Annu. Rev. Psychol. 1993. 44:383–425

INFORMATION PROCESSING MODELS: Microscopes of the Mind

Dominic W. Massaro

Program in Experimental Psychology, University of California, Santa Cruz, California 95064

Nelson Cowan

Department of Psychology, University of Missouri, Columbia, Missouri 65211

KEYWORDS: perception, experimental psychology, psychophysics, cognition, psychological theory

CONTENTS

0066-4308/93/0201-0383$02.00

INFORMATION-PROCESSING (IP) APPROACH

"Information," though difficult to define precisely, refers to representations derived by a person from environmental stimulation or from processing that influences selections among alternative choices for belief or action. "Information processing" (IP) refers to how the information is modified so that it eventually has its observed influence. "IP models" are theoretical descriptions of a sequence of steps or stages through which this processing is accomplished. In this chapter, we (*a*) reexamine the assumptions and rationale of the IP-modeling approach as it was conceived in the initial work on psychophysics, perception, attention, and memory (Atkinson & Shiffrin 1968; Broadbent 1958; Green & Swets 1966; Sternberg 1969), (*b*) review the theoretical literature in which clarifications or modifications of the approach are suggested, (*c*) illustrate some of these points in a discussion of the applications of IP models to various topics within contemporary cognitive psychology, and (*d*) evaluate the IP approach in comparison to related approaches.

IP models have played a major role in shaping the current dominant understanding of perception and action. In the last *Annual Review* article on IP models Posner & McLeod (1982) saw the IP approach as a search for elementary operations. Their vantage point was contrasted with that of the Newell (1980) and Simon (1979) school, which emphasized simulation of a wide range of mental activity by complex information processing models. Posner & McLeod chose instead to emphasize "fundamental operations that can be used to characterize the human mind" (1982:478).

One can sometimes gain insight into existing metatheory by considering what is taught to psychology students. Today, most courses and textbooks covering such experimental topics as human perception, memory, and thought are called "Cognitive Psychology" or "Cognition," but they most often profess allegiance to the IP approach (Anderson 1990a; Glass & Holyoak 1986; Massaro 1989a; Solso 1991). However, even though IP has had a solid and continuing tradition beginning with Broadbent's (1958) work and progressing through surveys by Neisser (1967), Norman (1969), and Lachman et al (1979), the IP paradigm generally has been neither clearly defined nor contrasted with other metatheories. We seek to fill this void by reviewing the recent literature that allows the metatheory of IP to be articulated and contrasted with other metatheories of psychological inquiry.

Characteristics of the IP Approach

In an important paper, Palmer & Kimchi (1986) described five properties of the IP approach. First, an *informational description* means that the environment and mental processing can be described in terms of the amount and types of information. *Recursive decomposition*, perhaps better described as hierarchical decomposition, denotes the breaking down of one stage of processing into substages. For example, a memory stage can be broken down into acquisition, retention, and retrieval stages; retrieval can be further broken down into memory search and decision; and memory search can be further broken down into access and comparison stages. The *flow continuity* principle states that information is transmitted forward in time. All inputs necessary to complete one operation are available from the outputs that flow into it. Central to the IP approach, as well, is the principle of *flow dynamics,* asserting that each stage or operation takes some time (i.e. that a mental process cannot be instantaneous). Finally, the *physical embodiment* principle is the assumption that information processing occurs in a physical system. Information is embedded in states of the system called representations, and operations used to transform the representations are called processes.

INFORMATION The use of the term "information" in the IP approach is not identical to the classic information measure. For Shannon (1948), the amount of information in a given message is positively related to how much the message reduces the number of possible outcomes. John von Neumann suggested that Shannon call the measure of information "entropy": Because no one knew what entropy was, Shannon would always have the advantage in debate. Shannon's measure does formally resemble the mathematical definition of entropy (Tribus & McIrvine 1964). Young (1987) tries valiantly to define information in mass-energy terms—the putative nature of all events and objects in a traditional scientific view. The form characteristics are the primary ingredient in information flow, and the energetic events are simply the substrate embodying the form characteristics. Information transmission between successive stages of processing can also be clarified using an example from Young. Consider sound waves setting the tympanic membrane into vibration. The sound waves do not leave the air and go into the membrane (the air molecules retain their identity, although their pattern is likely to be changed by rebounding off the tympanic membrane); rather, the form characteristics of the sound waves flow to the membrane the way waves move a boat in water. This action represents a resonance in which one oscillatory system influences another's activity. This observation is important in thinking about transmission of information from one stage of processing to another. The representation of the preceding stage maintains its integrity even after it has been "transformed" and transmitted to the following stage of processing. For example, the notion that categorization of a visual item necessarily supplants any visual representation is not reasonable. The maintenance of multiple representations has become a landmark of models of short-term memory (e.g. Baddeley 1986; Massaro 1975).

The classic measure of information often does not permit the number of possible outcomes, and hence the amount of information, to be calculated unambiguously. In practice, however, a precise definition is not essential because it becomes clear that one is discussing types of information such as feature values or category assignments that distinguish among potential stimuli or responses in a specific experimental situation (Neisser 1967). Psychologists did not adhere to the restrictive formal definition of Shannon.

It is important for our purposes to distinguish between data and information. Information for us is knowledge within the receiver, whereas data are in the environment. A classic illustration of this distinction is found in this telegram from Myron Tribus's daughter in Paris: PLEASE SEND ME FIFTY DOLLARS AMERICAN EXPRESS NICE LETTER OF EXPLANATION FOLLOWS LOVE LOU. One reader of this cable might expect to receive a nice letter; another would know that Nice is a city on the French Riviera. The reader's knowledge determines the interpretation of the telegram.

This example also weakens the accepted claim that the traditional measure of information is devoid of meaningful content, inasmuch as the meaning determines the nature and number of alternatives. The number of viable alternatives for "NICE" could not be deduced without some consideration of meaning. A similar point was made by MacKay (1969) and reiterated by Gregory (1986). Perhaps Gregory (1986) is correct in regretting that information theory has not played a more central role in the study of information processing.

INFORMATION PROCESSING The basic notion of IP is that one must trace the progression of information through the system from stimuli to responses. To construct an IP theory one must first postulate certain stages of processing. This is not always easy, and the method of doing so will vary from situation to situation. One generally starts by mapping out a logically necessary sequence of processes, which must include at least stimulus decoding and response selection stages. Then various experimental methods can be used to search for manipulations that differentially affect hypothesized stages. Although an IP model usually describes the mapping from one stage to another, it is generally the case that several different stages can operate at once. For example, in reading aloud, one can pronounce a word while silently reading ahead to identify the next word. Several stages can operate at once; but if a particular input were followed through the system, the operations carried out on it might occur in sequential order (thus the basis of the term "stage"). In this case, the IP model lends itself to powerful analytic devices, such as Donders's subtraction method, Sternberg's (1969) additive factor method (AFM), backward masking, and various mathematical models. In this situation, each of the hypothesized underlying mechanisms of psychological processing can be associated with a separate segment of time between a stimulus and the response to that stimulus. Furthermore, the operations within a stage might be characterized by a mathematical expression.

It should be stressed that not all researchers using the IP approach claim to explain the processes underlying behavior. For example, mental processes have been conceptualized more weakly as intervening variables that permit a parsimonious interpretation of research findings. In this view, IP is purely pragmatic in allowing descriptive and prescriptive accounts that would not be possible without mental processes as intervening variables. Van der Heijden & Stebbins (1990) claim this much less ambitious goal for the IP approach. For these authors, the only reasonable goal of the IP approach is to describe differences in behavior as a function of differences in external and/or internal conditions: A certain behavior under situation A may be expected to be more accurate or faster, for example, than this same behavior under situation B. To achieve even this more limited goal, however, the IP approach must explain the processes causing behavior (Hatfield 1991).

Note that Newell & Simon's metatheory (the Physical Symbol Systems view) is more restrictive than the IP approach we articulate here (see the section below on variations on the IP approach). We do not restrict representations to symbols, let alone discrete symbols, nor do we restrict processes to rule-like operations performed on these symbols. For example, the currency of most memory models usually consists of memory traces, feature vectors, or simply familiarity—continuous representations similar to activation in many connectionist models. Thus, IP psychologists do not necessarily subscribe to Fodor's (1975) notion of a "language of thought." Within IP, processes are not necessarily rules, nor are representations always discrete objects, concepts, or events. This distinction has played an important role in contemporary theory. The recent connectionist view (e.g. Rumelhart & McClelland 1986) is essentially an alternative to the Physical Symbol Systems view but falls within the general IP framework (Massaro 1988, 1990).

Having defined the IP approach, we must further articulate its goals. We therefore evaluate and seek to justify the IP approach as a metatheory for psychological inquiry. We review well-known constraints on psychological inquiry and their implications for research strategy.

Justifications of the Approach

Although many psychologists work within the IP framework, the approach has not often been justified explicitly. Although most existing justifications in our textbooks center around criticisms of behaviorism, IP has adopted many of the best features of the behaviorist's experimental paradigm (van der Heijden & Stebbins 1990). Even the object of inquiry did not change as dramatically during the "cognitive revolution" as many textbooks have suggested. For example, empirical work on attention did not diminish during the heyday of behaviorism, although the concept of attention carried less theoretical value (Lovie 1983). Van der Heijden & Stebbins (1990) review evidence that the IP approach provided few features absent from mainstream experimental psychology. We must nevertheless not underestimate the importance of the attempt by the IP approach to account for the mental processes intervening between stimulus and response. Where behaviorism aimed to understand behavior, the IP approach seeks to elucidate the processes that cause behavior

(Hatfield 1991). As in other natural sciences, the IP approach attempts to understand complex behavior in terms of the interaction of simpler processes.

CONSTRAINTS ON PSYCHOLOGICAL INQUIRY One way to justify the IP approach is to consider it in light of several constraints on psychological research. The first is that behavior is both variable and complex. By dissecting complex behaviors into simpler component stages, the IP approach may offer the parsimony critical to scientific inquiry. The IP approach is less daunted than other approaches by behavioral variability because it attends to information and information processing of component stages rather than to global behaviors. As an example, individual differences in speech perception may be caused by differences in information at a particular stage of information processing (Massaro 1992).

A second constraint on psychological research stems from our inadequacies as theorists and researchers. As noted several centuries ago by Francis Bacon, we tend to interpret the world as more orderly than it actually is. In addition, scientists, like all humans, have a strong confirmation bias: We actively search for evidence that supports our beliefs, often ignoring contradictory data.

Mitroff (1974) and Wenner & Wells (1990) documented confirmation bias in even the most experienced scientists. In addition, both political maneuvering (Mahoney 1976) and downright cheating sometimes occur in scientific inquiry (Broad & Wade 1982).

A third factor affecting psychological inquiry is the difficulty of determining which of many possible theories best explains a given phenomenon. Consider the competing current theories of language acquisition and use. At issue is whether a child's ability to produce language requires an internalization of rules, or whether it can be explained adequately by reference to associative and generalization mechanisms (MacWhinney & Leinbach 1991; Rumelhart & McClelland 1986; Pinker & Prince 1988; Plunkett & Marchman 1991). In Berko's classic experiment (1958), young children were able to generate plurals of pseudowords they had never heard before. Berko concluded that the children used a rule to achieve the "correct" outcome. As emphasized by Baron (1977), Brooks (1978), and Glushko (1979), however, the children might have performed correctly by generalizing from specific words they already knew. Knowing the plurals of rug, bug, and tug, children might simply generalize that the plural of wug would be wugs.

IMPLICATIONS FOR PSYCHOLOGICAL INQUIRY Three characteristics consistent with the IP approach can overcome the three constraints just mentioned. First, the research strategies of falsification (Popper 1959) and strong inference (Chamberlin 1965; Platt 1964) should be used. Given the constraints on inquiry, the investigator must develop opposing models and devise an experiment capable of distinguishing between the predictions of the different models.

Contradictory evidence disqualifies a theory—even a theory consistent with many other findings. For example, although the hypothesis that whole word shape had a function in reading was consistent with many findings, it was falsified by means of an IP approach (Adams 1979; Paap et al 1984).

The falsification strategy has been criticized (Feyerabend 1975). One criticism concerns the obvious boundary conditions for any test, but we do not see how this disqualifies a falsification strategy. Newell (1990) argues that researchers should nurture rather than falsify theories. Newell sees falsification as a weak research strategy because it leads not to rejection but only to modification of theories. We argue, on the contrary, that the process of modification allows large subclasses of models to be rejected. Moreover, "modification" is an inappropriate description of the outcome if the contrasting alternatives are specific enough. If an experiment decides between categorical and continuous perception (Massaro 1987), it is difficult to see how one alternative can be modified without being made identical to the other. As long as we are concerned with specific assumptions rather than global theories, falsification, strong inference, and fine-grained analysis should be profitable.

Second, we must develop specific, precise, and simple experiments in order to reveal fundamental regularities in the phenomena we study. Such regularities or laws cannot easily be discerned amid the myriad factors present in complex situations. A New Realist philosophy of science allows for predictability in the laboratory but not in the naturally varying environment (Manicas & Secord 1983). Theories will have predictive power only in the laboratory where complexity can be reduced, measured, and controlled. In many respects, the complexity of the prototypical psychology experiment still exceeds any theory's predictive power.

Third, the IP approach enables the investigator to perform the kind of thorough, systematic, and fine-grained analyses of observations that alone can enable the winnowing of alternative interpretations.

We must attempt actively to eliminate alternative models. In addition, parsimonious models are to be preferred. Collyer (1985) argued that a more complex model (with more free parameters) is not necessarily preferable even if it is more accurate than a simpler model. It is difficult to falsify models so general that they predict a wide range of alternative results. One research strategy permits only models with "discriminating taste" to survive. A model has discriminating taste if it predicts *only* actual results, not the universe of possible results.

Metatheoretical Issues and IP

Several metatheoretical issues must be addressed to help situate the IP approach. The first—identifiability—concerns the feasibility of discriminating among alternative explanations of a set of observations. The second—proximal vs distal causes—concerns the types of evidence most relevant to psychology (to IP models in particular).

IDENTIFIABILITY What may be a weakness of the IP approach concerns whether various explanatory models can be differentiated by experimental results. The so-called identifiability issue concerns whether a given model of an experimental result can be *identified* as the correct one. The issue arises from the theorems of E. F. Moore (1956) and from subsequent work in formal automata theory (see also Greeno & Steiner 1964). Moore was concerned with the behavior of sequential machines. Observers of machines or people can record only their inputs and outputs. It is not possible to look, so to speak, inside the black box. The question is: To what extent can the accuracy of one model of the inner workings of a black box be distinguished from that of another model, given only a set of input-output observations? Moore proved that any input-output function can be exactly mimicked by some other such function. No explanatory model of an experimental result can exclude all others.

Several prominent investigators have been convinced by this argument. Hintzman (1991), for example, points out how exemplar models of categorization can explain outcomes once thought explicable exclusively by prototype models. Hintzman (1991) advocates the use of formal models in scientific inquiry to overcome non-identifiability. For Anderson (1990b), the identifiability problem places an enormous constraint on "traditional" psychological research concerned with mechanism or process. How can we converge on a given process or mechanism when we can always compose another set of processes to make the same prediction? Anderson's solution is to limit the family of acceptable models to those that are behaviorally optimal. We deal with Anderson's strategy of adaptive rationality in the next section.

Scientific inquiry can potentially choose among apparently nonidentifiable models by extending the empirical data-base, evaluating the models on the basis of parsimony, and testing among viable models using the principles of falsification and strong inference described above. Extending the data-base to include additional measures of performance is a valuable strategy for distinguishing models that make identical input-output predictions of other measures. Consider a series of experiments on how children add two numbers. One model claims that children use a simple lookup table. A second model claims that, at one stage of development, the child recognizes the numbers, chooses the larger one, and then adds the smaller number by counting from the larger to the smaller in successive units (Groen & Parkman 1972). Thus "6 + 3" requires the series "6, 7, 8, 9," whereas "7 + 1" requires only the shorter series "7, 8". Both models predict correct answers to addition problems. However, experiments have been able to falsify the first model of addition by measuring reaction times (RTs) to different problems. The problem "6 + 3" takes about the same amount of time as the problem "4 + 3," as predicted by both the lookup table and counting models. However, these problems take longer than "7 + 1," a result consistent only with the counting model.

Within the IP approach, Townsend (1990) has repeatedly demonstrated that what look like the results of a serial search process in cognition may actually be the results of a parallel search process. A phenomenon such as limited cognitive capacity could cause parallel search to produce the same observable results as serial search. Such demonstrations do not leave the psychologist helpless to pursue the distinction between the two kinds of search, since certain experimental results remain more informative than others. A flat function showing no increase in RT with increases in the number of items is evidence against a serial search. Furthermore, there should be experimental manipulations that can address the role of limited capacity when increasing linear functions are found.

It is also possible to identify parallel processing without manipulating the memory or array set sizes (Schweickert 1978). Egeth & Dagenbach (1991) presented two-element displays in which the visual quality of the elements was independently manipulated, resulting in displays in which 0, 1, or 2 of the items were of high quality. The diagnostic was based on target-absent trials, in which both items would always have to be searched. It was assumed that a high-quality item could be searched in time T, whereas a low-quality item would take T + ΔT to search. If the search process occurred serially, then it would be expected to take an average of 2 × (T + ΔT) on low-low trials, T + (T + ΔT) on low-high or high-low trials, and 2T on high-high trials. If the two items could be searched in parallel, then the search would be completed when the slowest item-search was completed—on the average, in time (T + ΔT) when at least one of the items was of low quality, and in time T when both items were of high quality. Results agreed with the parallel processing predictions when subjects searched arrays consisting of the letters X and O, as well as arrays of T and L in canonical orientation. When the arrays consisted of T and L and the orientation of each letter varied, however, the predictions for serial processing were fulfilled. The difference may have occurred because subjects can perceive well-learned patterns (letters) in parallel but have no well-learned representation of rotated letters. Alternatively, our poorer sensitivity to oblique than to vertical and horizontal lines may account for the difference.

As pointed out by S. Sternberg (personal communication), within the serial model described by Egeth & Dagenbach, it must be assumed that (*a*) *all* operations influenced by legibility are serial and (*b*) there is no additional switching time on trials with two letters differing in visual quality. If either of these two assumptions does not hold, a mechanism that includes an underlying "serial" process might give results matching the parallel model. Conversely, Egeth & Dagenbach warned that their diagnostic is not conclusive for results indicating serial processing. The reason is that a subject using parallel search cannot respond in the target-absent condition until all of the items in the array have been perceived. With variability in the perception times, the more poor-quality items in the display, the slower would be the expected processing time. Thus, a parallel search could give mean RTs similar to those expected from a serial search. Notwithstanding the limits of this diagnostic method, it holds promise and can be broadened. For an example of how RT distributions might be analyzed, see Roberts & Sternberg (in press).

Roberts & Sternberg (in press) apply a falsification and strong-inference strategy within the context of the additive factor method (AFM) to overcome problems of identifiability. They describe three models—a successive stage model, an alternative pathways model, and a cascade model—each of which can predict additive effects of two factors on RTs. The standard stage model assumes serially arranged and separately changeable processes. Roberts & Sternberg (in press) strengthen this stage model by adding the assumption of stochastic independence—the durations of the stages in question are stochastically independent. The alternative pathways model assumes that one process is used on some proportion p of the trials and another process is used on the other $1 - p$ trials. The cascade model assumes that one process provides continuous output to a second process that occurs concurrently. Although these three models are not identifiably different with respect to mean RTs (Ashby 1982; McClelland 1979), they make different predictions about the RT distributions. The results from four diverse experiments were reanalyzed to test the new predictions of the models (see also Ashby & Townsend 1980). The analyses of the RT distributions and their variances falsified the alternate pathways model and cascade model, while supporting the stage model. The success of these analyses provides a boost for the IP approach. The research illustrates the value of a falsification strategy in inquiry, how problems with identifiability can be overcome, and the potential for broadening the domain of inquiry by extending the range of dependent measures and statistical tests used.

Other recent research has made progress in overcoming identifiability problems. In Massaro & Friedman's (1990) analysis, some models that cannot be identified as correct in a task with just two response alternatives made different predictions for four responses. Cohen & Massaro (in press) demonstrated that some models required more free parameters than others. Massaro (1989b) performed a fine-grained analysis on the joint contribution of stimulus information and context in order to distinguish between the Fuzzy Logical Model of Perception (FLMP) and the TRACE model of speech perception. McClelland (1991) then modified TRACE and the class of interactive activation models to bring them into line with the new empirical results. Although the FLMP and interactive activation models now made similar predictions for asymptotic performance, Massaro & Cohen (1991) were able to discriminate between the models by attending to their predictions about the dynamics of information processing. The interactive activation models had difficulty predicting (*a*) substantial context effects, given little processing time, and (*b*) a strong stimulus influence, given substantial processing time. The FLMP, on the other hand, provided a good quantitative description of these results.

In summary, the problem of identifiability is *not* insurmountable. Moore's theorem applies to the situation in which there is only a single experimental result. Rewarding progress can be made by examining additional predictions of the models and additional experimental situations (Townsend & Ashby 1983). Other examples of solutions to the identifiability problem are described below.

PROXIMAL AND DISTAL CAUSATION A second issue that helps situate the IP approach concerns the causes of behavior. An important distinction made in evolutionary biology is between proximal (occurring nearby—here construed as nearby in time) and distal (temporally distant) influences on behavior (Alcock 1989). Proximal influences (proximal causes) include psychological processes that affect behavior. Distal causes concern the adaptive significance of an observed behavior. As psychologists, we consider primarily proximal influences. For example, what visual features are used in letter recognition, how are these features combined, and how is a decision made on the basis of this information? Distal causes, such as how the ability of the visual system to detect edges evolved, we usually ignore. Research within the IP approach is concerned with ongoing mental processes, whose modulation by proximal causes is most readily observed.

Some psychologists, on the other hand, have considered distal causation as a constraint upon psychological theorizing. As a solution to the identifiability problem Anderson (1990b) proposes selection of the model that assumes optimal adaptation to the environment. In most domains, however, many models can meet optimality criteria (Gigerenzer et al 1988; Massaro & Friedman 1990). Furthermore, optimality and computational constraints are fuzzy concepts that seem to be used without consensus. Recent explanations within evolutionary theory seem to stretch traditional notions of optimal behavior (Anderson 1990b; Cosmides 1989; Real 1991). In addition, Schoemaker (1991) offers several cases where "optimality" does not explain the phenomenon of interest.

Finally, it is likely that not every behavior is optimal. Evolutionists note many behaviors that have no obvious purpose. The questions of interest to psychologists are fundamentally empirical ones not simply answerable in terms of the optimality of models (Nosofsky 1991).

The issue of proximal vs distal causation can be clarified by acknowledging different levels of understanding in inquiry. David Marr described three levels at which any machine carrying out an information-processing task must be understood. In computer terms, the computational level is an abstract description of the problem to be solved, the algorithmic level is the software program to solve the problem, and the implementation level is the computer it is being run on. The computational level concerns the nature of the problem being solved. This entails the information available and the mapping of this information to another kind of information. It is clear that evolutionary history can inform this computational level of analysis. The algorithmic level, which is most compatible with the IP approach, entails the operations that transform the information from one type to another. This level specifies the representations for the input and output and the mapping between them. Understanding proximal causation seems most productive in illuminating the algorithmic level. The implementation or hardware level describes the physical realization of the algorithmic level. Both distal and proximal causation would appear to be relevant to the hardware level.

CHARACTERIZING STAGES OF PROCESSING

Here we present recent evidence that information processing occurs in stages, and then we take up several important issues related to stages of processing. Given a single stimulus, we can distinguish between input and output representations, and between transformation and transmission processes. Each of these representations and processes can be characterized as discrete or continuous. Comparable distinctions apply when multiple stimuli are presented. If multiple codes can co-exist in a stage, they can be processed either in parallel or serially. Finally, strategic and attentional effects can modulate the character of information processing.

Recent Evidence for Stages of Processing

An analysis of information processing performance into separate stages can lead to specific, quantitative predictions for a particular task if one is willing to make certain strong assumptions about the nature of processing. One must, of course, assume that some processing stages between the presentation of a stimulus and the subject's response can be identified. Sanders (1990) identifies seven stages of processing and presents evidence for each of these stages. Massaro (1991) describes a variety of research on perception-action relationships in terms of a similar stage model.

Roberts (1987) provided dramatic evidence for two stages of processing in accounting for response rates under various levels of food deprivation and various schedules of reinforcement in rats, pigeons, and goldfish. Response rates in these animals show selective influences from deprivation time and schedule of reinforcement. These two influences combine multiplicatively to influence response rate. Roberts proposed that the first process generates pulses that are transmitted to the second process, a filter. A response occurs whenever a pulse passes through the filter. Either deprivation time changes the rate of pulses in the generator and reinforcement schedule changes the setting of the filter, or vice versa. This simple theory was made more complex by allowing the occurrence of operant responses that are not under stimulus control. In the tradition of strong inference, Roberts showed how this stage theory gave a better description of the results than alternative theories that violated the assumption underlying the stage theory. This support for stages with different organisms and behaviors is an impressive achievement in psychological inquiry.

Using an IP approach, Theios & Amrhein (1989) illuminated the representation and processes involved in reading words and naming pictures. An IP model sought to explain why subjects took longer to name pictures than to read words. According to this model, picture (and color) naming took longer because it involved two processes (determining the meaning and mapping this meaning into a response) while word naming involved only one. The same model described visual and conceptual comparisons among pictures and words—a successful use of Donders's Subtractive Method. Using this model it was possible to test whether pictures are also perceived more easily than words. According to this model, they were not: Previous findings to the contrary had apparently been based on the fact that the pictures were larger than the words.

Discrete vs Continuous Representation and Processing

As Miller (1988, 1990) has pointed out, if one assumes that processing occurs in stages one must consider separately (*a*) whether the representational codes input to or output from a particular stage of processing are discrete or continuous, (*b*) whether the transformation accomplished at a particular stage takes place in a discrete manner or gradually (i.e. continuously), and (*c*) whether the information is transmitted to the next stage in discrete steps or continuously. These are important issues within the IP approach because some IP models require types of discreteness and some require types of continuity (see below). Miller acknowledges that continuity and discreteness are not dichotomous, but matters of degree. For example, placing a stimulus feature in one of many ordinally arranged categories would probably produce experimental observations indistinguishable from those using continuous coding but far different from those using a binary classification system.

THEORETICAL POSSIBILITIES Stage models propose four ways in which discrete and continuous information and information processing can occur. In all cases, the *input* to a stage can be continuous or discrete. Second, either of these types of input can be *transformed* in a discrete or continuous fashion. Third, the *transmission* of information from one stage to the next can be discrete or continuous. [Miller (1988) argued that if the transformation is discrete the transmission must also be discrete. However, consider a discrete transformation of 0 to 1 with a transmission that sums the outcome of the transformation over some finite time. The average passed on by the transmission will be 0 until the transformation produces a 1. The transmission will then grow continuously, however, depending on the averaging period.] Finally, regardless of the type of transformation and/or transmission, the *output* of a stage can be discrete or continuous. Sixteen alternatives appear to be possible.

Discrete transmission is usually assumed to apply in the additive factor method (AFM); without it, RT would not necessarily equal the sum of durations of processing at all stages. In addition, application of the AFM is usually assumed to require errorless performance (in which coding is necessarily discrete). However, Schweickert (1985) proved that, with some additional assumptions, factors having additive effects on RT will also have additive effects on log percent correct. This method should encourage investigators to carry out their tests at several points on the speed-accuracy function. Studies with performance significantly below perfect accuracy can be more sensitive to effects of the independent variables of interest. There are other IP models of stages with continuous codes, transformation, and transmission that make specific testable predictions. For example, McClelland's (1979) cascade model assumes that information flows continuously from one stage to the next. Still, the idea of sequential stages is meaningful, and quantitative predictions for reaction time can be derived. Similarly, in the dynamic FLMP (Massaro & Cohen 1991), evaluation and integration transform and transmit information continuously, and yet quantitative predictions of response probability and RT can be made.

In many instances the outcome of identification is necessarily continuous or "fuzzy" because the available category labels describe some stimuli better than others (Massaro 1987). This situation would violate the constant-stage output assumption of the AFM. That is, the output of an identification stage would take on a range of values rather than be limited to one of the response alternatives in the task. This is not a problem for the IP approach in general because continuous outputs are compatible with serially arranged stages of processing that are separately influenced. In the FLMP, for example (Massaro & Friedman 1990), continuous information is obtained from each source and then transmitted to an integration stage. The outcome of the integration stage, also continuous, is transmitted to a decision stage. The stages in the FLMP are sequential and separately changeable even though the outputs of some of them are continuous.

An alternative scheme, intermediate between discrete and continuous, is Miller's (1988) asynchronous discrete coding model. Here successive stages can overlap in time—i.e. information can be transmitted to the next stage before the current stage is complete. However, the transmission of information about each separable code within the stimulus is discrete. As an example, discrete information about the color, shape, and size of an object would be separately transmitted as soon as the processing of each of these dimensions is completed. Several recent studies show discreteness in some situations and continuity in others, a result consistent with Miller's asynchronous discrete model. However, results from other paradigms, such as backward recognition masking and speech identification, falsify the asynchronous discrete model's central assumption that a single feature dimension is transmitted in discrete steps.

EMPIRICAL EVIDENCE At some stage of processing, information and information processing are best characterized as continuous. In an ingenious pioneering study, Allport (1968) showed that the sensory system makes information available to conscious perception continuously. Sets of lines on an oscilloscope, presented with a short asynchrony between lines, were perceived in a particular overlapping fashion that could only be explained by a continuously moving window of perceived simultaneity. Even though lines were presented one at a time, 11 of 12 were visible at any moment because of the subject's sensory memory. A discrete-moment hypothesis would predict perceived movement of the nonvisible line or "shadow" in a direction opposite of the line presentation sequence, but the shadow instead moved in the same direction as the line presentation sequence, as a continuously moving temporal window of experience would predict. A recent sophisticated study ruled out two general classes of discrete perceptual moment (stimulus-independent, stimulus-triggered) (Ulrich 1987). Most recent inquiries have searched for discreteness using RTs rather than perceptual reports.

Miller & Hackley (1992) have extended Miller's research on response preparation. Miller (1982) had previously showed that a subject's hand can be put in a response-ready state before the subject knows exactly how to respond. The subject was signaled about which hand to use in the response by means of an easily processed feature of the stimulus (e.g. its shape); which finger to use was signaled by a feature that took longer to process (e.g. subtle differences in size). In their work a decade later Miller & Hackley employed the lateralized readiness potential (LRP), a component of the movement-related brain potential, as an additional dependent measure and adopted a slightly different "go/no-go" procedure in which the easier stimulus feature signaled which hand to use if a response was to be made, while the more difficult feature signaled whether or not to make the response at all. Motor preparation, as measured by the LRP, was observed even when the response was ultimately aborted, suggesting that the two critical features of the stimulus were transmitted at different times. In a similar study, Osman et al (1992) found that the two stimulus characteristics were processed concurrently and that preparation of the appropriate response hand began before the "go/no-go" decision was completed. These results are consistent with both the continuous and asynchronous discrete models.

The continuous model predicts that information along a single dimension is transformed and transmitted continuously. The asynchronous discrete model, on the other hand, assumes that information from a single dimension is transmitted only after processing is complete (i.e. discretely). To distinguish between the models, a subsequent experiment used four values of a single attribute, size (with scale values of 8, 10, 16, and 19). An easy judgment about size was enough to decide which hand to make ready (e.g. for some subjects, size 8 would indicate a left-hand reaction and size 19 a right-hand reaction). The two intermediate values signaled that no response was to be made. Visual processing should indicate which hand should be used before indicating whether or not a response should be made. If this information is transmitted to the response stage, a LRP should be observed. That no advanced preparation was observed was interpreted to suggest that a given individual stimulus feature is transmitted all at once rather than continuously. Of course, the size experiment only supports the null hypothesis, and there was a nonsignificant trend toward a readiness potential in that experiment (Miller & Hackley 1992: Figure 6). There is also evidence from other paradigms, such as backward recognition masking, for continuous processing of single stimulus dimensions (see below).

Reeve & Proctor (1984) challenged the logic of Miller's demonstrations of advanced motor preparation. They used a speeded keypress response with the index or middle finger of the left or right hand. Advanced preparation cues in the form of "+" marks above some keys allowed the subject to ready the appropriate hand, to ready the same finger of each hand (e.g. both the left and right index fingers), or to ready two unrelated fingers (e.g. left index and right middle). A control condition carrying no information was also included. Subjects could ready any combination of fingers if given a long enough (e.g. 3-s) preparatory period. Moreover, in an experiment in which the left and right hands were placed on the keyboard in an overlapping fashion, with fingers in the order "right index, left middle, right middle, left index," it was shown that the speed advantage was for preparation of responses to be made in a particular (left or right) spatial portion of the response array, not for the left or right hand per se.

The types of preparation observed in these complex situations (long preparation intervals, lack of spatial separation of hands) might not apply to the simpler situations that Miller has used. Even if they do apply, however, they do not (contrary to Reeve & Proctor's suggestion) negate Miller's conclusions about the continuous passage of information to the motor system. Even if advanced preparation always occurs on the basis of spatial location and not the limb of the effector, it is motor preparation nonetheless, and an easily perceived feature of the stimulus still facilitates that preparation.

Meyer et al (1985) addressed the issue of whether information can be transmitted continuously to a response preparation. They used a choice RT task in which a left-hand response was signaled by an arrow facing in one direction, and a right-hand response by an arrow facing in the other direction. However, a preceding prime stimulus presented on some trials perfectly predicted the RT signal that was to follow (words were followed by right arrows, nonwords by left arrows). The prime was presented either 200 or 700 ms in advance of the arrow. A 700-ms interval consistently permitted the subject to be in a state of readiness, whereas a 200-ms interval did so about half the time. The distribution of responses in the 200-ms condition looked like a hybrid of the distributions in the 700-ms and no-prime conditions, with an early peak coinciding with that of the 700-ms condition, a later peak coinciding with that of the no-prime condition, and tails covering the entire range. This result was taken to suggest that information from the prime was transmitted to the response-preparation process in a discrete all-or-none fashion and that, in the 200-ms condition, subjects were in a prepared state on some trials but not on others. However, the results were different in a situation in which signals were presented at four different locations, signaling a response with the left or right middle or index finger (with a complex mapping of signal to finger). In this situation, the prime stimuli only indicated which hand to use in making the response, not which finger. In contrast to the previous experiment, the 200-ms condition produced an RT distribution with a single, intermediate peak, suggesting continuous transmission of information from the prime. Thus whereas stimulus information sufficient to plan the response was used in a discrete, all-or-none fashion, stimuli providing only partial information about the response produced a continuous buildup of readiness over time.

Support for continuous stimulus evaluation is found in the results of the "speed-accuracy decomposition technique" (Meyer et al 1988a,b), which relies upon a RT task in which subjects are induced sometimes to make hasty decisions following the appearance of a response signal at various intervals. The results were analyzed with a mathematical model in which it is assumed that the formation of complete stimulus information is in a race with a guessing process (based on partial information and response bias) that is initiated when the response signal occurs. Guessing accuracy was found to increase in a continuous fashion for a simple lexical decision task but in a stepwise (3-state) way for a more complex task in which the lexical status of two words were to be compared. However, in the latter situation, the general availability of continuous information for each stimulus item might be obscured when the two items must be compared. A comparison process might wait for discrete information about each of the two items before it begins. Thus, the continuous transformation and transmission of information of one stage might be obscured by a following discrete stage.

The Meyer et al (1988a,b) studies may challenge continuous theories less than would first appear. Ratcliff (1988) showed that even the results believed by Meyer et al to support the discrete model could be explained by a continuous model. An important aspect of the analysis concerns the decision time required under "partial" and "complete" stimulus information. Meyer et al assumed that the decision time would be equivalent in these two cases. Ratcliff, on the other hand, argued that more decision time would be required given "partial" rather than "complete" information. His argument is reasonable given the well-known negative correlation between decision time and stimulus information. Subjects are naturally slower when stimulus information is incomplete. This finding has been described within the context of signal-detection theory—response time is longer to the extent the perceptual observation is close to the criterion separating two responses (Norman & Wickelgren 1969; Thomas & Myers 1972). As an example, RT to a speech stimulus is positively correlated with ambiguity of the speech event (Massaro 1987:Ch. 5). Subjects' longer decision times when given "partial" than when given "complete" information could be responsible for the results taken to support a discrete model.

Abrams & Balota (1991) extended our understanding of IP by adding a behavioral measure of force of a handle movement in the RT paradigm (see also Schweickert, in press). Word frequency influences RT in a lexical decision task; this study showed that it also influences response force (although the effects were small). More forceful responses were found for high-frequency words. A similar result was found in the Sternberg memory search task. Response force reflected the relative evidence for "yes" and "no" responses. These results can be interpreted in several ways. Either the information arriving at the decision process was stronger in some cases than in others (a continuous-coding hypothesis—e.g. more word-like information for high-frequency words) or the information was transmitted to the decision process over different periods (a continuous-transmission hypothesis—e.g. temporally more compact transmission in the case of high-frequency words). Of course, both hypotheses could be true.

Additional evidence for both continuous transformation and transmission processes and continuous output codes is the backward masking of recognition. This task has been valuable for examining the temporal course of perceptual processing of visual (Breitmeyer 1984) as well as auditory (Hawkins & Presson 1986; Kallman & Massaro 1979, 1983) stimuli. A brief target stimulus is followed, after a variable stimulus onset asynchrony (SOA), by a second stimulus (the mask) and then a multiple-choice test of the target's identity. The amount of time for which the target information is available for recognition processing can be carefully controlled by manipulating the duration of the SOA. The accuracy of target identification increases as the SOA lengthens to about 250 ms (Breitmeyer 1984; Cowan 1984, 1988; Kallman & Massaro 1983; Massaro 1975; Turvey 1973), even though the presence of the unrecognized targets can still be detected.

Although two different explanations of backward recognition masking have been offered, both are consistent with the view that information is transmitted continuously from one stage of processing to the next. Consider presentation of a pure tone that must be identified as high or low in pitch. All would argue that here the stimulus code is continuous. Is the code transformed and transmitted in a discrete or continuous fashion? According to Massaro (1972, 1975), the target tone is transduced by the listener's sensory system and retained in a preperceptual sensory store that briefly holds a single event within the sensory modality. Processing of the target is necessary for perceptual recognition. The mask is said to replace the target in the preperceptual auditory store and therefore to terminate any further reliable perceptual processing of the target. An account offered by Hawkins & Presson (1986) differs from this one in that the mask is said to switch attention away from the target rather than replacing it in the sensory store.

In either case, a continuous response output would suggest that transformation and transmission were continuous. Such continuity is indeed observed. No matter whether the discrimination must be made on the basis of a single feature such as pitch (Massaro 1975) or on the basis of multiple features (Moore & Massaro 1973; Kallman & Massaro 1983), steplike masking functions have never been observed, even in individual-subject results. An alternative to this continuity-based explanation would be that the relevant information is all-or-none and that the probability that the discrete information is available on a particular trial is what varies with the SOA. However, backward recognition masking also occurs in experiments in which subjects report the perceived quality of the target using a graduated response scale rather than identifying the target in a multiple-choice situation (Cowan 1987; Idson & Massaro 1977). Such experiments indicate a gradual shift in perceived quality across SOA rather than a change in the proportion of an all-or-none response. For example, Cowan (1987) measured the perceived loudness of targets of 3 intensities and found a general growth of loudness across SOAs for all targets, superimposed on an increasing discriminability of the loudness of targets across SOAs. Still, a closer examination of the distribution of responses (which should be bimodal if transformation and transmission of a discrete code occurred in an all-or-none fashion) would prove helpful (Massaro & Cohen 1983).

As the above discussion suggests, it is important to distinguish between the metatheory of IP and the assumptions made within particular applications of the metatheory (e.g. Sternberg's AFM). It is important to evaluate models with reference to the specific objects processed, the types of processing implicated, and the strategies induced by the task. Continuously formed information sometimes may be transmitted in discrete packages when the task demands discourage the use of partial information and encourage delaying the response until more complete information is available (i.e. the "criterion" amount of information necessary for some level of accuracy).

Serial vs Parallel Processing

Whereas the issue of discreteness vs continuity is relevant to how a single stimulus item is processed in each stage, the issue of serial vs parallel processing concerns the processing of a stimulus array at each stage. Resolving this issue is important because the ultimate goal of a processing model is to determine how the entire stimulus field, not simply an item within that field, is processed. If it uses serial processing, a stage can handle items only one at a time; if it employs parallel processing, it can handle multiple items at the same time. The question of whether serial or parallel processing is used must be settled independently for each processing stage.

Determination of the manner of processing is not a simple task. Early on, in an experiment involving search of a set of items in memory to detect the presence or absence of a probe item, Sternberg (1966) found a linear increase in RT as a function of the number of items in the memory set. The slope of the function for trials in which the probe was a member of the set of items in memory was the same as that for the trials in which it was not a member. Sternberg sought to account for this finding with a model of exhaustive serial search. However, as Townsend (1974) pointed out, a parallel search of the items in memory could also account for the results, provided that the time for each item-search is affected by the number of concurrent searches in a linear fashion. In Ratcliff's (1978) random walk model, for example, evidence of the presence or absence of the probe in the memory set is accumulated relative to all items in the memory set at once, until a decision threshold is reached. A linear increase in RT is one possible outcome that can be predicted by this model.

Ruling out the possibility that a process is serial is easier than excluding the possibility that it is parallel. For example, Schneider & Shiffrin (1977) asked subjects to search a set of 1–4 items presented on the computer screen to determine whether it included any one of 1–4 items held in memory. In a "consistent mapping" condition, the memory set on each trial was drawn from a larger, fixed set of items, and the foils were drawn from a different set. After considerable practice in this condition, response rates were no longer affected by the number of items in either the stimulus set or the memory set. This could occur only if subjects searched for all of the stimulus items in parallel, and is called "automatic processing." This result does not imply that it would hold for larger or unlimited sets, however (Shaw 1984).

In another condition, termed "variable mapping," the items that were the potential targets on some trials (i.e. members of the memory set) could be foils on other trials. This sort of search resulted in a roughly linear increase in the RT as a function of the number of items in either the stimulus set or the memory set, regardless of the amount of practice. This could be accounted for either by a "serial search" or by a parallel search at a rate that is slower when more items are present. The latter is termed a "capacity-limited parallel search," in which the nature of the limitation is left unspecified until there is evidence identifying it.

The traditional serial vs parallel processing distinction may not be the distinction that best captures what is important for IP models. A more fine-grained analysis would be better. Such an analysis would distinguish among fully serial processing (in which items can be processed only one at a time), several degrees of capacity-limited parallel processing (in which multiple items can be processed at once but with some interference between items), and capacity-free parallel processing (in which there is no interference). As an analogy, consider the case of pedestrians crossing a bridge. More important than a determination of whether the pedestrians cross one abreast or several abreast is an estimate of the delay in one pedestrian's passage as a function of the number of others hoping to cross simultaneously. The situation in which two or three people can pass at once is far closer to the one-at-a-time situation than to a 100-at-a-time circumstance. This analogy reveals the value of determining in detail the efficiency of search over a wide range of conditions. Exemplifying this approach, Duncan & Humphreys (1989) observed that both (*a*) decreasing the similarity between the target and foils and (*b*) increasing the similarity among the foils increase efficiency. In terms of the bridge analogy, it is as if only some people are permitted to cross the whole bridge, such that passage is more efficient for those whose passage matters most. It would be useful for the bridge travel authority to have a simple means to distinguish the privileged class from others attempting to gain access to the bridge. The Duncan & Humphreys study suggests that the different categories of bridge crossers wear different uniforms, so to speak, to set them apart.

The limited-capacity model may apply even to automatic processing; the width of the bridge may simply vary with the degree of automaticity. Thus, Shaw (1984:111) suggested that automatic and controlled search differ quantitatively, not qualitatively. Fisher et al (1988) presented evidence supporting a different type of limited-channel model. Their assumption was that the resources devoted to each channel remain constant with changes in the number of stimuli in the display (load) up to a limit of about 4 channels. This model suits the bridge analogy even better than a limited-capacity model because here the most critical factor is thought to be rate of item presentation rather than a capacity to be divided among items.

Clearer distinctions are needed among the limited-capacity and limited-channel explanations. In one experiment, for example, Fisher et al required subjects to search for the digit 5 among letter foils in a series of matrixes of 8 items (high-load condition) or 4 items (low-load condition). In each condition, one matrix rapidly followed another, and in the low-load condition the display alternated between square and diamond arrangements of 4 items. A large load effect was obtained. However, what would happen if the square and diamond arrangements were randomly mixed? Fisher's limited-channel model would predict no added difficulty, because the rate of presentation has not changed and is still relatively low. A capacity-based model, on the other hand, would predict added difficulty, because the location at which the target may appear becomes less predictable.

Once a capacity-limited parallel process has been identified, further work can be focused on determining the nature of the capacity limitation. For example, in the case of consistently vs variably mapped search, Logan (1988) has proposed that the availability of sufficient knowledge in memory plays a critical role. His results from various experiments suggest that, in speeded tasks, there is a race between an algorithm and direct retrieval of the needed information from memory. When a sufficient number of exemplars with the correct stimulus-response mapping exist in memory, direct retrieval usually wins. This occurs after sufficient practice in a task with consistent mapping.

Strayer & Kramer (1990) clarified the nature of processing in Sternberg search tasks. They found that a secondary memory load influenced search rate in a variably mapped condition. When the emphasis on the search task relative to the secondary memory task was manipulated there were tradeoffs in performance in the two tasks. Neither secondary memory load effects nor tradeoffs were obtained following considerable practice in a consistently mapped condition. These results suggest that the search that is presumably used in a variably mapped condition involves holding the search set in a working memory with a limited storage capacity. Considerable practice in a consistently mapped condition changes this type of processing.

Attentional and Strategic Effects

As Shiffrin (1988) noted, "Attention has been used to refer to all those aspects of human cognition that the subject can control ... or aspects of cognition having to do with limited resources or capacity, and methods of dealing with such constraints." While under certain circumstances stimulus search and memory search are apparently examples of unconstrained parallel processing, a processing bottleneck typically occurs later in the processing sequence when a response to each target must be selected and executed. Several studies (e.g. Duncan 1980; Sorkin et al 1973) suggest that, in contrast to the type of search in which only one target is present on each trial, there is considerable interference among multiple targets presented simultaneously.

In a study by Pashler (1989), subjects received a tone to be identified as quickly as possible, followed after a variable interval (50, 150, or 650 ms) by a visual array to be searched for a target (of a type that differed between experiments) at the subject's leisure. The array was followed by a mask to make target search and detection difficult. The interval between the tone (stimulus for Task 1) and array (stimulus for Task 2) had little effect on RT in Task 1 or on accuracy in Task 2. In other experiments, the array was not masked, and a speeded response was obtained in both tasks. In that situation, a dramatic impact of intertask interval on the RT for Task 2 was found, even when the first response was manual and the second response vocal. There was little influence of intertask interval on Task 1 RT. These results show that responding to Task 1 delays responding to Task 2, even though the perceptibility of Task 2 stimuli has not been affected.

Of course, it is unlikely that attentional effects are limited to one stage of processing. Johnston & Heinz (1978) showed that subjects can select on the basis of either physical or semantic characteristics, although with a greater cost in terms of performance on a secondary RT task in the case of semantic selection. A response bottleneck did not preclude perceptual limitations in the Pashler (1989) study, either. When Task 1 was a complex visual task (identification of a right-pointing slash among left-pointing slashes), perceptibility was affected along with response processes. Severe interference with Task 2 accuracy was observed. An IP explanation of these observations would identify detection of the physical cue and semantic recognition of the semantic cue as two different stages, and would differentiate the attentional limitations of these two stages.

Pashler (1990), too, argued for several types of attentional limitation. One is a general "bottleneck" or serial constraint on response selection. Superimposed on this general response bottleneck are limitations in more specific resources. One such limitation is a resource used to perceive elements within a complex visual array (Pashler 1989). Another is a short-term memory constraint that may come into play when subjects must make two similar, consecutive responses. Pashler (1990) found that unpredictability in the order of two responses greatly increased the amount of interference between them if both were manual responses, whereas predictability had little effect when one response was manual and one vocal. The residual interference that occurred regardless of response modality or predictability was taken to reflect a response bottleneck, whereas the effect that was dependent on predictability presumably reflected intrusion errors in short-term memory for the two manual responses. These three processing limitations may be localized in what would appear to be distinct stages of information processing (perceptual encoding of the stimulus, response buffering in memory, and response selection).

When attention does *not* have a consistent effect on an assumed stage of processing, that stage may have to be broken into smaller stages, according to the principle of decomposition described by Palmer & Kimchi (1986). For example, considerable evidence has suggested that while perception is partly preattentive, some important perceptual products emerge only in the presence of attention (Cowan 1988; Posner & Snyder 1975; Treisman 1991). Thus the stage termed "perception" may comprise at least two substages (e.g. featural encoding and featural combination) that are affected in different ways by manipulations of attention (also see Massaro 1985).

These considerations indicate a strategy for investigating attention within an IP approach. One should try to identify aspects of processing that are preattentive, and try to identify attentional effects that can be localized in a particular stage. A good example of this approach is a study by Shulman (1991), who found effects of attention (to one of two figures rotating in opposite directions) on the degree of motion aftereffect. A physiological example is the finding that attention to one of several auditory stimulus channels affects the latency of the eyeblink startle reflex (Hackley & Graham 1987). These studies suggest that attention can affect not only response processing but even fairly early stages of stimulus processing. As Kinchla (1992) concludes, selectivity occurs in many processes ranging from early to late stages of IP.

Strategies are controllable aspects of cognition in which control is exercised in order to maximize task performance. Attention and strategies are central to IP research (and to all of experimental psychology) in an important, often unacknowledged way. In our experiments we often require subjects to follow a precise protocol. In tasks with visual displays, we tell subjects to focus on (attend to) the fixation point; in memory tasks, to remember strategically a set of items; and so on. We take it for granted that subjects can follow these instructions to the letter, but we sometimes forget to ensure that they are actually doing so. For example, according to Logan (1988), subjects in the consistently mapped search task might carry it out not by using the presented search set on each trial, as Shiffrin & Schneider (1977) assumed, but by overlearning the entire superset from which each search set was drawn.

Summary

It is possible to use performance accuracy and speed to chart the flow of information through a number of distinct stages of information processing. Investigators have usefully shifted among several levels of analysis in order to capture different aspects of information flow. At a sequential-stage level, investigations have focused on factors that affect distinctly different segments in the chain of processing, treating the basic types of information transfer (e.g. discrete or continuous). Zooming in to a substage level, investigations have examined the responses of the processing system when multiple, concurrent demands are placed on a particular processing stage. Zooming out to a superstage level, they have examined the strategic and attentional factors that modulate the way stages act and interact. The processing system appears capable of shifting from parallel processing and continuous transmission of information under low-demand conditions, to more serial processing and discrete transmission when necessary to meet special demands such as unfamiliar stimuli or the need for great accuracy of response.

These conclusions are speculative, but substantial empirical and theoretical progress is being made. For example, Yantis et al (1991) have provided a detailed discussion of various ways to analyze response distributions to determine whether subjects are using a single strategy or a mixture of strategies. Schweickert (in press) has summarized recent theoretical work that extracts from the intuitively appealing early concepts about interactions and non-interactions in parallel and serial processing (Sternberg 1975) a more logically complete set of dependencies. For example, factors affecting sequential stages can interact, but stages can be combined into superstages that do not interact. Researchers can usefully apply these concepts to new data, provided they take seriously the need to check the assumptions underlying the theoretical analyses.

ILLUSTRATIVE DOMAINS OF INQUIRY

Psychophysics

It took an engineering approach to appreciate the value of a stage model in psychophysics. Two stages occur between stimulus and response: sensory and decision. The sensory stage's input is the stimulus; its output is some sensory event. This information is made available to the decision stage before a response is made. The output of the decision operation determines the response, but its input is the information given by the outcome of the sensory process rather than by the stimulus itself. Transformation of the stimulus into a sensory experience is determined by both stages. The output of the sensory system—the sensation value—should be a direct function of the characteristics of the stimulus and the state of the sensory system. The operations of the decision stage should be affected by variables that determine the appropriateness of a response—for example, knowledge of the experimental situation, payoffs, attitudes, and motivations. Each process is influenced only by an exclusive set of variables. The rule of the decision process is independent of the operation of the sensory process. Similarly, the sensitivity of the sensory process should be independent of any decision bias induced by the decision process. The value of this distinction goes well beyond psychophysics. We discuss its use in visual perception, speech perception, reading, memory, and judgment. Before doing so, we consider evidence for a belief bias in addition to a decision bias.

BELIEF BIAS AND DECISION BIAS The preceding two-stage analysis permits independent measures of sensitivity and bias. The former measures operation of the sensory process and the latter operation of the decision process. Although the distinction between sensitivity and bias is important, we must also distinguish between two types of bias. One we term a "belief bias," referring to the way the subject interprets the stimulus. The other we term a "decision bias," referring to the way the subject is inclined to respond, given the payoff matrix that is in place. A subject might go against his/her beliefs for a certain payoff. For example, subjects willingly learn to respond in concordance with false information feedback rather than with their actual experience. In one study, subjects learned to call a loud tone soft and a soft tone loud in order to match the feedback given after each trial (Massaro 1969).

Perhaps because it is difficult to distinguish between the two types of bias, signal-detection theory blurs belief bias and decision bias. However, the distinction can be made in some situations. If a signal-detection analysis is performed on an optical illusion, such as the Müller-Lyer figure, there is reason to believe that the illusion would be primarily reflected in bias and not sensitivity. That is, we would see the Müller-Lyer figure with outgoing wings as longer than a control figure but our ability to discriminate line length would remain intact. In fact, Nevin (1991) found that the illusion is adequately explained by a bias effect without changes in line-length discrimination. Although Nevin calls the bias "response bias," we interpret it as belief bias. The perceiver actually "sees" one line as longer than another, even though they are the same length (a true optical illusion).

Although the distinction between sensitivity and bias is reasonable and has been confirmed by experimental analysis, identification of the type of bias would remain speculative. Luckily, empirical techniques and theoretical analyses now exist that enable us to distinguish belief bias from decision bias. Connine & Clifton (1987) replicated the finding that lexical context can influence a phoneme judgment in speech (Ganong 1980). For example, subjects were more likely to report /t/ in the context *-ype* and /d/ in the context *-ice*. The lexical contribution occurred only within the ambiguous range of the stimulus property distinguishing /t/ and /d/. To address the bias issue, an ingenious follow-up study involved only nonwords. A monetary payoff scheme was imposed to bias the subjects to respond with one alternative or the other. The results of the test revealed a bias similar in magnitude to that found with lexical context. If the lexical effects are belief bias and the payoff effects are decision bias, the response probabilities cannot be taken to reflect this fact. Luckily, the pattern of RTs for the two tasks differed even though the response probabilities did not. When the bias was produced lexically, the RTs of word judgments were faster only for speech stimuli that gave a context effect on response probability. When the bias was produced with a monetary payoff, the RTs were always faster for the bias-consistent alternative even for speech stimuli that gave no effect of payoff on response probability. Given this evidence for two types of bias, it seems reasonable that the lexical effect induced a belief bias and the monetary payoff a decision bias.

Visual Perception

Stages of information processing have been taken most seriously in the field of visual perception (see Banks & Krajicek 1991 for a recent review). Perception is hierarchical; it is traditionally assumed, for example, that there are at least three stages of processing: retinal transduction, sensory cues, and perceived attributes (DeYoe & Van Essen 1988). There is a one-to-many and many-to-one relationship between sensory cues and perceived attributes. As an example, physical motion provides information about both the shape of an object and its movement. Similarly, information about the shape of an object is enriched not only by physical motion, but also by linear perspective, binocular disparity, and shading (e.g. chiaroscuro).

Bennett et al's (1989) "observer theory" has three basic assumptions compatible with most of our current understanding of perception: (*a*) perception is a process of inference, (*b*) perceptual inference is not deductively valid, and (*c*) perceptual inferences are biased. Of course, the first two assumptions go back at least to Helmholtz and require little explanation. The third simply means that the perceptual system reaches some interpretations in preference to others—seeing many two-dimensional projections as three-dimensional, for example. Included in perceptual bias is the minimality principle of Gestalt psychology: We see the simplest possible interpretation of a pattern (Leeuwenberg & Boselie 1988). The three assumptions mentioned here confront the inverse mapping problem: The perceiver's goal is to determine what environmental situation exists, given the current conflux of sensory cues.

Speech Perception

Stimulus information and context contribute to speech perception (Bagley 1900). An important controversial issue is the nature of their joint contribution. Two competing explanations are the TRACE model (McClelland & Elman 1986) and the Fuzzy Logical Model of Perception (FLMP; Massaro 1987, 1991). The first assumes that context modifies lower-level representations (a sensitivity effect); the second assumes independent contributions to a higher-level representation (a bias effect). The distinction between sensitivity and bias has enabled several tests between these alternatives. Simulations have shown that TRACE predicts sensitivity effects whereas the FLMP does not (Massaro 1989b). Using a signal-detection analysis of behavioral results, the contribution of phonological context on phoneme identification was shown to be located in bias and not sensitivity (Massaro 1989b). This result supports the FLMP and contradicts the predictions of the TRACE model. Similar conclusions are warranted for another set of context effects in speech perception.

In the original type of phonemic-restoration study (Warren 1970), a phoneme in a word is removed and replaced with some other stimulus, such as a tone or white noise. Subjects perceive the word as intact and have difficulty indicating what phoneme is missing. Samuel (1981) asked whether failure to spot the missing phoneme is a sensitivity effect or a bias effect, as these effects are defined in signal-detection theory. Signal and noise trials were tested. For noise trials, the phoneme was replaced with white noise. Signal trials contained the same noise superimposed on the original phoneme. Subjects indicated whether or not the original phoneme was present. Sensitivity is reflected in the degree to which the two types of trials can be discriminated, and can be indexed by d' within the context of signal-detection theory. Bias, indexed by β, would be reflected in the overall likelihood of saying that the original phoneme is present.

Samuel compared performance on phonemes in test words to performance on the phoneme segments presented in isolation. A bias was observed in that subjects were more likely to respond that the phoneme was present in the word than in the isolated segment. In addition, subjects discriminated the signal from the noise trials much better in the segment context than the word context. The d' values averaged two or three times larger for the segment context than for the word context. Thus, top-down context from the word appears to have a large negative effect on sensitivity. However, the segment vs word comparison in the Samuel study confounds stimulus contributions with top-down contributions (Massaro 1989b). Forward and backward masking may degrade the perceptual quality of a segment more when presented in a word than when presented alone. In addition, the word context may provide co-articulatory information about the critical phoneme that is not available in the isolated segment. Thus one ought not to conclude that the difference in d' values results from top-down context.

A second study compared a word context to a pseudoword context—e.g. *modern* was compared to *madorn*. The presentations were primed because subjects would not know what sequence of segments makes up a pseudoword. Each word or pseudoword was spoken in intact form (primed) before it was presented as a test item. A d' advantage of primed pseudowords over primed words was observed. Unfortunately, natural speech was used in this experiment, and the pseudowords' durations averaged about 10% longer than those of the words. When words are spoken with a longer duration, as in citation speech, they usually provide a higher-quality speech signal.

In a final experiment, Samuel placed test words in a sentence context. The same test word was either predicted or not by the sentence context. The influence of sentence predictability appears to be a valid comparison because the test stimuli were the same in the predictable and unpredictable contexts. The predictability of the test word significantly influenced bias but not sensitivity. To summarize, sensitivity effects were found only when the stimuli were confounded. Repp (1992) found similar evidence against sensitivity effects in phonemic restoration.

Selective adaptation in speech perception also appears to influence bias and not sensitivity. In selective adaptation, listeners are exposed to a number of repetitions of an "adapting" syllable and then asked to identify syllables from a speech continuum between two speech categories. Relative to the baseline condition of no adaptation, the identification judgments of syllables along the speech continuum are pushed in the direction opposite that of the adapting syllable (a contrast effect). Roberts & Summerfield (1981) employed different adaptors to evaluate unimodal and cross-modal adaptation and found no evidence for cross-modal adaptation. The visual adaptors presented alone produced no adaptation along the auditory continuum. Similarly, equivalent levels of adaptation were found for an auditory adaptor and a bimodal adaptor with the same phonetic information. The most impressive result, however, was the adaptation obtained with the conflicting bimodal adaptor. An auditory /be/ paired with visual /ge/ adaptor is sometimes perceived as /the/, /ve/, or /de/, but seldom as /be/. Even so, this condition produced adaptation equivalent to the auditory adaptor /be/. Thus, the adaptation followed the auditory information and was not influenced by the visual information or the phenomenal experience of the bimodal syllable. This result suggests that there is an acoustic feature-evaluation stage that is unaffected by visual feature processing. The feature-evaluation stage is also immune to a following integration stage that combines features from different modalities to give a phenomenal experience of the speech event.

More generally, top-down effects on sensitivity have yet to be convincingly demonstrated. The context effects can be explained by supposing that stimulus information and context make independent but joint contributions to word recognition. Contrary to interactive activation models (McClelland & Elman

1986), the concept of top-down activation of lower-level representations appears to be unnecessary and wrong (Massaro 1989b).

Reading Written Words

Context effects occur in reading as well as in speech perception. Some of the first experiments in experimental psychology established a word superiority effect (WSE) (Cattell 1886). Letters in a word were more accurately reported than letters in a nonword. Although a variety of uninteresting reasons could be responsible for this result, the IP approach has convincingly demonstrated that word context influences word perception.

Two classes of explanations of the WSE can be distinguished. One explanation, within the context of the FLMP, states that readers given a word have two sources of information, while readers given a nonword have only a single source. Integrating letter information and word context will lead to more accurate performance relative to a nonword or single-letter condition. The interactive activation model (IAM) has a different explanation (McClelland & Rumelhart 1981): The WSE occurs because of activation from the word level to the letter level.

Using an IP approach, Massaro (1979) evaluated these two explanations. A reader was asked to read lowercase letter strings containing an ambiguous test letter with a shape intermediate between that of *c* and *e*. In addition, the test letter was placed in orthographic contexts that supported *e* and *c* to various degrees. The test string was presented for a short duration followed after some short interval by a masking stimulus. Subjects were instructed to identify the test letter on the basis of what they saw. Both the test letter and the context influenced performance in the expected direction. Furthermore, the effect of context was larger for the more ambiguous test letters along the stimulus continuum.

A signal-detection analysis indicated that orthographic context strongly affected bias. The influence of orthographic context was *independent* of the influence of the ambiguous test letter. That is, context did not influence the reader's sensitivity to the differences in the ambiguous letters. Similarly, the test letter did not influence the perceiver's sensitivity to the differences in the orthographic context. Oden (1984) obtained similar results.

This study also evaluated context effects as a function of processing time controlled by backward masking. Both the test letter and the context influenced performance at all masking intervals. The test letter had less effect at the short than at the long processing times. That is, the identification functions covered a larger range across the *e*–*c* continuum with increases in processing time. Context has a significant effect at all masking intervals. In fact, the context effect was larger for the unambiguous test letters at the short than at the longer masking intervals. This result follows naturally from the trade-off between stimulus information and context in the FLMP. Context has a smaller influence to the extent the stimulus information is unambiguous.

These results falsified stochastic interactive activation (SIAC) models (McClelland 1991). In the FLMP a strong context will not override a relatively weak stimulus, as it can in the SIAC models. Given the latter's assumption of interactive activation, context can sometimes overwhelm stimulus information about the target as additional processing occurs. This prediction is contradicted by both experience and experimental results. We are more likely to notice a misspelling in a word we read carefully. In experiments varying target information, context, and processing time, stimulus effects are larger as processing time increases. SIAC models cannot predict the observed stimulus and context effects with increasing processing time.

Memory

Priming—the influence of earlier information on later performance—is a central interest in memory research. A stage analysis, in the context of signal-detection theory, has illuminated the nature of priming. In a typical task, subjects are presented with a list of study items. Some time later, they are asked to identify test words, each flashed briefly on a screen and masked. The likelihood of correctly identifying the word is greater for words previously studied than for words not previously studied. Ratcliff et al (1989) asked if this advantage would be located in the sensitivity (d') or the criterion (β) component of the signal-detection formulation.

Subjects read sentences containing priming words—e.g. *died.* Priming was positive if the subject was more likely to identify *died* correctly when it was presented later as a test word. A forced-choice task with two alternatives, *died* or *lied,* was used. If the context sentence enhances the *discrimination* of the words within the sentence from other words, then subjects should be better able to discriminate *died* from *lied.* Enhanced discriminability would lead to an increase in d'. The context sentence might also simply increase the overall likelihood of the subject's responding *died.* Both results might also occur.

Ratcliff et al (1989) replicated previous results, showing that the likelihood of correct identification of a test word was greater given priming of the test word by a context sentence. However, the forced-choice task revealed that the effect was on β, not d'. In contrast to the positive effect on β, prior occurrence in a sentence context did not influence d'. The results, therefore, showed that a sentence context biases the perceptual system in the direction of deciding that it sees a particular test item that has been seen in the recent past; the sentence context does not enhance the ability of the perceptual system to discriminate that word from similar words. This perceptual bias is probably advantageous in daily life where words are often repeated in text (or conversation) in contexts where similar words are unlikely. Such a bias can increase the efficiency of processing without significantly increasing processing cost.

Decision Making

Signal-detection theory has been used to clarify the role of prior probability in decision making. Such a use is controversial, owing primarily to the work of Kahneman & Tversky (1972) and the studies of Leon & Anderson (1974), Birnbaum & Mellers (1983), and Gigerenzer et al (1988). Situations can be created in which subjects will ignore prior probability in their decisions. As an example, Kahneman & Tversky create a hypothetical male who is shy, withdrawn, meek, tidy, methodical, and orderly. Subjects rank the likelihood that he is a farmer, salesman, airline pilot, librarian, or physician. Librarian is ranked higher than farmer despite the fact that there are many more farmers than librarians and that there are relatively few male librarians. This result might be interpreted to mean that subjects do not distinguish between prior probabilities—prior probability is uninformative. However, interpretation requires a test to see whether subjects can detect a difference between two *different* prior probabilities when they are also given an influential description.

Nevin (1991) evaluated this question by factorially combining prior probability and description. Presented with descriptions of a male or female as either stereotypically (*a*) shy and withdrawn or (*b*) quiet but strong, subjects were asked to assign as occupations either librarian or airline pilot. As expected, the personality description influenced the likelihood that the person was judged to be a librarian. However, gender also influenced the judgment. Subjects are sensitive to gender differences in occupations, and a strong stereotypic personality description does not overwhelm this sensitivity. Furthermore, when measured within the framework of signal-detection theory, the influence of prior probability was relatively constant across the two different descriptions. Based on these results, we can conclude that the description and the gender of the individual being described provided independent sources of evidence for the person's occupation. Both sources of evidence were influential and one source did not override the other.

Summary

Psychophysical theory has helped us to determine at which information-processing stage a particular variable has an effect. Visual perception is hierarchically structured in terms of stages and influenced by multiple sources of information. In speech perception and reading, a fundamental question concerns how multiple sources of information influence processing. In a simple two-stage model, each source of information is evaluated and passed forward to a stage that integrates these evaluations. Two important questions concern the extent to which crosstalk occurs between the evaluations and whether integration feeds back and influences evaluation. The evidence in both domains indicates a negative answer to both questions. Similarly, the contribution of a priming stimulus in memory can bias later perceptual processing but does not appear to feed down and modify sensory processing. Finally, a biased description of a person does not override information about prior probability of occurrence.

VARIATIONS ON THE IP FRAMEWORK

We support the IP paradigm as a general metatheory for psychological inquiry. We envision other metatheories as variations on the IP framework rather than as alternatives. In many respects, these variations constrain the general framework of IP. We consider four of these variations: Physical Symbol Systems (PSS), Connectionism, Modularity, and Ecological Realism. The first three propose unique architectures that house processing. The fourth denies as much as possible any internal structure. In many respects the differences among variations are modest.

Physical Symbol Systems (PSS)

At the heart of Physical Symbol Systems (PSS) theory are *symbolic architectures,* whose structures underlie computer science (Newell 1990; Newell et al 1989). In this framework, as Harnad (1990) notes, arbitrary *physical tokens* are manipulated by *explicit rules* that are also composed of tokens. The manipulations are based solely on the physical properties of the tokens, not their meaning. All processing entails *rule-based combination* of symbol tokens and strings of tokens. The entire system is *semantically interpretable*; that is, its components all stand for objects or describe states of affairs.

Several assumptions of the PSS framework are more restrictive than those of the IP approach. In the latter, units of information need not be limited to semantically interpretable symbols, inasmuch as subsymbolic and even continuous featural values might be processed. Also, inasmuch as the stages of processing need not conform to the steps that one would observe in the rule-based manipulation of symbols, the IP approach places a greater emphasis on understanding the temporal course of processing in humans. Few mainstream experimental psychologists who adhere to the IP approach work in the PSS framework.

Newell (1990) notes that computation is necessarily local—internal; symbols are therefore needed to represent the external world. Newell's SOAR architecture is the most ambitious PSS to date. He proposes 10 operating principles of the model human processor. This processor is clearly nonmodular, since the operating principles are constant across different processing domains. SOAR uses a production system as its foundation, and functions as a recognize-act system. A fundamental process is search through a problem space that is constructed on the fly. The learning mechanism, called chunking, resolves conflicts among productions by combining (*a*) the conditions involved in the conflict with (*b*) the outcomes to store a new production in memory. Given adherence to the traditional PSS architecture, there doesn't seem to be much room in SOAR for continuous information (fuzziness)—a shortcoming also noted by Lindsay (1991). It should also be noted that pattern recognition is not one of the 10 operating principles. This stands in marked contrast to the apparently central role played by pattern recognition, even in cognitive domains such as chess that have been investigated by advocates of PSS (Chase & Simon 1973).

Connectionism

Connectionism (Rumelhart & McClelland 1986) includes the following features: a set of *processing units* (simulated neurons, computing elements) that are interconnected by *connection weights*. Inputs to the system cause activations that are modulated by the connection weights. Inputs to a unit are combined with its current state according to an *activation rule*. The connection weights are manipulated on the basis of *learning rules,* which are mathematical *functions*. The manipulation is based purely on the value of the activations and the connection weights (not their "meaning"). Inputs to the system are transformed into outputs by the activation rules, which consist of *combining* and recombining activation values. Although some have argued that connectionism does not have explicit representations (Ramsey et al 1991) there is evidence that the system and all its parts are semantically interpretable (P. Verschure 1992). More generally, connectionism does not seem to have departed from the "representationalist strategy" (Cummins 1991). The approach has both semantically structured representations and learning procedures that are characteristic of GOFAI—good old-fashioned artificial intelligence.

Connectionism shares several attributes with the general IP approach (Massaro 1990). Both information processing and connectionism contain both parallel and sequential processing. Traditional IP models have assumed that feature analysis of letters occurs in parallel, and letter recognition is dependent on the output of the feature analysis [e.g. Selfridge's (1959) seminal pandemonium model]. In IP models, certain processes are assumed to be sequential. For example, a short-term memory search might not begin until the test item is recognized (Sternberg 1975). Sequential processing also occurs in connectionist recurrent network models in that top-down activation of lower-level units might not occur until activation of lower-level units activates higher-level units (Anderson et al 1977).

The distinction between local and distributed processing in connectionism parallels to some extent the difference between feature and node theories of memory representation (e.g. Schvaneveldt & Meyer 1973). Node theories place all the information about a pattern within a single representation, whereas feature theories distribute the information across several different representations. Local representation of some concept exists when all of the information about the concept is represented in a single location, whereas distributed representation refers to a representation of the concept that is distributed across several locations or a representation that is not located in one place. In connectionism, local representation corresponds to the case in which information about a pattern is stored in the connections of a single unit reserved for that pattern. Representation is distributed when information about a pattern is stored in the interconnections among many processing units. One class of distributed models can be mimicked by local models, which may blur somewhat the local-distributed distinction (Smolensky 1986). Thus, the local-distributed distinction does not differentiate connectionist and IP paradigms of inquiry. Estes (1991) believes that it will be difficult, if not impossible, to distinguish between local and distributed representation.

Connectionist models have been used to explain behavior at a functional level of description by assuming processes *analogous* to those occurring at a physiological level. Within IP, the feature detectors for letters were viewed as instances of neural units uncovered in electrophysiological research (e.g. Lindsay & Norman 1972). Models are metaphors; the metaphorical aspect of connectionist models is less detectable than that of other models because the connectionist metaphor is glossed neurologically rather than psychologically (Gentner & Grudin 1985). When models are formulated in a connectionist paradigm using "neurological" terms, they may not attract the analytic scrutiny that is necessary for precision, systematization, and empirical evaluation.

The defining characteristic of IP models, if there is one, is their inclusion of a multitude of mental processes operating jointly to produce behavior. Connectionist models, to date, do not assume distinct stages of processing but simply a direct mapping between input and output. To us the distinction between different stages of processing appears necessary to explanations of even the simplest behavior. In reading aloud English text, for example, the visual form of the letters does not predict the pronunciation as well as do the letter names (i.e. categories). Solving this mapping with hidden units camouflages the possibility that distinct processing stages are involved. Although the desired behavior may be produced by a connectionist model, the model does not necessarily elucidate the behavior (McCloskey 1991).

Some claim that connectionism offers an alternative to the physical-symbol-system paradigm (Derthick & Plaut 1986). The latter often uses symbols to embody sensory experience and rules to map experience into action. Connectionism uses activations to embody sensory experience and the modification and transmission of these activations to map experience into action. This aspect of connectionism is consistent with the IP approach: IP theory has a history of nonsymbolic representations including discriminability, familiarity, memory strength, and even activation and inhibition among representations.

Connectionist models with more than two layers of units may be too unconstrained to be informative. Models of this type may be Turing-equivalents capable of mimicking any computable function—any possible result. Hidden-unit models can predict not only observed results but also results that do not occur (Massaro 1988). That is, connectionist models with hidden units can simulate results that have not been observed in psychological investigations and results generated by incorrect process models of performance (Massaro

1988). More recently, it has been shown that a wide class of input-output mappings can be simulated as long as the theorist uses a sufficient number of hidden units and the operational system does not get trapped in a local minimum (Massaro 1988; Hornik et al 1989).

This superpower of connectionist models with hidden units allows the investigator to avoid the traditional framework of psychophysics (specifying the environmental characteristics that are utilized by subjects). If any input can be mapped to the desired output, then the characterization of the input does not matter. Massaro (1988) illustrated how the superpower of connectionist models with hidden units can also camouflage the observation of different stages of processing. Hidden units can simulate the outcomes of intervening stages of processing, but they do not shed light on how the intervening processes work (McCloskey 1991).

Finally, feedforward models with hidden units, as currently instantiated, make no predictions about the time course or dynamics of IP. Kawamoto (in press) has extended these models by adding cascading assumptions as formulated by McClelland (1979). Activation at each layer of units grows gradually and continuously passes its activation on to the next layer. This extension has the potential to predict response choice and RT as a function of the available processing time. An important discovery was that even the simplest networks have several configurations of weights that can solve some problem, such as exclusive or (XOR). Kawamoto found that these different solutions make significantly different predictions about the dynamics of behavior. This result is analogous to the superpower of the static predictions of these same models. Thus, dynamic hidden-layer models also have the potential for being superpowerful in that they can predict several different types of results. The variety of predicted results will most likely exceed the types of actual results, creating a situation in which the theory is too powerful and not falsifiable. We believe that connectionist theory will have to become more stage-like to be falsifiable and to solve mappings between input and output in an informative manner. That is, networks created for each of the prototypical stages uncovered in the IP approach should be informative and testable.

Modularity

Modularity has been offered as another framework for studying the functional level of behavior (Fodor 1983; Gardner 1985). Modularity assumes independent input systems (such as the one responsible for object perception) and more general cognitive processes called central systems. Input systems are *domain specific* in the sense that each uses different information and processes it differently. Input systems resemble stages of information processing in terms of selective influence by variables. Input systems are also *mandatory* because they must operate given the appropriate stimulus. Their most important property is *encapsulation*: Processing is influenced only by information within the input module's domain. Thus, a speech module is influenced only by speech input, not by situational and linguistic context. Input modules are *cognitively impenetrable*—i.e. not subject to volitional control. Input systems have shallow outputs; for example, a language module for lexical access outputs only word meanings. Input systems operate independently of one another and do not communicate. Finally, an input module is associated with specific neural structure.

In contrast to input systems, central systems are influenced by many different variables. Central systems have access to the outputs of all of the input systems and all knowledge in memory. In this dichotomy, input systems are considered to be computational systems, whereas central systems correspond to what the organism "believes." Input systems can be studied as computational systems, whereas central systems cannot. Because so many factors influence central processes, Fodor believes that a scientific account of the latter is not possible (see also Gardner 1985). This conclusion stands in marked contrast to proponents of PSS. Simon & Kaplan (1989) believe that "deep thinking" has proved easier to understand and simulate than hand-eye coordination.

Fodor & Pylyshyn (1981) argue that the process of inference must stop somewhere. That is, the system must be grounded at some level in which inference is not necessary. They propose a trichotomy of levels: transducers, input systems, and central systems. A transducer putatively does not infer: It detects a property P whenever it occurs, and this detection gives rise to state S. State S does not arise if property P does not occur. Bennett et al (1989) provide an illuminating critique of this assumption. Even for the simplest property such as light, however, transducers do not meet this criterion. State S can occur without light as, for example, when the eye is pressed and produces phosphenes. The accepted theory of signal detection is also based on the assumption that inference is central to the detection of outputs from transducers. Thus, inference is also involved in simple "transducer" functions such as light detection. Inference is not deductively valid at all three levels and the buck of inference does not stop. Using this same logic, Bennett et al also demonstrate that the distinction between input systems and central systems is unjustified. Additional evidence is Massaro's (1987) finding of similar processes in speech perception (an input system) and fixation of belief (a central system).

In summary, the modularity approach shares many of the same premises as the IP approach, particularly the assumptions of separable systems for perception and action (Massaro 1987, 1989). However, modularity's distinctions between transducers and input systems and between input systems and central systems appear to be erroneous.

Ecological Realism

Ecological realism appears to be best described in terms of what it denies (Gibson 1966, 1979). This approach rejects the notion that perception is a form of knowing. Ecological realists reject the idea of an ambiguous environment embellished by processes of the perceiver. For them, the perceiver simply extracts *invariants* from the sensory flux. Processing occurs, but intermediate processes are absent, as it is assumed that the important properties of the environment are invariant relations that can simply be "picked up" by the observer. In addition, ecological realists consider processing to be *non-algorithmic*. Gibson suggested that the perceptual system is attuned to pick up environmental properties without computation in a manner resembling the mechanical resonance between one piano string and another. This metaphor makes oscillatory systems and nonlinear dynamic systems attractive explanatory constructs applicable to information processing.

Many ecological realists in psychology view their work as contrary to that of IP-oriented investigators. As we have noted, IP psychologists focus on internal representations and processes whereas neo-Gibsonians place a high priority on understanding the mapping of invariant properties of the environment to perceptual responses (Carello & Turvey 1991). On the other hand, ecological realists do not seem immune to hidden processes and feel free to use such concepts as attention (Gibson 1979). However, ecological realism has not solved the inverse mapping problem in which the environmental situation must be induced from information available to the perceiving system. Thus, ecological realism must confront the role that inductive inference necessarily plays in perception and action. To accomplish this goal, we believe they will find it necessary to adduce internal mental processes.

RETROSPECTIVE

Here we have attempted to (*a*) characterize the IP approach, (*b*) justify it, (*c*) describe important metatheoretical and theoretical issues, and (*d*) compare it to other metatheories. We hope to have demonstrated that the IP approach goes far beyond the mere use of a single technique such as the AFM. The IP approach has its roots in mathematical psychology as well as the experimental method, and the value of its fine-grained analyses is becoming apparent. We have not been able to survey the impressive progress made in the use of anatomical measures as converging indexes of IP (Meyer et al 1988b; Posner & Carr 1992).

Posner & McLeod (1982) distinguished between more global theoretical syntheses and simpler experimental studies. We look forward to a merging of these two enterprises. Theorists should become more sensitive to empirical constraints and experimentalists should direct their inquiries toward tests of theories. That said, has the field advanced over the past decade? While we have documented here how our understanding of information processing has increased over the 10-year period, we suggest that the field's most important progress has been in learning how to ask the right questions and how most fruitfully to pursue the answers. We look forward to the next decade of progress.

ACKNOWLEDGMENTS

The authors' research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314 to Massaro and R01 HD 21338 to Cowan), the National Science Foundation (BNS 8812728 to Massaro), and the graduate division of the University of California, Santa Cruz. The authors thank Steve Hackley, Steve Kitzis, Nancy McCarrell, Dave Meyer, Jeff Miller, Hal Pashler, Gregg Oden, Allen Osman, Roger Ratcliff, Seth Roberts, Rich Schweickert, Saul Sternberg, Jim Townsend, Lex van der Heijden, and Paul Verschure for helpful comments on a draft of this chapter.

Literature Cited

Abrams, R. A., Balota, D. A. 1991. Mental chronometry: beyond reaction time. *Psychol. Sci.* 2:153–57

Adams, M. J. 1979. Models of word recognition. *Cogn. Psychol.* 11:133–76

Alcock, J. 1989. *Animal Behavior.* Sunderland, MA: Sinauer

Allport, D. A. 1968. Phenomenal simultaneity and the perceptual moment hypothesis. *Br. J. Psychol.* 59:395–406

Anderson, J. A., Silverstein, J. W., Ritz, S. A., Jones, R. S. 1977. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychol. Rev.* 84:413–51

Anderson, J. R. 1990a. *Cognitive Psychology and Its Implications.* New York: Freeman

Anderson, J. R. 1990b. *The Adaptive Control of Thought.* Hillsdale, NJ: Erlbaum

Ashby, F. G. 1982. Deriving exact predictions for the cascade model. *Psychol. Rev.* 89:599–607

Ashby, F. G., Townsend, J. T. 1980. Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *J. Math. Psychol.* 21:93–123

Atkinson, R. C., Shiffrin, R. M. 1968. Human memory: a proposed system and its control processes. In *The Psychology of Learning and Motivation: Advances in Research and Theory,* ed. K. W. Spence, J. T. Spence, 2:89–195. New York: Academic

Baddeley, A. D. 1986. *Working Memory.* Oxford: Oxford Univ. Press

Bagley, W. C. 1900. The apperception of the spoken sentence: a study in the psychology of language. *Am. J. Psychol.* 12:80–130

Banks, W. P., Krajicek, D. 1991. Perception. *Annu. Rev. Psychol.* 42:305–31

Baron, J. 1977. Mechanisms for pronouncing printed words: use and acquisition. In *Basic Processes in Reading: Perception and Comprehension,* ed. D. LaBerge, S. J. Samuels. Hillsdale, NJ: Erlbaum

Bennett, B. M., Hoffman, D. D., Prakash, C. 1989. *Observer Mechanics: A Formal Theory of Perception.* San Diego: Academic

Berko, J. 1958. The child's learning of English morphology. *Word* 14:150–77

Birnbaum, M. H., Mellers, B. A. 1983. Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *J. Pers. Soc. Psychol.* 45:792–804

Breitmeyer, B. G. 1984. *Visual Masking: An Integrative Approach.* New York: Oxford

Broad, W., Wade, N. 1982. *Betrayers of the Truth.* New York: Simon & Schuster

Broadbent, D. E. 1958. *Perception and Communication.* New York: Pergamon

Brooks, L. 1978. Nonanalytic concept formation and memory for instances. In *Cognition and Categorization,* ed. E. Rosch, B. B. Lloyd, pp. 169–211. Hillsdale, NJ: Erlbaum

Carello, C., Turvey, M. T. 1991. Ecological units of analysis and baseball's "illusions." In *Cognition and the Symbolic Processes,* ed. R. R. Hoffman, D. S. Palermo, pp. 371–86. Hillsdale, NJ: Erlbaum

Cattell, J. M. 1886. The time it takes to see and name objects. *Mind* 11:53–65

Chamberlin, T. C. 1965. The method of multiple working hypotheses. *Science* 148:754–59

Chase, W. G., Simon, H. A. 1973. Perception in chess. *Cogn. Psychol.* 4:55–81

Cohen, M. M., Massaro, D. W. 1992. On the similarity of categorization models. In *Probabilistic Multidimensional Models of Perception and Cognition,* ed. F. G. Ashby, pp. 395–447. New York: Academic.

Collyer, C. E. 1985. Comparing strong and weak models by fitting them to computer-generated data. *Percept. Psychophys.* 38:476–81

Connine, C. M., Clifton, C. 1987. Interactive use of lexical information in speech perception. *J. Exp. Psychol.: Hum. Percept. Perform.* 13:291–99

Cosmides, L. 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187–276

Cowan, N. 1984. On short and long auditory stores. *Psychol. Bull.* 96:341–70

Cowan, N. 1987. Auditory sensory storage in relation to the growth of sensation and acoustic information extraction. *J. Exp. Psychol.: Hum. Percept. Perform.* 13:204–15

Cowan, N. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychol. Bull.* 104:163–91

Cummins, R. 1991. The role of representation in connectionist explanations of cognitive capacities. In *Philosophy and Connectionist Theory*, ed. W. Ramsey, S. P. Stich, D. E. Rumelhart, pp. 91–114. Hillsdale, NJ: Erlbaum

Derthick, M., Plaut, D. C. 1986. Is distributed connectionism compatible with the physical symbol system hypothesis? *Proc. Cogn. Sci. Soc.*, pp. 639–44

DeYoe, E. A., Van Essen, D. C. 1988. Concurrent processing streams in monkey visual cortex. *Trends Neurosci.* 11:219–26

Duncan, J. 1980. The locus of interference in the perception of simultaneous stimuli. *Psychol. Rev.* 87:272–300

Duncan, J., Humphreys, G. W. 1989. Visual search and stimulus similarity. *Psychol. Rev.* 96:433–58

Egeth, H., Dagenbach, D. 1991. Parallel versus serial processing in visual search: further evidence from subadditive effects of visual quality. *J. Exp. Psychol.: Hum. Percept. Perform.* 17:551–60

Elman, J., McClelland, J. 1986. Exploiting lawful variability in the speech wave. In *Invariance and Variability in Speech Processes*, ed. J. S. Perkell, D. H. Klatt. Hillsdale, NJ: Erlbaum

Estes, W. K. 1991. Cognitive architectures from the standpoint of an experimental psychologist. *Annu. Rev. Psychol.* 42:1–28

Feyerabend, P. K. 1975. *Against Method*. London: NLB

Fisher, D. L., Duffy, S. A., Young, C., Pollatsek, A. 1988. Understanding the central processing limit in consistent-mapping visual search tasks. *J. Exp. Psychol.: Hum. Percept. Perform.* 14:253–66

Fodor, J. 1975. *The Language of Thought*. New York: Crowell

Fodor, J. A. 1983. *Modularity of Mind*. Cambridge, MA: Bradford

Fodor, J. A., Pylyshyn, Z. W. 1981. How direct is visual perception?: some reflections on Gibson's ecological approach. *Cognition* 9:139–96

Ganong, W. F. III. 1980. Phonetic categorization in auditory word recognition. *J. Exp. Psychol.: Hum. Percept. Perform.* 6:110–25

Gardner, H. 1985. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books

Gentner, D., Grudin, J. 1985. The evolution of mental metaphors in psychology: a 90-year retrospective. *Am. Psychol.* 40:181–92

Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin

Gigerenzer, G., Hell, W., Blank, H. 1988. Presentation and content: the use of base rates as a continuous variable. *J. Exp. Psychol.: Hum. Percept. Perform.* 14:513–25

Glass, A. L., Holyoak, K. J. 1986. *Cognition*. New York: Random House

Glushko, R. J. 1979. The organization and activation of orthographic knowledge in reading aloud. *J. Exp. Psychol.: Hum. Percept. Perform.* 5:674–91

Green, D. M., Swets, J. A. 1966. *Signal Detection Theory and Psychophysics*. New York: Wiley

Greeno, J. G., Steiner, T. E. 1964. Markovian processes with identifiable states: general considerations and application to all-or-none learning. *Psychometrika* 29:309–33

Gregory, R. L. 1986. *Odd Perceptions*. London: Methuen

Groen, G. J., Parkman, J. M. 1972. A chronometric analysis of simple addition. *Psychol. Rev.* 79:329–43

Hackley, S. A., Graham, F. K. 1987. Effects of attending selectively to the spatial position of reflex-eliciting and reflex-modulating stimuli. *J. Exp. Psychol.: Hum. Percept. Perform.* 13:411–24

Harnad, S. 1990. The symbol grounding problem. *Physica D* 42:335–46

Hatfield, G. 1991. Representation in perception and cognition: connectionist affordances. See Ramsey et al 1991, pp. 163–95

Hawkins, H. L., Presson, J. C. 1986. In *Handbook of Perception and Human Performance: Vol. 2. Cognitive Processes and Performance*, ed. K. R. Boff, L. Kaufman, J. P. Thomas, pp. 26-1–26-64. New York: Wiley

Hintzman, D. L. 1991. Why are formal models useful in psychology? In *Relating Theory and Data: Essays on Human Memory in Honor of Bennett B. Murdock*, ed. W. E. Hockley, S. Lewandowsky, pp. 39–56. Hillsdale, NJ: Erlbaum

Hornik, K., Stinchcombe, M., White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2:359–66

Idson, W. L., Massaro, D. W. 1977. Perceptual processing and experience of auditory duration. *Sensory Processes* 1:316–37

Johnston, W. A., Heinz, S. P. 1978. Flexibility and capacity demands of attention. *J. Exp. Psychol.: Gen.* 107:420–35

Kahneman, D., Tversky, A. 1972. Subjective probability: a judgment of representativeness. *Cogn. Psychol.* 3:430–54

Kallman, H. J., Massaro, D. W. 1979. Similarity effects in backward recognition masking. *J. Exp. Psychol.: Hum. Percept. Perform.* 5:110–28

Kallman, H. J., Massaro, D. W. 1983. Backward masking, the suffix effect, and preperceptual storage. *J. Exp. Psychol.: Learn. Mem. Cogn.* 9:312–27

Kawamoto, A. H. 1992. Time course of processing in feed-forward connectionist networks. *J. Math. Psychol.* In press

Kinchla, R. A. 1992. Attention. *Annu. Rev. Psychol.* 43:711–42

Lachman, R., Lachman, J. L., Butterfield, E. C. 1979. *Psychology and Information Processing.* Hillsdale, NJ: Erlbaum

Lee, D. N., Raddish, P. E. 1981. Plummeting gannets: a paradigm of ecological optics. *Nature* 293:293–94

Leeuwenberg, E., Boselie, F. 1988. Against the likelihood principle in visual form perception. *Psychol. Rev.* 95:485–91

Leon, M., Anderson, N. H. 1974. A ratio rule from integration theory applied to inference judgments. *J. Exper. Psychol.* 102:27–36

Lindsay, P. H., Norman, D. A. 1972. *Human Information Processing: An Introduction to Psychology.* New York: Academic

Lindsay, R. K. 1991. Symbolic-processing and the SOAR architecture. *Psychol. Sci.* 2:294–302

Logan, G. D. 1988. Toward an instance theory of automation. *Psychol. Rev.* 95:492–527

Lovie, A. D. 1983. Attention and behaviourism—fact and fiction. *Br. J. Psychol.* 74:301–10

MacKay, D. M. 1969. *Information, Mechanism, and Meaning.* Cambridge, MA: MIT Press

MacWhinney, B., Leinbach, J. 1991. Implementations are not conceptualizations: revising the verb learning model. *Cognition* 40:121–57

Mahoney, M. J. 1976. *Scientist as Subject: The Psychological Imperative.* Cambridge: Ballinger

Manicas, P. T., Secord, P. F. 1983. Implications for psychology of the new philosophy of science. *Am. Psychol.* 38:399–413

Massaro, D. W. 1969. The effects of feedback in psychophysical tasks. *Percept. Psychophys.* 6:89–91

Massaro, D. W. 1972. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychol. Rev.* 79:124–45

Massaro, D. W. 1975. *Experimental Psychology and Information Processing.* Chicago: Rand-McNally

Massaro, D. W. 1979. Letter information and orthographic context in word perception. *J. Exp. Psychol.: Hum. Percept. Perform.* 5:595–609

Massaro, D. W. 1985. Attention and perception: an information-integration perspective. *Acta Psychol.* 60:211–43

Massaro, D. W. 1987. *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry.* Hillsdale, NJ: Erlbaum

Massaro, D. W. 1988. Some criticisms of connectionist models of human performance. *J. Mem. Lang.* 27:213–34

Massaro, D. W. 1989a. *Experimental Psychology: An Information Processing Approach.* San Diego, CA: Harcourt Brace Jovanovich

Massaro, D. W. 1989b. Testing between the TRACE model and the Fuzzy Logical Model of Perception. *Cogn. Psychol.* 21: 398–421

Massaro, D. W. 1990. The psychology of connectionism. *Behav. Brain Sci.* 13:403–6

Massaro, D. W. 1991. Language processing and information processing. In *Contributions to Information Integration Theory.* Vol. I: *Cognition,* ed. N. H. Anderson, pp. 259–92. Hillsdale, NJ: Erlbaum

Massaro, D. W. 1992. Broadening the domain of the fuzzy logical model of perception. In *Cognition: Conceptual and Methodological Issues,* ed. H. L. Pick Jr., P. Van den Broek, D. C. Knill, pp. 51–84. Washington, DC: Am. Psychol. Assoc.

Massaro, D. W., Cohen, M. M. 1983. Phonological context in speech perception. *Percept. Psychophys.* 34:338–48

Massaro, D. W., Cohen, M. M. 1991. Integration versus interactive activation: the joint influence of sensory information and context in perception. *Cogn. Psychol.* 23:558–614

Massaro, D. W., Friedman, D. 1990. Models of integration given multiple sources of information. *Psychol. Rev.* 97:225–52

Massaro, D. W., Friedman, D. 1991. Book review of *The Adaptive Character of Thought* by John R. Anderson. *Am. J. Psychol.* 104: 467–74

McClelland, J. L. 1979. On the time relations between mental processes: a framework for analyzing processes in cascade. *Psychol. Rev.* 86:287–330

McClelland, J. L. 1991. Stochastic interactive processes and the effect of context on perception. *Cogn. Psychol.* 23:1–44

McClelland, J. L., Elman, J. L. 1986. The TRACE model of speech perception. *Cogn. Psychol.* 18:1–86

McClelland, J. L., Rumelhart, D. E. 1981. An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychol. Rev.* 88:375–407

McCloskey, M. 1991. Networks and theories: the place of connectionism in cognitive science. *Psychol. Sci.* 2:387–95

Meyer, D. E., Osman, A. M., Irwin, D. E., Kounios, J. 1988a. The dynamics of cognition and action: mental processes inferred from speed-accuracy decomposition. *Psy-*

chol. Rev. 95:183–237
Meyer, D. E., Osman, A. M., Irwin, D. E., Yantis, S. 1988b. Modern mental chronometry. *Biol. Psychol.* 26:3–67
Meyer, D. E., Yantis, S., Osman, A. M., Smith, J. E. K. 1985. Temporal properties of human information processing: tests of discrete versus continuous models. *Cogn. Psychol.* 17:445–518
Miller, J. 1982. Discrete versus continuous stage models of human information processing: in search of partial output. *J. Exp. Psychol.: Hum. Percept. Perform.* 8:273–96
Miller, J. 1988. Discrete and continuous models of human information processing: theoretical distinctions and empirical results. *Acta Psychol.* 67:191–257
Miller, J. 1990. Discreteness and continuity in models of human information processing. *Acta Psychol.* 74:297–318
Miller, J., Hackley, S. A. 1992. Electrophysiological evidence for temporal overlap among contingent mental processes. *J. Exp. Psychol.: Gen.* 121:195–209
Mitroff, I. I. 1974. *The Subjective Side of Science.* Amsterdam: Elsevier
Moore, E. F. 1956. Gedanken-experiments on sequential machines. In *Automata Studies,* ed. C. E. Shannon, J. McCarthy. Princeton, NJ: Princeton
Moore, J. J., Massaro, D. W. 1973. Attention and processing capacity in auditory recognition. *J. Exp. Psychol.* 99:49–54
Neisser, U. 1967. *Cognitive Psychology.* New York: Appleton-Century-Crofts
Nevin, J. A. 1991. Signal-detection analysis of illusions and heuristics. In *Signal Detection: Mechanisms, Models, and Applications,* ed. M. L. Commons, J. A. Nevin, M. C. Davison, pp. 257–74. Hillsdale, NJ: Erlbaum
Newell, A. 1980. Physical symbol systems. *Cogn. Sci.* 4:135–83
Newell, A. 1990. *Unified Theories of Cognition.* Cambridge, MA: Harvard Univ. Press
Newell, A., Rosenbloom, P. S., Laird, J. E. 1989. Symbolic architecture for cognition. In *Foundation of Cognitive Science,* ed. M. I. Posner, pp. 93–131. Cambridge, MA: MIT Press
Norman, D. A. 1969. *Memory and Attention: An Introduction to Human Information Processing.* New York: Wiley
Norman, D. A., Wickelgren, W. 1969. Strength theory of decision rules and latency in retrieval from short term memory. *J. Math. Psychol.* 6:192–208
Nosofsky, R. M. 1991. Relation between the rational model and the context model of classification. *Psychol. Sci.* 2:416–21
Nosofsky, R. M. 1992. Similarity scaling and cognitive process models. *Annu. Rev. Psychol.* 43:25–53
Oden, G. C. 1984. Dependence, independence, and emergence of word features. *J. Exp. Psychol.: Hum. Percept. Perform.* 10:394–405
Osman, A., Bashore, T. R., Coles, M. G. H., Donchin, E., Meyer, D. E. 1992. On the transmission of partial information: inferences from movement-related brain potentials. *J. Exp. Psychol.: Hum. Percept. Perform.* 18:217–32
Paap, K. R., Newsome, S. L., Noel, R. W. 1984. Word shape's in poor shape for the race to the lexicon. *J. Exp. Psychol.: Hum. Percept. Perform.* 10:413–28
Palmer, S. E., Kimchi, R. 1986. The information processing approach to cognition. In *Approaches to Cognition: Contrasts and Controversies,* ed. T. J. Knapp, L. C. Robertson, pp. 37–77. Hillsdale, NJ: Erlbaum
Pashler, H. 1989. Dissociations and dependencies between speed and accuracy: evidence for a two-component theory of divided attention in simple tasks. *Cogn. Psychol.* 21:469–514
Pashler, H. 1990. Do response modality effects support multiprocessor models of divided attention? *J. Exp. Psychol.: Hum. Percept. Perform.* 16:826–42
Pinker, S., Prince, A. 1988. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition* 29:73–193
Platt, J. R. 1964. Strong inference. *Science* 146:347–53
Plunkett, K., Marchman, V. 1991. U-shaped learning and frequency effects in a multilayered perceptron: implications for child language acquisition. *Cognition* 38:43–102
Popper, K. 1959. *The Logic of Scientific Discovery.* New York: Basic Books
Posner, M. I., Carr, T. H. 1992. Lexical access and the brain: anatomical constraints on cognitive models of word recognition. *Am. J. Psychol.* 105:1–26
Posner, M. I., McLeod, P. 1982. Information processing models—in search of elementary operations. *Annu. Rev. Psychol.* 33:477–514
Posner, M. I., Snyder, C. R. R. 1975. Attention and cognitive control. In *Information Processing and Cognition,* ed. R. L. Solso. Hillsdale, NJ: Erlbaum
Ramsey, W., Stich, S. P., Rumelhart, D. E., eds. 1991. *Philosophy and Connectionist Theory.* Hillsdale, NJ: Erlbaum
Ratcliff, R. 1978. A theory of memory retrieval. *Psychol. Rev.* 85:59–108
Ratcliff, R. 1988. Continuous versus discrete information processing: modeling accumulation of partial information. *Psychol. Rev.* 95:238–55
Ratcliff, R., McKoon, G., Verwoerd, M. 1989. A bias interpretation of facilitation in perceptual identification. *J. Exp. Psychol.: Learn. Mem. Cogn.* 15:378–87
Real, L. A. 1991. Animal choice behavior and the evolution of cognitive architecture. *Science* 253:980–86
Reeve, T. G., Proctor, R. W. 1984. On the ad-

vance preparation of discrete finger responses. *J. Exp. Psychol.: Hum. Percept. Perform.* 10:541–53

Repp, B. H. 1992. Perceptual restoration of a "missing" speech sound: auditory induction or illusion? *Percept. Psychophys.* In press

Roberts, S. 1987. Evidence for distinct serial processes in animals: the multiplicative-factors method. *Animal Learn. Behav.* 15:135–73

Roberts, S., Sternberg, S. 1992. The meaning of additive reaction-time effects: test of three alternatives. In *Attention and Performance.* Vol. 14: *Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience—A Silver Jubilee,* ed. D. E. Meyer, S. Kornblum. Cambridge, MA: MIT Press

Roberts, M., Summerfield, Q. 1981. Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Percept. Psychophys.* 30:309–14

Rumelhart, D. E., McClelland, J. L., eds. 1986. *Parallel Distributed Processing.* Vol. 1: *Foundations.* Cambridge, MA: MIT Press

Rumelhart, D. E., McClelland, J. L. 1986. On learning the past tense of English verbs: implicit rules or parallel distributed processes? In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* ed. D. E. Rumelhart, J. L. McClelland, the PDP Research Group, 2:216–71. Cambridge, MA: MIT Press

Samuel, A. G. 1981. Phonemic restoration: insights from a new methodology. *J. Exp. Psychol.: Gen.* 110:474–94

Sanders, A. F. 1990. Issues and trends in the debate on discrete vs. continuous processing of information. *Acta Psychol.* 74:123–67

Schneider, W., Shiffrin, R. M. 1977. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84:1–66

Schoemaker, P. J. H. 1991. The quest of optimality: a positive heuristic of science? *Behav. Brain Sci.* 14:205–45

Schvaneveldt, R. W., Meyer, D. E. 1973. Retrieval and comparison processes in semantic memory. In *Attention and Performance,* ed. S. Kornblum, pp. 395–409. New York: Academic

Schweickert, R. 1978. A critical path generalization of the additive factor method: analysis of a Stroop task. *J. Math. Psychol.* 18:105–39

Schweickert, R. 1985. Separable effects of factors on speed and accuracy: memory scanning, lexical decision, and choice tasks. *Psychol. Bull.* 97:530–46

Schweickert, R. 1992. Information, time, and the structure of mental events: a twenty-five year review. In *Attention and Performance.* Vol. 14: *Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience—A Silver Jubilee,* ed. D. E. Meyer, S. Kornblum. Cambridge, MA: MIT Press

Selfridge, O. G. 1959. Pandemonium: a paradigm for learning. In *Mechanization of Thought Processes,* pp. 511–26. London: Her Majesty's Stationery Office

Shannon, C. 1948. The mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423, 623–58

Shaw, M. L. 1984. Division of attention among spatial locations: a fundamental difference between detection of letters and detection of luminance increments. In *Attention and Performance.* Vol. 10: *Control of Language Processes, ed. H. Bouma, D. G. Bouwhuis. Hillsdale, NJ: Erlbaum*

Shiffrin, R. M. 1988. Attention. In *Stevens' Handbook of Experimental Psychology,* ed. R. C. Atkinson, R. J. Herrnstein, G. Lindzey, R. D. Luce, 2:739–811. New York: Wiley

Shiffrin, R. M., Schneider, W. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* 84:127–90

Shulman, G. L. 1991. Attentional modulation of mechanisms that analyze rotation in depth. *J. Exp. Psychol.: Hum. Percept. Perform.* 17:726–37

Simon, H. A. 1979. Information processing models of cognition. *Annu. Rev. Psychol.* 30:363–96

Simon, H. A., Kaplan, C. A. 1989. Foundation of cognitive science. In *Foundation of Cognitive Science,* ed. M. I. Posner, pp. 1–47. Cambridge, MA: MIT Press

Smolensky, P. 1986. Neural and conceptual interpretation of PDP models. In *Parallel Distributed Processing.* Vol. 2: *Psychological and Biological Models,* ed. J. L. McClelland, D. E. Rumelhart, pp. 390–431. Cambridge: MIT Press

Smolensky, P. 1988. On the proper treatment of connectionism. *Behav. Brain Sci.* 11:1–73

Solso, R. L. 1991. *Cognitive Psychology.* New York: Allyn & Bacon

Sorkin, R. D., Pohlmann, L. D., Gilliom, J. D. 1973. Simultaneous two-channel signal detection. III. 639 and 1400 Hz signals. *J. Acoust. Soc. Am.* 53:1045–50

Sternberg, S. 1966. High-speed scanning in human memory. *Science* 153:652–54

Sternberg, S. 1969. The discovery of processing stages: extensions of Donders' Method. *Acta Psychol.* 30:276–315

Sternberg, S. 1975. Memory scanning: new findings and current controversies. *Q. J. Exp. Psychol.* 27:1–32

Strayer, D. L., Kramer, A. F. 1990. Attentional requirements of automatic and controlled processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 16:67–82

Theios, J., Amrhein, P. C. 1989. Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: naming, and visual and conceptual comparisons. *Psychol. Rev.* 96:5–24

Thomas, E. A. C., Myers, J. L. 1972. Implications of latency data for threshold and nonthreshold models of signal detection. *J. Math. Psychol.* 9:253–85

Townsend, J. T. 1974. Issues and models concerning the processing of a finite number of inputs. In *Human Information Processing: Tutorials in Performance and Cognition*, ed. B. H. Kantowitz. Hillsdale, NJ: Erlbaum

Townsend, J. T. 1990. Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychol. Sci.* 1:46–54

Townsend, J. T., Ashby, F. G. 1983. *Stochastic Modeling of Elementary Psychological Processes.* London: Cambridge Univ. Press

Treisman, A. 1991. Search, similarity, and the integration of features between and within dimensions. *J. Exp. Psychol.: Hum. Percept. Perform.* 17:652–76

Tribus, M., McIrvine, E. C. 1964. Energy and information. *Sci. Am.* 211:179–88

Turvey, M. T. 1973. On peripheral and central processes in vision: inferences from an information processing analysis of masking with patterned stimuli. *Psychol. Rev.* 80:1–52

Ulrich, R. 1987. Threshold models of temporal-order judgments evaluated by a ternary response task. *Percept. Psychophys.* 42:224–39

van der Heijden, A. H. C., Stebbins, S. 1990. The information-processing approach. *Psychol. Res.* 52:197–206

Venezky, R. L., Massaro, D. W. 1987. Orthographic structure and spelling-sound regularity in reading English words. In *Language Perception and Production: Shared Mechanisms in Listening, Speaking, Reading and Writing,* ed. D. A. Allport, D. G. MacKay, W. Prinz, E. Scheerer, pp. 111–29. London: Academic

Verschure, P. F. M. J. 1992. Taking connectionism seriously: the vague promise of subsymbolism and an alternative. Proc. 14th Annu. Conf. Cogn. Sci. Soc., Bloomington, Indiana, pp. 653–58

Warren, R. M. 1970. Perceptual restoration of missing speech sounds. *Science* 167:392–63

Wenner, A. M., Wells, P. H. 1990. *Anatomy of a Controversy: The Question of a "Language" among Bees.* New York: Columbia Univ. Press

Yantis, S., Meyer, D. E., Smith, J. E. K. 1991. Analyses of multi-nomial mixture distributions: new tests for stochastic models of cognition and action. *Psychol. Bull.* 110:350–74

Young, P. 1987. *The Nature of Information.* New York: Praeger