COGNITIVE PSYCHOLOGY 7, 44–64 (1975)

On Data-limited and Resource-limited Processes

DONALD A. NORMAN

University of California, San Diego

AND

DANIEL G. BOBROW

Xerox Palo Alto Research Center

This paper analyzes the effect on performance when several active processes compete for limited processing resources. The principles discussed show that conclusions about the interactions among psychological processes must be made with caution, and some existing assumptions may be unwarranted. When two (or more) processes use the same resources at the same time, they may both interfere with one another, neither may interfere with the other, or one may interfere with a second without any interference from the second process to the first. The important principles are that a process can be limited in its performance either by limits in the amount of available processing resources (such as memory or processing effort) or by limits in the quality of the data available to it. Competition among processes can affect a resource-limited process, but not a data-limited one. If a process continually makes preliminary results available even before it has completed all its operations, then it is possible to compute performance-resource operating characteristics that show how processes interact. A number of experiments from the psychological literature are examined according to these processing principles, resulting in some new interpretations of interactions among competing psychological processes.

DATA AND RESOURCE LIMITS ON PROCESSES

The general principles by which information processing systems operate have implications for the study of human processes. The processing resources for any system are limited, and when several processes compete for the same resources, eventually there will be a deterioration of performance. When human processes become overloaded,

We thank Edward E. Smith and Daniel Kahneman for their discussions with us on these topics. The paper was written while D.A. Norman was a fellow at the Center for Advanced Studies in the Behavioral Sciences, Stanford, CA, and we are grateful to the Center for the facilities which they provided. Research support was provided by grant NS 07454 from the National Institutes of Health and by grant GB 32235X from the National Science Foundation. Daniel G. Bobrow is a principal scientist in the Computer Science Laboratory of the Xerox Palo Alto Research Center, Palo Alto, CA. Requests for reprints should be addressed to Donald A. Norman, Department of Psychology, University of California, San Diego; La Jolla, CA 92037.

there often appears to be a smooth degradation on task performance rather than a calamitous failure. This is an important property of the human processing system, one that we call "The principle of graceful degradation." This principle, in turn, implies a basic principle of operation: "The principle of continually available output." By this, we mean that processes must continually provide outputs over a wide range of resource allocation, even when their analyses have not yet been completed. These principles are two of the basic properties that we use in our examination of the ways that interactions among processors can affect performance.

Any information processing device has programs and some mechanisms for executing those programs. When a program is executed, it requires input data and it consumes resources. A set of programs that is being executed for a common purpose and for which resources are allocated as a unit is called a process. Resources are such things as processing effort, the various forms of memory capacity, and communication channels. Resources are always limited. If several processes request a portion of the same available resource, this resource must be allocated among them. The results that the processes produce depend upon the nature of the data which they receive and the amount of resources that have been allocated to them. In general, it is this property that leads to the principal of graceful degradation. However, if there is some critical amount of a resource which is required for the results of a process to be successful, then when the resource available to that process is decreased enough, the gradual degradation will become an observed catastrophic failure in performance. We believe such abrupt performance changes to be the exception rather than the rule.

Because processes can simultaneously compete for a number of different resources, a full analysis of interprocess competition requires examination of each resource competition separately, including an analysis of the trade offs among the various resources and the criteria for scheduling of resources. In this paper we explore the interprocess interaction that results when there is competition for a single resource.

The analyses presented here are related to several previous discussions of processing limitations. Kahneman (1973) has shown the importance of one type of resource—processing effort—in determining how well a task can be performed. Garner and his colleagues (Garner, 1970, 1974; Garner & Morton, 1969) have shown how different aspects of the quality of the data can lead to limitations on performance. We have elaborated upon these and related ideas to introduce the concept of data-limited and resource-limited processes and to use these concepts to analyze a set of experiments selected from the literature on perception and attention.

Resource-limited Processes

Consider the problem of performing a complex cognitive task. Up to some limit, one expects performance to be related to the amount of resources (such as psychological effort) exerted on the task. If too little of some processing resource is applied (perhaps because processing resources are limited by competition from other tasks being performed at the same time) then one would expect poor performance. As more resources are applied to the task, then presumably better and better performance will result. Whenever an increase in the amount of processing resources can result in improved performance, we say that the task (or performance on that task) is *resource-limited.*

The principle of continually available output allows an increased use of computational resources to be reflected in an improvement in performance. If a process using a fixed strategy did not provide an output until it was finished, then increasing resources would simply shorten the time required to get some output But, if the process continually makes available its preliminary results, higher level processes can continually be making use of them. As increased resources allow the process to upgrade the quality of its output, the improvement can be immediately used by any other processes for which the output is relevant. In a similar fashion, processing overloads need not cause calamitous failure, but simply a decrease in performance.

Data-limited Processes

Consider the task of detecting a superthreshold sound: for example, the sound made by striking a piano key in a quiet room. The detection task is straightforward: the processing is limited by the simplicity of the data structure. Consider now the task of determining whether or not a particular signal has occurred within a background of noise. Suppose the recognition mechanism uses all the most powerful techniques at its disposal—matched filters, correlational techniques, and so on. In either of these two tasks, once all the processing that can be done has been completed, performance is dependent solely on the quality of the data. Increasing the allocation of processing resources can have no further effect on performance. Whenever performance is independent of processing resources, we say that the task is *data-limited.*

In general, most tasks will be resource-limited up to the point where all the processing that can be done has been done, and data-limited from there on. There are two forms of data limitations to consider: those resulting from the signal and those from memory.

Signal data-limits. When the task is something like the detection of a weak signal in a noisy environment, the limit to performance depends primarily upon the signal-to-noise ratio. When performance is directly dependent on the quality of the input data signal, we call the process *signal data-limited.* Most psychophysical tasks and many discrimination tasks are signal data-limited.

Memory data-limits. When the task is something like performing an absolute identification of a clearly audible signal, or perhaps identifying which of two almost identically oriented, clearly presented lines has just been seen, neither an improvement in the quality of the input data nor the allocation of more resources will improve performance. The bottleneck is in the quality of the representation of the stored paradigm. To improve performance one must improve the memory. We call this a *memory data-limited* process.

Garner's state and process limits. Our distinction between signal data-limited processes and memory data-limited processes is identical to that proposed by Garner (1970, 1974) as state- and process-limited operations. Different experimental manipulations can separately affect the signal data-limitations and the memory data-limitations. Flowers and Garner (1971) show how to affect signal data-limitations, and in his book (Garner, 1974; especially Chap. 7), Garner discusses those factors which affect memory data-limited processes.

The Performance-Resource Function

In general, the function that relates performance to resource allocation should be monotonically nondecreasing. In order to determine any particular performance-resource function, one must know about the details of operations of the processes in question. A performance-resource function may be continuous, or performance may sometimes increase in discrete, quantized increments which would in turn require discrete, quantized amounts of resources. Often, some minimum threshold value of resource must be allocated before there is even the initial processing output (call that value $R_{\min}$). When the function reaches an asymptote in performance it is data-limited: beyond that point, increases in the allocation of resources can have no effect. Call the level of resource allocation where performance becomes data-limited R_{dl}. When r, the resources actually allocated to a process are less than R_{dl} and greater than $R_{\min}$ ($R_{\min} < r < R_{dl}$) then the process is, by definition, resource-limited. (The performance-resource function may have zero slope in the resource-limited portion of its operation if performance can only take on discrete values.)

Figure 1 illustrates these points with one composite performance-resource function. It shows a function with a resource threshold, $R_{\min}$. When the resource allocated exceeds this value, performance immediately attains some initial level. From there, the function is resource-limited, with one discrete step in its performance, which remains con-

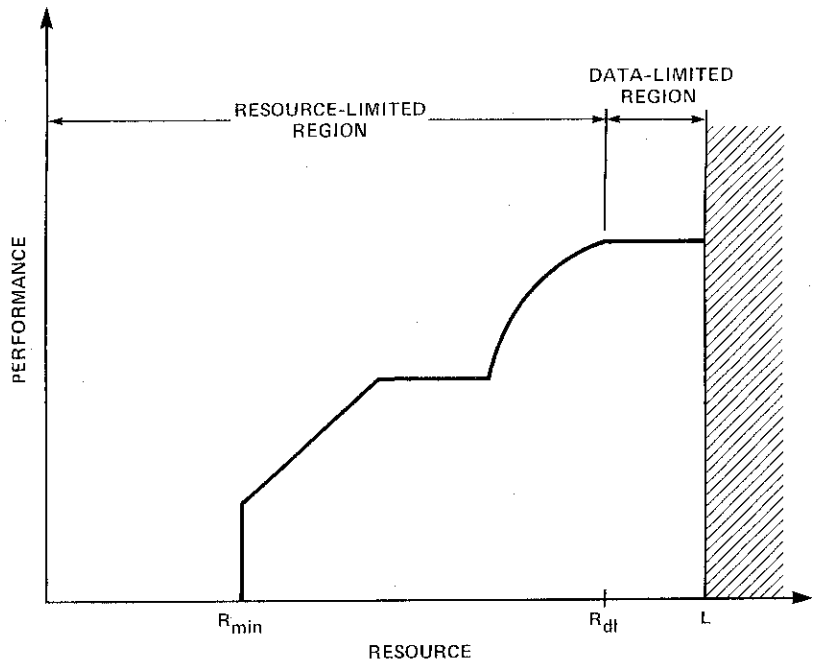

FIG. 1. The performance-resource function. Performance is a monotonically nondecreasing function of the amount of processing resources that are allocated, with the upper limit on available resources given by L. Performance within the data-limited region of operation is independent of the expenditure of processing resources. The exact form of the relationship between performance and resource allocation within the resource-limited region depends upon the details of operation of the processes which are involved.

stant until the application of a discrete minimum of additional resource. Finally, when the resource allocated exceeds R_{dl}, the performance is data-limited.

In actuality, performance-resource functions need not have all the features shown in Fig. 1. A pure, continuous, resource-limited function would rise smoothly from the origin up to the limit of processing resources, L. A pure, data-limited function would be a horizontal line, with performance completely independent of resources. A special case of the pure, continuous, resource-limited function would be one that follows the square-root law: performance is proportional to the square root of processing resource consumed. This result will occur for any process in which increasing the resource is equivalent to increasing the number of independent data samples, leading to a corresponding decrease in sampling variance. If the number of samples changes linearly with resource, many measures of performance which depend upon sample variance will then increase with the square root of the resource

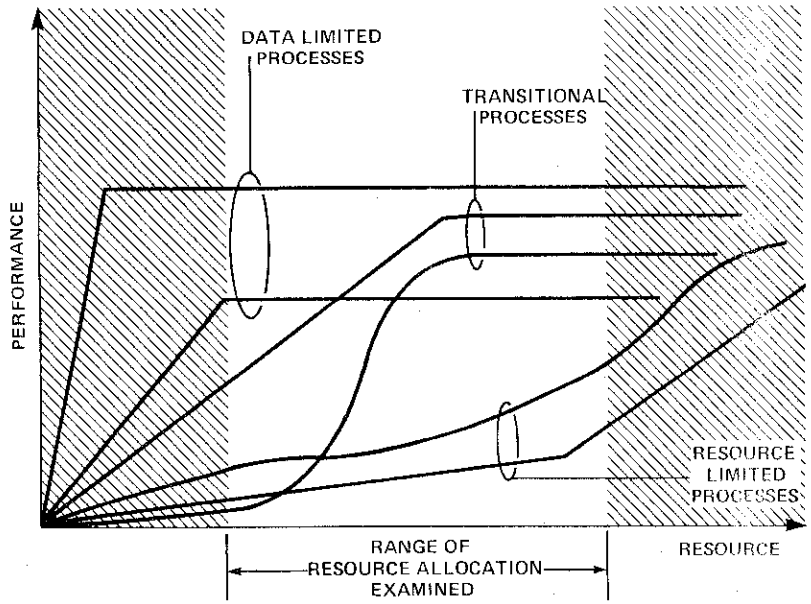

FIG. 2. Observable classes of performance-resource functions. When processes are examined only over a limited range of resource allocation, some will appear to be independent of resources (because they are data-limited in the region under consideration), others will appear to require indefinite amounts of resources (because they are resource-limited within this region), and others will be in a transition between data-and resource-limited operation.

allocated (usually, the resource that is involved will be processing effort). The square-root law has been observed with the d′ measure of detection theory for performance (Taylor, Lindsay & Forbes, 1967).

Most processes have both data-limited regions and resource-limited regions. Whenever some medium range of resources are allocated to a group of processes (the situation that probably applies to most experiments), some processes will appear to be always data-limited (therefore not evidently affected by the availability of additional processing power), others will appear to be always resource-limited (therefore improving with the availability of additional processing power), and others will appear to be in the transition stage between resource-limited and data-limited, their exact status depending upon how much processing resource is allotted (see Fig. 2).

LIMITED CAPACITY CENTRAL PROCESSING

When several active processes compete for the same limited resource, then the performance-resource functions of these processes become critically important in determining just what effects will be observed. We

assume that there is a fixed upper limit on available processing resources: Let the limit be signified by L. Operations which share the same limited capacity mechanism will not interfere with one another until the total processing resources required by all exceeds L. Moreover, in any given range of resource allocation, one process may interfere with others, but the others need not interfere with it. Just what kind of interference effects are found depends upon the particular form of the performance-resource function for each process. Interference can only be observed when a process is operating within its resource-limited region. Note, therefore, that the effects of interference need not be symmetrical. If task A interferes with task B, but not the reverse, then it would be incorrect to conclude that one of these tasks does not require processing capacity from the same central pool as the other. On the contrary, interference in either direction implies that both tasks draw resources from the same common pool. The asymmetry in effect results when one task is data-limited while the other is resource-limited. The symmetry or asymmetry of interference between two tasks is likely to depend in large part upon task instructions and subject strategy—upon which of the competing tasks receives first priority. The high-priority task will tend to be data-limited, and the low-priority task resource-limited.

Wherever two tasks show an asymmetry in interference effect, it should be possible to demonstrate interfering effects on both by a sufficient change in the availability of processing resources. One can change the available resources either by increasing or reducing the demands of existing tasks or by adding or removing tasks. As usual, some caution must be used in deciding whether or not one has managed to change resource allocation. If $R_{dl} < R_{\min}$ for a task, then that task is always data-limited, and the only way to change its demand upon resources is either to remove it or to add it anew: no partial allocation of effort is possible.

The assumption that all processes draw from a common pool of resources does not imply that all interfere with one another in the same manner. For example, tasks which invoke different sensory modalities (see Brooks, 1967, 1968, for example) may not compete for processing effort, but for the use of common data structures and common memory. This is what Kahneman (1973) calls *structural interference*. Multiple resource competition can be factored into analyses of simple resource competition or of trade-offs among resources. (Kerr, 1973, discusses the various forms of resource competition that have been studied in the literature. Greeno & Simon, 1974, show how the same task can be performed with different trade-offs among the necessary resources, especially between those resources of memory load and processing effort.)

Performance Operating Characteristics

Whenever several processes must share the total capacity of available processing resources, then knowing the performance-resource functions and the division of resources for each process allows one to determine the resulting division of performance. Moreover, as the resource allocations change, the resulting changes in performance can easily be computed. In the simplest case where two processes are competing, it is possible to trace out a performance operating characteristic which shows how the performance of one process varies with the performance of the other. (Of course, the operating characteristic can be computed with any number of processes, but one orthogonal dimension is required for each; the resulting n-dimensional functions are difficult to plot.)

Performance operating characteristics are illustrated in Fig. 3. They are computed by assuming complete complementarity of processing resources for the two processes. When process 1 uses some amount of resources, r, process 2 uses an amount $L - r$. To trace out a curve, r is allowed to vary from 0 to L. In general, the performance operating characteristic will show a monotonically nonincreasing relation between the

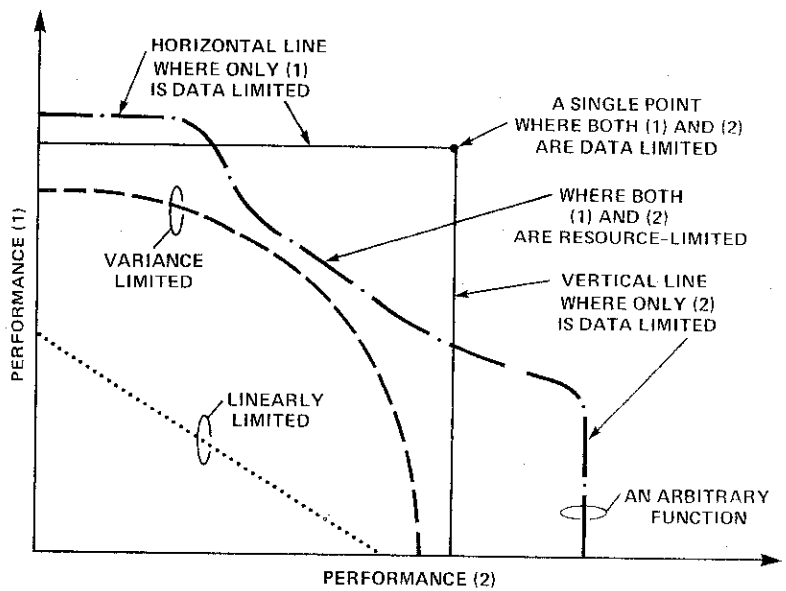

FIG. 3. Performance operating characteristics. These operating characteristics show how performance on two processes co-vary as the division of processing resources between them varies. If total available resources is L, and if the resources expended on process 1 is r, then the resources available for process 2 is $L - r$. Each characteristic function is obtained by letting r vary from 0 to L.

performance on one process and the performance on the other. In regions where one process is data-limited whereas the other is not, the function will be either a horizontal or a vertical line. Where both are data-limited, the resulting function is a single point. These are all illustrated in Fig. 3. The functions are of most value when both processes are resource-limited. Figure 3 shows two special cases of resource-limited functions: first, when both are variance-limited so the resulting characteristic follows a circle (assuming normalization of the two performance scores); second, when performance on both are linear with resource so the operating characteristic is linear. The circular operating characteristics have the form found by Taylor, Lindsay and Forbes (1967): they found that for some classes of competing tasks, regardless of the division of resources, the sum of the performance measures for each remained constant [the sum of the $(d')^2$ was constant].

Measuring the Performance-Resource Function

To determine a task's performance-resource function we need to vary processing resource systematically while measuring some aspect of the task performance. This is not easy to do. In fact, we have been unable to find any instances of experiments in the literature that allow us to illustrate actual performance-resource functions. The major difficulty comes with the control of the resource allotment.

Perhaps the best way to control resource is to require subjects to perform two tasks simultaneously. One task is the interfering task. It should require a fixed amount of resource. The other is the primary task. It is assumed to use all the remaining available resource. We measure performance on this second task as we vary the amount used by the first. To do this set of experiments, we need to have an interfering task whose resource requirements can be systematically and consistently varied over a wide range. Few tasks meet this requirement.

Consider any task in which we can control the amount of resource required. The best way to exert this control is to establish a narrow range for errors and require that performance fall within this range, neither higher nor lower. Thus, in a tracking task, we might measure the mean square error and require the subject to maintain performance within a narrow range. Similarly, in a shadowing task, we might monitor the shadowing behavior and require subjects to maintain a fixed error rate. It is critically important that subjects be controlled both on the upper and the lower limits of their accuracy at the task. The common instructions to subjects usually emphasize the maximum error rate that is permitted, but allow any level that falls below that. This specification only establishes a lower bound on the amount of resource exerted at the task. Once we are able to control the error rate of the subject, we can vary the resource expended upon the task, and therefore the amount available for other tasks.

To determine a performance-resource function we need to know how to transform the error rate on the competing task to resource allotment for the task of interest. In general, we do not know the appropriate transformation.[1] In principle, by using the procedures of either functional measurement (Anderson, 1974, in press) or conjoint measurement (see Krantz, Luce, Suppes & Tversky, 1971), it ought to be possible to determine performance-resource functions, using several primary tasks, all of which are calibrated against the same interfering task.

Reaction Time and Accuracy as Performance Measures

The two most widely used measures of task performance are reaction time and accuracy. Unfortunately, these two measures differ. Thus, in many psychophysical tasks and tests of recognition memory, fast responses tend to be more accurate and also to receive higher confidence judgements than slow responses (see, for example, Norman & Wickelgren, 1969). In other tasks, however, it is the slower responses that are more accurate.

Whenever one finds a speed-accuracy tradeoff, the underlying assumption is that a fast response occurs when a subject has curtailed the normal processing and, as a result, is both faster and less accurate. In situations where the performance-resource function has a resource threshold ($R_{min} > 0$), then curtailing processing below R_{min} will no longer provide decreased processing time. This is the situation assumed by "deadline models" of reaction time (see Ollman & Billington, 1972; Yellot, 1971).

To determine the relationship between speed and accuracy measures, one usually needs to know something of the structure of the underlying processes. However, we argue that in general in a data-limited process, we expect reaction time to relate inversely to accuracy. If the input data happen to be of relatively high quality, then the analysis is simpler and a relatively more accurate response can be made quickly; for low quality input data, we would expect a slower and less accurate response. When a process is resource-limited, then we expect reaction time to be directly related to accuracy, because better resulting output is dependent on more processing resources being allocated to the process. This usually will require more processing time, hence the positive correlation

[1] Often, theoretical considerations allow one to determine the relationship between error rate and resource. Thus, in a tracking task, one might expect performance to be variance-limited, so that mean square deviations in tracking might decrease linearly with processing resource. Such assumptions, of course, would have to be tested before being applied to the present analysis.

between time and accuracy. Thus, only in resource-limited processes is there a speed accuracy trade-off.

The analysis presented here is quite similar to that presented by Thomas (1973) who showed how "the relationship between the speed and accuracy of a response depends on the quality of sensory information on which the response is based" (from the abstract, page 613). Thomas' analysis is restricted primarily to data-limited situations, and he assumes that the subject tends to operate with an allocation of resources equal to R_{dl}, the transition between resource- and data-limited operation. This is an optimum place to operate, for it represents a maximization of performance without wasting processing resources. Any further processing would only delay the response, but would not increase accuracy. Detailed examination of this issue is not appropriate here, and the interested reader is referred to the paper by Thomas for further details.

ANALYSIS OF EXPERIMENTAL RESULTS

Almost all processes will have regions that are resource-limited and regions that are data-limited. A failure to recognize this distinction lies at the apparent discrepancy in many reported experiments: when one experimenter reports interfering effects of one task upon the performance of another and a second experimenter finds no interference, the difference can most simply be traced to the fact that one worked within the region where both functions were resource-limited whereas the other did not. Alternatively, as discussed above, measures taken may reflect varying data properties, and not be true performance measures.

Focal Attention—Beck and Ambler

Beck and Ambler (1973) performed an experiment that appears to be ideal for illustrating the effects of data-limited and resource-limited processing. In their experiment, Beck and Ambler presented the subject with 8 letters arranged in a circle. This display was flashed for 50 msec and was followed by a masking field. (Subjects were required to maintain fixation in the center of the 36° display, so letters were always viewed peripherally.) The task of the subject was to determine whether the letters were all upright T's or whether the display contained a disparate letter. The disparate letter could be either an L or a tilted T (the T was tilted 33° from the vertical). The subject was about equally able to detect a disparity caused by the L or the tilted T when he was prewarned of the critical location 150 msec. before the display: the error rates were 15 and 19%, respectively. When the possible positions for the disparate letter increased from the known position to any one of two positions or any one of 8 positions, ability to detect the disparity remained essentially unchanged when it was caused by a tilted T (the error rates were

19, 22, and 19%) but the ability to detect the disparity caused by an L steadily decreased (the error rate increased from 15 to 31 to 41%).

Beck and Ambler concluded that their results "demonstrate that focal attention increases the sensitivity of the visual system to peripherally presented differences in line arrangement" (1973, p. 229). This would appear to be true, but we describe the results differently in terms of the processes we speculate are necessary to detect an L, an upright T, and a tilted T. To determine the difference between a tilted T and an upright T, all that is needed is the detection of a diagonal line—any diagonal line. The tilted T contains two lines, one at an angle of 33° from the vertical, the other 123° from the vertical. If the tilted T process detects either line, it is successful. (Note that the process needed to detect the tilted T is critically dependent upon the experimental context. In this experiment it is simple. It would be more complex if the task were to detect a tilted T among an array of X's.)

The process necessary to discriminate an L from a T is more complex. First, it must detect the presence of both a vertical and a horizontal line segment. Second, it must ascertain that the two line segments have the proper relationship to one another. Not only is the L process more complex than the tilted T process, but it must take longer to perform, for it requires a series of tests, one of which must await receipt of information from two others. Thus, the performance-resource function for a tilted T will become data-limited much earlier than will the function for detecting an L.

The resulting performance-resource functions are shown in Fig. 4. Here we see that in the experimental condition with focused attention, both processes are essentially data-limited by the discriminability limits set by the exposure duration, contrast, letter, font, and mask. As attention became more and more diffuse (going from 1 known position to 2 possibilities to 8), the amount of resources available for a process that examines the letters at a possible target position continually decreases. This decrease of available processing resources brings the operation of the L process from its data-limit region into the resource-limited portion of its operation. For these experimental conditions, the tilted T process stays within the data-limited portion of its operation.

This analysis provides an explanation for another finding of Beck and Ambler. Note that the L process is more complex than the tilted T one and it has decision components that depend upon the results of prior processes. Hence, even when performance is equated on both processes (when both are in the data-limited portion of their operation), it should take longer to detect an L than to detect a tilted T, as was reported. In addition, we would predict that the finding is highly sensitive to the set of letters used in the task. Were the task to detect a disparate L or tilted

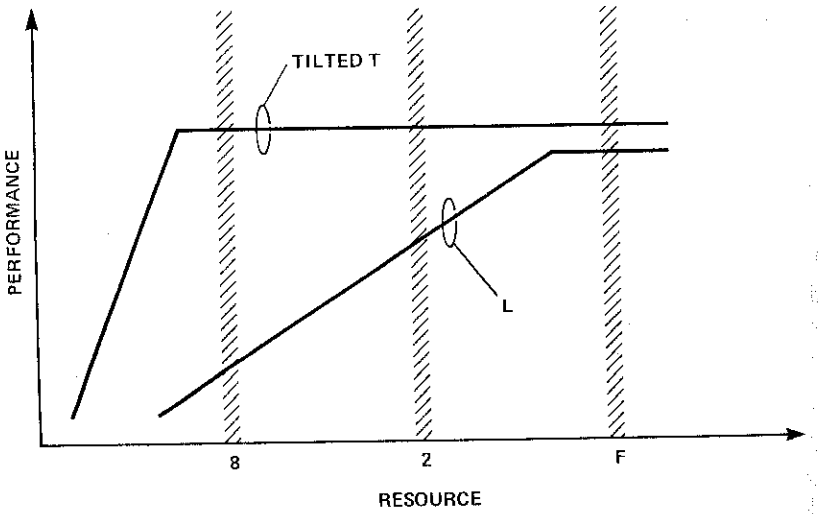

FIG. 4. The performance-resource function for the Beck and Ambler experiment. The three vertical lines indicate the resources available for each process in the conditions of focused attention (F), where the target position is in one of 2 places (2), and where the target position is in one of 8 places (8). The diagram shows why performance remained unchanged for the detection of tilted T's as the condition was changed from F to 2 to 8, but continually decreased for the detection of L's.

T in a circle of X's, the results would probably be exactly reversed. Finally, because the process for detecting a tilted T in this experiment need only contain a check for any diagonal line, we would expect that the results would be essentially unchanged were a tilted L or an X substituted for the tilted T.

The analysis just performed can be applied in essentially the same form for a number of other experiments in the literature, including Leibowitz and Appelle (1969) and Mackworth (1965). A number of experiments reported by Erikson and his collaborators fit very similar patterns (see, for example, Colegate, Hoffman & Eriksen, 1973).

Channel Independence in Visual Processing—Gardner and Estes

A number of different investigators have examined the independence of the processing of information from different channels, all of which are presented simultaneously. (A useful review of these results is given by Smith & Spoehr, 1974.) In general, it would seem the results are consistent with the experiment by Beck and Ambler (1973) which we have just reviewed. When there is one target item to be detected amidst other nontargets, then the number of other items does not seem to affect detectability if they are not confusable with the target. Detectability of the target decreases if the other items can be confused with the target.

Our analysis of this general result is basically analogous to the way in which we handled the results from Beck and Ambler. When the nontargets are not confusable with the target, they are data-limited processes: not much processing resource is expended upon them. When the nontargets are similar to the targets, however, then much more processing resource must be expended to analyze them and the entire system becomes resource-limited. A good example of this type of finding can be found in the experiments reported by Gardner (1973).

Estes (1972) has examined the interactions of signals, memory load, and background on visual processing and suggests that two different processes seem to be occurring in his experiments: A primary detection process which is characterized by high accuracy and short latency, and a secondary detection process which may require a fuller sensory analysis and deeper coding and decision processes than need be performed for the primary process. For our purposes, we identify his primary processes with ones that are essentially data-limited and his secondary processes with ones that are resource-limited.

Selective Attention—Treisman & Geffin and Lawson

The literature on selective attention provides a rich set of data to be analyzed. Consider the experiment in which a subject is presented with two channels of spoken information by having two voices played to him over earphones, one voice to each ear. He is asked to repeat aloud the words that he hears on one channel (the procedure is called "shadowing") while the experimenter manipulates what happens on the other channel. In the literature the channel that is to be shadowed is called the *primary* channel and the other the *secondary* one.

In one such experiment, performed by Treisman and Geffin (1967), special target words were inserted into either the primary or secondary channels and subjects were instructed to tap the desk with a ruler whenever they detected a target word, no matter on which channel the target had occurred.

For the purposes of this experiment, we must compare the recognition of words on the primary channel with the recognition of words on the secondary channel. Processing resource is divided into two parts, with R_p going to the primary channel and $R_s = L - R_p$ going to the secondary channel. The relevant performance-resource functions are shown in Fig. 5. The performance on the primary task—shadowing—is determined by the instructions to the subject. Its level is high, yielding a reasonably high accuracy. The shadowing level tells us something of the performance level and thereby allows us to determine the primary processing resource, R_p. That being known, the maximum value that the secondary

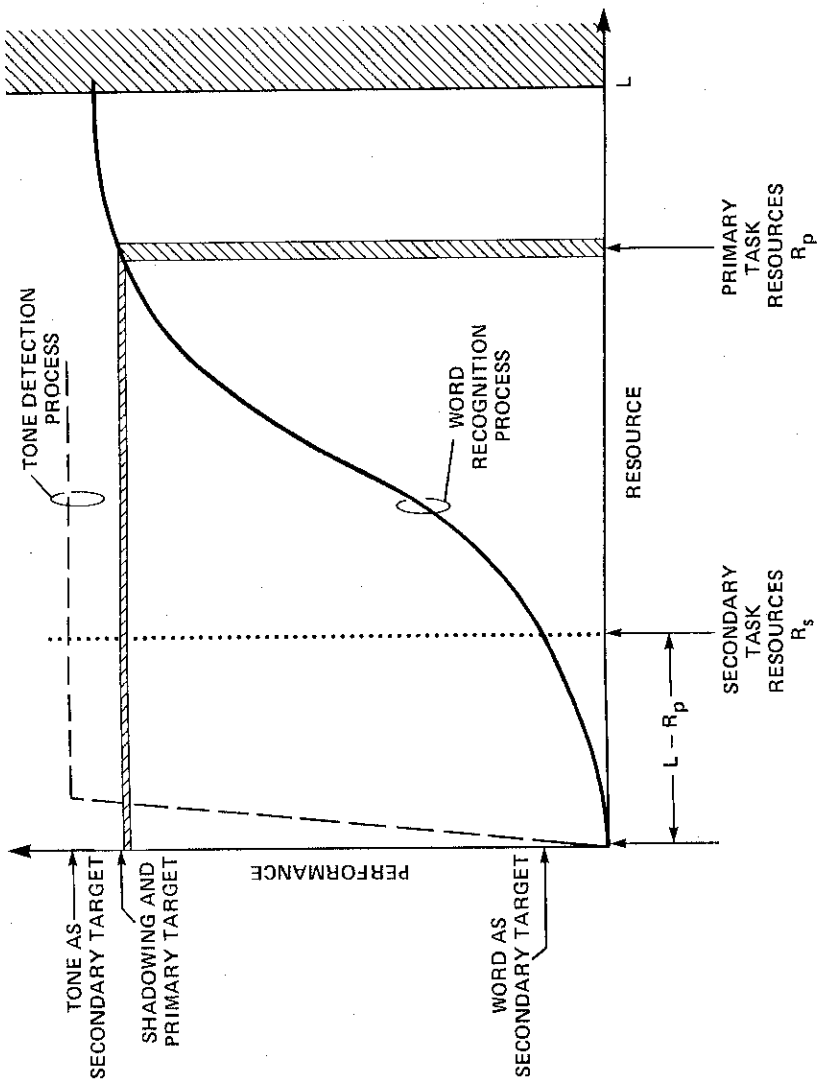

FIG. 5. The performance-resource function for a shadowing task. When the primary task is to shadow, sufficient resources must be allocated that performance is relatively high. As a result, performance on a secondary task which is governed by a similar performance-resource function must appear to be "attenuated" (most secondary tasks which are verbal meet this criterion). A task which operates upon a much simpler performance-resource function will still be performed at a high level despite the simultaneous performance of the primary task. This is illustrated with a data-limited function corresponding to the task of detecting an auditory tone.

processing resource, R_s, can take on is $L - R_p$. Figure 5 shows that this guarantees a low level of performance for the detection of words on the secondary channel.[2]

Suppose the target signal were an auditory tone instead of a word. Presumably the process necessary to detect a superthreshold tone and to discriminate it from speech sounds is rather simple, implying that it becomes data-limited at low resource values. Thus, as shown in Fig. 4, a tone should always be detectable on the secondary channel, even with very high performance levels on the primary channel. This is essentially the result found by Lawson (1966).[3]

The analysis presented here is quite consistent with the idea that unattended inputs are "attenuated." The difference is that we are stating how a division of processing resource might force a process lower down on its resource-limited function, thereby essentially "attenuating" its analysis. The notion of "attenuation" is thus seen to depend critically upon the principle of continually available output in the operation of the relevant processes.

Simultaneous Attention—Moray and Sorkin

In recent years, Moray and Sorkin (see Moray, 1974; Moray & Fitter, 1973; Sorkin & Pohlmann, 1973; and Sorkin, Pohlmann & Gilliom, 1973) have demonstrated that when subjects are not required to make responses, they seem able to monitor signals arriving simultaneously on two (and possibly more) channels without any apparent interference. However, when responses are required, there can be severe decrements in two-channel tasks. As Moray, Favreau and Nagy put it: "When timesharing their attention between two messages observers appear to process them in parallel provided they do not believe that they have de-

[2] In the Treisman and Geffin experiment, the probability of detecting the target word on the primary channel usually differed from the probability of performing a shadowing response to the same word. To the processes shown in Fig. 5 we must also add two response processes: one that executes a spoken response; one that performs a tapping response. The shadowing process is the more complex of the two, and it must have a performance-resource curve that lies below the corresponding curve for the tapping response process. In addition, some target words were embedded with a sentence context and some were not. Because the process for recognizing a word in a sentence differs from that for recognizing a word out of context, different processes are probably involved here as well. The major point can be made, however, without need for these elaborations.

[3] There are other aspects of the attentional system that require explanation, such as the fact that secondary messages that fit within the context of the primary message are frequently responded to, as is the subject's own name, and so on. These results are all compatible with the present formulation. To describe these phenomena requires a description of the scheduling algorithm for the processes, as well as some elaborations on the inner operations which they perform. These discussions are beyond the range of the present paper.

tected a target. At moments when they believe a target to be present, they appear to alter their criterion for making a response to the contralateral message, and usually behave as if the signals on the contralateral channel to that in which they believe a target to be present have become less detectable" (Moray, Favreau & Nagy, 1973).

We interpret these findings to indicate that the amount of processing resource required to monitor a channel is not large relative to the amount of available processing capacity. Thus, monitoring is usually a data-limited process. Initiation of a response, however (which perhaps includes the final stages of decision about the presence of the response), is a resource-limited function. Thus, whenever a response process is initiated it would appear to require sufficient processing resource to prevent any other response process from performing well. It is also possible that the required resource is sufficient to cause the monitoring performance on other channels to drop to the resource-limited range.

Sorkin and Pohlmann (1973) make the point that a signal need not actually occur for this decrement to arise: all that is needed is a sufficiently large observation (perhaps caused by noise) that the observer initiates the final decision and response processes.

Perceptual Learning—LaBerge

A common finding in the study of motor or skill learning, is that as a task is repeatedly practiced, its performance appears to become "automated," apparently requiring less and less conscious processing (see Woodworth, 1938). For present purposes, we need not describe exactly the changes that take place in processes as they are learned and practiced except to note that they become more and more efficient, apparently eliminating processing steps, or learning to process only the relevant input data and to ignore the rest, or refining the accuracy of their outputs. Whatever the nature of the change, it is reflected in increased performance for given resource. That is, as shown in Fig. 6, with increased learning, processes reach their data-limited portion sooner.

If this is so, then we should be able to demonstrate the effects of learning by adding interfering tasks to force the processes into the resource-limited region of operation. Essentially just this demonstration has been performed by LaBerge. In a series of experiments, LaBerge (see, for example, LaBerge, 1973) has shown that with practice, new perceptual figures appear to become more automated in processing, requiring that less and less attention need be paid to the stimulus in order to complete the analysis. LaBerge showed, moreover, that subjects often appear to perform equally well to newly learned patterns and to well learned patterns if their attention is focused on the stimulus, but the performance on newly learned patterns deteriorates if subjects are

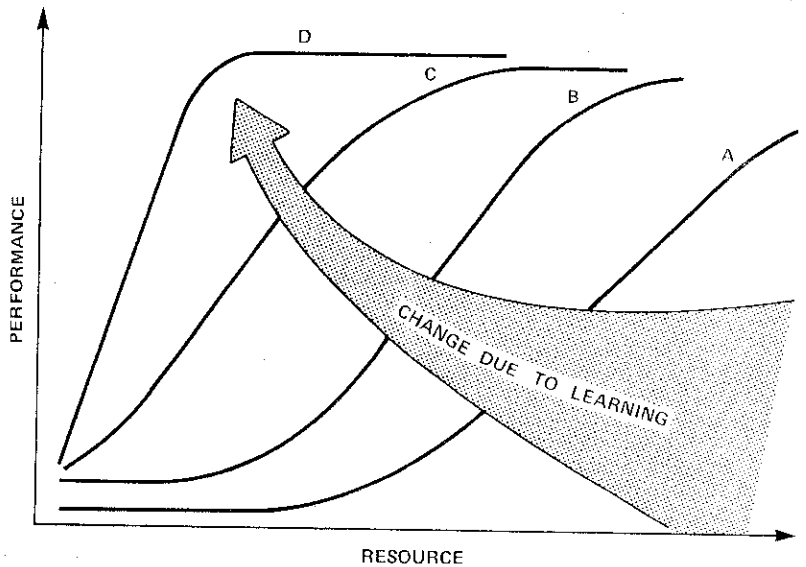

FIG. 6. The effect of practice on the performance-resource function. Continued learning changes the performance-resource curve from A to B to C to D.

presented with the pattern at moments when they do not expect it. When only a single, expected task is tested, then both well learned and newly learned processes will be in the data-limited portions of their operations: hence, both will appear to give equal performance. Under conditions of distraction, however, the newly learned process can be driven to the resource-limited region, whereas the well learned process will often stay within the data-limited region. Presumably, severe attentional distraction will force even the well learned process towards the resource-limited portion of its operation.

CONCLUSION

When an information processing task is performed, the result depends both upon the quality of the data and upon the processing resources that are used. By considering the effects of the data quality and processing resources on performance, and by invoking the principle that the underlying processes should have continually available output, we have been able to reevaluate a number of different issues from the literature on attention and perception. In brief, we have shown that one cannot conclude that processes are independent of one another or that they do not require processing resources from the same limited capacity source unless one has explored the entire range of the performance-resource function for those processes. Processes which share a processing

resource will not show any interference with one another until they have been forced to operate within the resource-limited region of their performance-resource function. If either process can be shown to affect the other, even though the reverse interference does not seem to apply, then it can be safely concluded that both share a common processing resource. Thus, it may be false to conclude, as did Posner and Boies, that "the nearly perfect time sharing between preparation and encoding suggests that at least one of these operations does not require central processing capacity. If they both did, they would be expected to interfere" (Posner & Boies, 1971, p. 407).

How does one demonstrate that performance is indeed in a resource-limited portion? Not by measuring error rate alone. Thus, the often followed procedure of insuring that performance is above chance and below perfection in order to avoid "floor" and "ceiling" effect is well guided but not sufficient. The process could still be data-limited, regardless of the error rate. The only way to insure that performance is resource-limited is to demonstrate that it is affected by changes in resource.

The picture we describe is one that differs somewhat from the normal multistage view of information processing. Nowhere did we need to speak of stages of processing, and nowhere did we need to worry about the level at which processing was blocked. Rather, we suppose that there exists a pool of possible processes which can actively pursue their analyses at a rate determined, in part, by the division of processing resources allocated among them. All that is needed to use these ideas is the ability to distinguish between resource- and data-limited operations and to know at any time which one is taking place.

One power of this analysis is that it requires only rather weak assumptions about the mechanisms which underlie the initial stages of information processing. Many processes are either data- or resource-limited, whereas others are changed from one to the other by a changing allocation of resources. This classification alone seems useful to our analyses of psychological phenomena. Experimenters must use caution in describing just when one process is independent of others. Unless the full performance-resource functions are known, statements about the independence of processes may simply reflect performance in a very restricted range of operation.

REFERENCES

ANDERSON, N. H. Algebraic models in perception. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of Perception.* Vol. 2. New York: Academic Press, in press.

ANDERSON, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce and P. Suppes (Eds.). *Contemporary Developments in Mathematical Psychology.* Vol. 2. San Francisco: Freeman, 1974.

BECK, J., & AMBLER, B. The effects of concentrated and distributed attention on peripheral acuity. *Perception and Psychophysics,* 1973, **14,** 225–230.

BROOKS, L. The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology,* 1967, **19,** 289–299.

BROOKS, L. Spatial and verbal components of the act of recall. *Canadian Journal of Psychology,* 1968, **22,** 349–368.

COLEGATE, R. L., HOFFMAN, J. E., & ERIKSEN, C. W. Selective encoding from multielement visual displays. *Perception and Psychophysics,* 1973, **14,** 217–224.

ESTES, W. K. Interactions of signal and background variables in visual processing. *Perception and Psychophysics,* 1972, **12,** 278–286.

FLOWERS, J. H. & GARNER, W. R. The effect of stimulus element redundancy on speed of discrimination as a function of state and processing limitation. *Perception and Psychophysics,* 1971, **9,** 158–160.

GARDNER, G. T. Evidence for independent parallel channels in tachistoscopic presentation. *Cognitive Psychology,* 1973, **4,** 130–155.

GARNER, W. R. *The Processing of Information and Structure.* Potomac, MD: L. Erlbaum Associates, 1974.

GARNER, W. R. The stimulus in information processing. *American Psychologist,* 1970, **25,** 350–358.

GARNER, W. R. & MORTON, J. Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin,* 1969, **72,** 233–259.

GREENO, J. G. & SIMON, H. A. Processes for sequence production. *Psychological Review,* 1974, **81,** 187–198.

KAHNEMAN, D. *Attention and Effort.* Englewood Cliffs, NJ: Prentice Hall, 1973.

KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. *Foundations of Measurement.* Vol. 1. New York: Academic Press, 1971.

KERR, B. Processing demands during mental operations. *Memory and Cognition,* 1973, **1,** 401–412.

LABERGE, D. Attention and the measurement of perceptual learning. *Memory and Cognition,* 1973, **1,** 268–276.

LAWSON, E. A. Decisions concerning the rejected channel. *Quarterly Journal of Experimental Psychology,* 1966, **18,** 260–265.

LEIBOWITZ, H. W. & APPELLE, S. The effect of a central task on luminance thresholds for peripherally presented stimuli. *Human Factors,* 1969, **11,** 387–392.

MACKWORTH, N. H. Visual noise causes tunnel vision. *Psychonomic Science,* 1965, **3,** 67–68.

MORAY, N., FAVREAU, D., & NAGY, V. Quantitative studies in attention #3: Why is timeshared attention inefficient? Manuscript, 1973.

MORAY, N. & FITTER, M. A theory and the measurement of attention. In S. Kornblum (Ed.), *Attention and Performance, IV.* New York: Academic Press, 1973.

MORAY, N. A data base for theories of selective listening. In S. Dornic and P. Rabbitt (Eds.), *Attention and Performance, V.* London: Academic Press, 1974.

NORMAN, D. A. & WICKELGREN, W. A Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology,* 1969, **6,** 192–208.

OLLMAN, R. T. & BILLINGTON, M. J. The deadline model for simple reaction times. *Cognitive Psychology,* 1972, **3,** 311–336.

POSNER, M. I. & BOIES, S. J. Components of attention. *Psychological Review,* 1971, **78,** 391–408.

SMITH, E. E. & SPOEHR, K. T. The perception of printed English: A theoretical review. In B. H. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance and Cognition.* Potomac, MD: L. Erlbaum Associates, 1974.

SORKIN, R. D. & POHLMANN, L. D. Some models of observer behavior in two-channel auditory signal detection. *Perception and Psychophysics,* 1973, **14,** 101–109.

SORKIN, R. D., POHLMANN, L. D., & GILLIOM, J. D. Simultaneous two-channel signal detection. III. 630- and 1400-Hz. signals. *Journal of the Acoustical Society of America,* 1973, **53,** 1045–1050.

TAYLOR, M. M., LINDSAY, P. H., & FORBES, S. M. Quantification of shared capacity processing in auditory and visual discrimination. In R. F. Sanders (Ed.), *Attention and Performance, IV.* Amsterdam: North Holland, 1967.

THOMAS, E. A. C. On expectancy and the speed and accuracy of responses. In S. Kornblum (Ed.), *Attention and Performance, IV.* New York: Academic Press, 1973.

TREISMAN, A. & GEFFIN, G. Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology,* 1967, **19,** 1–17.

WOODWORTH, R. S. *Experimental Psychology.* New York: Holt, Rinehart, and Winston, 1938.

YELLOT, J. I., JR. Correction for fast-guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology,* 1971, **8,** 159–199.

(Accepted September 9, 1974)

COGNITIVE PSYCHOLOGY 7, 65–81 (1975)

The Perceptual Span and Peripheral Cues in Reading

KEITH RAYNER

University of Rochester

Skilled readers read passages that were displayed on a Cathode Ray Tube controlled by a computer. The readers' eye movements were monitored and certain critical words were changed by the computer as the eye was in motion. The experimental technique utilized in the study provided data on how wide the area is from which a reader acquires information during a fixation in silent reading. The results also delineate different types of visual information that are acquired from various areas within the perceptual span. It was found that a reader was able to make a semantic interpretation of a word that began 1–6 character spaces from his fixation point. When he fixated 7–12 character spaces prior to a word, he was able to pick up such gross visual characteristics as word shape and initial and final letters. It was concluded that the skilled reader is able to take advantage of information in the periphery. However, the size of the area from which he does is rather small.

Determining the size of the area from which a person picks up information during a fixation in reading has long intrigued psychologists (Woodworth, 1938; Huey, 1908). In the past, five general types of research have been used to identify the perceptual span in reading. However, each of these techniques has particular problems associated with it that have led to equivocal results and differing estimates of the perceptual span. The first and simplest type of research has been to divide the number of letters per line by the number of fixations per line (Taylor, 1957; Taylor, 1965). This method of estimating the perceptual span is based on the assumption that on successive fixations the perceptual spans do not overlap or they overlap the same amount. This assumption is probably false.

This paper is based on part of a dissertation submitted to Cornell University in partial fulfillment of the requirements for the Ph. D. degree. The author is grateful to Harry Levin and George J. Suci, who were members of the Special Committee, for their comments and criticisms. He is particularly grateful to George W. McConkie, Chairman of the Special Committee, for his suggestions, support, and criticisms. This study was supported by Grant OEG-2-71-0531 from the United States Office of Education to George W. McConkie. Appreciation is expressed to the members of the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology where the study was carried out and especially to George W. McConkie and David Silver for programming assistance. Requests for reprints should be sent to Keith Rayner at the Center for Development, Learning, and Instruction, University of Rochester, Rochester, NY 14627.