it difficult to reach unambiguous interpretations of patterns of additivity and interaction in reaction-time experiments.

The effects of errors and criterion relaxation. We have seen that reaction-time results are often ambiguous. The potential for ambiguity is increased by the fact that subjects probably do not always adhere to a fixed criterion. Instead, it often appears that subjects relax their criterion as time goes on, essentially accepting poorer and poorer information as the basis for responding at later times (e.g., Reed, 1976). Such a tendency is probably partially responsible for the usual positive correlation between reaction time and error rate across experimental conditions. Depending on the slope and curvature of the relation between the accuracy criterion and time since stimulus

[3] In work done independently and at about the same time as the present research, Townsend and Ashby (Note 2) have noted an interesting related result. Defining stage duration as the time between a pulsed input to a stage and the final output, they have shown that when two stages are arranged in cascade, the total duration of the pair of stages (i.e., time from the presentation of a pulsed input to the first and the final output of the second) is equal to the sum of the durations of the separate stages. Further, they find that additivity of factor effects holds for the joint manipulation of factors affecting the durations of separate stages.

Table 4
Effects of Simple and Joint Manipulations of Characteristics of Processes in Cascade on the Parameters of the Time–Accuracy Curve

Manipulation	Result
Simple effects	
Reduce relative asymptotic activation	Reduce asymptote
Decrease rate of relatively fast process	Increase delay
Decrease rate of relatively slow process	Increase delay and decrease rate
Effects of insertion	
Insert relatively fast process	Increase delay (possibly affect asymptote)
Insert relatively slow process	Increase delay and decrease rate (possibly affect asymptote)
Combined effects	
Reduce rates of two different fast processes	Additive delay effects
Reduce rate of two relatively slow processes	Slightly underadditive delay and rate effects
Reduce rate of same process twice	Overadditive effects on delay with rate effects if process becomes rate-limiting
Reduce rate of one process and relative asymptotic activation	Independent effects
Reduce relative asymptotic activation twice	Independent effects (additive or proportional)

presentation, this can turn an additive effect into an interaction or vice versa. Compounding the problem still further, high error rates for conditions producing long reaction times introduce potential distortion of the observed mean reaction time from the mean reaction time that would be observed had errors not removed trials that might otherwise have produced long reaction times. As a result, it may be very difficult to reach a firm conclusion about the locus of the effect of a manipulation in a reaction-time experiment. It might be possible, with additional assumptions, to take the effects of errors and criterion relaxation into account in reaction-time experiments. However, the interpretation of the results of reaction-time experiments would then depend on the accuracy of these additional assumptions, which might be difficult to check. For this reason, I think the method described below may well offer our best hope for determining the specific nature of the effect of an experimental manipulation on the underlying processes.

Toward a Method for Analyzing Processes in Cascade

In the context of the cascade model, standard reaction-time methods have serious drawbacks. Many patterns of results that are unambiguous under the discrete stage model are ambiguous under the cascade model, even accepting the fixed criterion hypothesis, and additional ambiguity is introduced by the dubious status of the fixed criterion hypothesis itself. Thankfully, we need not be trapped in this ambiguity. We may use the deadline and response-signal methods for analyzing the shape of the time–accuracy curve to avoid many of the ambiguities.

As Table 4 illustrates, the relation between the effects of manipulations on the underlying processes and the resulting time–accuracy curve is well specified under the cascade model. In fact, it is possible to use the parameters of the best-fitting curve given by Wickelgren's equation as the basis for inferences about the effects of certain parameter manipulations on the underlying processes. Some ambiguities do remain. We cannot distinguish between a manipulation that inserts a process and one that alters the rate of a process, as we found in the discussion of the effect of imagery on retrieval of paired associates (Corbett, 1977). Nor can we specify the locus of the effect of a manipulation that alters the asymptote of the time–accuracy curve, even when the asymptote manipulation is carried out in conjunction with manipulations of other factors. However, we can specify more precisely the type of effect of a manipulation on the time–accuracy curve, and by performing these simple manipulations we can determine

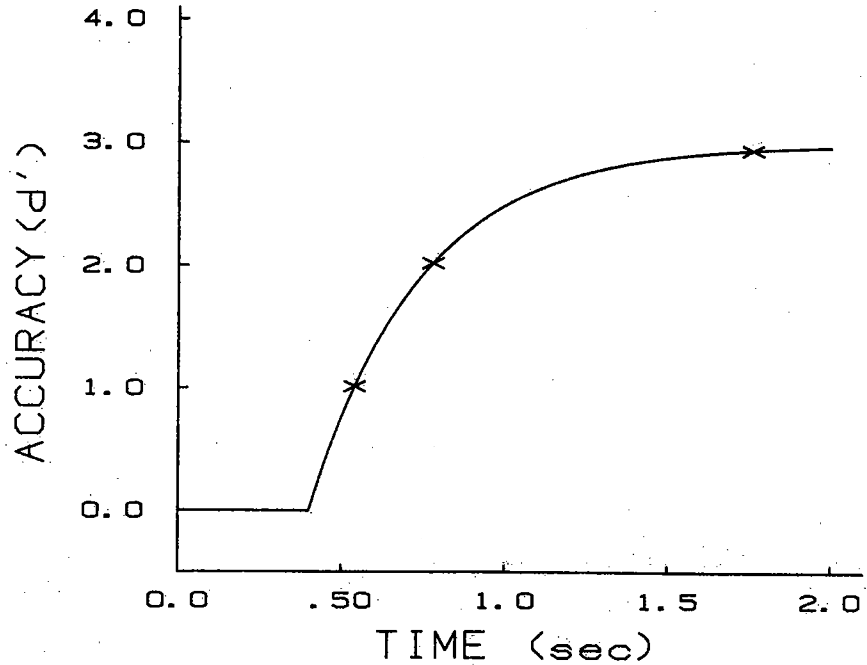

Figure 19. Three well-chosen points are sufficient to determine the parameters of Wickelgren's equation.

whether a manipulation affects the rate or asymptote of a process. If a rate parameter is affected, we can go on to ask whether a process whose rate is affected by a manipulation is the rate-limiting process.

A difficulty with using the parameters of the time–accuracy curve as the basis for inferences about the underlying system of processes is that this method seems to require the collection of a large amount of data from individual subjects. To do these studies right, subjects must be trained to conform their behavior to the deadline or response signal that is in force. Apparently some training is required for most subjects (Shouten & Bekker, 1967; Wickelgren, 1977). In addition, it is not appropriate to average data before computing the time–accuracy curve, since averaging can distort its shape, especially if there are large discrepancies between individuals. Therefore, very large numbers of trials are usually run on each subject (e.g., Reed, 1976, ran 12,180 trials per subject).

Some of these problems can be alleviated, however, since it may not be necessary to obtain more than three different points on the time–accuracy curve for each subject in each condition, given that we know the general shape of the curve. In fact, it is possible to determine the placement of the asymptote by simply obtaining a point far enough out on the time–accuracy curve to ensure that it is effectively asymptotic. Presumably, this can be accomplished by simply giving subjects instructions that stress accuracy without regard to speed, paying them off well for accurate performance.

Once the asymptote has been determined for a particular experimental condition, it remains only to determine the slope and intercept of the approach to asymptote. To do this, it is only strictly necessary to obtain two points along the rising portion of the time–accuracy curve, set at about one third and two thirds of asymptotic accuracy (Figure 19). It may be difficult to choose just the right temporal placement of the deadline or response signal to obtain these points, but perhaps some sort of titration procedure can be used to zero in on it (see Jackson, 1978, for a preliminary attempt). Furthermore, it is not necessary that the placement of the points along the curve be exactly the same in all conditions, so long as the time–accuracy curve conforms to an exponential approach to asymptote following a delay, as given by Wickelgren's equation. If we know the asymptote and have two points on the rising portion of the curve, there is only one curve conforming to Wickelgren's equation that will go through all the points.

An alternative method for extracting the underlying activation function may be simply to allow the subject to respond freely and compute accuracy conditional on reaction time (Grice, Nullmeyer, & Spiker, 1977; Lappin & Disch, 1972a, 1972b, 1973; Shouten & Bekker, 1967). Assuming that variations in reaction time in these studies can be attributed entirely to variation in the placement of a response activation criterion (Grice et al., 1977), it is possible to reconstruct the underlying activation function. Unfortunately, it seems unlikely that this will be the sole source of variation in reaction time; in addition to the possibility of stimulus variability and variability in base activations of response units, response execution time presumably varies in some fashion from trial to trial as well (McGill, 1963).

However we determine the parameters of the time–accuracy curve, we can then compare these across conditions to determine what parameters were affected by the manipulation that differed between them, using Table 4 as a guide. Under the assumptions of the cascade model, it is clear from the table that the joint effects of two manipulations do not always tell us anything more than we can learn from studying the effects of the two manipulations taken separately. Therefore, we may wish to hold off on performing a factorial experiment until we have identified two factors whose joint manipulation could potentially give us more information. For example, if we know that one manipulation only influences the asymptote of the time–accuracy curve, then we know that we will not be able to learn exactly where in the system of processes this effect is introduced by combining this manipulation with another one. On the other hand, if we have two manipulations that affect the dynamics of the time–accuracy curve, then we can combine these manipulations factorially to determine whether they affect the same or different processes, following the logic which guided Sternberg (1969a) when he formulated the additive factors method.

A serious problem arises from the fact that the inferences we can make about the organization of processes in cascade depends on the measurement of the asymptote. Often, howhowever, we may be interested in investigating a process at a high asymptotic accuracy level. It may always be possible to lower the asymptote (e.g., by visual degrading), but lowering the asymptote may affect subjects' choices of processing strategies and remove effective ceilings that might come into play at higher accuracy levels. It is not clear how to get around these problems.

Table 4 provides us with a basis for making inferences about the characteristics of processes that are assumed to be operating in accord with the assumptions of the cascade model. In addition and perhaps just as importantly, it provides a set of predictions derivable from the model that can be used as a basis for testing and possibly rejecting or elaborating the cascade model. For example, in analyzing performance in some task, we may find that two manipulations that affect the delay parameter of the time–accuracy curve will have noticeable underadditive effects on the delay parameter when they are combined. If so, we would know that at least one of the assumptions underlying the cascade model must be incorrect for the system of processes in question.

Discussion

Looking back at the assumptions of the cascade model, it is clear that many of them are oversimplifications, at least for many interesting cases. For example, there is no particular reason why all of the component processes that contribute to the identification of a word need to operate strictly in accord with a unidirectional flow of information from one process to the next. There may be some bypassing of levels (e.g., for some words at least, the outline shape of the word may directly signal its identity, supplementing the usual path of preliminary letter analysis; McClelland, 1977). There may also be within-level interactions (Anderson, 1977; Anderson et al., 1977). Although the asymptotic effects of within-level interactions can perhaps be simulated by between-levels interactions, their dynamic effects may be somewhat different (Furnas, Note 3). Finally, we cannot rule out the possibility of feedback effects in which the results of processing at higher levels affect the activation of units at a preceding level (Rumelhart, 1977).

Even if processes operate strictly in cascade, they may not conform to all of the specific assumptions of the cascade model. For one thing, the simple first-order differential equation given to characterize the dynamics of the system may be a great oversimplification. In fact, a much more complicated nonlinear differential equation appears to be necessary to characterize the behavior of the early stages of visual information processing (Ganz, 1975; Sperling & Sondhi, 1968). It is probably an equally serious oversimplification to imagine that the rate constants of all units at the same processing level are all equal. If this assumption breaks down, the mathematical simplicity of the cascade model is destroyed.

Finally, the assumptions required to derive the timing and accuracy of responding from relative activation curves are simplifications also. For example, there is some indication that response execution time in response signal experiments may depend on the ease of the decision (as given by d'; Corbett & Wickelgren, 1978), violating the assumption that execution time is a constant.

In spite of these oversimplifications, the model seems to have done a good job in accounting for the shape of the time–accuracy curve and for the effects of experimental manipulations from a variety of time–accuracy experiments. Whether these accounts attest to the veracity of the model, to the robustness of its results, or to its indistinguishability from other models is not entirely clear. But several points are worth making. For one thing, if the rate-limiting process in a system of processes in cascade conforms to the simple first-order differential equation given by Equation 2, the exact dynamics of other processes in the system will make very little difference to the general shape of the time–accuracy curve. The Appendix illustrates, for example, that the effects of driving the output of a process against its ceiling can be approximated adequately by an exponential approach to the ceiling as long as the system contains at least one other process that is somewhat slower in rate. The Appendix also illustrates that processes containing units driven by the largest of their inputs also generate curves with the same general shape and show the same effects of parameter manipulations as the curves generated by the cascade equation. McClelland (Note 1) showed that the same applies to the curves generated by units that compare two sets of inputs by computing a weighted product. These facts suggest that the cascade equation (as well as the shifted exponential that approximates it) will serve as a rather close approximation to the function that might be derived from a wide range of more complex models.

It also appears that some of the other findings derived from the cascade model may be quite general as well. First, the fact that the dynamics of all elements of the same processing level are identical does not depend on the simple first-order form of the differential equation given by Equation 2. The dynamic function may be considerably more complex as long as all of the elements within a given level share the same equation with the same constant or constants, and as long as the system remains linear (Furnas, Note 3). Second, it is also likely that manipulations of the dynamics of relatively fast processes will only have a shifting effect on the activation function regardless of their exact dynamics, and manipulations of the dynamics of relatively slow processes will continue to determine the slope of the activation function.

It appears, then, that the cascade model has wide applicability as a first-order approximation and heuristic guide for research. I have only shown how the model may be applied to the analysis of speed–accuracy curves and mean reaction times. However, it should be obvious how the model might be elaborated to generate predictions about the shape of reaction-time distributions and probability of errors in different conditions, following the lead of Ratcliff (1978). Furthermore, the cascade model may also be applicable to experiments studying accuracy of perception under a variety of different target and masking conditions. Thus, the model has the potential for tying together not only reaction-time and speed–accuracy research but also research on the accuracy of perception. The examination of the range of its applicability and of the situations in which it breaks down provides an exciting prospect for future research.

References Notes

1. McClelland, J. L. *On the time-relations of mental processes: A framework for analyzing systems of processes in cascade* (Chip Rep. 77). La Jolla, California: University of California, San Diego, Center for Human Information Processing, October, 1978.
2. Townsend, J. T., & Ashby, F. G. *On the stochastic modeling of elementary psychological processes.* Manuscript in preparation, 1978.
3. Furnas, G. *Notes on a draft of this article.* Stanford University, Stanford, California, June 1978.

References

Anderson, J. A. A theory for the recognition of items from short memorized lists. *Psychological Review*, 1973, *80*, 417–438.

Anderson, J. A. Neural models with cognitive implications. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension.* Hillsdale, N.J.: Erlbaum, 1977.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 1977, *84*, 413–451.

Becker, C. A. Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 556–566.

Collins, A. M., & Loftus, E. F. A spreading activation theory of semantic processing. *Psychological Review*, 1975, *82*, 407–428.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 240–247.

Corbett, A. Retrieval dynamics for rote and visual image mnemonics. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 233–246.

Corbett, A., & Wickelgren, W. Semantic memory retrieval: Analysis by speed–accuracy trade-off functions. *Quarterly Journal of Experimental Psychology*, 1978, *30*, 1–15.

Divinsky, N. *Linear algebra.* Palo Alto, Calif.: Page Flickin, 1975.

Donders, F. C. Over de snelheid van psychische processen. Onderzoekingen gedann in het Physiologish Laboratorium der Utrechtsche Hoogeshool, 1868–1869, Tweede Reeks, II, 92–120. In W. G. Koster (Ed. and transl.), *Attention and performance II.* Amsterdam: North-Holland, 1969. (Reprinted from *Acta Psychologica*, 1969, *30*, 412–431.)

Dosher, B. A. The retrieval of sentences from memory: A speed–accuracy study. *Cognitive Psychology*, 1976, *8*, 291–310.

Egeth, H. E. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1966, *1*, 245–252.

Fitts, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, *71*, 849–857.

Ganz, L. Temporal factors in visual perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 5). New York: Academic Press, 1975.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Grice, G. R. Stimulus intensity and response evocation. *Psychological Review*, 1968, *75*, 359–373.

Grice, G. R., Nullmeyer, R., & Spiker, V. A. Application of variable criterion theory to choice reaction time. *Perception & Psychophysics*, 1977, *22*, 431–449.

Henderson, L. Word recognition. In N. S. Sutherland (Ed.), *Tutorial essays in experimental psychology* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1977.

Hunt, E. Mechanics of verbal ability. *Psychological Review*, 1978, *85*, 109–130.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7). New York: Academic Press, 1973.

Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? *Cognitive Psychology*, 1975, *7*, 194–227.

Jackson, M. D. *Memory access and reading ability.* Unpublished doctoral dissertation, University of California, San Diego, 1978.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293–323.

Lappin, J. S., & Disch, K. The latency operating characteristic: I. Effects of stimulus probability on choice reaction time. *Journal of Experimental Psychology*, 1972, *92*, 419–427. (a)

Lappin, J. S., & Disch, K. The latency operating characteristic: II. Effects of visual stimulus intensity on choice reaction time. *Journal of Experimental Psychology*, 1972, *93*, 367–372. (b)

Lappin, J. S., & Disch, K. The latency operating characteristic: III. Temporal uncertainty effects. *Journal of Experimental Psychology*, 1973, *98*, 279–285.

Link, S. W. The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, 1975, *12*, 114–135.

Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

McClelland, J. L. Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *1*, 80–91.

McClelland, J. L. Letter and configuration information in word identification. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 137–150.

McGill, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

McGill, W. J., & Gibbon, J. The general gamma distribution and reaction times. *Journal of Mathematical Psychology*, 1965, *2*, 1–18.

Meyer, D. E., & Schvaneveldt, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 1971, *90*, 227–234.

Meyer, D. E., & Schvaneveldt, R. W. Meaning, memory structure, and mental processes. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco: Freeman, 1975.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word recognition. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press, 1975.

Miller, J. O. *Effects of stimulus probability on encoding mechanisms in information processing tasks*. Unpublished doctoral dissertation, University of Michigan, 1976.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, *76*, 165–178.

Norman, D. A., & Bobrow, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, *7*, 44–64.

Ollman, R. Choice reaction time and the problem of distinguishing task effects from strategy effects. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, N.J.: Erlbaum, 1977.

Pachella, R. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. New York: Halstead Press, 1974.

Pachella, R. C., & Fisher, D. F. Effect of stimulus degradation and similarity on the trade-off between speed and accuracy in absolute judgments. *Journal of Experimental Psychology*, 1969, *81*, 7–9.

Pew, R. W. The speed-accuracy operating characteristic. In W. C. Koster (Ed.), *Attention and performance II*. Amsterdam: North-Holland, 1969. (Reprinted from *Acta Psychologica*, 1969, *30*, 16–26.)

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, *74*, 392–409.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Mass.: MIT Press, 1968.

Ratcliff, R. A theory of memory retrieval. *Psychological Review*, 1978, *85*, 59–108.

Reed, A. Speed–accuracy tradeoff in recognition memory. *Science*, 1973, *181*, 574–576.

Reed, A. V. List length and the time course of recognition in immediate memory. *Memory & Cognition*, 1976, *4*, 16–30.

Rubinstein, H., Garfield, L., & Millikan, J. A. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 57–62.

Rumelhart, D. E. Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, N.J.: Erlbaum, 1977.

Selfridge, O. G. Pandemonium: A paradigm for learning. In, *The mechanisation of the thought process*. London: Her Majesty's Stationery Office, 1959.

Sharney, L., Wasserman, L. R., Gevirtz, N. R., Schwartz, L., & Tendler, D. Significance of the time lag in tracer movement: Representation of unidirectionally-connected pool sequences by time lag. *American Journal of Medical Electronics*, 1965, *4*, 95–99.

Shouten, J. F., & Bekker, J. A. M. Reaction time and accuracy. *Acta Psychologica*, 1967, *27*, 143–153.

Simmons, G. F. *Differential equations*. New York: McGraw-Hill, 1972.

Sperling, G. Linear theory and the psychophysics of flicker. *Documenta Ophthalmologica*, 1964, *18*, 3–14.

Sperling, G., & Sondhi, M. M. Model of visual luminance discrimination and flicker detection. *Journal of the Optical Society of America*, 1968, *58*, 1133–1145.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, *153*, 652–654.

Sternberg, S. Two operations in character recognition: Some evidence from reaction time measurements. *Perception & Psychophysics*, 1967, *2*, 45–53.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II*. Amsterdam: North-Holland, 1969. (Reprinted from *Acta Psychologica*, 1969, *30*, 276–315.) (a)

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, *57*, 421–457. (b)

Stone, M. Models of choice reaction time. *Psychometrica*, 1960, *25*, 251–260.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychological Review*, 1961, *68*, 301–340.

Townsend, J. T. Issues and models concerning the processing of a finite number of inputs. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, N.J.: Erlbaum, 1974.

Turvey, M. On peripheral and central processes in version: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 1973, *80*, 1–52.

van Valkenburg, M. E. *Network analysis* (3rd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Wickelgren, W. Network strength theory of storage and retrieval dynamics. *Psychological Review*, 1976, *83*, 466–478.

Wickelgren, W. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 1977, *41*, 67–85.

Wickelgren, W. A., & Corbett, A. T. Associative interference and retrieval dynamics in yes–no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, *3*, 189–202.

Appendix

Derivation of the Cascade Equation

The intention of this section is to sketch the mathematics behind the derivation of the cascade equation. The presentation should be reasonably accessible to anyone who has a rough familiarity with the basic concepts covered in a freshman course in calculus. The discussion begins with a presentation of a notation for representing the output of a processing level as a whole and for representing the action of the processing level on its inputs to produce the outputs. The notation comes from the work of Anderson (1977; Anderson et al., 1977). Background for this material may be found in standard texts on linear algebra (e.g., Divinsky, 1975). The next section presents the derivation of the cascade equation. The derivation is not novel, and in fact the dynamic portion of the cascade equation has a long history, reviewed by McGill and Gibbon (1965). A fuller discussion of the techniques used in the derivation is presented in any standard electronics text on network analysis (e.g., van Valkenburg, 1974). More abstract presentations may be found in textbooks on differential equations (Simmons, 1972).

Vectors and matrices. Consider the set of linear integrator units at the nth processing level in a system of processes in cascade. Let $a_{nj}(t)$ represent the activation of unit j at level n at time t, and let a_{nj} represent the asymptotic activation of the unit to an input initiated at time $t = 0$. We can denote the asymptotic output of the process as a whole by $\mathbf{A}_n$, defined by

$$\mathbf{A}_n = (a_{n1}, \ldots, a_{nj}, \ldots, a_{nN}), \quad \text{(A1)}$$

where N is the number of units at level n. $\mathbf{A}_n$ is then a vector in which the value of each element of the vector represents the activation of one of the units. The input to level n is just the output of level $n - 1$, which can be represented by the vector $\mathbf{A}_{n-1}$.

The activation level to which a linear integrator unit will be driven by its inputs is equal to the weighted sum of the inputs. As the system approaches equilibrium, then, the activation of a given unit will approach the weighted sum of the asymptotic activations of its inputs:

$$a_{nj} = \sum_{j'} w_{j'j} a_{(n-1)j'}, \quad \text{(A2)}$$

where j' indexes the units at level $n - 1$ and $w_{j'j}$ is the weight constant expressing the effect of a unit j' at level $n - 1$ on unit j of level n. Consider the entire set of weight constants $w_{j'j}$. We can arrange these constants into a matrix in which each column is the set of weight constants for the effects of each of the units of level $n - 1$ on a single unit at level n, and each row is the set of weight constants for the effects of a single unit at level $n - 1$ on each of the units of level n.

The matrix permits us to translate a vector $\mathbf{A}_{n-1}$ representing the asymptotic output of one level, into a vector $\mathbf{A}_n$, representing the asymptotic output of the next level by simply multiplying the input times the matrix. That is,

$$\mathbf{A}_n = \mathbf{A}_{n-1}\mathbf{M}_n, \quad \text{(A3)}$$

where $\mathbf{M}_n$ stands for the matrix of weight constants associated with processing level n.

The visual input to the processing system can be represented by a vector in which the luminance of each point is given by the value of the corresponding element of the vector. This vector characterizes the input to the first processing level in the system, so it is appropriate to call it $\mathbf{A}_0$.

Each of several processing levels in a system of processes in cascade receives the output of one process and processes it according to its weight constants, passing the result on to the next process. The asymptotic output of level n is then simply given by

$$\mathbf{A}_n = \mathbf{A}_0 \prod_{i=1}^{n} \mathbf{M}_i. \quad \text{(A4)}$$

Vector form of the differential equation. The model assumes that the rate of change of activation of each unit at level n is given by a constant (the rate constant) times the difference between the level the element is being driven to and the activation level the unit has already achieved. Since we have assumed that the k_{nj}s are all equal at level n, the expression for the activation of the level of processing as a whole is

$$\frac{d}{dt}(\mathbf{A}_n) = k_n(\mathbf{A}_{n-1}(t)\mathbf{M}_n - \mathbf{A}_n(t)). \quad \text{(A5)}$$

Laplace transforms. The differential equation represented by Equation A5 characterizes the dynamic behavior of processing level n. It turns out that there is an alternative characterization in terms of the Laplace transform of the characteristic differential equation. Roughly, Laplace transforms are to differential equations what logarithms are to arithmetic expressions: Taking logs lets us do addition instead of multiplication to solve arithmetic problems; taking Laplace transforms lets us do algebra instead of calculus to solve differential equations. First we take the Laplace transform of the differential equation, we manipulate it algebraically, then we take the inverse transform to get the solution.

Let $f(t)$ be a function of t for $t \geq 0$. Then the Laplace transform of $f(t)$ is given by

$$L[f(t)] = \bar{f}(s) = \int_0^\infty f(t)e^{-st}dt. \quad \text{(A6)}$$

Thus, the Laplace transform takes the given function $\bar{f}(t)$ and associates with it another function, $\bar{f}(s)$. The s in $\bar{f}(s)$ is a dummy variable. It can take any value we want to give it, but for now we just leave it as s. $\bar{f}(s)$ is unique to $f(t)$ and vice versa, so that if we know either one, we can determine the other. In fact, tables of Laplace transforms exist to make translation from one to the other quite painless.

The real benefit of Laplace transformation for our purposes comes from the fact that the transform of the derivative of a function is related in a simple algebraic way to the transform of the function itself:

$$L\left[\frac{d}{dt}f(t)\right] = s\bar{f}(s) - f(0), \quad \text{(A7)}$$

where $f(0)$ is the value of $f(t)$ at $t = 0$.

The transfer function. Now we are ready to take the transform of Equation A5. It is simply

$$s\bar{\mathbf{A}}_n(s) - \mathbf{A}_n(0) = k_i(\bar{\mathbf{A}}_{n-1}(s)\mathbf{M}_n - \bar{\mathbf{A}}_n(s)). \quad \text{(A8)}$$

Let us treat the case in which there are no initial activations at any level. (We will consider below the effects of base activations.) Then $\mathbf{A}_n(0) = 0$, so the preceding expression reduces to

$$s\bar{\mathbf{A}}_n(s) = k_n(\bar{\mathbf{A}}_{n-1}(s)\mathbf{M}_n - \bar{\mathbf{A}}_n(s)). \quad \text{(A9)}$$

Rearranging and dividing through by $(s + k_n)$ we get

$$\bar{\mathbf{A}}_n(s) = \bar{\mathbf{A}}_{n-1}(s)\left[\frac{k_n}{s + k_n}\mathbf{M}_n\right]. \quad \text{(A10)}$$

The expression in brackets succinctly expresses the relation between the input to process n and its output. It is called the *transfer function* of process n.

The procedure that produced Equation A10 can be used again to derive an expression for $\bar{\mathbf{A}}_{n-1}(s)$ in terms of its transfer function times $\bar{\mathbf{A}}_{n-2}(s)$. Repeating this process successively and substituting into Equation A10, we get

$$\bar{\mathbf{A}}_n(s) = \bar{\mathbf{A}}_0(s)\left[\prod_{i=1}^{n}\frac{k_i}{s + k_i}\prod_{i=1}^{n}\mathbf{M}_i\right]. \quad \text{(A11)}$$

We now have a transfer function for the entire system of processes given by the bracketed expression. In principle, we can determine $\bar{\mathbf{A}}_n(s)$, and from this $\mathbf{A}_n(t)$, given that we know the form of the input $\mathbf{A}_0(t)$.

Incorporating the temporal form of the input. This article is concerned with stimuli that are instantly illuminated at $t = 0$ and then left on until after all of the events of interest have transpired. In this case, $\mathbf{A}_0(t)$ is given by $u(t)\mathbf{A}_0$, where $u(t)$ is the *unit step function*, the function that has the value 0 for $t < 0$ and 1 for $t \geq 0$. Since $L[u(t)] = 1/s$, Equation A11 becomes

$$\bar{\mathbf{A}}_n(s) = \frac{1}{s}\prod_{i=1}^{n}\frac{k_i}{s + k_i}\left[\mathbf{A}_0\prod_{i=1}^{n}\mathbf{M}_i\right]. \quad \text{(A12)}$$

We need only take the inverse transform of the expression and we are done.

Taking the inverse transform. To find the inverse transform, we have to break the expression into parts whose inverse transforms are known. The problem is solved using Heaviside's method of partial fractions (see van Valkenburg, 1974, for a discussion of the method). The method reveals that

$$\begin{aligned}\frac{1}{s}\prod_{i=1}^{n}\frac{k_i}{s + k_i} &= \frac{1}{s} - \prod_{l\neq 1}^{n}\frac{k_l}{k_l - k_1}\frac{1}{s + k_1} \\ &\quad - \cdots - \prod_{l\neq i}^{n}\frac{k_l}{k_l - k_i}\frac{1}{s + k_i} \\ &\quad - \cdots - \prod_{l\neq n}^{n}\frac{k_l}{k_l - k_n}\frac{1}{s + k_n} \\ &= \frac{1}{s} - \sum_{i=1}^{n}\prod_{l\neq i}^{n}\frac{k_l}{k_l - k_i}\frac{1}{s + k_i}. \quad \text{(A13)}\end{aligned}$$

The inverse transform of $1/s$ is 1 (for $t \geq 0$) and the inverse transform of $1/(s + k_i)$ is $e^{-k_i t}$. Substituting Equation A13 back into Equation A12, we get

$$\mathbf{A}_n(t) = \mathbf{A}_0\prod_{i=1}^{n}\mathbf{M}_i\left(1 - \sum_{i=1}^{n}\left(\prod_{l\neq i}^{n}\frac{k_l}{k_l - k_i}\right)e^{-k_i t}\right). \quad \text{(A14)}$$

For any given unit at level n indexed by j, its asymptotic activation given the presentation of stimulus S $(a_{nj/S})$ is equal to the value of the jth element of the vector $\mathbf{A}_{0/S}\prod^{n}_{i=1}\mathbf{M}_i$. The activation function for the unit is then

$$a_{nj/S}(t) = a_{nj/S}\left(1 - \sum_{i=1}^{n}\left(\prod_{l\neq i}^{n}\frac{k_l}{k_l - k_i}\right)e^{-k_i t}\right), \quad \text{(A15)}$$

which is the cascade equation.

Derivation of the Relation Between Time and Accuracy

Base activations. The base activations represented in the cascade model act as a background that stays approximately constant within a given trial, so that activations produced by stimuli are superimposed on them. Of course, base activations at an early processing level will be propagated to higher levels, but these propagation effects would be expected to stay reasonably static along with their sources. It might seem that these base activations, if incorporated into the cascade equation, would complicate the picture considerably. However, if the system is linear, the effects of inputs coming up through the perceptual system simply add to the base activations. Therefore, we can represent the total activation of response unit j given stimulus S and a base activation that is normally distributed with mean zero and unit variance as

$$a_{rj/S,N}(t) = a_{rj/S}\left(1 - \sum_{i=1}^{n}\left(\prod_{l\neq i}^{n}\frac{k_l}{k_l - k_i}\right)e^{-k_i t}\right) + d(0, 1), \quad \text{(A16)}$$

where $d(0, 1)$ is a normal deviate drawn from the standard normal distribution.

The formula for d'. Consider a trial in a yes–no experiment. Let the activation at time t of the yes response unit, given a stimulus appropriate to that response, be given by $a_{y/y}(t)$, and let the activation of the same unit by a stimulus appropriate to the no response be given by $a_{y/n}(t)$. Let $\hat{a}_{y/y}(t)$ and $\hat{a}_{y/n}(t)$ represent the mean values of these quantities over repeated trials. Also, let $\sigma_{y/y}(t)$ and $\sigma_{y/n}(t)$ represent the standard deviations of the activations of the yes response unit when stimuli are presented from the yes and no stimulus sets, respectively. For responses initiated on the basis of the activations at time t, signal detection theory (Green & Swets, 1966) tells us that

$$d'_I(t) = \frac{\hat{a}_{y/y}(t) - \hat{a}_{y/n}(t)}{\sigma_T(t)}, \quad \text{(A17)}$$

where $d'_I(t)$ stands for the accuracy of responses initiated at time t and

$$\sigma_T(t) = \sqrt{\frac{\sigma_{y/y}(t)^2}{2} + \frac{\sigma_{y/n}(t)^2}{2}}.$$

The cascade equation tells us that $\hat{a}_{y/y}(t) = \hat{a}_{y/y}\Gamma_n(t)$ and that $\hat{a}_{y/n}(t) = \hat{a}_{y/n}\Gamma_n(t)$, so the numerator of A17 becomes

$$(\hat{a}_{y/y} - \hat{a}_{y/n})\Gamma_n(t). \quad \text{(A18)}$$

What is the value of the denominator? In the text, it was assumed that the asymptotic activation of the yes response unit by the stimulus shown on a particular trial was normally distributed about the mean asymptotic effect of the set of stimuli associated with the yes response. That is,

$$a_{y/y} = \hat{a}_{y/y} + d(0, \sigma_{y/y}), \quad \text{(A19)}$$

where $\sigma_{y/y}$ is just the standard deviation of this normal distribution. Note that we have yet to include the deviation due to base activations. Since the effect of the stimulus on the unit varies with the time since stimulus onset by the formula $a_{y/y}(t) = a_{y/y}\Gamma_n(t)$, and since $\hat{a}_{y/y}(t) = \hat{a}_{y/y}\Gamma_n(t)$, we find that

$$d(0, \sigma_{y/y})(t) = d(0, \sigma_{y/y})\Gamma_n(t). \quad \text{(A20)}$$

Thus, we can see that the noise added by the stimulus has standard deviation $\sigma_{y/y}\Gamma_n(t)$. The noise due to the base activations of the response units has a standard deviation of 1, so the standard deviation of the total noise is simply found by pooling the standard deviations of the two additive sources of noise:

$$\sigma_{y/y}(t) = \sqrt{1 + (\sigma^2_{y/y})\Gamma_n(t)^2}. \quad \text{(A21)}$$

An analogous expression can be written for $\sigma_{y/n}(t)$. Substitution of Equations A18 and A21 into Equation A17 produces

$$d'_I(t) = (\hat{a}_{y/y} - \hat{a}_{y/n}) \times \frac{\Gamma_n(t)}{\sqrt{1 + \left(\frac{\sigma^2_{y/y}}{2} + \frac{\sigma^2_{y/n}}{2}\right)\Gamma_n(t)^2}}. \quad \text{(A22)}$$

If we let σ stand for the pooled standard deviation of the two response units (which is equal to $\sqrt{[(\sigma^2_{y/y} + \sigma^2_{y/n})/2]}$) and we take into account the 100-msec delay between the time of response initiation and the time of the actual response, we obtain the expression for $d'(t)$ in a yes–no task given in the text of this article.

Dynamics of Maximum Units

A maximum unit is simply a unit whose output is driven to the activation level of the

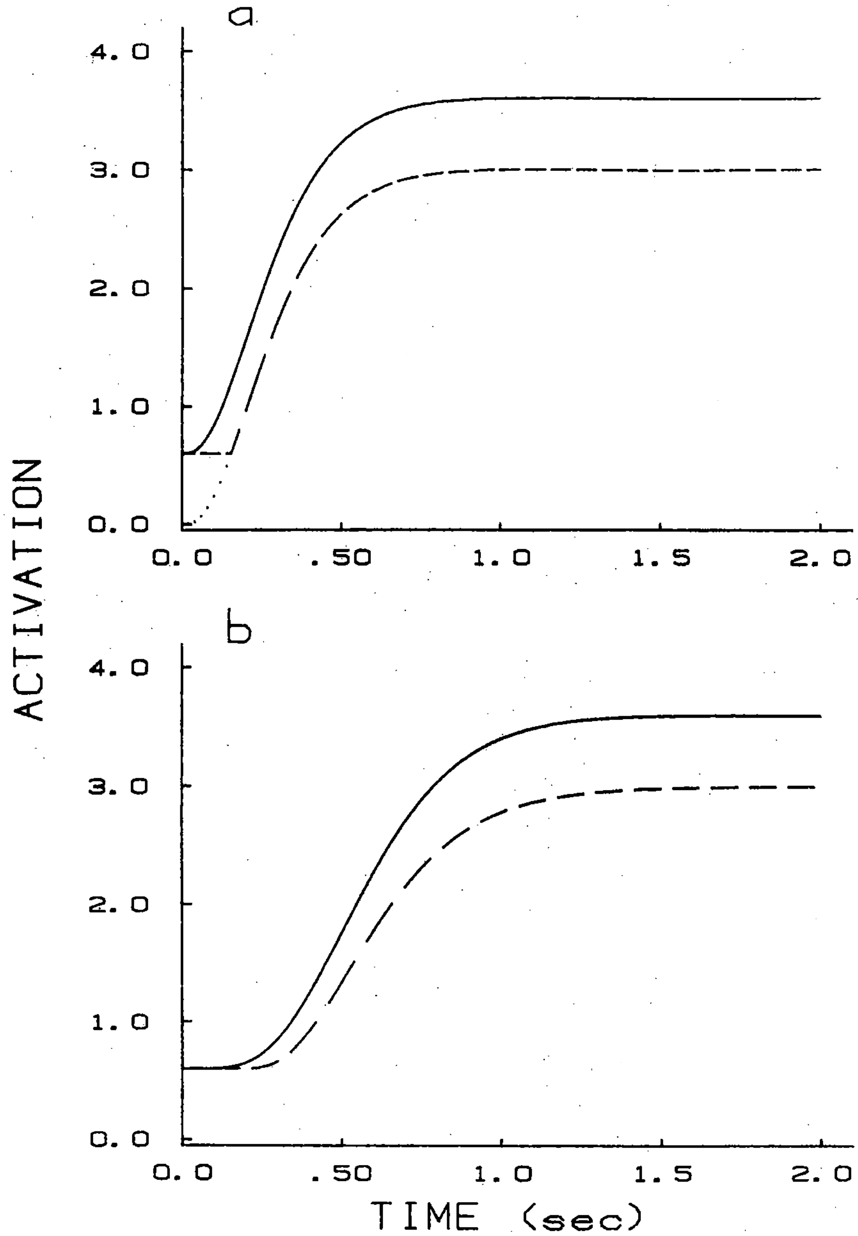

Figure A1. (a) The largest activation in a set of word detectors is shown for two different cases: The solid curve illustrates the case in which the word primed is the word shown, and the broken curve illustrates the case in which the word primed is not the word shown. (The curves are generated assuming three processes preceding the maximum unit.) (b) The outputs of units two processes beyond the process containing the maximum unit for the same two cases. (The rate constants of all processes are approximately equal to 10.)

largest of the inputs. If we assume that the base activations are 0 at all levels on the input side of a maximum unit, the unit will behave exactly like a regular linear integrator with a single input. The input unit with the highest asymptotic activation will have the highest activation for all $t > 0$, so the maximum unit will be driven by the same input unit throughout any given trial.

Complications arise if there are nonzero base activations on the input side of a maximum unit, because the same input unit will not necessarily have the largest activation level through the course of a trial. For example, consider two trials in a lexical decision task. Assume that preceding both trials, the subject is shown the word DOCTOR as a prime. On one trial the target is NURSE, and on the other it is BUTTER. The prime will presumably produce a base activation in the detector for NURSE but not in the detector for BUTTER. At the beginning of the experimental trial, then, the most strongly activated word detector will presumably be the one for the word NURSE, regardless of the actual target. This base activation will of course produce a base activation of the maximum unit and of any other units further along in the system that are activated by the maximum unit or its successors. When the word NURSE is presented, the output of the corresponding detector will grow, and since this detector starts out as the most strongly activated unit, it will continue to be most strongly activated. As a result, the

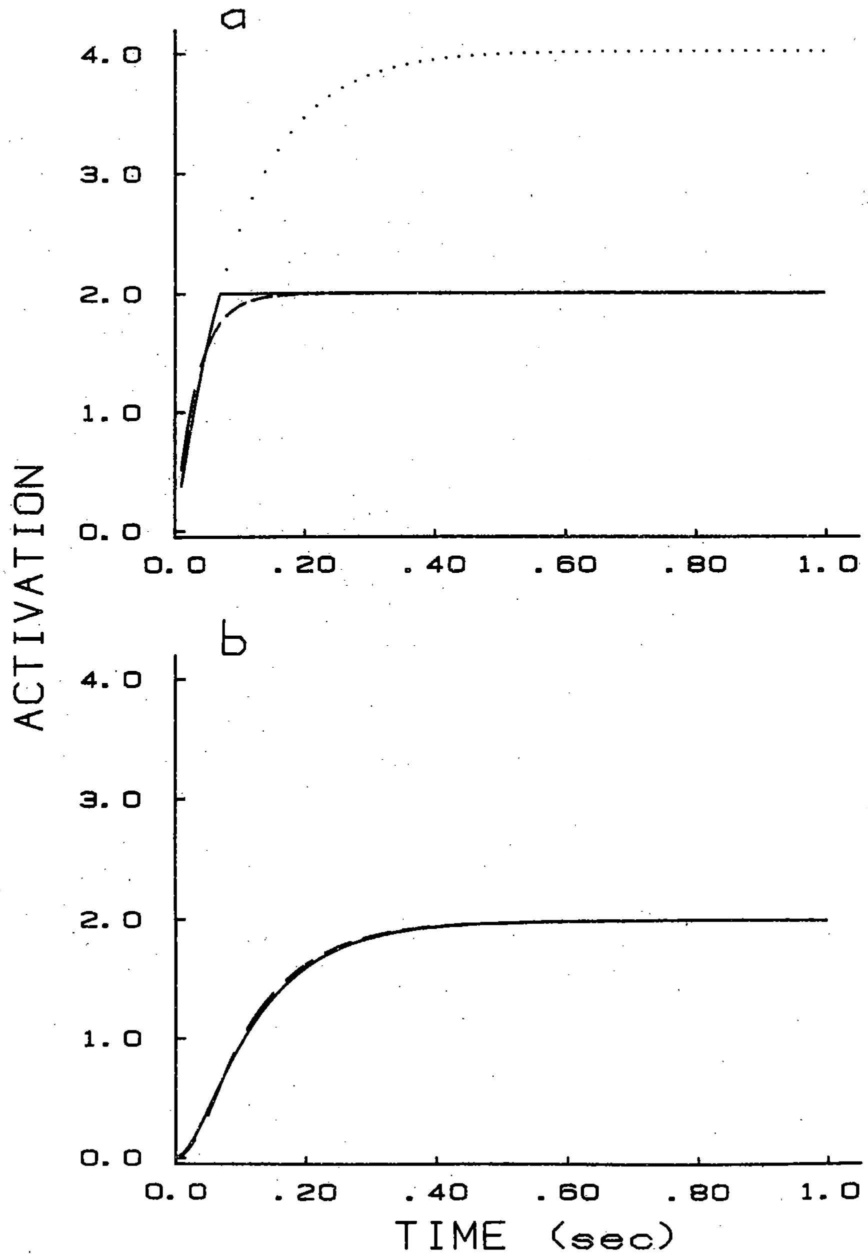

Figure A2. (a) The effect of a ceiling activation level on the activation function for a unit driven by a step-function input. (The input would tend to drive the activation of the unit to twice its ceiling level with a rate constant of 10 [dotted curve], but the growth in activation stops at the ceiling [solid curve]. The dashed curve illustrates an approximation to the solid curve in the form of an ordinary exponential approach to the ceiling with a rate constant of 30.) (b) The output of the unit in a is input to a second unit with a rate constant of 10. (The actual activation function [solid curve] and the simulation based on the fast exponential approach to asymptote [dashed curve] are nearly identical, indicating the adequacy of the approximation.)

effect of the stimulus will simply be superimposed on the base activation of the maximum unit, as shown in the solid curve in Figure A1a. When the target is BUTTER, on the other hand, the largest output of the word level will stay constant until the activation of the detector for BUTTER surpasses the activation of the detector for NURSE. The effect of the unprimed target, then, will not be exerted on the maximum unit until the activation of the detector for the target reaches the level of activation of the primed nontarget, as illustrated by the dotted and dashed curves in Figure A1a. The temporal form of the input to the maximum unit can be approximated by an exponential approach to an asymptote following a delay. Using this approximation, we can easily determine the form of the output of any unit at any number of levels downstream from the maximum unit. Applying this method to determine the output, we can see in Figure A1b that the delay will be propagated throughout the system. This same approximation method was used in generating the curves shown in the discussion in the text of the Meyer et al. (1975) experiment.

Effect of Saturation

What happens when the activation of a unit reaches a maximum level beyond which it cannot be driven by its input? This *saturation* effect is represented in Figure A2 for a unit driven by a step-function input. In the illustrated case, the input would drive the unit to twice its ceiling level if the ceiling were not in place. The further the input would tend to drive the response beyond the ceiling, the faster the activation of the unit would reach its ceiling level. Strictly speaking, the function relating the activation of the unit to time is no longer an exponential approach to asymptote. However, it is possible to simulate a saturating process with a process that rises more quickly to an asymptote placed at the point of saturation. Although Figure A2a indicates that there is some discrepancy between the actual state of affairs and the approximation, the discrepancy washes out when an additional relatively slow process is added, as illustrated in Figure A2b. The slower the rate constant of the added process or processes, the smaller the discrepancy will be.

One effect of saturation is that it makes the rate of approach to the ceiling different for units that would be driven to different asymptotes were it not for the saturation. For example, a ceiling might cause portions of a visual display that are of high contrast to produce saturation before lower contrast portions. In such cases, the assumption that all the units of the same process have the same rates breaks down. The distorting effects of the ceiling could be propagated all the way through a system of processes if there were a consistent mapping from one level to the next so that the units that saturate slowly activate one set of units at higher levels and units that saturate quickly activate a different set of units. However, the assumption of equal rates could still yield an adequate approximation to the output of a multilevel system if units at higher levels integrated inputs from units with a range of different saturation times. Although the exact shape of the time–accuracy curve might be altered in this manner, the independence of rate and asymptote parameters of the time–accuracy curve might still hold up for units at higher levels. In addition, a slow process at a higher level would ameliorate the distortion. In summary, then, it appears that the cascade model may provide an adequate approximation, even for cases in which some of the assumptions of the model are violated.

Received October 20, 1978 ■