CHAPTER 34

From Momentary Maximizing to Serial Response Times and Artificial Grammar Learning

Charles P. Shimp, Walter Herbranson, *and* Thane Fremouw

Abstract

Our overarching priority has been to develop method and theory to clarify the ideas of James and Skinner on the importance of streams of thought and of behavior. We describe experimental methods to quantitatively control and theoretical methods to explain the local continuity, or moment-to-moment nature, of thought and behavior in time.

Keywords: optimality in behavior streams, optimality in categorizing, local-global attentional switching, artificial grammar learning, local and global statistical learning in serial response times, behaving theories

"The stream of thought" (James, 1890) emphasizes that mental life is dynamic and continuously changing, and "the behavior stream" makes the same point with regard to behavior (Schoenfeld & Farmer, 1970). These metaphors remind us of what seems obvious but nevertheless is often neglected in contemporary behavioral science: Anything we do or think comes after one thing and before something else, and this local temporal patterning is usually critical to understanding behavior. For example, the order of words in a sentence changes the meaning of an individual word, and the order of notes in a melody changes the meaning of an individual note. Individual words or notes may therefore carry little meaning outside of the local patterning of words or notes. We believe if scientific psychology is to understand mental life and behavior, it must look to the dynamic local temporal contexts of thoughts and behaviors. An artificial hand or leg would be a poor substitute for the real thing if it could not continuously move, and method and theory for understanding how a pitcher throws a baseball would do a poor job if it applied only to a static, average hand position. A snapshot of a person sitting in a chair, or even a blurry composite image of a person walking across a street might convey some useful information but would scarcely tell us what it is like to walk—that is, to actually behave. This point seems so obvious that it says much about the enormous power of tradition to shape scientific behavior when we see many standard methods, concepts, and theories in behavioral science that scarcely acknowledge it (Shimp, 1992, 2009). We are fascinated to observe that even most theories of forgetting, attending, or timing do not attempt to place these processes in the context of continuous behavior streams. In short, we believe there is great need to facilitate the development of "behaving theories" that address the continuity of behavior in time (Shimp, 1992, 2009). As a result, in this chapter we have chosen to describe what we believe is progress in developing some newer experimental methods and "behaving theories" that explicitly address the local continuity, or moment-to-moment nature, of behavior in time (Shimp, 1992, 2009).

We have chosen to describe the stream of behavior and the stream of thought by focusing on our own comparative research. We have done so to make the presentation more nearly manageable, but in doing so we regrettably have had to omit much essential work of others. Fortunately, much of this work appears elsewhere in this volume. We start by describing some of the intellectual priorities and themes that have guided our work.

An overarching priority has been to develop method and theory to clarify the ideas of James and Skinner on the importance of streams of thought and of behavior. James devoted an entire chapter in his classic *The Principles of Psychology* (1890) to the stream of thought, and Skinner spent much of his career closely examining local patterns of behavior in cumulative records. Skinner later lamented the extinction of cumulative records and their replacement primarily by overall properties of behavior. Looking only at overall average behavior is an eminently sensible activity if local properties of the stream that produce it are random. Otherwise, overall output becomes problematic because the same overall output can obviously be caused by different interacting local processes. Local behaving seems scarcely ever to be random, so we have been motivated to study the local organization of behavior.

Three additional priorities have influenced much of our research. First, we have been enamored of discovering methods that actually control the quantitative local performance of individual organisms, on the conventional scientific grounds that discovering variables that control something, in this case the stream of behavior, facilitates understanding it. Second, we have happily stolen method and theory from human cognitive psychology to study nonhuman animal performances. Our approach has been to exploit human methodology when it has been shown to give insight into human cognition and to invoke the possibility of mental continuity when intuitively it has seemed that nonhuman analogous methods could be developed. Third, we have felt strongly the need for better theoretical understanding of the empirical literature on nonhuman animal performances. This in turn has led us to consider history, sociology, psychology, philosophy of science, and philosophy of language to gain perspective on what it means to understand and evaluate a theory and has led us to search for implicit assumptions and beliefs underlying claims on behalf of objective method and results (Benham & Shimp, 2004; Shimp, 1990, 2001).

Sequential Behaviors in Probabilistic Tasks

When we began in the 1960s, probability learning was seen as a useful tool to study intelligence from a comparative perspective and to facilitate the development of general theories of learning, such as stimulus sampling theory. The comparative question was a special case of how organisms behave rationally or irrationally. As this chapter is being written, this question has recently arisen with some psychological violence in the "everyday" financial world and in the more ethereal world of economic theory. In both cases, how humans deal rationally or irrationally with risk has apparently been widely misunderstood. In the much smaller world of comparative cognition, the question is sometimes seen in terms of the relative intelligence of different species, and one can ask whether nonhuman animals probability match or maximize in probability learning tasks. That is, do they produce suboptimal steady-state choice probabilities that approximately equal or match corresponding reinforcement probabilities, or do they behave more rationally and tend exclusively to choose the alternative with the greater reinforcement probability? In short, the comparative question when we began our research was whether nonhuman animals suboptimally "matched" or optimally "maximized." This question is not of only esoteric laboratory interest, because behavioral economics sometimes looks for the evolutionary bases for human behavior, and accordingly, comparative laboratory studies of probabilistically reinforced choice behavior become potentially relevant to real-world human economic behavior. The comparative question was not the only reason why researchers were interested in this empirical question. Stimulus sampling theory was the best-articulated available quantitative theory of choice behavior, and it predicted probability matching rather than maximizing. For both comparative and learning-theory reasons, research therefore explored whether animals matched or maximized. The answer turned out to depend on technical issues, including whether correction or noncorrection procedures were used (Shimp, 1966). While stimulus sampling theory sometimes used trial-by-trial sequential behavior to estimate theoretical parameters, most of the nonhuman animal empirical literature, being focused on overall choice as a measure of intelligence, did not, and researchers instead focused almost exclusively on overall choice proportions. In addition, researchers looked primarily at overall choice proportions because in a specific free-operant choice procedure (a particular kind of "concurrent variable interval schedule"), overall choice proportion matched overall reinforcement proportions (Herrnstein, 1961).

Our initial analysis of choice behavior examined both overall choice proportions *and* sequential properties of the behavior stream because sequential features of a behavior stream were important to the development of stimulus sampling theory. An early outcome of our research showed that pigeons tended locally to choose alternatives in a sequential manner that approximately maximized local reinforcement probability ("momentary maximizing") in a way that seemed to explain the matching obtained by Herrnstein and others (Shimp, 1966). This demonstration implied pigeons might generally learn local reinforcement probabilities in complex dynamic contexts, so we proceeded to examine more directly whether pigeons could discriminate local reinforcement probabilities that changed over time. We found that they could indeed discriminate among rapidly changing reinforcement probabilities for different choices, and we identified some task parameters that modulated the degree of precision with which they could do so (Shimp, Long, & Fremouw, 1996).

All these tasks were variations of probability learning tasks, and accordingly were of a discrete-trials type that could give only a sequence of static snapshots of the continuous behavior stream. We began to imagine how the behavior stream might consist of a succession of different behavioral units, each having some temporal extension, so that they themselves would involve temporal patterning. To explore this possibility, we made reinforcement contingent on extremely simple temporal patterns, interresponse times (IRTs), patterns consisting of an initial key peck followed by a prescribed temporal duration and terminated by a second key peck. We found that IRTs and even sequences of IRTs conformed to lawful quantitative functions, and did so in ways that highlighted the crucial importance of temporal parameters of tasks, as would be expected from the perspective according to which the behavior stream consists of sequences of temporal patterns (Shimp, 1968). We then further generalized behavioral units to involve temporal durations between successive pecks on multiple keys (inter-changeover times) and found that they too depended lawfully not just on reinforcement probabilities but on the temporal durations as well (Shimp, 1979). These and other demonstrations (Hawkes & Shimp, 1975, 1998) that complex local patterns of responding can be established and maintained by directly reinforcing them led to the question of whether animals can actually remember the temporal order of their own behaviors. We therefore examined the simple possibility that differential reinforcement established these complex patterns because an animal remembered having made them when a reinforcer was delivered. We asked if pigeons could remember the sequential order in which they had made recent responses and found that they could (Shimp, 1976a). Other results described below on the relation between implicit and explicit knowledge indicate that this result does not have universal applicability, but it suggests that at least on some occasions, there is a direct correlation between the local sequential organization of events a pigeon has recently encountered and how a pigeon "reports" or "describes" that serial organization.

We interpreted these results to imply that perhaps a great many different kinds of local temporal patterns of responding could be directly reinforced and shaped to function as higher-order units of behavior, the behaviors in terms of which phenomena, principles, and processes can be expressed. Serving as such units, these serial patterns of behavior would depend in elegant quantitative ways on their own temporal organization, as well as on reinforcement parameters associated with them. Our results supported the speculation that at least some such local units consisted of behavioral patterns a subject could remember having made when a reinforcer was delivered. We saw these results on local organization as generally compatible with a growing interest in the role of organization in human memory (Shimp, 1976b), and we saw them as encouraging mental continuity as a conceptual basis for the further exploration of the applicability of human method and theory to nonhuman animals.

Optimality in Categorization

Two features of the phenomenon of momentary maximizing attracted our special attention: local optimality and local serial organization. First let us consider optimality. It is easy to see optimality where there is none (Voltaire, 1959 [1759]). In our research, we have therefore tried to restrict our claims that performance was optimal to cases where optimal performance could be clearly defined and compared to nonoptimal performances.

Multidimensional Categorization

Multidimensional categorization can be conceptualized as a generalization of choosing between available alternatives such as left or right, or red, yellow, or green, in a probability learning task. Accordingly, categorization tasks can investigate the degree to which nonhuman animals choose optimally.

The literature on complex, multidimensional categorization in humans grew exponentially when categories ceased to be viewed primarily in terms of the binary logic of truth tables and began to be viewed in more naturalistic ways involving natural language and permitting some ambiguity in the definitions of categories. We see a parallel transition between early and late philosophies of Wittgenstein (1922, 1953). A procedure that has proven useful to the study of optimality in multidimensional categorization is the randomization procedure (Ashby & Maddox, 1998). In this procedure, a two-dimensional category is represented as a bivariate normal distribution. Such distributions are well suited to research on categorization because potentially limitless numbers of individual exemplars can be sampled from them, matching the limitless exemplars that constitute real-world categories such as "tree," "pigeon," or "rock." Prototypical examplars are located toward the peak of the distribution, and are thus more likely to be sampled. Atypical exemplars are located further away, and are less likely to be sampled. Figure 34.1 is taken from Herbranson et al. (1999) and summarizes the procedure. The left panel depicts two category distributions, A and B. The space over which the distributions are defined is typically referred to as the stimulus space, in which each point represents a particular two-dimensional stimulus (for example, a rectangle with width *x* and height *y*). The third coordinate, *z*, is the likelihood with which that stimulus will occur given a particular category. The right panel in Figure 34.1 shows two equal-likelihood contours, each of which efficiently summarizes a bivariate normal distribution by showing points corresponding to stimuli that are equally likely to occur given a particular category. The right panel also shows the optimal decision bound, the line formed by the points corresponding to stimuli that are equally likely to occur given either category. Optimality of responding is easily diagnosed with this procedure because a participant maximizes the likelihood of correct categorization if stimuli on one side of the optimal decision bound are categorized as belonging to Category A and stimuli on the other side are categorized as belonging to Category B.

As it has been our custom to explore mental continuity, we adapted this procedure for use with nonhuman animals. We showed pigeons rectangles of varying lengths and widths. After a rectangle was presented, a bird could categorize the rectangle by pecking one of two locations, each of which corresponded to one of the two categories, A or B. If a response corresponded to the category from which the presented stimulus had been drawn, a bird was presented with food. Across conditions, we sampled stimuli from categories that yielded different optimal decision bounds (corresponding, for example, to rules such as "go left if a rectangle is wider than tall; otherwise, go right," or "go left if a rectangle is wider than the average width regardless of its height"). In each condition, the great majority of pigeons' categorizations were optimal, with averages ranging from around 77% to 91%, depending on

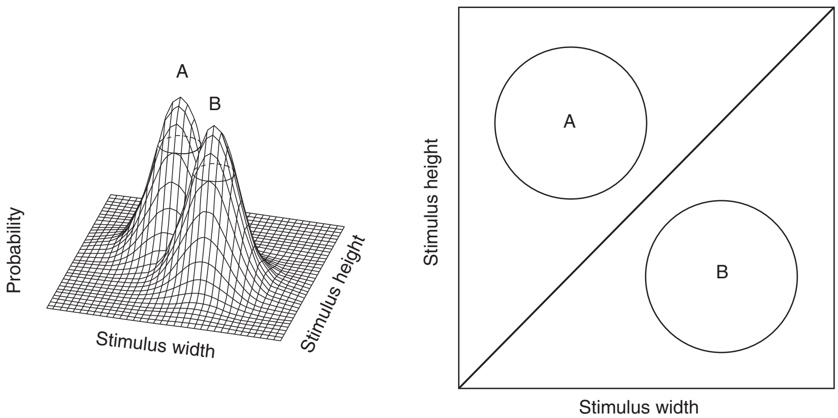

Fig. 34.1 *Left panel*: Bivariate normal distributions representing the likelihoods with which specific rectangles could be sampled from either of two overlapping categories, A and B. Each point in the stimulus space corresponded to a rectangle having a width and height equal to the *x* and *y* coordinates, respectively. One arbitrary contour of equal likelihood is shown for each category. Each contour consisted of all points corresponding to rectangles equally likely to be sampled from the distribution. *Right panel*: Arbitrary contours of equal likelihood for each category and the corresponding linear optimal decision bound (from Herbranson, Fremouw, & Shimp, 1999).

the specific categories used. Furthermore, we found that if accurate categorization in a condition required information from both stimulus dimensions, pigeons generally divided attention between the two dimensions and made choices that were roughly optimal. Similarly, if categories were defined based on a single dimension (with the second stimulus dimension varying randomly), pigeons selectively attended to the relevant dimension and again made choices that were roughly optimal. These results mirrored those that have been obtained from human participants in similar situations (Ashby & Maddox, 1998). Note that the stimuli used in these categorization tasks, indeed in virtually all categorization tasks, with humans as well as nonhuman animal subjects, are static, whereas naturalistic stimuli are often dynamic. We accordingly adapted the categorization task so that the two dimensions were the speed and direction with which a virtual object moved across a computer screen (Herbranson, Fremouw, & Shimp, 2002). We surmised that pigeons might perform well on this task because in their natural environment, the ability to categorize some moving objects (say, predators) in terms of their dynamic characteristics could be quite useful. We did in fact discover that pigeons used speed and direction to categorize moving objects with astonishing precision and, on the average, with almost perfect optimality (Fig. 34.2). We believe this optimal categorization of an object moving on different trials at different speeds and in different directions greatly generalizes our original phenomenon of momentary maximizing. Pigeons can learn how local reinforcement probability rapidly changes as a function of several different kinds of local environmental stimuli, including stimuli in tasks involving different kinds of static and dynamic multidimensional stimuli, features of their own behavior, and patterns of recent events.

We think these highly diverse forms of locally controlled behaviors suggest that nonhuman animals accurately estimate local reinforcement probabilities and temporal task parameters within the context of a behavior stream, and can do so with a level of precision that is generally underestimated. Our results suggest that it is especially underestimated when conventional empirical procedures do not permit a researcher to directly estimate local control so that the possibility is ignored altogether (Shimp, 1973, 1979). We think a more careful approach, given the range of precise local control that can be demonstrated, is to explicitly show local control does not apply before assuming that it does not.

Attentional Dynamics

A stream of mental life may quickly change course if an organism encounters an hierarchically structured stimulus like a forest and its component trees, to which one can attend in either a global or a local manner. The ability to quickly switch between

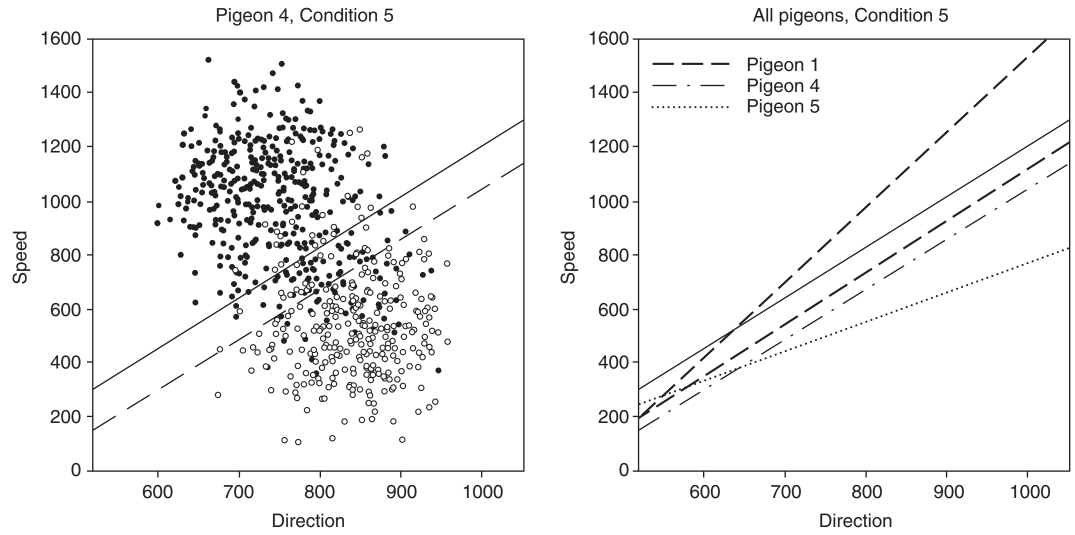

Fig. 34.2 *Left panel*: The obtained stimulus space for one bird in Herbranson, Fremouw, and Shimp (2002). Filled and open circles correspond to individual responses categorizing stimuli as members of categories A and B, respectively. The dashed line shows the estimated decision bound for this bird, and the solid line shows the optimal decision bound. *Right panel*: Individual estimated decision bounds for three birds, along with average estimated (bold dashed) and optimal (bold solid) decision bounds.

global or local analyses, as in attending either to a forest or its trees, is important if, as often seems to be the case, reinforcement depends on the level of the stimulus. Humans can, of course, switch back and forth between these levels of perceptual analysis, and this ability permits humans to respond over short periods of time in more nearly optimal ways. We have used hierarchically organized stimuli to show that pigeons, much like humans, can shift attention either to local or global features of stimuli.

We presented pigeons with complex stimuli having both a global and a local level (Navon, 1977). Pigeons were "primed" at either the local or global level, and rewarded for responding to targets that could occur at either level. On each trial, pigeons were shown an hierarchical stimulus (i.e., a global character that was created from a number of smaller, local characters). For each set of hierarchical stimuli there were two possible target stimuli and two possible irrelevant "distractor" stimuli. Both the target stimuli and the distractor stimuli could occur at either the local or the global level, resulting in a total of eight hierarchical stimuli, four with a target stimulus at the global level and a distractor stimulus at the local level and four with a target stimulus at the local level and a distractor stimulus at the global level. One such hierarchical stimulus set is shown in Figure 34.3. Pigeons were rewarded for pecking the left key if one target was present in the hierarchical stimulus and the right key if the other target was present in the hierarchical stimulus, regardless of the level at which that target occurred. In short, the pigeon had to search for a target that could be present at either the local or global level.

Initially, we asked if pigeons could change the level to which they attended based on the frequencies with which targets at different levels occurred. That is, we used a base-rate procedure to train (prime) the pigeons to expect a target at a particular level (Fremouw, Herbranson, & Shimp, 1998). Specifically, we presented successive blocks of trials within which targets at the primed level occurred 85% of the time and the targets at the non-primed level occurred the remaining 15% of the time. Over the course of the experiment we alternated blocks of trials with either global or local levels primed. The notion of mental continuity led us to believe that pigeons, like humans, would respond faster to targets at the primed level than to the same targets at the non-primed level. That is what we found. Response times were significantly faster to local targets than to global targets during the blocks in which targets appeared more frequently at the local level (local level primed), and response times were significantly faster to global targets than to local targets during the blocks in which targets appeared more frequently at the global level (global level primed). These results suggest that pigeons can flexibly switch attention between local and global levels of perceptual analysis.

These first experiments did not identify the time frame over which pigeons can switch attention. Attention might have built up slowly as the base rates were learned, and once built up at the primed level, it might have simply remained "active" at that level until the base rate changed. However, we believed that in nonhuman animals, as is the case in humans, shifts of local–global attention can occur

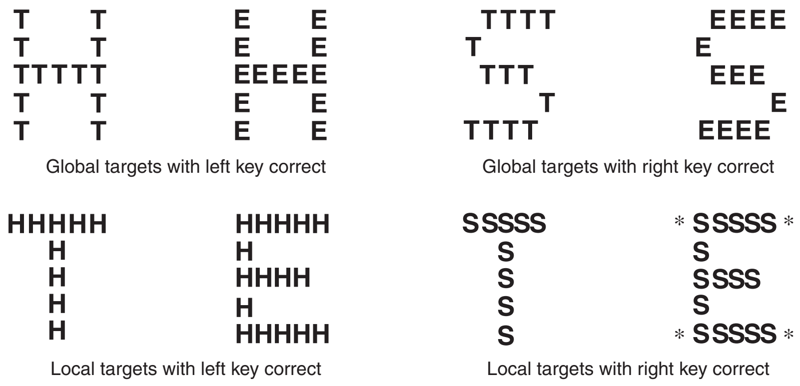

Fig. 34.3 Stimuli similar to those used by Fremouw et al. (1998, 2002). Each hierarchical stimulus had a target stimulus (here either an H or an S) at either the local or global level and an irrelevant distractor stimulus (here either an E or a T) at the other level. The four stars used as a prime in the trial-by-trial priming version of the task (Fremouw et al., 2002), are shown surrounding the bottom right hierarchical stimulus. The stars were presented 1 second or less before a hierarchical stimulus was presented and the color of the stars (red or green) primed a pigeon to the perceptual level at which the target was more likely to occur.

much more quickly. In accordance with our intuitive notions about the stream of consciousness, we believed that moment-to-moment experiences often drive moment-to-moment dynamic shifts in attention. To explore this notion in the case of local–global attention we ran a series of new experiments in which we used a trial-by-trial cuing procedure to train (prime) the pigeons to expect a target at a particular level (Fremouw, Herbranson, & Shimp, 2002). On each trial, we presented a brief priming cue a second or less prior to the presentation of the hierarchical stimulus. In one of the experiments the priming cue consisted of four stars, either all green or all red, that formed the corners of a box slightly larger than the stimuli. The color of the stars predicted, with 85% accuracy, the level at which the target would occur. Targets occurred at the global level a random 85% of the time and at the local level the other 15% of the time if the stars were red (a global prime), and vice versa if the stars were green (a local prime). Local and global targets occurred with equal probabilities overall. As was the case for the base-rate priming, trial-by-trial priming caused a significant change in response time, suggesting that pigeons can flexibly switch attention between local and global levels of analysis, in this case on a moment-to-moment time frame, comfortably compatible with the metaphor of a continuous stream of mental life.

We see the outcomes of these experiments on shifts between local and global levels of attention as entirely compatible with our previous work on shifts between spatial attention to one location or another (Shimp & Friedrich, 1993): in both cases, pigeons can flexibly and quickly switch attention. We accordingly think, again simply because of the possibility of mental continuity, that it is reasonable to anticipate a dynamic attention system across many species. In this we are guided especially by Gestalt ideas that imply the possibility that the stream of mental life can involve rapid switches between local and global levels of analysis as examples of reversals between figure and ground. We believe that these demonstrations of dynamic shifts of attention in nonhuman animals suggest that local dynamics shape to some degree, as it does in humans, what animals see at any moment. That is, we agree with the cognitive position according to which dynamic attentional shifts rapidly change representations of stimuli in streams of thinking. We accordingly believe that long-term theoretical goals will have to include explaining both these rapid shifts of attention and their effects on streams of thinking and of behaving. We believe the importance of these goals is generally underestimated when analyses focus exclusively on long-term average performance.

We think it is sufficiently important to ponder potential mechanisms for attentional shifts to warrant noting that recent neurophysiological findings from bats offer some intriguing possibilities for how feedback between cortex and lower sensory nuclei might play a role in attentional phenomena on both slower time scales, such as in the blocking task, and on faster time scales, such as in the cuing task. Suga et al. (Ma & Suga, 2003; Suga, Gao, Zhang, Ma, & Olsen, 2000) showed that repetitive stimulation of auditory cortex can refine and strengthen neuronal firing in the inferior colliculus, a nucleus that occurs earlier in the auditory processing stream than auditory cortex. For example, stimulating an area of cortex that responds best to a particular frequency range or to a particular delay between sounds seems to strengthen the response of neurons in the inferior colliculus that also respond to that particular frequency range or delay. Inactivation had the opposite effect: The response in the inferior colliculus weakened. This neuronal modulation developed over time, from 2 to 30 minutes, and lasted from minutes to hours. Casseday, Fremouw, and Covey (2002) speculated that this process might help select, enhance, and maintain processing of specific auditory features over the time course of a bat's evening hunt.

We wonder if a similar mechanism, perhaps working on spatial frequency, might play a role in creating local–global attention seen in the base-rate blocking experiments where the dynamics of attention may be relatively slow. Perhaps the high base rate of a particular target level leads to repetitive and prolonged activity of neurons tuned to the appropriate spatial frequency for the corresponding perceptual level. Perhaps once a target level is perceived on a trial, the neurons involved in encoding that level remain active longer, and at a higher level than the neurons that encode the non-target level. Such increased activity might then strengthen and fine-tune the response of neurons to that level in both visual cortex and earlier structures. The enhanced neuronal response might in turn produce faster or more accurate perception of subsequent targets at that perceptual level.

Activity in auditory cortex can also enhance specific auditory features in the inferior colliculus on a much faster, stimulus-by-stimulus time frame (Jen, Chen, & Sun, 1998; Zhou & Jen, 2000). Perhaps a similarly fast-acting mechanism plays a role in the visual system and at least partially mediates the local–global attention shifts seen in the priming cue task. We describe these speculative possibilities to illustrate how research on neurophysiology might inform understanding of continuous behavior streams and streams of mental life, and vice versa.

In summary, we showed that pigeons display flexibility in switching attention between local and global levels of perceptual analysis in a manner similar to that of humans attending either to the forest or the trees. Thus, pigeons can flexibly display a kind of rapid figure–ground reversal that forms part of the core meta-theoretical perspective of Gestalt psychology.

Artificial Grammar Learning

As we conducted pigeon experiments on the effects of local context on such diverse phenomena as optimality in choice behavior, the behavior stream, local–global attentional switching, spatial attention, and multidimensional category learning, (as well as on serial response times as we describe in the next section), and found in each case that local context had profound effects on behavior, we began to speculate on the possibility that pigeons might be sensitive to local context even in the form of grammatical context. Grammar has been identified as one of the key issues in the "cognitive revolution" (Gardner, 1985). Chomsky's (1959) attack on Skinner's (1957) analysis of verbal behavior was largely focused on the linear chaining that seemed implicit in Skinner's analysis and that became explicit in the "Jack and Jill" demonstrations of two pigeons communicating (Epstein, Lanza, & Skinner, 1980). The role of grammar was prominent by its virtual absence in Skinner's analysis of language. While Chomsky's attack may have been more relevant to 1930s methodological behaviorism and to the reductionism inherent in logical positivism than to Skinnerian radical behaviorism, Chomsky certainly assigned to grammar a far greater role in language than Skinner did. We knew that the problem of animal "language" was controversial but felt we could approach it from a new and constructive perspective by using Artificial Grammar Learning (AGL), a method that Chomsky and Miller (1958) had described years earlier, and that Reber (1967) had used to excellent purpose in the study of implicit versus explicit knowledge in humans. AGL research continues to be actively pursued with human participants and is informing our understanding of the evolution of component mechanisms of natural language, if not directly of natural language itself. AGL seemed to us to offer a potentially powerful tool for examining how local context affects visual categorization, specifically how pigeons categorize letter strings generated by formal rules. We saw AGL as a way to move our comparative work on local sequential structure toward increasingly complex stimuli. We hoped that sharply focusing on whether pigeons could learn artificial grammars would facilitate clarifying mechanisms that might be involved in language while letting us avoid some of the more intangible and inscrutable problems that would arise if we asked generally if pigeons could learn language (for discussion, see Savage-Rumbaugh, Shanker, & Taylor, 1998; Rumbaugh & Washburn, 2003; and Terrace, 1979).

An artificial grammar (Reber, 1967) is a set of rules for generating strings of characters. An example of one such grammar is depicted in Figure 34.4. A character string is generated by entering the grammar at the left, with each transition from one state to another adding a character to the string, until exiting via the "out" arrow at the right. In this manner, each unique path through the grammar produces a different character string. By virtue of its recursive loops, the grammar shown can generate an infinite number of unique character strings. Limiting the length of character strings naturally decreases this number, but still results in a large number of unique strings. The simple grammar depicted in Figure 34.4, for instance, can generate 43 character strings between three and eight characters in length. Note that local sequential context is critical to the grammar's definition, in the sense that grammaticality is not determined by individual letters or total numbers of letters, but by the sequential order in which letters appear.

In a prototypical artificial grammar experiment such as that of Reber (1967), undergraduates were

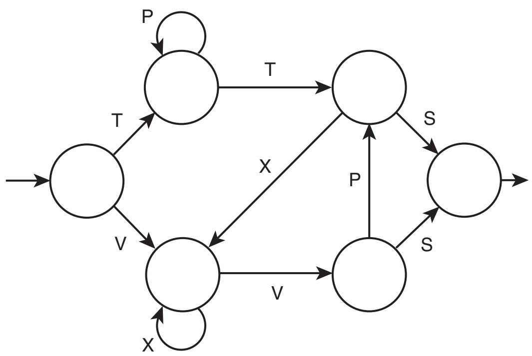

Fig. 34.4 The artificial grammar used by Reber (1967) and by Herbranson and Shimp (2003, 2008).

shown grammatical character strings and asked to memorize them. Later, they were presented with novel strings and were told the letter strings they had memorized earlier were generated by a grammar. They were not, however, shown the grammar or told anything about it. They were simply asked whether the novel strings conformed to the same grammar. Despite the participants' inability to accurately describe the rules of the grammar, performance at diagnosing grammaticality was reliably above chance. This result parallels an important aspect of normal language acquisition in young children, who recognize grammatical sentences when they hear them, and do so without being able to describe grammatical rules. It also parallels the learning of naturalistic visual categories where categories such as "tree" are quickly learned, even though the basis for the discrimination may be unclear even to the experimenter himself, let alone to the accurate categorizer. From the perspective of the participants, artificial grammars have the virtue of preserving the "family resemblance" characteristic of naturalistic categories, where the basis for the discrimination seems complex and ambiguous (Wittgenstein, 1953). At the same time, from the perspective of the experimenter, artificial grammars have the virtues of simplicity and precision: The experimenter actually knows the rules, the true structure, of the category (which is not the case with most naturalistic categories, such as "tree").

Artificial grammar learning in humans may involve nonlinguistic precursors of component processes of human language and therefore deserves a comparative analysis (see Gebhart, Newport, & Aslin, 2009; Gentner, Fenn, Margoliash, & Nussbaum, 2006; Seidenberg, MacDonald, & Saffran, 2002; and Zeigler & Marler, 2008, for related discussion). Pigeons can learn an artificial grammar (Herbranson & Shimp, 2003). We trained birds to discriminate between grammatical and nongrammatical character strings. Birds were rewarded for pecking one key when a character string was displayed that conformed to the rules of the grammar in Figure 34.4. They were rewarded for pecking a different key when the displayed character string violated the grammar. After extensive training (mean of 179 days of training), birds reached a stable level of above-chance performance (62.3% correct) on the training set of 62 character strings (31 grammatical and 31 nongrammatical), suggesting that they may have learned something about the grammar, or at least had learned something correlated with some of the rules of the grammar. To examine the possibility that pigeons were simply memorizing specific training exemplars, at least some of which presumably seemed familiar to the pigeons by the end of training, we subsequently presented novel probe strings (12 novel grammatical and 12 novel nongrammatical strings) that the pigeons had not encountered during training. Performance on these novel strings was also reliably above chance (60.7%), supporting the notion that pigeons acquired a flexible conception of the grammar that went beyond the specific stimuli presented during training.

Most important for our purposes here is that the grammatical status of a character string depended on its local spatial organization. This spatial organization was linear, and in that sense spatially sequential, but of course a pigeon did not necessarily process the linear string in a corresponding sequential order. (An interesting question for future research would be whether a pigeon can be trained to process character strings in a particular sequential order, and if so, whether the order of processing affects judgments of grammaticality depending on the information serially provided by different sequences.) Both grammatical and nongrammatical strings consisted of the same component characters, and the only difference that allowed a pigeon to accurately discriminate between them was the local sequential organization of the characters. We speculate that a striking asymmetry in the data of this experiment further supports the importance of sequential organization. Pigeons' categorizations of grammatical character strings were reliably above chance (70.0% correct), while their categorizations of nongrammatical character strings were not (51.4% correct). We interpret this asymmetry as having been caused by grammatical strings having consistent local sequential organization that nongrammatical strings lacked.

This first AGL experiment motivated us to conduct a second set of experiments (Herbranson & Shimp, 2008) designed to clarify the role of local sequential organization. One experiment used the same general procedure outlined above, but rather than discriminating between grammatical and nongrammatical character strings, pigeons were required to discriminate between two sets of grammatical character strings generated by two different grammars. That is, both categories of character strings derived from consistent rules, and as a result, the asymmetry in performance was eliminated: Pigeons learned both categories with performance above chance for each (78% and 75%). Learning was also much faster than in the first AGL experiment, so much faster that in fact the learning rate was well within those of many visual discrimination tasks. Some investigators find rapid learning rate to be an important criterion for identifying “natural” cognitive processes, and our first AGL experiment may have left some researchers wondering how general our results could have been. The second experiment left little doubt, we believe, that by the standards of plain English, pigeons can learn sequential rules that generate large numbers of character strings.

We believe these AGL experiments importantly extend what is known about what at least some nonhuman animals can learn about arbitrary, abstract, sequential rules. The artificial grammars in these experiments were indeed artificial; the rules were abstract relations involving sequential order of arbitrary characters. If food reward is sufficient to teach pigeons such arbitrary, “meaningless,” sequential relations, we think it becomes all the more likely pigeons can and do learn sequential relations in other contexts where, because no sequential relations are *required* for reinforcement, it may be incorrectly assumed that no learning of sequential context takes place (Shimp, 1976b). That is, we speculate that the metaphors of behavior streams and streams of mental life should be the default interpretations: Attention to, and control by, local structure should be assumed unless otherwise shown to be irrelevant to performance. This suggestion is equivalent to suggesting that a Gestalt-like interpretation be carefully examined before automatically adopting a more atomistic, reductionistic interpretation or one involving only long-term average performances. In the specific case of language, we thereby ally ourselves with Chomsky and others who emphasize the role of sequential structure. Finally, we think it is important that our experiments, especially our second set of experiments, show that pigeons can learn relatively complex sequential structure rapidly, without the kind of social interaction that has been part of other demonstrations of complex sequential patterns in avians (Pepperberg, 2000) and without the kind of conceptual processing that may be involved in the case of perception of musical or artistic style (e.g., Porter & Neuringer, 1984; Watanabe, Sakamoto, & Wakita, 1995).

Local Temporal Context: Serial Response Times

It would not surprise us if it turned out that AGL could be interpreted in terms of statistical learning processes—that is, of mechanisms that learned the statistical likelihoods of various categories, including ill-defined categories, of complex sequences. That is, we suspect that probability learning and the kind of quantitative rule learning that develops in the randomization task may turn out to also identify processes responsible for learning sequential dependencies in AGL. Recent evidence already indicates, for example, that human infants quickly learn statistical relations in natural language (Gebhart, Newport, & Aslin, 2009; Safran, 2003). We do not claim that statistical learning mechanisms acting independently could explain natural language, but they might provide essential input to a larger set of dynamically interacting mechanisms that could do so.

We have developed a response time task that is uncovering statistical learning mechanisms that may have contributed to the evolution of language and may still play a role in language learning in humans. Response times have generally had a bad reputation in behavior analysis (but with noteworthy exceptions; see Blough, 2006) and a good one in cognitive psychology. Skinner derided response times because they had formed a key component of mentalism since the earliest days of experimental psychology, and they involved discrete-trials methodology rather than the continuous free-operant methodology he advocated. Nevertheless, we began to use them because these complaints seemed more philosophical or meta-theoretical than empirical. Furthermore, from the perspective of ebb and flow in streams, a response time of as little as several hundred milliseconds might involve considerable mental dynamics and causal processing, and from our mental-continuity perspective, being allied to human cognitive psychology is as much a virtue as it is a problem.

We became interested in a serial response time task that is very familiar in human neuropsychology but is less so in comparative cognition and still less so in behavior analysis. The procedure turns out to be an astonishingly easy task for pigeons to learn so that it permits the efficient study of effects of local temporal and statistical information. We have been further impressed by the elegance of the data it has produced and by the parsimonious theoretical interpretation to which the data submit.

Froehlich, Herbranson, Loper, Wood, and Shimp (2004) developed a serial response time procedure patterned after a classic human-participant procedure of Nissen and Bullemer (1987) (also see Vickrey & Neuringer, 2000). We required a pigeon to peck sequences of target keys successively lit across three spatial locations. In some versions of the task, the sequentially lit locations were random, and in others, they followed a repeating list, such as LRCCLRLRC. . ., where L, C, and R stand for left, center, and right locations, with the nine-item list starting over at the beginning after the terminal C. Response times were measured from the onset of a lit key until a pigeon pecked it. The task permitted recording several hundred response times from each pigeon in each day's session because pecks were reinforced only intermittently. The question was what a pigeon learned about the sequential structure of the list. We answered it in terms of the first-order and second-order local likelihoods with which one spatial location followed another. We computed the first-order local likelihood as shown in the following example. The first-order local likelihood that C followed R in the above list was 2/3 (the third and ninth items in the list involved center-lit keys following right-lit keys) and there was one occasion on which R was not followed by C (it was followed by L), so two of three occasions of C followed R. Second-order local likelihoods of a spatial location given the previous two locations were computed similarly. For instance, in the same list, LRCCLRLRC, the second-order local likelihood of R, given a previous C and L, was 2/2. We described what a pigeon knew about the first-order and second-order structure of a list by plotting response time as a function of the first- and second-order local likelihoods computed in this manner. Over experimental conditions, we varied the sequential structures of the nine-item lists, the intertrial interval, and the base rate of occurrences of the three different locations within a list.

Pigeons rapidly learned the first- and second-order local likelihoods with which one location followed a previous one or two locations, in the sense that response time, averaged over several pigeons, decreased in accordance with a straight line as the likelihood of a location increased; the more likely a location, the faster a pigeon responded to it. We interpreted this function to show how a pigeon used local statistical information to "anticipate" the next location. We found that the slope of the function relating response time to local information was similar to that obtained with human participants. In a second experiment, we varied intertrial interval and found that the optimal intertrial interval was approximately the same as with human participants. In a third experiment, we varied the global likelihood with which a spatial location appeared within random conditions in which the location of one item provided no local sequential information about the location of the next item. We expected some degree of rational use of overall base-rate information: we expected that a pigeon in an unstructured condition with unequal base rates would respond more quickly to a spatial location that occurred more frequently than another. Instead, we found a not uncommon phenomenon in the human literature, base-rate neglect, a form of irrational and less-than-optimal behavior in the sense that overall statistical information was not used (Tversky & Kahneman, 1990). Thus, we found on the one hand in the first experiment that local statistical information was learned and used, and on the other hand, in the third experiment overall statistical information in the absence of local information was not used. These results showed that in this task, local processes controlled behavior more precisely and more rationally than overall processes.

These results from pigeons corresponded closely to those from human participants (Hunt & Aslin, 2001). Accordingly, the serial response time task appears to be a marvelously efficient procedure for the study of "anticipation" in nonhuman animals and for discovering properties of a general statistical-learning mechanism. As Froehlich et al. (2004, p. 44) concluded, "It no doubt would be asking too much to expect universal similarity in statistical learning mechanisms across an extremely wide range of species. The present results, however, encourage the view that the universality of the likelihood estimation problem animals face in nature may have generated surprisingly similar likelihood estimation mechanisms." This conclusion is virtually a defining exemplar of what "mental continuity" means, and in this case, the continuity is in terms of local features of the behavior stream.

We found these results to be so encouraging that we subsequently conducted two more experiments using the same basic serial response time task (Shimp, Froehlich, & Herbranson, 2007). Both experiments examined how incentive, in the form of anticipated reinforcement, affected information processing. The first experiment varied the overall probability of reinforcement in a task where the spatial location of the target lit key varied with equal probabilities over the three possible locations. Overall response time was a linear function of overall probability of reinforcement, in this context of a random task without any local sequential structure. So, unlike in the case in Froehlich et al., where overall base rate of a target location had no effect on response time in the random task where one target location did not provide information about the subsequent location, and overall reinforcement was held constant, overall reinforcement probability in the equal-probability version of the same task did affect overall response time. The latter outcome is an example of an overall causal relation in the absence of local contingencies. A virtue of the serial response time task is that it permits the separate manipulation of effects of local and overall statistical parameters so that this kind of overall control can be identified. Much more work obviously needs to be conducted on the generality of these two phenomena, control by overall reinforcement probability and base-rate neglect.

The second experiment further explored the extent to which the serial response time task could reveal how incentive in the form of information about anticipated reinforcement affects performance. The second experiment varied incentive in the form of likelihood of reinforcement, on an item-by-item, moment-to-moment, basis, instead of on an overall basis as in the first experiment. That is, in the second experiment, different serial positions in a nine-item list were associated with different reinforcement probabilities. We were startled by the elegance with which information theory could describe the outcome. Response time was a linear function of momentary amount of information, in accordance with the Hick-Hyman law (Hick, 1952; Hyman, 1953), implying that pigeons processed momentary statistical information at a constant rate, with more information taking longer processing time (Fig. 34.5). We think these results, combined with other avian results interpreted in terms of information theory (Vickrey & Neuringer, 2000; Young & Wasserman, 2001), strongly encourage for two reasons further examination of the dynamics of behavior streams in terms of the local processing of statistical information. First, the successful application of the Hick-Hyman law to both the linear functions relating response time to overall and local reinforcement probability is to our knowledge the first conceptual unification of overall or "molar" analyses and local or "molecular" analyses in terms of information processing. Our results suggest pigeons may process both kinds of information, local incentive and overall incentive in the form of anticipated likelihood of reinforcement, at constant rates. Second, the overall results, ours and those of others, using the serial response time task with nonhuman animals strongly encourage the use of the mental continuity idea as a heuristic to discover new similarities in mental life across species.

We are not aware of corresponding results in the human literature, where it is relatively rare to manipulate a variable such as reinforcement probability, either local or overall. Our results suggest that anticipated reinforcement probability, like any other probability, may be viewed in terms of amount of information (see equation 1 in Shimp et al., 2007). We find it particularly interesting that there may as yet be no human equivalent to the present results obtained from pigeons, because the idea that information is an incentive is not new. The current era is not infrequently referred to as the "information age," and the term "information arms race" also appears. In economic theory, it plays a large and prominent role. Indeed, the importance of this idea is such that perhaps we may be forgiven for speculating that the present results provide some important comparative insight into the workings of the

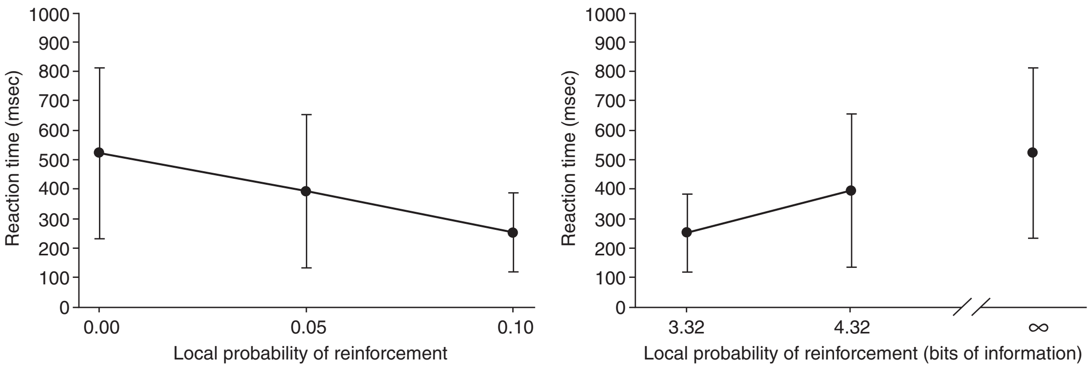

Fig. 34.5 Results from Experiment 2 of Shimp, Froehlich, and Herbranson (2007). *Left panel*: Local mean median reaction time averaged over four birds and the last 5 days of a condition, plotted as a function of the local probability of reinforcement. *Right panel*: The same data, plotted in terms of amount of information (number of bits). Information theory predicts a negative slope in the left panel and a correspondingly positive slope in the right panel.

stock market and other human activities where there is a premium placed on the value of information. Perhaps there is a similarity between an investor finding incentive to acquire information about the likelihood of a business venture succeeding or failing, and a pigeon finding incentive to acquire information about the likelihood of reinforcement. In both cases, perhaps an organism learns statistical information because it is an incentive that facilitates anticipating subsequent reinforcement.

The serial response time task might serve as a useful model for the study of fundamental activities other than economic behavior if it can reveal properties of a basic statistical learning mechanism. This mechanism has been speculated to have evolved to enable organisms to adapt to environments in which predation or predator avoidance, food availability, mating, and other basic activities require estimation of statistical likelihoods. It has been speculated that such a mechanism may have contributed to the evolution of natural language because human infants display statistical learning about sequential events with very little laboratory training.

We think the rapid speed with which pigeons learn the serial response time task suggests it may be a method of unsurpassed convenience for studying the behavior stream in terms of local sequential behaviors and dynamic mental information processing; within several dozens of trials and not too many minutes, a pigeon can be demonstrated to be learning important statistical information. In addition, we speculate the task might help us to understand the difference between implicit and explicit knowledge of sequential information. We believe it would not be too difficult to learn whether pigeons in the serial response time task "know" or are "aware of" the statistical knowledge they learn in this task. We think procedures could be developed to ask animals questions about "what they know" about the statistical information they have learned, or about what they anticipate happening next, analogous to procedures that have been developed to ask them what they know about what they have recently done (Shimp, 1984a). Thus, we think the serial response time task, along with the AGL task described above, are potentially two valuable methods for the study of implicit learning in pigeons, just as they are with human participants.

A Dynamic Interactive Systems Theory

What kind of theory can describe a stream of behavior and the moment-to-moment cognitive processes that interact with it? We chose computational-processing, computer-simulation methods as the most likely to have the required power and flexibility. Our goal was to develop a computer simulation model that, placed in control of the behavior of a suitable robot, could generate behavior streams resembling those of real experimental subjects.

Assumptions

We chose assumptions that we knew were oversimplifications but that were individually well known and generally supported by great bodies of empirical evidence. For this brief summary, we concentrate on how different versions have been basically the same, and delete the relatively minor details of how they have differed.

1."Mental representation" of a stimulus. When the "organism" defined by the theory perceives a stimulus, it samples a corresponding set of theoretical stimulus elements, and some or all of its elements are "activated." How these elements are functionally related and organized is assumed to be simpler than is probably the case. In different versions they have been assumed either to be independent with respect to the sampling, forgetting, and retrieval processes described below, or to be completely dependent in the sense that the unique pattern of activated elements acts as a memory unit.

2. Short-term forgetting (memory for recent stimuli and behavior). As soon as a stimulus is removed, its representation starts to decay in a purely time-dependent manner. A representation is subjected to decay every small unit of time. How exactly it decays has depended on the version of the theory.

3. Base rate of responding. The simulated organism is assumed to respond randomly at a low rate when it first confronts a task. This assumption ensures that the simulated organism, or simulated robot, will in fact contact the reinforcement contingency.

4. Response rule (how an organism responds per unit of time given the associative status of the representations it "experiences"). The theoretical organism is assumed to respond with a high probability if it encounters in any small unit of time a stimulus the representation of which has become associated with reinforcement. Otherwise, if the representation is not associated with reinforcement, the probability of a response is low.

5. Associative learning (how stimuli change their associative meanings as a function of

behavioral experience with reinforcement). The theoretical organism is assumed to associate the currently activated stimulus elements with reinforcement when it emits a response and is reinforced. Reinforcement simply consists of setting an entry in a table equal to one. This table keeps track of which sets of elements are associated with reinforcement and which are not. If there is more than one available response, the table of associations keeps track of which sets of elements are associated with which responses.

6. Unlearning (how stimuli change their associative meanings as a function of behavioral experience with non reinforcement). When the theoretical organism responds to a set of elements associated with reinforcement but is not reinforced, the representation's association with reinforcement is assumed to change with some probability and become associated instead with non reinforcement, in which case the corresponding entry in the table of associative memories is set to zero.

Further details of different versions may be found in Shimp (1978, 1979, 1981, 1992, 1994), Shimp, Childers, and Hightower (1990), and Shimp and Friedrich (1993). All the versions generate simulated behavior streams that can be compared to the performances of real organisms.

Theoretical Successes

This family of behaving theories has generated a wide range of performances that match real performances at least qualitatively and often quantitatively. It has learned new behavioral units consisting of temporal patterns (Shimp, 1978, 1979, 1981, 1984b), it has produced the correct functions relating these units to local and overall rates of reinforcement (Shimp, 1978, 1979, 1984c, 1994), and it has demonstrated the kind of overall "undermatching" that real organisms often display in choice situations. It has demonstrated correct switching performances in choice situations (Shimp, 1984b, 1992). It handles several kinds of outcomes in temporal psychophysics, including temporal discrimination, temporal bisection, and the constancy of the Weber fraction (Shimp, 1978, 1981). It has described various spatial attention phenomena, including the validity effect and the alerting effect and of course it learned to perform the spatial attention task in the first place (Shimp & Friedrich, 1993). And, it has explained how an organism allocates time to different behaviors, each of which takes up a different amount of time (Shimp, 1979).

In all these and several other cases, the "behaving theory" generated a behavior stream from elementary, local, dynamic processes interacting in time, and this stream when analyzed was seen to have characteristics similar to those of behavior of real organisms. As is the case with behaving theories in general (e.g., see Catania, 2005, and MacDonall, Goodell, & Juliano, 2006), "behavior" emerges from interacting basic processes and is not assumed to directly describe any theoretical process.

We think this breadth compares favorably to that of other computational processing theories, especially taking into account that the theory integrates local and overall phenomena *and* integrates animal and human phenomena (Shimp, Childers, & Hightower, 1990; Shimp & Friedrich, 1993). The theory integrates these research literatures that are so different that in several cases, one literature does not even acknowledge through cross-references the existence of the other, as in the case of temporal bisection and spatial attention literatures, or the behavioral unit literature and the literature on switching performances in choice tasks. Yet it will not take but a moment for a reader to think of additional challenges the theory should be made to face. However, we emphasize that making a theory face a challenge is not the same as "testing" it. We have not developed theories in order to test them, on the grounds that the entire deductive "hypothesis testing" and "theory testing" program is highly problematic (Benham & Shimp, 2004). We have developed them instead to show how specific performances can be conceptualized and interpreted as the outcome of more general local processes. In our judgment, the challenge our theory faces is not a "test" but how it can be revised to retain a set of simple assumptions while handling a broader range of the phenomena we have described here, including the differences between implicit and explicit performances (self-reports of one's own performances), categorizations in the randomization task, AGL, and performances in the serial response time task. We think an inspiring goal is to develop a computational-processing, computer-simulation, behaving theory that would conceptually integrate these diverse phenomena, clarify the causal mechanisms underlying the behavior stream, and more fully legitimatize behavior analysis and comparative cognition as sciences.

As we indicated above, we believe there is great need to facilitate the development of "behaving theories" that address the continuity of behavior in time (Shimp, 1992, 2009), and as outlined above, we have tried over the years to do so. In closing, we briefly describe a few of the issues that we believe are critical for continued progress on "behaving theories."

First, we suggest that neither mathematical approaches involving closed-form expressions nor connectionist approaches will suffice to characterize the continuity of behavior streams. We think the former is too methodologically constraining, although that opinion might merely reflect our own limited mathematical expertise. We suspect the latter is too theoretically constraining and that it will become too computationally unwieldy when it is required to deal with the continuity of a stream of any sizeable temporal duration, meaning of as little as a few seconds. If this approach is to succeed, we suspect it will be due to the adoption of an iterative approach to determine what a connectionist theory predicts, similar to how we have used iteration in our computer simulations to discover how the theory described above behaves. As to self-organizing systems theory, fractals, and other approaches, we would like to think of ourselves as open-minded but confess to being perhaps a bit old-fashioned; we prefer to base the functional components of a computer model on known cognitive, behavioral, or neurophysiological processes rather than assume they are the same as those in ecosystems, astrophysics, and viruses. The grandiose nature of claims for such universality across different scientific disciplines can feel thrilling but in our judgment is likely to lead to disappointment.

Second, we suggest that it will prove difficult but essential to conceptually unify the continuity of behavior and of mental life, on the one hand, with the discontinuity produced by segmenting and chunking behavior streams into successive behavioral units, on the other hand. We expect future progress in behavioral science to repeat analogous previous progress in the more established sciences, once there is better understanding of how behavioral units emerge and interact continuously over time.

Finally, local and global analyses should not be thought of in any sense as on different "levels" corresponding to the difference between physics and chemistry, or the difference between chemistry and biology, unless the processes and standards that define different levels in these other sciences are first shown to apply to behavior (Shimp, 2009). The mean of a distribution of numbers does not emerge from the distribution in the manner in which wave phenomena emerge from water molecules, so that the overall mean rate of responding is not an emergent phenomenon on some level different from local response rates. We think it is reasonable to reverse the usual line of thinking that an overall empirical outcome stands on its own unless it can be shown that a local explanation is needed. We think it might as well be the other way around: If overall empirical results are to be taken seriously as complete, self-contained accounts of behavior, they need to be shown not to be derived from local phenomena. Ideally, when local effects are described, it would be helpful if they were linked to overall ones, and when overall effects are described, it would be helpful if it were acknowledged how they might derive from local ones. Thinking in terms of the continuity of behavior streams might help remind researchers of the essential need to know how to methodologically and theoretically unify these kinds of analyses.

References

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making: Handbook of perception and cognition* (pp. 251–301). San Diego: Academic Press.

Benham, B., & Shimp, C. P. (2004). Falsification in social science method and theory. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (Vol. 2, pp. 9–14). San Diego: Academic Press.

Blough, D. S. (2006). Reaction-time explorations of visual perception, attention, and decision in pigeons. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition; Experimental explorations of animal intelligence* (pp. 89–105). Oxford: Oxford University Press.

Casseday, J. H., Fremouw, T., & Covey, E. (2002). The inferior colliculus: A hub for the central auditory system. In D. Oertel, R. R. Fay, & A. N. Popper (Eds.), *Integrative functions in the mammalian auditory pathway* (pp. 238–318). New York: Springer-Verlag.

Catania, A. C. (2005). The operant reserve: A computer simulation in (accelerated) real time. *Behavioural Processes*, *69*, 257–278.

Chomsky, N. (1959). Review of Skinner's Verbal Behavior. *Language*, *35*, 26–58.

Chomsky, N., & Miller, G. A. (1958). Finite-state languages. *Information & Control*, *1*, 91–112.

Epstein, R., Lanza, R. F., & Skinner, B. F. (1980). Symbolic communication between two pigeons (*Columba livia domestica*). *Science*, *207*, 543–545.

Fremouw, T., Herbranson, W. T., & Shimp, C. P. (1998). Priming of attention to local or global levels of visual analysis. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 278–290.

Fremouw, T., Herbranson, W. T., & Shimp, C. P. (2002). Dynamic shifts of avian local/global attention. *Animal Cognition*, *5*, 233–243.

Froehlich, A. L., Herbranson, W. T., Loper, J. D., Wood, D. M., & Shimp, C. P. (2004). Anticipating by pigeons depends on local statistical information in a serial response time task. *Journal of Experimental Psychology: General*, *133*, 31–45.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, *16*, 486–490.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*, 1204–1207.

Hawkes, L., & Shimp, C. P. (1975). Reinforcement of behavioral patterns: Shaping a scallop. *Journal of the Experimental Analysis of Behavior*, *23*, 3–16.

Hawkes, L., & Shimp, C. P. (1998). Linear responses. *Behavioural Processes*, *44*, 19–43.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (1999). The randomization procedure in the study of categorization of multi-dimensional stimuli by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 113–135.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (2002). Categorizing a moving target in terms of its speed and direction. *Journal of the Experimental Analysis of Behavior*, *78*, 249–270.

Herbranson, W. T., & Shimp, C. P. (2003). Artificial grammar learning in pigeons: A preliminary analysis. *Learning & Behavior*, *31*, 98–106.

Herbranson, W. T., & Shimp, C. P. (2008). Artificial grammar learning in pigeons. *Learning & Behavior, 36*, 116–137.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267–273.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial response time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*, 658–680.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *65*, 39–46.

James, W. (1890). *The principles of psychology*. New York: Henry Holt.

Jen, P. H., Chen, Q. C., & Sun, X. D. (1998). Corticofugal regulation of auditory sensitivity in the bat inferior colliculus. *Journal of Comparative Physiology* [A], *183*, 683–697.

Ma, X., & Suga, N. (2003). Augmentation of plasticity of the central auditory system by the basal forebrain and/or somatosensory cortex. *Journal of Neurophysiology*, *89*, 90–103.

MacDonall, J. S., Goodell, J., & Juliano, A. (2006). Momentary maximizing and optimal foraging theories of performance on concurrent VR schedules. *Behavioural Processes*, *72*, 283–299.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.

Pepperberg, I. M. (2000). *The Alex studies: Cognitive and communicative abilities of grey parrots*. Cambridge: Harvard University Press.

Porter, D., & Neuringer, A. (1984). Music discriminations by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*, 138–148.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, *6*, 855–863.

Rumbaugh, D. M., & Washburn, D. A. (2003). *Intelligence of apes and other rational beings*. New Haven: Yale University Press.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.

Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, language, and the human mind.* Oxford: Oxford University Press.

Schoenfeld, W. N., & Farmer, J. (1970). Reinforcement schedules and the behavior stream. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 215–245). New York: Appleton-Century Crofts.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*, 553–554.

Shimp, C. P. (1966). Probabilistically reinforced choice behavior in pigeons. *Journal of the Experimental Analysis of Behavior*, *9*, 443–455.

Shimp, C. P. (1968). Magnitude and frequency of reinforcement and frequency of interresponse times. *Journal of the Experimental Analysis of Behavior*, *11*, 525–535.

Shimp, C. P. (1973). Synthetic variable–interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *19*, 311–330.

Shimp, C. P. (1976a). Short-term memory in the pigeon: Relative recency. *Journal of the Experimental Analysis of Behavior*, *25*, 55–61.

Shimp, C. P. (1976b). Organization in memory and behavior. *Journal of the Experimental Analysis of Behavior*, *26*, 113–130.

Shimp, C. P. (1978). Memory, temporal discrimination, and learned structure in behavior. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 39–76). New York: Academic Press.

Shimp, C. P. (1979). The local organization of behavior: Method and theory. In M. D. Zeiler & P. Harzem (Eds.), *Advances in analysis of behavior, Vol. 1: Reinforcement and the organization of behavior* (pp. 262–298). Chichester, England: Wiley.

Shimp, C. P. (1981). Local structure of steady-state operant behavior. In C. M. Bradshaw, E. Szabadi, & C. F. Lowe (Eds.), *Quantification of steady-state operant behavior* (pp. 189–203). Amsterdam: Elsevier/North Holland.

Shimp, C. P. (1983). The local organization of behavior: Dissociations between a pigeon's behavior and self-reports of that behavior. *Journal of the Experimental Analysis of Behavior*, *39*, 61–68.

Shimp, C. P. (1984a). Self reports by rats of the temporal patterning of their behavior: A dissociation between tacit knowledge and knowledge. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 215–229). Hillsdale, NJ: Erlbaum.

Shimp, C. P. (1984b). Timing, learning and forgetting. In J. Gibbon & L. Allan (Eds.), *Timing and time perception* (Vol. 423, pp. 346–360). New York: New York Academy of Sciences.

Shimp, C. P. (1984c). Relations between memory and operant behavior, according to an associative learner (AL). *Canadian Journal of Psychology*, *38*, 269–284.

Shimp, C. P. (1990). Theory evaluation can be unintentional self-portraiture: A reply to Williams. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 217–221.

Shimp, C. P. (1992). Computational behavior dynamics: An alternative description of Nevin (1969). *Journal of the Experimental Analysis of Behavior*, *57*, 289–299.

Shimp, C. P. (1994). Computational behavior and behavior analysis: An interpretation of Catania and Reynolds (1968). In E. Ribes Inesta (Ed.), *B. F. Skinner, In memoriam* (pp. 69–83). Guadalajara, Mexico: University of Guadalajara Press. [In Spanish]

Shimp, C. P. (2001). Behavior as a social construction. *Behavioural Processes*, *54*, 11–32.

Shimp, C. P. (2009). “Behavior streams” versus “behavior extended in time.” *Behavior & Philosophy*, *37*, 157–163.

Shimp, C. P., Childers, L. J., & Hightower, F. A. (1990). Local patterns in human operant behavior and a behaving model to interrelate animal and human performances. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 200–212.

Shimp, C. P., & Friedrich, F. J. (1993). Behavioral and computational models of spatial attention. *Journal of Experimental Psychology: Animal Behavioral Processes*, *19*, 26–37.

Shimp, C. P., Froehlich, A. L., & Herbranson, W. T. (2007). Information processing in pigeons: Incentive as information. *Journal of Comparative Psychology*, *121*, 73–81.

Shimp, C. P., Long, K. A., & Fremouw, T. (1996). Intuitive statistical inference: Categorizations of binomial samples depends on sampling context. *Animal Learning & Behavior*, *24*, 82–91.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Suga, N., Gao, E., Zhang, Y., Ma, X., & Olsen, J. F. (2000). The corticofugal system for hearing: Recent progress. *Proceedings of the National Academy of Science USA*, *97*, 11807–11814.

Terrace, H. S. (1979). *Nim*. New York: Knopf.

Tversky, A., & Kahneman, D. (1990). Judgment under uncertainty: Heuristics and biases. In P. K. Moser (Ed.), *Rationality in action: Contemporary approaches* (pp. 171–188). New York: Cambridge University Press.

Vickrey, C., & Neuringer, A. (2000). Pigeon reaction time, Hick’s law, and intelligence. *Psychonomic Bulletin & Review*, *7*, 284–291.

Voltaire. (1959) *Candide* (Bair, Lowell, Ed.). New York: Bantam Dell. Originally published 1759.

Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons’ discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, *63*, 165–174.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

Young, M. E., & Wasserman, E. A. (2001). Entropy and variability discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27,* 278–293.

Zeigler, H. P., & Marler, P. (2008). *Neuroscience of birdsong*. Cambridge: Cambridge University Press.

Zhou, X., & Jen, P. H. S. (2000). Brief and short-term corticofugal modulation of subcortical auditory responses in the big brown bat, *Eptesicus fucus*. *Journal of Neurophysiology*, *84*, 3083–3087.