NIH-PA Author Manuscript
NIH-PA Author Manuscript

Published in final edited form as:
Behav Processes. 2010 February ; 83(2): 139–153. doi:10.1016/j.beproc.2009.12.003.

Defining the Stimulus - A Memoir

Herbert Terrace
Columbia University & New York State Psychiatric Institute

Abstract

The eminent psychophysicist, S. S. Stevens, once remarked that, "the basic problem of psychology was the definition of the stimulus" (Stevens, 1951, p. 46). By expanding the traditional definition of the stimulus, the study of animal learning has metamorphosed into animal cognition. The main impetus for that change was the recognition that it is often necessary to postulate a representation between the traditional S and R of learning theory. Representations allow a subject to *re*-present a stimulus it learned previously that is currently absent. Thus, in delayed-matching-to-sample, one has to assume that a subject responds to a representation of the sample during test if it responds correctly. Other examples, to name but a few, include concept formation, spatial memory, serial memory, learning a numerical rule, imitation and metacognition. Whereas a representation used to be regarded as a mentalistic phenomenon that was unworthy of scientific inquiry, it can now be operationally defined. To accommodate representations, the traditional discriminative stimulus has to be expanded to allow for the role of representations. The resulting composite can account for a significantly larger portion of the variance of performance measures than the exteroceptive stimulus could by itself.

> "There is only one problem in psychology and that is the definition of the stimulus." (Stevens, S.S., 1951)

During the past 50 years, we have witnessed a remarkable integration of operant conditioning and animal cognition. As one of B. F. Skinner's graduate students, I was initially wary of that integration because I was taught that any mention of operant conditioning and animal cognition in the same breath was an oxymoron. Skinner eschewed all references to cognitive processes on the grounds that they were unobservable and that they were always rooted in observable behavior. That made them superfluous as explanations of behavior.

Like many of Skinner's students, I swallowed his approach to psychology with uncritical enthusiasm. It was hard not to given the large vocabulary of explanatory concepts he introduced e.g., contingencies and schedules of reinforcement, generalization, the discriminative operant, superstition and verbal behavior. Skinner and his students also showed how the principles of operant conditioning could be applied to behavioral therapy (Bandura, 1977), teaching children to acquire various kinds of knowledge with the help of a teaching machine (Skinner, 1959), training pigeons to play ping-pong ("Pigeon ping pong clip," 1987) training pigeons to guide missiles (Skinner, 1960), training concepts (Blough, 1984; Herrnstein, 1985; Cook, Wright, et al., 1988) etc.

Skinner and his students showed how the discriminative operant could serve as the basic unit in the experimental analysis of behavior. It consisted of 3 items: S^D: R → $S^{\mathbf{R/r}}$. The first is a discriminative stimulus. Unlike a stimulus that elicits a response (the CS in classical

---

conditioning), a discriminative stimulus simply sets the occasion for reinforcement. Responses made in its presence (S+) are reinforced; those made in its absence (S-) are not. The 2nd item of a discriminative operant is a voluntary response—a response that is not elicited and that is sensitive to its outcome. The 3rd item is reinforcement, either primary or secondary.

The discriminative stimulus has exquisite flexibility. It could take the form of a spoken or written word, the position of an opposing basketball player under the basket, a green traffic light, a proprioceptive stimulus that is used unconsciously to maintain a certain posture, a smile, an exemplar of a concept, a signal from a catcher to the pitcher, a musical phrase, etc. The reinforcer, could take the form of a raise in salary, a cookie, a smile, the elimination of an aversive noise, a grade of A, the postponement of an exam, etc. Both types of reinforcer could be presented after each correct response or on some intermittent schedule.

Questions about the definition of a discriminative stimulus that, according to Skinner, could only include exteroceptive stimuli, arose even in the research I performed as a graduate student[1]. Indeed, such questions came up with increasing regularity during the course of my research, all of which raised questions about the nature of the controlling stimulus: errorless discrimination learning (Terrace, 1963), auto-shaping (Terrace, Gibbon, et al., 1975), ape language (Terrace, Petitto, et al., 1979) the serial memory of arbitrary and numerical stimuli (Son, Kornell, et al., 2003; Terrace, Son, et al., 2003), imitation (Subiaul, Cantlon, et al., 2004) and metacognition (Kornell, Son, et al., 2006). As a result, I began to rely more and more on terms from the “new cognitive psychology”, notably, the concept of a representation (Gardner, 1985)[2]. This shift in my outlook did not go unnoticed by Skinner who, in 1983, sent me a stern letter saying that, “You've finally come out of the closet, you're a cognitive psychologist after all”.

Defining the Stimulus

The route I took from operant conditioning to animal cognition began innocently enough when I was a first-year as graduate student at Harvard. Like all new students, I was required to take the pro-seminar given by S. S. Stevens, who was arguably the most prominent psychophysicist of the 20th century. Our first assignment was Stevens' chapter on measurement, which appeared at the beginning of his classic text the *Handbook of Experimental Psychology* (Stevens, 1951). One sentence from that chapter, which remains etched in my memory, turned out to be as true today as it was then: “There is only one problem in psychology and that is the definition of the stimulus.” At the time, I assumed that Stevens' maxim was intended only for psychophysicists, in particular, psychophysicists interested in the multi-dimensional properties of stimuli[3]. For example, in one of Stevens' best-known experiments, he showed that judgments of pitch, while mainly influenced by the auditory frequency of the stimulus, were also influenced by auditory intensity (Stevens & Davis, 1938).

---

[1]For Skinner, external stimuli meant external to the CNS. On that view, proprioceptive and interoceptive qualify as external stimuli.

[2]The new cognitive psychology was, to a large extent, inspired by the widespread use of computers after WW II. Various stages of information processing occurred between the input and output of a computer, terms that are analogous to the S and the R of operant conditioning. In many instances, the transformation of information within a computer was functionally similar to psychological processes postulated by psychologists, e.g., memory, perception and decision making to name a few. Although there were clear differences between the mechanisms of a computer and psychological processes, computer mechanisms were nonetheless real and could not be dismissed as subjective.

[3]It should be noted that the concept of the discriminative stimulus fits very well with Stevens' emphasis on the definition of the stimulus. Examples include experiments on sensory processes (Stevens & Davis, 1938), generalization (Terrace, 1966) and concept formation (Herrnstein, Loveland, et al., 1976).

The Influence of Herb Jenkins

It didn't take long before I began to see the wisdom behind Stevens' characterization of the basic problem of psychology. After my first year as a graduate student, I spent the summer working for Herb Jenkins at Bell Telephone Labs in Murray Hill, New Jersey. Those of you who were lucky enough to work with Jenkins will appreciate his sharp and skeptical mind and his ability to recognize phenomena that would often be missed by others. I will digress briefly by describing one such example of Jenkins' brilliance in an experiment on auditory stimulus control in pigeons, an experiment that also illustrates the importance of defining the stimulus. Jenkins' experiment was modeled after Norm Guttman's classic experiment on wavelength generalization in which he trained responses to S+ on an intermittent schedule of reinforcement (Guttman and Kalish, 1956) and then presented test stimuli of other wavelengths in extinction[4]. As shown in Figure 1, Guttman obtained generalization gradients that were symmetrical and centered around S+.

In an attempt to replicate Guttman's experiment with an auditory stimulus, Jenkins trained pigeons to respond to a blank key in the presence of a 1000 Hz tone. As shown in Figure 2, the gradient he obtained was flat (Jenkins & Harrison, 1960).

Why the difference in generalization gradients obtained following training on visual and auditory stimuli? Jenkins attributed the flat gradient to the fact that pigeons rarely, if ever, learn to discriminate the presence or the absence of a tone. Accordingly, Jenkins first trained pigeons to discriminate the presence and absence of a 1000 Hz tone. Lo and behold, he was then able to obtain a symmetrical generalization gradient with a peak at 1000Hz. *Homework question:* Did the definition of the S+ as a 1000 Hz tone remain the same after the discrimination between the presence and the absence of S+ was trained?

Discrimination Learning With and Without Errors

During the summer I worked with Jenkins, he was investigating factors that influenced resistance to extinction (Jenkins, 1961). While intermittent reinforcement was regarded as the major factor that influenced resistance to extinction, Jenkins persuaded me that the extinction of responding to S- during the acquisition of discrimination might also be a factor. When I returned to Harvard, I began an experiment in which I sought to test Jenkins' hypothesis by varying the degree of unreinforced responding to S- during discrimination training. The idea was to compare the acquisition of one group of pigeons on a relatively difficult color discrimination in which, throughout training, S+ was red and S- was green, with the acquisition of a second group in which S- was initially much briefer and much dimmer than S+ (Terrace, 1963).

By slowly decreasing the large difference between S+ and S-, I expected to reduce the number of errors that would normally occur to S-. After training a response to S+ (red), I introduced S- as a brief (1 sec) exposure of a black stimulus and then gradually increased its duration until it reached a value of 30 sec. During the 2nd phase of training I slowly added color to S- by increasing its brightness until it was a fully saturated green. This method worked beyond my wildest expectations. Out of 3 pigeons, the range of responses to S- was 5-9. By contrast, the range of the 3 pigeons that learned the same discrimination, without the benefit of the fading procedure, was 1922-4153 errors.

---

[4]In previous experiments on generalization, subjects were rewarded for every response they made to S+, during training. Little responding occurred during the generalization test because of the abrupt shift from reward on every trial to extinction. Intermittent reinforcement resulted in a substantial increase in responding, during extinction. That added to the reliability of the data obtained during the generalization test.

In a related experiment, I used the fading method to train pigeons to discriminate white vertical (S+) and horizontal lines (S-) that were presented against a black background. All of the subjects in this experiment had previously learned to discriminate a red S+ and a green S-. During this experiment, however, one group of pigeons began discrimination training with composite stimuli. S+ was first presented as a white horizontal line superimposed on a red background, and S- as a white horizontal line superimposed on a green background. Unsurprisingly, subjects continued to respond correctly to S+ and S- when they were superimposed on red and green backgrounds. During successive trials, the brightness of the red and green backgrounds was gradually diminished until the white vertical and horizontal lines appeared on a black background. For a second group of subjects, S+ and S- remained constant throughout discrimination training. Again the results were surprising. Neither of the 2 pigeons trained by the fading procedure made any errors. The range of the errors made by the 2 subjects that didn't have the benefit of fading, was 404-2609 while learning the same discrimination.

By-products of Discrimination Learning

In other experiments, I showed that certain “by-products of discrimination learning” (Terrace, 1972), such as behavioral contrast (Terrace, 1964, 1968) the peak shift (Terrace, 1964) and inhibition in the presence of S- (Terrace, 1966), did not occur if pigeons learned to discriminate S+ & S- without errors. As shown in Figure 3A, behavioral contrast is an increase in the rate of responding to S+ during discrimination learning re. It's baseline value during training in which responding to S+ and S- resulted in the same frequency of reinforcement. Behavioral contrast has been interpreted as a by-product of the frustration that the pigeon experiences while responding to S-. As shown in Figure 3B, the peak shift refers to a displacement of the maximum rate of responding in a direction away from its usual value (S+), to a test stimulus whose distance from S- is greater than its distance from S+. Thus, if S+ is a 550 nm stimulus (green), and S- is 570 nm (yellow), the peak occurs in the vicinity of 540 nm (blue). That displacement has been attributed to a gradient of inhibition that is established around the value of S-.

Another clever insight by Jenkins paved the way for a procedure that provided direct evidence of inhibition following discrimination training. Instead of using stimuli from the same dimension, e.g., 550 vs. 570 nm or 1000 Hz vs. 1300 Hz, stimuli were selected from different dimensions (Jenkins & Harrison, 1962). That allowed S+ and S- to be varied independently of one another. Consider, for example, a discrimination in which S+ is a green stimulus, say, 555nm, and S- is a vertical white line presented against a black background. After this discrimination was trained, generalization tests were administered along each dimension. As expected, the gradient based on stimuli that varied in hue, was centered at 555 nm. The gradient based on lines of different orientations, -180°→ +180°, was centered at 90°, where responding was minimal. The relevant data are shown in Figure 3C.

Jenkins' measure of inhibition raised an interesting question. What type of inhibitory gradient would be obtained following discrimination learning without errors? Unsurprisingly, pigeons that learned the discrimination with errors responded with increasing frequency to test stimuli as the distance between S- and the test stimuli was increased. By contrast, the U-shaped function was absent after discrimination learning without errors. The frequency of responding to S- and the test stimuli was the same. The flat gradient implied that inhibition to S- occurred only after discrimination training with errors.

That result fit very nicely with the Spence-Hull theory of discrimination learning, which postulated an excitatory gradient centered at S+ and an inhibitory gradient centered at S-[5]. As shown in Figure 4, the peak of the excitatory gradient, which is normally centered at S+, is displaced away from S- when the inhibitory gradient is subtracted from the excitatory gradient. Unfortunately, the Spence-Hull theory does not predict behavioral contrast. Indeed it predicts that the frequency of responding to S+ would decrease.

These results were only partially confirmed in replications by Rilling and his colleagues (Kendrick, Rilling, et al., 1986). Their data suggested that errorless learning might not be as qualitatively different from conventional training as I claimed initially. For example, Rilling and his colleagues demonstrated that by-products of discrimination learning can occur after errorless learning, but that they were not as large as those obtained following the conventional training. Their results led to a decline in interest about errorless learning during the early 70's[6,7]. However, errorless learning did attract the interest of researchers in applied psychology, e.g., numerous studies that have been conducted with children in educational settings (Skinner, 1959; Crist, 1969) and adults suffering from Parkinson's disease (Connor, Wing, et al., 2002; Clare & Jones, 2008).

Auto-Shaping

Jenkins' discovery of auto-shaping influenced yet another phase of my research. Jenkins showed that pigeons would start pecking at a stimulus that was presented approximately every 15 seconds, even though reinforcement was not contingent upon the pigeon's response (Brown & Jenkins, 1968). For obvious reasons, Jenkins referred to this phenomenon as auto-shaping. To investigate this phenomenon, John Gibbon, and I performed an experiment in which we varied the inter-trial interval (Gibbon, Baldock, et al., 1977). What we found was that the time needed for the acquisition of the key-peck increased as the value of the ratio of trial duration to the duration of the inter-trial duration increased. The linear function we obtained turned out to be the basis of the scalar timing theory that was developed by Gibbon and Church (Gibbon, 1977; Gibbon, Church, et al., 2006).

Shortly after Jenkins' discovery of auto-shaping, David Williams demonstrated a surprising effect of omission training on an auto-shaped response (Williams & Williams, 1969). In omission training, reward was given for *not* responding to a particular stimulus. Williams found that omission training did little to discourage responding. Indeed pigeons missed as many as 90% of the reinforcers they would have received if they *didn't* respond when a stimulus appeared on the response key. Contrary to Jenkins' view that auto-shaping was a simpler method of training a pigeon to peck at a response key than the method of rewarding successive approximations to a key peck, it turned out that an auto-shaped response was a respondent, not an operant.

The phenomenon of auto-shaping turned out to be an embarrassment for operant conditioning. During auto-shaping, the stimulus elicits, rather than occasions, the conditioned response. A physical description of the stimulus used during auto-shaping is woefully inadequate as an explanation of the auto-shaped response. At the very least a definition of that stimulus would have to include its function (eliciting vs. occasioning) and the mechanism that gives rise to an auto-shaped response. Though hardly exhaustive, these examples should make clear that the physical value of an external stimulus is often incomplete as a definition of the controlling stimulus.

---

[5]The Spence-Hull model assumes that the maximum of the excitatory gradient is higher than the maximum of the inhibitory gradient (Spence, 1937).

[6]When attempting to replicate an experiment about errorless learning, it is difficult to reproduce the fading procedure very accurately because it is based on the experimenter's intuition about when to change the value of S-. Data about such changes are virtually impossible to obtain without videotapes that showed the behavior of the pigeon before it was given a reward.

[7]It's unclear why a relative difference in performance following discrimination learning with and without errors is less interesting than an absolute difference. Indeed a relative difference might be more interesting than an absolute difference because it reduces the number of variables one would have to contend with while trying to model the difference between discrimination learning with and without errors.

Taking Stock

At first glance, Skinner's imaginative use of the concept of a discriminative stimulus fit very well with Stevens' emphasis on searching for the stimulus that is the best predictor of a particular response. In the spirit of dustbowl” empiricism, an approach that dominated psychology during the first half of the 20th century (Hanson, 1959; Hansen, 1998; Schoenfeldt, 1999)[8], both psychologists limited their search to exteroceptive stimuli,

Generalization gradients provide a good example of this type of S-R psychology. A generalization gradient shows the frequency of responding to S+ and to various test stimuli. As in psychophysics, the underlying assumption is that the differences in the observed frequencies are fully determined by the physical specification of the stimulus; in this case, the wavelength of each stimulus and their distance from S+[9,10].

In the case of behavioral contrast, in which the frequencies of responding to S+ and S- diverge from their baseline values, the most widely accepted explanation is the frustration that a subject experiences while responding to S- (Amsel, 1958; Amsel & Ward, 1965). Evidence of frustration can be seen in the behavior of pigeons during S-, e.g., agitation and wing flapping (Terrace, 1971). Frustration, however, is a variable that cannot be varied independently of S-, suggesting that Skinner's explanation of discrimination performance is limited because it doesn't take into account any physiological factors.

The peak shift poses a slightly different problem (Hanson, 1959). We have seen how the Spence-Hull model of discrimination predicts a peak shift but that, in addition to the displacement of the psychological values of S+ and S-, behavioral contrast is needed to account for the elevated level or responding to S+ and adjacent stimuli. These problems vanish when the S-R unit of analysis is expanded to include another factor, the organism (O). The result is S-O-R psychology, where O is any stimulus inside the organism (Woodworth, 1924)[11].

Project Nim

Towards the end of the '60s, I was distracted from research with pigeons by the reports of two projects that claimed that chimpanzees could learn language. Allan and Beatrice Gardner announced that they had trained a female chimpanzee, which they named Washoe, to learn signs of American Sign Language (ASL) (Gardner & Gardner, 1969). David Premack reported that he trained a female chimpanzee named Sarah to learn a vocabulary composed of plastic chips that varied arbitrarily in shape, color and size (Premack, 1971). There was, however, a sub-text in these reports. Once you can teach a non-human primate a vocabulary, you could attempt to teach it to use a grammar.

The Gardners and Premack's reports resonated strongly with my interest in evolution, an interest that developed in courses in biology that I took as an undergraduate at Cornell

---

[8]Dustbowl empiricism's only concern is data, “Just the facts ma'am, only the facts”. It emphasizes the collection of data and eschews theory. The only theory the results from dustbowl empiricism are patterns in the data that can replace the raw data, e.g., Steven's power law (Stevens and Galanter, 1957)

[9]The schedule of reinforcement during training to respond to S+, continuous or intermittent, is also a factor. But, although intermittent schedules of reinforcement increase the frequency of responding during a generalization test, it does not do so differentially with respect to S+ or S-.

[10]The book, *Schedules of Reinforcement* (Ferster & Skinner, 1957), is a good example of dustbowl empiricism. It contains hundreds of pages of cumulative records that are presented uncritically for their own sake. Interpretations are limited to metaphors, e.g., responding on a fixed ratio is like doing piecework; on a fixed interval schedule, like waiting for a train, etc.

[11]S-O-R psychology was advocated by Woodworth, an eminent psychologist whose broad knowledge of psychology enabled him to write the last handbook of psychology that was authored by a single individual (Woodworth 1938). Woodworth's point was simple. Why end the definition of a stimulus at the subject's skin? If additional information (cognitive and/or physiological) adds the predicted power of S, why not incorporate it into S?

University. As a psychology major, it slowly dawned on me that psychology was the only life science that hadn't fully assimilated the theory of evolution and that the main obstacle was the human mind—language, in particular. Many people were willing to accept the argument that nonhumans evolved according to the Darwinian principles of “descent with modification” as honed by “natural selection” (Darwin, 1859), but they balked at applying those principles to something as complex as the human mind. Two major exceptions were Pavlov's and Watson's explanations of language as a form of higher order conditioning (Benjamin, 2006). However, neither Pavlov nor Watson explored the mechanisms needed to learn vocabulary and to use a grammar. I was therefore pleasantly surprised, when I read Skinner's *Verbal Behavior* (1957), in which he described, in considerably detail, how the concept of the discriminative operant could account for the acquisition of vocabulary[12]. Less clear was Skinner's treatment of grammar. Indeed, it was Skinner's attempt to use chaining theory to account for grammar that was the focus of Noam Chomsky's devastating review of *Verbal Behavior*” (Chomsky, 1959)[13].

If vocabulary training were the only accomplishment of the chimp language projects, that news would have been followed by a yawn. After all, pigeons can be trained to acquire “words” by the same operant techniques that were used to train the apes[14]. In both instances the result was a vocabulary whose only function was a means to obtain some basic reward, e.g., a banana, an apple, a game of tickle, etc. Those words were examples of what Elizabeth Bates referred to as “proto-imperatives” (Bates, 1976). Totally lacking in an ape's vocabulary were words whose function was conversational. Such words, which Bates referred to as “proto-declarative,” functioned as a means for a speaker and a listener to exchange information[15].

The amazing claim of these projects was that apes could not only learn “words”, but that they could also combine words from their vocabulary to create particular meanings (Gardner & Gardner, 1969). That claim was significant for two reasons. Descartes and other philosophers have argued that language is exclusively human (Descartes, 1637). More recently, Noam Chomsky (Chomsky, 1968), who is arguably the greatest linguist of the 20th century, developed descriptive grammars that made explicit what counts as human language. As opposed to a formal, *prescriptive* grammar that students learn in elementary school (one that tells them not to say “ain't” and, more generally, rules for the proper way to combine words), a *descriptive* grammar is an objective, nonjudgmental description of the rules for generating sentences. It shows how words, proper or not, relate to other words in a sentence.

I shared with the Gardners the goal of training a chimpanzee to learn a large vocabulary of signs and then asking whether it could combine those signs to generate different meanings. That approach was consistent with the general view of psycholinguists that human language makes use of two levels of structure: the word and the sentence. In contrast to the fixed character of various forms of animal communication, e.g., bird songs that function as mating or “stay-out-my-territory” calls, or bees that perform “dances” that specify the location of a food source and its distance from the hive, the meaning of a word is arbitrary. Also, because of the fixed functions of these messages, it is not possible to add new meanings such as “meet me at the tall tree,” “chase me”, or “blue flowers have a sweeter nectar than red flowers,” etc.

---

[12]The concept of the discriminative operant was modified for word acquisition so that the response to a word that was presented as a spoken or written stimulus produced a secondary rather than a primary reward. Thus, in the presence of a dog, the child said “doggie” and a parent said “good”. That exchange would be represented as a discriminative operant: physical dog: “doggie” → “good”. Noticeably lacking in a chimpanzees vocabulary were words that would support a conversation; words that a speaker and a listener would use to comment about someone from their group or some object in their environment.

[13]Chomsky also attacked Skinner's treatment of vocabulary but, as far as I'm aware, no one else, Chomsky included, has formulated a better one.

[14]A seldom-appreciated fact about all of the ape language projects is the use of intermittent primary reinforcers such as food, drink, and tickle games, etc. as rewards for uttering particular words. The implications of that practice will be discussed below.

[15]As far as I'm aware, proto-declaratives have never been observed in non-human primates or, for that matter, in any non-human animal.

No such restrictions exist in the case of human languages. It matters not if you refer to a particular color as *red*, *rouge*, *roit*, or any of the thousands of equivalents that can be found in other languages. However, the range of meanings of individual words pales in comparison to the essentially innumerable meanings that can be created by combining words. Chimpanzees and other animals may share with humans the ability to learn individual "words" but there was no evidence that chimpanzees use words declaratively[16]. Most important was the lack of evidence that a non-human primate could produce and understand sentences—an ability that requires knowledge of a grammar, a second level of structure that specifies the rules for combining words to create particular meanings. Chomsky and others have argued that it is grammatical knowledge that creates an unbridgeable gap that separates human and animal communication.

Washoe's most famous utterance was described by Roger Fouts, one of Washoe's main trainers, in an early diary report (Fouts, 1975). Washoe reportedly signed *water bird* after Fouts, asked her "What's that?" in the presence of a swan. What made this observation even more remarkable was the fact that Washoe didn't have signs for specific water birds such as swans and ducks. It therefore seemed to Fouts that Washoe invented a way of conveying a combination of signs for what she saw. Fouts' report prompted Roger Brown, a highly respected psycholinguist, to comment, "It was rather as if a seismometer left on the moon had started to tap out 'S-O-S'" (Brown, 1970).

I was also surprised when I read Fouts' report, but I felt skeptical about the Gardners' claim that a chimpanzee was using a grammar to generate sentences (Gardner & Gardner, 1969). I thought that any of four simpler interpretations of *water bird* could apply. 1. Washoe may have been prompted by Fouts to sign *water bird*. 2. Washoe may have signed *bird water*, but Fouts may have recorded Washoe's utterance in the order in which we naturally speak. 3. Washoe may have signed *water* and *bird* as two separate utterances. 4. Fouts had previously trained Washoe to make the signs *water* and *bird* for food reward before the swan appeared. Washoe may have therefore signed *water bird* for more food reward without any specific understanding of what the signs *water* or *bird* referred to.

I decided that the only way to eliminate such killjoy interpretations of the "sentences" that chimps were purported to have produced was to start my own project. The goal of that project was to record as many of the chimp's signs *and* the contexts in which they occurred. My plan was to require each of Nim's teachers to whisper that information into small dictating machines during the time they spent with him.

In December 1973, I acquired a 2-week-old infant chimpanzee from a chimpanzee colony in Oklahoma and named him Nim Chimpsky. With the help of a family who lived on the upper West Side of Manhattan, Nim was raised in an environment in which he was taught to use ASL. Why sign rather than vocal language? The human and chimpanzee vocal apparatuses differ significantly (Lieberman, 1968). Accordingly, a chimpanzee's inability to articulate human sounds might explain earlier, failed attempts to teach chimpanzees English or Russian by imitation (Hayes, 1951; Ladyna-Kots, 1935). The idea was to use a human language that didn't depend on the human vocal apparatus.

By the late 1970s, much evidence had accumulated purporting to show that apes could create sentences—specifically, that an ape could create new meanings by combining words according to grammatical rules. However, careful analyses of videotapes of the "sentences" that Nim and

[16]I'm not aware of any project that trained an ape to use symbols for any purpose other than to obtain a primary reward. The intelligence needed to use symbols for that purpose is on a par with that needed to operate a vending machine. Using one or more symbols declaratively implies a speaker that is motivated to communicate information to a listener simply to share its meaning.

other apes produced showed that they could be explained without reference to grammatical competence. My students and I analyzed 20,000 of Nim's combinations of two or more signs as recorded on videotapes of his signing in the classroom and at home[17]. Superficially, many of Nim's combinations appeared to be generated by the rules of a simple finite-state grammar, for example, the sequences, *more → x*, *transitive verb → me*, etc. As such, Nim's combinations provided the strongest evidence that apes could create a sentence. Indeed, many of Nim's multi-sign utterances resembled a child's initial multiword utterances (Nelson, 1981).

Toward the end of 1978, I began to write an article about Nim's linguistic achievements and anticipated a strong interest in my claim that, for the first time, a chimpanzee had learned the rudiments of a human language. Unfortunately, I soon discovered that my evidence was tainted because I had committed the sin of not defining the stimulus properly. While teaching a new research assistant how to transcribe Nim's signs from a videotape, I explained to her as I've explained to other research assistants that they could be watching history being made by a signing chimpanzee and that therefore they should give their undivided attention to Nim's signing.

After my assistant settled into her task, I allowed myself the luxury of broadening the scope of my view of the video monitor on which Nim was signing. That allowed me to observe Nim and his teacher simultaneously, within the same frame. Within a few seconds, I had the horrific epiphany that Nim was simply mirroring the teacher's signs. About 500 msec before Nim signed, his teacher prompted him with a sign that he imitated. In some instances Nim also produced signs that he had learned as "wild cards" in the sense that they were appropriate for many contexts, e.g., "me," "Nim," "eat," "drink," "tickle," "hug," etc.

A frame-by-frame analysis of our videotapes revealed that Nim responded mainly to the urgings of his teacher and that the majority of his signs were full or partial imitations of his teacher's prior utterances (Terrace, Petitto, et al., 1979). Although young children also imitate many of their parents' utterances, the relative frequency of those utterances is substantially lower than Nim's. Further, as shown in Figure 5, Nim never moved beyond the imitative phase of language development. Analyses of the available films of other signing apes showed that the patterns of their signing were similar to Nim's, specifically, that the majority of their signs were prompted by their trainers and they were also highly imitative (Terrace, Petitto, et al., 1979).

The conclusions of Project Nim were criticized on various methodological grounds by other investigators. (Gardner, 1981; Van Cantfort & Rimpau, 1982; Hess, 2008). For example, it was said that Nim was dim [Nim → dim → dumb], that his language development was diminished because he had too many trainers and that his classroom was oppressive because it had cinder block walls, etc[18]. Importantly, however, my critics did not recognize that my conclusions did not depend on the data I collected from Nim or how he was cared for. I could have reached the same conclusions by looking at videotapes of other chimps using ASL, or, for that matter, as an armchair psychologist who had no experience trying to teach an ape to use language. Most importantly, my critics could have refuted my conclusions by providing a single unedited videotape of a chimpanzee signing with a human companion in which both

---

[17]Nim, and some of his trainers, lived in a large mansion on 23 acres in Riverdale, NY.

[18]No evidence was produced to substantiate these criticisms. Although Nim's intelligence was never tested, the consensus of dozens of people that met or observed Nim (including members of a site visit by NIMH) was that he was highly intelligent. Further, I'm not aware of any valid tests of intelligence for a chimpanzee. I published the full list of Nim's trainers to recognize their efforts on the project (Terrace, 1979). There was, however, no implication that each teacher spent an equal amount of time with Nim. My conservative estimate is that one of a core group of 5 trainers was with Nim during 85% of his waking hours. In that sense, Nim was raised by an extended family, a practice about which I never heard any criticism, in particular that it retards language acquisition. Finally, I am unaware that there is any evidence that cinder block walls in a classroom cause students to be depressed.

parties were visible in each frame. During the 40 years that have elapsed since I published my conclusions about how chimps were induced to produce sentence-like-sequences, there hasn't been a single videotape that provided evidence that challenged my conclusions[19].

Primate cognition

While writing an article about Project Nim, it occurred to me that the negative results of Project Nim contained a silver lining. Although there was no evidence that Nim, or any of the other so-called “linguistic apes”, could learn a grammar, I felt that many aspects of their performance could not be explained by traditional chaining theory because the ape language projects presented the relevant discriminative stimuli *simultaneously*. As noted earlier, traditional chaining theory only applied to sequences on which stimuli were encountered *successively*. In the case of Sarah and Lana, the chimps trained by Premack and by Duane & Sue Savage-Rumbaugh, all of the “words” of the sequence the chimps had to execute were presented simultaneously at the beginning of each trial. In the case of signed or spoken language, humans have simultaneous access to the words of their respective vocabularies. There is no reason to think that the same was not true of apes. Thus, any explanation of the sequences produced by apes trained to use words required a new model of chaining, one that was based upon simultaneous access to all of the relevant stimuli.

Beginning with Ebbinghaus (Ebbinghaus, 1964), various forms of chaining theory have assumed that an understanding of learned sequences in animals and humans would follow directly from an understanding of how particular stimuli become associated with particular responses. On this view, all instances of serially organized behavior are reducible to discrete stimulus-response (S-R) units, each linked to the next by virtue of extensive practice. It mattered not whether the sequence was a sentence or tying a shoelace.

In one of the most influential articles in modern psychology, Karl Lashley challenged chaining theory on the grounds that it could not explain a person's knowledge of relationships between items that are *not* adjacent to one another, for example, between words from different parts of a sentence (Lashley, 1951). Because Lashley's arguments were based on examples of human behavior, his critique had less influence on ideas about animal behavior than it had on human cognitive theory and research. Indeed, some critics have argued that Lashley's ideas do not apply to animals because their learned behavior does not approach the complexity of human skills and because their communication is simpler and less arbitrary than human language (Lewandowsky & Murdock, 1989). That view is no longer tenable. Recent advances in our understanding of serially organized behavior in animals have confirmed that Lashley's criticisms of chaining theory apply with the same force to animal behavior as it does to human behavior.

In response to claims that chimpanzees could create grammatical sequences, e.g., Mary → give → Sarah → apple or, Please →machine → give → apple,

I developed a new chaining paradigm that I referred to as the *simultaneous chaining paradigm* (Terrace, 1984). Sequences trained by this paradigm differ fundamentally from the kind of sequences on which animals have been traditionally trained. On a simultaneous chain (SimC), all of the list items are presented simultaneously, on a touch-sensitive-video-monitor

[19]Different considerations led to a rejection of the view that Sarah's and Lana's sequences were sentences. In one analysis of a corpus of approximately 14,000 of Lana's combinations that were collected by a computer, researchers concluded that those combinations could be accounted for almost entirely as conditional discriminations (Thompson & Church, 1980). For example, Lana first learned by rote to use “lexigrams” such as *apple, music, banana,* and *chocolate* in sequences like *Please → machine → give → X*. Typically, the symbol for a particular reward was inserted into the last position of the stock sentence. Although Lana clearly understood the meaning of the lexigrams that referred to a particular reward, there is no evidence that she understood the meanings of the other lexigrams in the stock sequences she produced e.g., *please, machine* and *give, put,* or *piece of.*

on which their configuration could be varied from trial to trial. This was done to insure that the sequence could not be learned as a motor sequence.

According to traditional chaining theory, what a subject learns in a maze with *n* choice points, are *n* discrete S-R associations (Skinner, 1934; Hull, 1935). Thus, as shown in Figure 6A, if *n* = 7, the subject has to learn the following 7 associations: [S1:R1], [S2:R2], [S3:R3], [S4: R4], [S5:R5], [S6:R6] and [S7:R7]. To run through the maze shown in Figure 5A, a subject need only learn which way to turn at each choice point: S1: R1 → S2: R2 → S3: R3 → S4: R4 → S5: R5 → S6: R6 → S7:R7 → SR. Responding correctly to S_n removes the subject from S_n and ensures that the subject will encounter S_{n+1}, and only S_{n+1} at the next choice point. For example, when the subject arrives at, say the 3rd choice point, S3 does not compete for the subject's attention with discriminative stimuli from other choice points (S1, S2, S4, S5, S6 or S7).

The essential feature of a simultaneous chain is that all list items are displayed throughout each trial and that no differential feedback is provided following each response (Terrace, 2005). As can be seen in Figure 6B, all correct responses allow a trial to continue but they provide no information about the next correct response. Similarly, all errors end a trial immediately without any indication of what the correct response was at that point in the sequence. These constraints make it necessary for the subject to (1) construct a representation of the required sequence and to (2) keep track of its position on that representation as it moves from one item to the next.

Consider, for example, a trial on which a subject is required to produce a 7-item simultaneous chain: [S1 S2 S3 S4 S5 S6 S7] R1 → R2 → R3 → R4 → R5 → R6 → R7 → SR. The 7 stimuli are presented simultaneously until the subject makes an error or until it earns a reward by responding to those stimuli in the correct order. A second critical difference between a successive and a simultaneous chain is the spatial location of the choice points on each trial. On the former, the spatial location of choice points remains fixed throughout training. That allows a subject to learn the maze as a sequence of specific motor responses. It follows that the execution of a successive chain does not require a representation to guide it through the sequence.

Representations

Curiously, the logic of a representation, one of the core concepts of cognitive psychology, was defined almost 100 years ago by the distinguished behaviorist, Walter Hunter, in a study of delayed responses in animals.

> … If comparative psychology is to postulate a representative fact…it is necessary that the stimulus represented be absent at the moment of the response. If it is not absent, the reaction may be stated in sensory-motor terms (Hunter & Nagge, 1931)

Hunter's point was that if there were no external stimulus that account for some response, the subject had to *re*-present the absent stimulus to itself. Hunter's logic is illustrated by the delayed-matching-to-sample (DMTS) paradigm whereby a delay is inserted between the sample and the test stimulus. If the subject responds correctly during test, it is necessary to postulate a representation of the sample as the stimulus that occasioned that response (Kendrick, Rilling, et al., 1981)[20].

[20]Hunter formulated the conditions for invoking a representation with the hope of providing evidence of such in his experiments on delayed responding in dogs, raccoons and rats (Hunter, 1913). Although those experiments did not provide the evidence Hunter sought, his concept of the representation he proposed has stood the test of time. Curiously, even though Hunter's formulation of the conditions for appealing to representations as an explanation of behavior is foundational for cognitive psychology, he has rarely been given credit for that contribution.

When I designed the SCP, I was aware that I was crossing the line between behavioral and cognitive psychology. I was also aware that the definition of the stimulus had to be broadened to include a self-generated internal component. Skinner objected to any reference to a representation and other mentalistic terms for two reasons. (1) They were not amenable to operational definitions and they added nothing to behavioral definitions. What Skinner failed to recognize, however, was how the cognitive revolution overcame both objections by characterizing the mind as a computer. (2) In many instances, particularly vision, the computational approach paid handsome differences in that it accounted for significant portions of the variance in experiments on that topic. Less progress has been made in other areas, e.g., in some areas, such as learning and memory; in particular, conditioned fear (Beylin, Gandhi, et al., 2001; Fanselow & Poulos, 2005; Lamprecht, Farb, et al., 2006; Lee, Kim, et al., 2009).

Ironically, Skinner also failed to recognize that he used the logic of a representation in his formulation of concept formation. For example, in *Verbal Behavior*, Skinner argues that the concept of Mozart's music is acquired by listening to lots of his music and abstracting particular features that individually, or in combination with one another, occasions the response “Mozart” (Skinner, 1957).

Consider, for example, a thought experiment in which subjects have to make yes-no judgments as to whether a particular piece of music was written by Mozart and that subjects' accuracy was greater than that predicted by chance. Although the concept of Mozart's music can't be defined physically, the fact that it can be identified correctly, justifies the conclusion that subjects relied on a representation of Mozart's music while making their judgments.

Stimulus control over time?

To maintain Skinner's style of behavior analysis, some behaviorists avoid any mention of any cognitive terms, representations in particular. Instead of referring to representations of events that are no longer physically present, as an explanation of behavior, they have appealed to the concept of *stimulus control* that has been extended over time (Wixted, Bellack, et al., 1990). The problem with such purity is that it stretches the laws of the physical world beyond recognition.

Stevens' concern about defining the stimulus is especially relevant to a definition of stimulus control over time. What are the spatial coordinates of such stimuli and how do they exert their effects? Does it stimulate a receptor in the same manner as a stimulus that is physically present? How does one measure the physical properties of a stimulus that exerts its control over time? The bottom line is that behaviorists can't have it both ways. In the case of representations, they object to explaining behavior by stimuli they can't observe and, instead, are willing to attribute fictional powers to stimuli they can observe.

Thinking without language

On the one hand, my negative conclusions about the linguistic abilities of chimpanzees support Descartes' view that animals are incapable of communicating with language (Descartes, 1637). On the other hand, the ability of monkeys to remember lengthy sequences challenges Descartes' argument that animals can't think because they can't learn a language. In the remainder of this article, I will describe some experiments I performed that provide unequivocal evidence that animals can think without language. For example, in one experiment, four monkeys had to learn by trial and error the correct order in which to respond to items from four different lists, each composed of seven arbitrarily selected photographs. Each list was composed of novel photographs. The monkeys had to determine the correct order in which to respond to the items from each list by trial and error where the odds of guessing the correct sequence on a 7-item list was less than 1 in 5000 (Terrace, Son, et al., 2003).

To appreciate the difficulty of learning a single 7-item list, imagine trying to enter your 7- digit personal identification number (PIN), say *9-2-1-5-8-4-7*, at a cash machine on which the positions of the numbers were changed at random each time you tried to operate it. You could not enter your PIN by executing a sequence of distinctive motor movements, as you might when making a phone call. Instead, you would have to locate each number on the number pad on each trial and mentally keep track of your position in the sequence as you pressed different buttons on the keypad. Any error terminated a trial. On the next trial, the same digits appeared, albeit in a novel configuration that was selected at random from more than 80,000 possible configurations. To determine your PIN, you would have to recall the consequences of any of the 36 types of logical errors you could might make while attempting to produce the required sequence (21 types of forward errors and 15 types of backward errors). Further, you would have to determine the first 6 digits without getting any money from the cash machine. This is precisely the type of problem the monkeys had to solve at the start of training on each of the four 7-item lists on which the list items were photographs, rather than the Arabic numerals. Instead of cash, the monkey's reward was banana pellets. The four 7-item lists are shown in Figure 7A

All four monkeys learned all 4 of the 7-item lists and also demonstrated that they could execute *all* of those lists when they were presented at random during a single session. As shown in Figure 8, each monkey became progressively more efficient at deducing the correct order in which to respond during the course of learning four 7-item lists. On the final 7-item list, they barely exceeded the minimum number of logical guesses needed to identify the first 2 items.

The monkeys were also tested on their ability to apply their knowledge of 4 different 7-item lists on a novel task. They were shown all of the 336 pairs that could be derived from the 28 items used to construct the four 7-item lists on which they were trained[21]. Examples of these subsets are shown in Figure 7B. Monkeys were rewarded for responding to the items in each pair in the order specified by their ordinal positions on the original list. As shown in Figure 9, monkeys responded at the same high level of accuracy (91%) on the *first* occasion on which each pair was presented, whether the items were drawn from the same or from *different* lists. For example, they were equally adept at responding in the correct order to the 2nd item of the third 7-item list and the 6th item of the first 7-item list (a between-list pair) as they were to the 2nd and the 6th items of the fourth 7-item list (a within-list pair).

The sequences that the monkeys learned in these experiments are arguably the most difficult lists ever mastered by any nonhuman primate, including those trained in experiments on the linguistic and numerical abilities of apes (Premack, 1976; Rumbaugh, 1977; (Biro & Matsuzawa, 1999). The ease with which they learned 7-item lists and the steady decrease in the number of sessions they needed to master new lists suggests that they could master longer lists. The monkeys' performance on the pair-wise test is significant because it showed that they could compare representations of the ordinal position of each item from each list and then apply that knowledge to solve a novel problem. Specifically, their performance shows that each monkey represented, in long-term memory, the ordinal position of items from each of the four 7-item lists they had learned and that they were able to compare, in working memory, the ordinal positions of any two items from any of the 7-item lists.

Rule-governed simultaneous chains

Having observed that monkeys perform so well on 2-item subsets that were composed of items from different lists, each composed of arbitrary items I felt encouraged to train them on other

[21]Pairs composed of items occupying the same position on different lists were not used because there was no correct answer for those subsets (e.g., $B_{list}1$ and B_{list3}).

serial tasks. For example, Liz Brannon and I have trained monkeys to learn lists based on numerical quantity (Brannon & Terrace, 1998). On lists composed of arbitrary items, it is necessary to discover, by trial and error, the correct order in which to respond to each item. There is, however, no reason why list items must be perceived as arbitrary stimuli. They could just as well conform to a particular rule. Consider, for example, a simultaneous chain on which the correct sequence is defined as a series of responses to the next larger (or smaller) stimulus. In this experiment the stimuli consisted of numerically defined list items, each composed of geometric stimuli that varied the number of elements of different shapes, sizes and colors. Such a list could be learned by trial and error or by applying an ascending rule. If lists that were composed of numerically defined items were learned on the basis of the specific elements used to compose list items, new lists should be equally difficult and the monkeys should be unable to order novel sets of stimuli. By contrast, a subject that learned an ascending rule should be able to execute new lists correctly on the first trial on which they were presented.

Precisely such results were obtained in an experiment in which monkeys were trained to respond to geometrically defined exemplars of the numerosities 1, 2, 3 and 4 in an ascending order. After training on 35 four-item lists, subjects were tested on 150 novel lists (all composed of trial-unique stimuli) to see if they abstracted a numerical rule during their initial training. Examples of the numerical exemplars used in this experiment are shown in Figure 10A.

As shown in Figure 10B, after subjects learned to respond at a high level of accuracy on the 35 different lists, they continued to do so, without any decrement, on 150 novel lists. Since it was impossible to memorize the order of list items that are trial unique, it follows that subjects abstracted a numerical rule during their initial training on numerically defined lists. It does not, however, follow that these subjects learned an ordinal rule. They could have simply assigned exemplars of each numerosity to one of four nominal categories, as might be the case with photographs of fish, trees, fruit and rocks.

If subjects were using a nominal rather than an ascending rule, it should be just as easy for them to learn a list composed of numerical stimuli in a non-monotonic order, e.g., $3 \rightarrow 1 \rightarrow 4 \rightarrow 2$, as it would be to learn a list in which a subject had to respond in an ascending order ($1 \rightarrow 2 \rightarrow 3 \rightarrow 4$). Contrary to that hypothesis, a subject trained on a $3 \rightarrow 1 \rightarrow 4 \rightarrow 2$ sequence showed no signs of improvement during training on 13 stimulus sets. Accuracy improved rapidly, however, once that subject was required to respond to list items in an ascending numerical order (Brannon & Terrace, 2000). The ease with which monkeys can learn a monotonic rule (as compared to a non-monotonic rule) provides strong evidence that they can perceive the ordinal relations between the numerosities on which they were trained.

Even stronger evidence of a monkey's ability to use an ordinal rule was provided by their performance on a test on which they were shown all 36 numerical pairs that could be derived from the numerosities 1-9. These are shown in Figure 10C. Ten of those pairs were composed of the novel numerosities 5, 6, 7, 8 & 9. Figure 10D shows some examples of each type of stimulus. As shown in Figure 11, subjects responded correctly on more than 70% of the trials on which both items were novel. This provides clear evidence of a monkey's ability to extrapolate an ordinal rule to novel numerical values.

Cognitive imitation

All of the experiments I've described thus far have centered on the ability of individual monkeys to represent exteroceptive events. However, most of our everyday behavior is social and it is therefore of interest to investigate the extent to which a rhesus monkey can represent a con-specific's behavior to facilitate its acquisition of novel problems.

Imitation is a common format for studying how one individual's behavior can influence another's. However, in virtually every experiment performed on imitation, imitation was defined with respect to motor tasks. That has lead to a lot of confusion because measures of what a naïve student sees while observing an expert perform a task are poorly defined, as are the criteria for determining which actions count as imitative.

To avoid these problems, we focused on *cognitive imitation*, a form of social learning that doesn't require motor imitation. It involves nothing more or less than the ability to copy one or more abstract rules. To investigate cognitive imitation, we trained monkeys to execute *simultaneous chains* (SCs). Other things being equal, SCs have to be learned by trial and error. In this experiment, our subjects were two monkeys, both of which were proficient at learning SCs from their experience in other experiments that used the SCD. The idea was to provide a "naïve" monkey, who had not been trained on a particular list with an opportunity to observe an "expert" monkey execute the list in question. To return to our previous example of entering your pin at an ATM, the current paradigm would be analogous to learning someone's password at an ATM by looking over her shoulder, with the important difference that, on an ATM, the spatial positions of the number buttons never change).

Subjects were trained in sound attenuated chambers, in which two adjacent walls were made of tempered glass. When an opaque partition was placed between the chambers, each glass wall functioned as a mirror. Monkeys that were run under that condition had no knowledge of what the other monkey was doing. When the partition was removed, subjects had a full view of one another.

Subjects were trained on thirty 4-item lists. Fifteen of those lists were trained in isolation, that is, with the partition between the chambers in place. Performance on those lists provided a baseline measure of trial-and-error learning. On the remaining 15 lists, the partition was removed (social-learning condition). This arrangement allowed the naïve monkey (the "student"), who had never seen items from the list it was about to learn, observe an experienced monkey (the "expert") perform that list. Those lists will be referred to as *social-learning* lists. Baseline and social-learning lists were alternated with one another throughout the experiment to balance any list-learning expertise that subjects might develop while learning new lists under each condition, i.e., baseline $list_1$, social-learning $list_1$, baseline $list_2$, social-learning $list_2$, etc.

Under the social-learning condition, the opaque barrier that separated the monkey's chambers was removed. The expert and the student were placed in their respective chambers at the same time prior to the start of each session. The expert performed a list on which it had been over-trained to insure a high level of accuracy. The student was introduced to a new list during two successive blocks of 20 trials each. During the first block, the student's monitor was dark and inactive (observation period). That arrangement allowed the student to observe, but not perform, the sequence that the expert was executing in the adjacent chamber. During the second block of 20 trials (test period), the student's monitor was turned on and activated, thereby providing the student with its first opportunity to respond to the same list items that appeared on the expert's screen. The student and the expert worked side-by-side, in full view of each other, throughout the test period.

As can be seen in Figure 12A, each monkey made significantly fewer responses before it completed their first trial correctly under the social-learning condition than under the baseline condition. This difference suggests that students learned new lists vicariously by observing an expert perform them under the social-learning condition.

To justify that conclusion, however, it was necessary to control for social facilitation and learning that was based entirely on the feedback displayed on the computer monitor as the sequence is executed. To rule out social facilitation, we trained each monkey as a student while it could observe the “expert” respond on a different list. To rule out non-social factors, we trained the student on a new list, with no monkey present in the second chamber. The list that the student was trying to learn was displayed on the screen of the second chamber. Responses that would have normally been made by the expert monkey were instead made by the computer, which also produced the feedback that normally follows a correct response.

As shown in Figure 12 B & C, neither control condition enhanced list learning. The number of responses students needed to complete their first correct trial under the social-facilitation and the non-social conditions did not differ from the number of responses needed during the relevant block of baseline training sessions.

The success of this experiment can be attributed to a variety of factors. Of greatest importance, our paradigm allowed for the separation of motor and cognitive factors, each of which may contribute independently to imitation. The use of a familiar motor task throughout training, in this instance the SCP, made it possible to obtain multiple measures of imitation from the same subject.

Defining the stimulus

One ubiquitous consequence of learning a particular sequence is that subjects place them on a continuum that allows them to compare their ordinal positions. We have seen an example of this ability in the experiment on two-item subsets that followed training on 7-item lists. As shown in Figure 11, subjects' accuracy increased and their reaction times (RTs) decreased as the ordinal distance between items increased. This *symbolic* distance effect[22], which was discovered in an experiment by Moyer and Landauer (1967), in which they asked human subjects to report which member of a pair of Arabic numerals was larger (or smaller), has also been observed in experiments on monkeys, chimpanzees and human children and adults in which subjects judged the ordinal positions of arbitrary or numerical stimuli. Some examples of those distance functions are shown in Figure 13 A & B.

One important implication of distance functions is that, in addition to learning item-item associations, subjects acquire associations between an item and its *ordinal position*. Item-item associations can be specified by a physical description of the items. But once again we encounter a problem of defining the stimulus. For example, to explain the distance effect, it is necessary to assume that subjects represent the ordinal position of each stimulus. It is also of interest to learn if the subject responds to the relative value of each stimulus, say, small, medium or large or to its absolute value, say 1st, 2nd, 3rd, etc. Those questions have been investigated in a recent experiment that provided clear evidence that both humans and monkeys represent ordinal position as relative values while learning a transitive inference task (Merritt & Terrace, 2004).

Metacognition

We have seen that human and non-human primates can represent stimuli that they can use to solve some problem. During the 1990's, cognitive psychologists have provided ample evidence that humans can also represent their confidence in the accuracy of their answers to questions that require them to use representations of various objects and events. Consider, for example your ability to answer the question, “What is the capital of Australia?” The correct answer is Canberra, a city that is not well known outside of Australia. If you answered Canberra, you would most likely feel highly confident of the accuracy of your answer. Now suppose that

[22]Moyer and Landauer's reference to a symbolic distance effect was meant to highlight the fact that a distance effect with respect to Arabic numbers was based on the psychological, rather than the physical, characteristics of the stimuli about which the subject made judgments.

instead of Canberra you tried to choose between two better-known cities: Sydney and Melbourne. Most likely, you would answer “not sure” that the capital of Australia is one of those cities.

In an experiment on metacognition, the dependent variable is your confidence about the accuracy of factual knowledge. Accordingly, if you answered “not sure” when asked about your confidence in the accuracy of your answer to the question about the capital of Australia, that answer would be classified as correct, even though your answer to the factual question (Sydney or Melbourne) was incorrect.

Having shown previously that monkeys can form representations of ordinal position, we were in a position to ask whether a monkey can form accurate meta-representations that are based on its confidence in the accuracy of its responses on various tasks on a problem that required subjects to make judgments of ordinal position. With Lisa Son and Nate Kornell, I conducted an experiment on two monkeys that was based on the answer that humans often use when asked about the certainty of their knowledge about some fact, “put you money where your mouth is” (Kornell, Son, et al., 2006).

Ebbinghaus and Lashley were first trained on a task on which they had to choose the largest (or smallest) line from an array of lines that were presented on a touch sensitive video monitor. On that task, the monkeys were given a banana pellet for each correct answer. After mastering the line length task, we changed the contingency for obtaining reward. Instead of rewarding the subject with food following an accurate response on the line length task, subjects were rewarded, with tokens, for reporting their confidence in the accuracy of their response on each line length trial. They did so by choosing one of two icons that appeared on their touch sensitive monitors. One icon meant high confidence; the other, low confidence. Tokens were accumulated in a bank that was displayed on the right side of a subject's monitor. Figure 14 outlines the two contingencies used in this experiment.

The number of tokens that a subject received during the second phase of this experiment was based on the *relationship* between the accuracy of its response on a given line-length trial and the subject's confidence in the accuracy of that response. Confidence was measured by the number of tokens the subject was willing to bet that its response on the line length task was correct. If the monkey was certain about its line length response and, if it were indeed correct, a ‘high-risk’ bet earned 3 tokens that were dropped into its bank. If, however, the subject placed a high-risk bet after an incorrect response on the line-length task, it *lost* 3 tokens. Small (low-risk) bets resulted in the delivery of one token, regardless of the monkey's performance on the memory task[23]. A strong correlation between high-risk bets and correct responses and low-risk bets following incorrect responses showed that monkeys could indeed form metacognitive representations of confidence.

During a subsequent phase of this experiment, the monkeys were trained on two additional magnitude discriminations (area and numerosity) and on a serial memory task that was qualitatively distinct from the magnitude estimation tasks on which they were originally trained. There was immediate transfer of the monkeys' metacognitive skills to the other magnitude estimation tasks and to the serial memory task.

What this experiment showed was that, on a test of serial memory, monkeys could represent a sample that was shown in a series of photographs at the beginning of a trial, in order to respond accurately during test to a “yes”/“no” question as to whether the sample appeared in an array of 4-6 photographs. They then had to represent the accuracy of their response on the memory

[23]This was done to avoid extinction if the subject placed too many high-risk bets after an incorrect response on the line length task.

task in order to report if their response on the serial memory task was correct or incorrect. A remarkable feat from an animal that doesn't have language!

Conclusions

Our journey has taken us from orthodox behaviorism to animal cognition, and from exteroceptive stimuli that can be fully described by their physical characteristics to representations that, at present, can only be described by their functional properties. What's needed is the concept of a superordinate stimulus that includes both physically defined stimuli, which originate from well-defined external events, and representations of those stimuli that take into account factors such as the size and the capacity of working memory and how items are retrieved from long-term memory. Working memory is responsible for most aspects of our everyday behavior. Although Stevens didn't have any reason to believe that the definition of the stimulus he sought would include cognitive components, his formulation of the basic problem of psychology can readily incorporate both cognitive and physiological events. Stevens' definition of the basic problem of psychology has also stood the test of time as a goal for research on animal cognition. Of course, that definition applies equally well to research on human cognition but reminding investigators in that field about Stevens' maxim would be like preaching to the converted. For the moment our concern should be about developing paradigms and theories that focus on cognitive factors that fill the gap between an exteroceptive stimulus and the responses that subsequently occur in its presence.

References

Amsel A. The role of frustrative nonreward in noncontinuous reward situations. Psychological Bulletin 1958;55:102–119. [PubMed: 13527595]

Amsel A, Ward JS. Frustration and persistence: resistance to discrimination following prior experience with discriminanda. Psychological Monographs: General and Applied 1965;79(4):1–41. [PubMed: 14300513]

Bandura, A. Social learning theory. New York: General Learning Press; 1977.

Bates, E. The Emergence of Symbols. New York, NY: Academic Press; 1976.

Benjamin, LT. A brief history of modern psychology. 1. New York: Wiley-Blackwell; 2006.

Beylin A, Gandhi C, Wood G, Talk A, Matzel L, Shors T. The role of the hippocampus in trace conditioning: temporal discontinuity or task difficulty. Neurobiology of Learning and Memory 2001;76(3):447–461. [PubMed: 11726247]

Biro D, Matsuzawa T. Numerical ordering in a chimpanzee (Pan troglodytes): Planning, executing and monitoring. Journal of Comparative Psychology 1999;113(2):178–185.

Blough, D. Form Recognition in Pigeons. In: Roitblat, H.; Bever, T.; Terrace, H., editors. Animal Cognition: Proceedings of the Harry Frank Guggenheim Conference; June 2-4 1982; Hillsdale, NJ: Erlbaum; 1984. p. 277-290.

Brannon E, Terrace H. Representation of the numerosities 1-9 by rhesus monkeys. Journal of Experimental Psychology: Animal Behavioral Processes 2000;25(1):31–49.

Brannon EM, Terrace HS. Ordering of the numerosities 1-9 by monkeys. Science 1998;282:746–749. [PubMed: 9784133]

Brown PL, Jenkins HM. Auto-shaping of the pigeon's keypeck. Journal of Experimental Analysis of Behavior 1968;11:1–8.

Brown, R. Psycholinguistics: Selected papers by Roger Brown. New York, NY: The Free Press; 1970.

Chomsky N. A review of Skinner's Verbal Behavior. Language 1959;35:26–58.

Chomsky, N. Language and mind. New York, NY: Harcourt Brace; 1968.

Clare L, Jones RSP. Errorless learning in the rehabilitation of memory impairment: a critical review. Neuropsychology Review 2008;18(1):1–23. [PubMed: 18247118]

Connor, BB.; Wing, AM.; Humphreys, GW.; Bracewell, RM.; Harvey, DA. Errorless learning using haptic guidance: research in cognitive rehabilitation following stroke. Paper presented at the 4th

International Conference on Disability, Virtual Reality and Associated Technologies; Veszaprém, Hungary. 2002.

Cook RG, Wright AA, Kendrick DF. Principles of Categorization in Pigeons and Humans. 1988

Crist, RL. Programmed instruction as a means of establishing "errorless" learning with kindergarten level children. Final Report. Illinois State University; 1969. Normal

Darwin, C. On the Origin of Species. London, UK: Mercury; 1859.

Descartes, R. Discourse on method and related works. London: Penguin Books; 1637.

Ebbinghaus, H. Memory: A contribution to experimental psychology (Originally published 1885; translated 1913 ed). New York, NY: Dover; 1964.

Fanselow M, Poulos A. The neuroscience of mammalian associative learning. Annual Review of Psychology 2005;56:207–234.

Fouts, RS. Capacity for language in great apes. In: Tuttle, R., editor. Socioecology and Psychology of Primates. Chicago: Mouton Publishers; 1975. p. 371-390.

Gardner BT, Gardner RA. Teaching sign language to a chimpanzee. Science 1969;162:664–672. [PubMed: 5793972]

Gardner BT. Project Nim: who taught who? Contemporary Psychology 1981;26:425–426.

Gardner, H. The Minds New Science: A History of the Cognitive Revolution. New York, NY: Basic Books Inc.; 1985.

Gibbon J. Scalar expectancy theory and Weber's law in animal timing. Psychological Review 1977;84:279–325.

Gibbon J, Baldock M, Locurto C, Gold L, Terrace H. Trial and intertrial durations in autoshaping. Journal of Experimental Psychology: Animal Behavior Processes 1977;3:264–284.

Gibbon J, Church RM, Meck WH. Scalar Timing in Memory. Annals of the New York Academy of Sciences 2006;423:52–77. Timing and Time Perception. [PubMed: 6588812]

Hansen JIC. Cognitions of a dustbowl empiricist. The Counseling Psychologist 1998;26:499–513.

Hanson HM. Effects of discrimination training on stimulus generalization. Journal of Experimental Psychology 1959;58:321–334. [PubMed: 13851902]

Herrnstein RJ. Riddles of Natural Categorization. Philosophical Transactions of the Royal Society of London 1985;308 B:129–144.

Hess, E. Nim Chimpsky: the chump who would be human. New York: Bantam Dell; 2008.

Hull CL. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution. Psychological Review 1935;42:219–245.

Hunter WS, Nagge JW. The white rat and the double alternation temporal maze. The Pedagogical Seminary and Journal of Genetic Psychology 1931;39:303–319.

Jenkins HM, Harrison RH. Effect of discrimination training on auditory generalization. Journal of Experimental Psychology 1960;59:246–253. [PubMed: 14407066]

Jenkins HM. The effect of discrimation training on extinction. Journal of Experimental Analysis of Behavior 1961;61:111–121.

Jenkins HM, Harrison RH. Generalization gradients of inhibition following auditory discrimination learning. Journal of the Experimental Analysis of Behavior 1962;5:435–441. [PubMed: 14029015]

Kendrick D, Rilling M, Stonebraker T. Stimulus control of delayed matching in pigeons: Directed forgetting. Journal of the Experimental Analysis of Behavior 1981;36(2):241–251. [PubMed: 16812243]

Kendrick, DF.; Rilling, ME.; Denny, MR. Theories of Animal Memory. Hillsdale: Lawrence Erlbaum Associates, Publishers; 1986.

Kornell N, Son L, Terrace H. Are monkeys metacognitive. Psychological Science. 2006 In press.

Ladyna-Kots, N. Infant Chimpanzee and Human Child. Vol. III. Moscow: Museum Darwinianum; 1935.

Lamprecht R, Farb C, Rodrigues S, LeDoux J. Fear conditioning drives profiln into amygdala dendritic spines. Nature Neuroscience 2006;9:481–483.

Lashley, KS. The problem of serial order in behavior. In: Jeffries, LA., editor. Cerebral mechanisms in behavior. New York, NY: John Wiley & Sons; 1951. p. 112-136.

Lee T, Kim J, Wagner A. Discriminitive conditioning with different CS-US intervals produces temporally differentiated conditioned responses in the two eyes of the rabbit (Oryctolagus cuniculus). Behavioral Neuroscience 2009;123(5):1085–1094. [PubMed: 19824775]

Lewandowsky S, Murdock BB. Memory for Serial Order. Psychological Review 1989;96:25–57.

Merritt, D.; Terrace, H. Inferential order judgments in rhesus monkeys and humans. Paper presented at the Meetings of the Psychonomic Society Meetings; 2004.

Nelson K. Individual differences in language development: implications for development and language. Development of Psychology 1981;17:170–187.

Pigeon ping pong clip (1987). In B. F. S. Foundation (Producer): The Cambridge Center for Behavioral Studies.

Premack D. Language in a chimpanzee? Science 1971;172:808–822. [PubMed: 5572906]

Schoenfeldt LF. From dust bowl empiricism to rational constructs in biographical data. Human Resource Management Review 1999;9(2):147–167.

Skinner, BF. The extinction of chained reflexes. Proceedings of the National Academy of Sciences; 1934. p. 234-237.

Skinner, BF. Verbal Behavior. New York, NY: Appleton-Century-Crofts Inc.; 1957.

Skinner, BF. Teaching Machines Cumulative Record. New York: Appleton-Century-Crofts, Inc.; 1959. p. 158-182.

Skinner BF. Pigeons in a pelican. American Psychologist 1960;15(1):28–37.

Son, L.; Kornell, N.; Terrace, H. Confidence Judgments by rhesus macaques on a serial memory task (Paper). Paper presented at the Meetings of the Psychonomic Society; 2003.

Spence KW. The differential response in animals to stimuli varying within a single dimension. Psychological Review 1937;44:430–444.

Stevens, S.; Davis, H. Hearing: its psychology and physiology. New York: J. Wiley & Sons, inc.; 1938.

Stevens, S. Handbook of Experimental Psychology. New York: John Wiley & Sons; 1951.

Subiaul F, Cantlon J, Holloway R, Terrace H. Cognitive imitation in rhesus macaques. Science 2004;305 (5682):407–410. [PubMed: 15256673]

Terrace H. Discrimination learning with and without errors. Journal of Experimental Analysis of Behavior 1963;6:1–27.

Terrace H. Wavelength generalization after discrimination learning with and without errors. Science 1964;144:78–80. [PubMed: 14107472]

Terrace H. Discrimination learning and inhibition. Science 1966;154(3757):1677–1680. [PubMed: 5924203]

Terrace H. Discrimination learning, the peak shift, and behavioral contrast. Journal of the Experimental Analysis of Behavior 1968;11(6):727–741. [PubMed: 16811320]

Terrace H. Escape from S- Learning and Motivation 1971;2(2):148–163.

Terrace, H. By-products of discrimination learning. In: Spence, H., editor. The psychology of learning and motivation. New York, NY: Academic Press; 1972. p. 195-265.

Terrace H, Gibbon J, Farrell L, Baldock M. Temporal factors influencing the acquisition and maintenance of an autoshaped keypeck. Animal Learning and Behavior 1975;3:53–62.

Terrace H, Petitto L, Sanders R, Bever T. Can an ape create a sentence? Science 1979;206:891–902. [PubMed: 504995]

Terrace, H. Simultaneous chaining: The problem it poses for traditional chaining theory. In: Commons, M.; Herrnstein, R.; Wagner, A., editors. Quantitative Analyses of Behavior: Discrimination Processes. Cambridge, MA: Ballinger Publishing Co.; 1984. p. 115-138.

Terrace H, Son L, Brannon E. Serial expertise of rhesus macaques. Psychological Sciences 2003;14:66–73.

Terrace H. The simultaneous chain: A new approach to serial learning. Trends in Cognitive Science 2005;9:202–210.

Van Cantfort TE, Rimpau JB. Sign language studies with children and chimpanzees. Sign Language Studies 1982;34:15–72.

Williams DR, Williams H. Auto-maintenance in the pigeon: Sustained pecking despite contingent nonreinforcement. Journal of Experimental Analysis of Behavior 1969;12:511–520.

Wixted, J.; Bellack, A.; Hersen, M. Behavior therapy. In: Bellack, A.; Hersen, M., editors. Handbook of comparative treatments for adult disorders. New York: John Wiley & Sons; 1990. p. 17-33.

Woodworth RS. Four varieties of behaviorism. Psychological Review 1924;31:257–264.

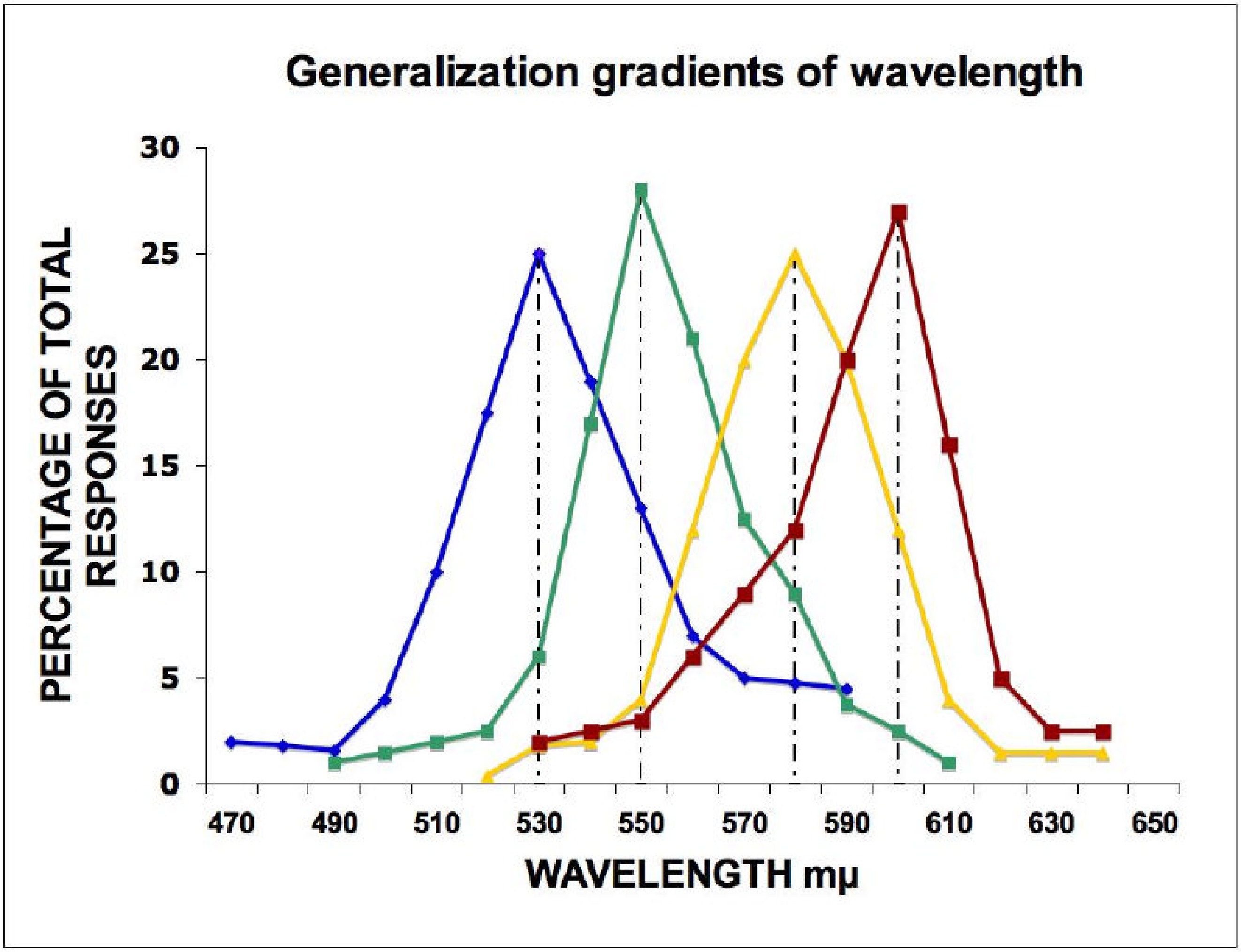


Figure 1.
Generalization gradients of wavelength obtained from pigeons trained to respond to different wavelengths. (Guttman & Kalish, 1956).

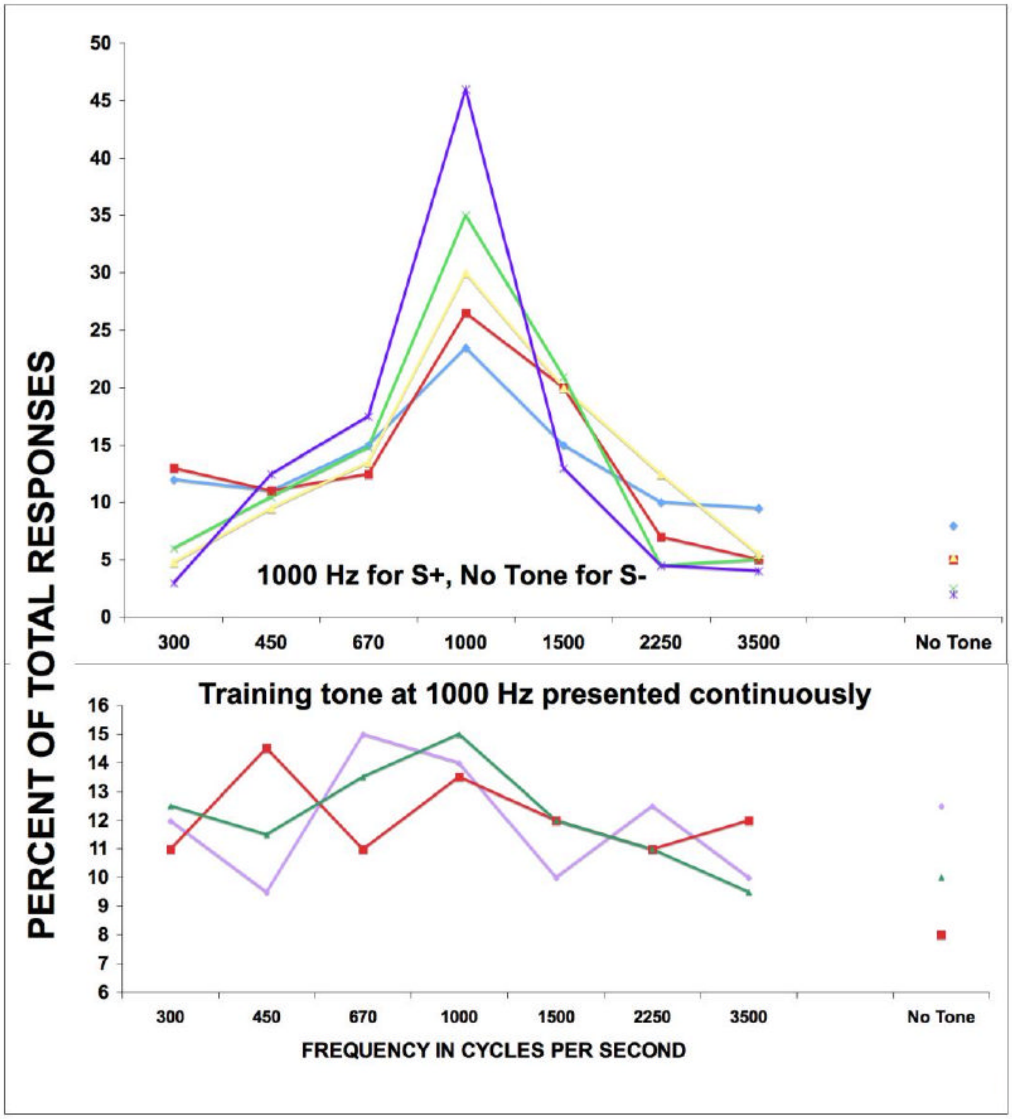


Figure 2.

Individual generalization gradients of tonal frequency obtained from pigeons following two training conditions. *Lower figure:* Training tone (1000 cps) was continuously present. *Upper figure:* These functions were obtained following discrimination training in which S+ was 1000 cps and S- was no tone (Jenkins & Harrison, 1960).

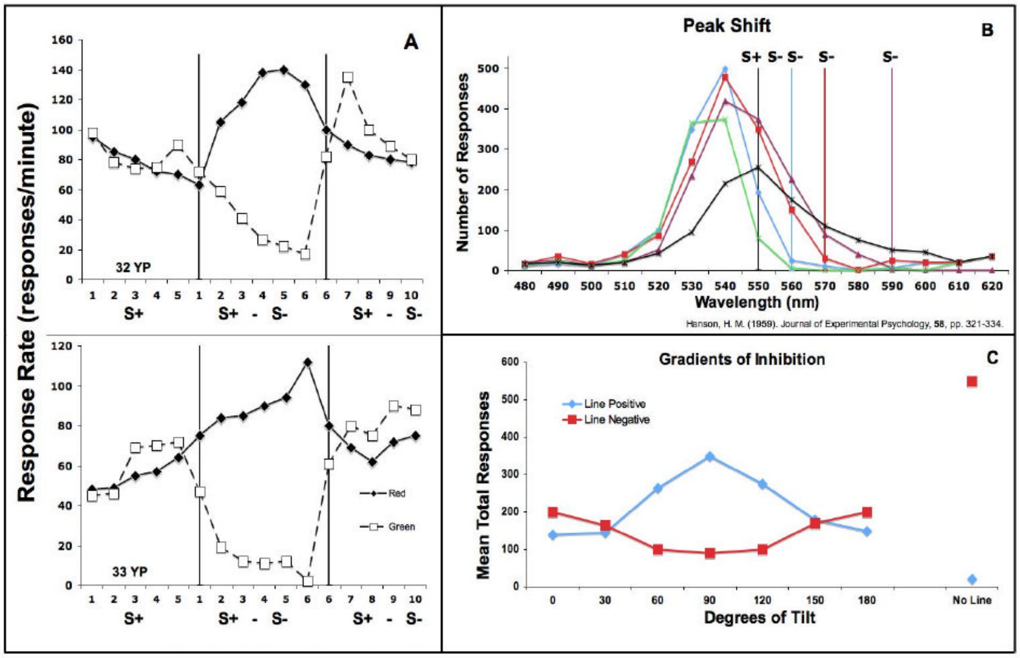


Figure 3.

Figure 3A. The development of behavioral contrast in individual subjects. The rate of responding by two subjects to each of two alternating discriminative stimuli are shown during the [VI 3′ VI 3′] and [VI 3′ Extinction] training (Adapted from Reynolds, 1961.)

Figure 3B. Effects of intradimensional discrimination training on the stimulus-generalization gradient. For all groups, S+ was 550 nm. Four groups received discrimination training with S- at 555, 560, 570, or 590 nm, respectively, as indicated by the vertical arrows. The maximum of a control group (solid curve) that received non-differential reinforcement at 550 nm was 550nm. The groups given discrimination training showed a positive peak shift with the maximum displaced from S+ to 540 nm. (Hanson, 1959. © 1959 by the American Psychological Association. Reprinted by permission.)

Figure 3C. Generalization gradients of responding (“excitation”) and not responding (“inhibition”) to a vertical line obtained from pigeons. For different subjects, the vertical line served as S+ and S-, respectively, in discrimination between the vertical line and a blank key (Honig *et al.*, 1963).

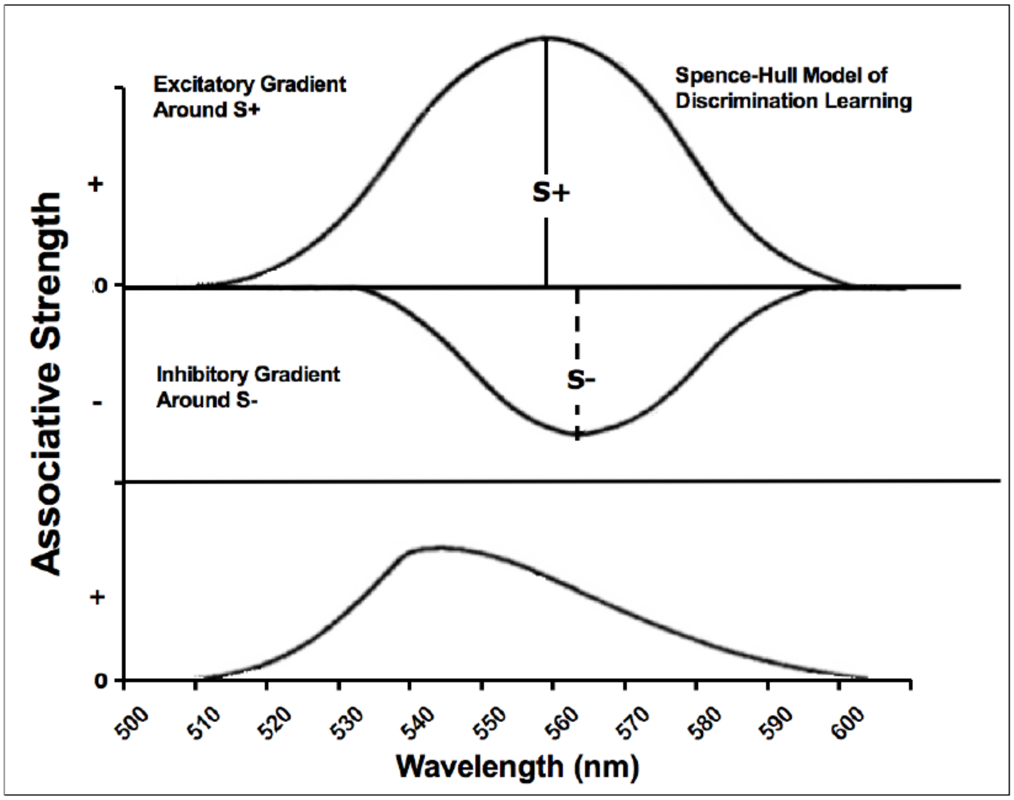


Figure 4.
Top panel: Gradients of excitation and inhibition, centered, respectively at S+ and S-. *Bottom panel:* If the gradient of inhibition is subtracted from the gradient of excitation the result is a gradient whose peak is shifted in a direction away from S-.

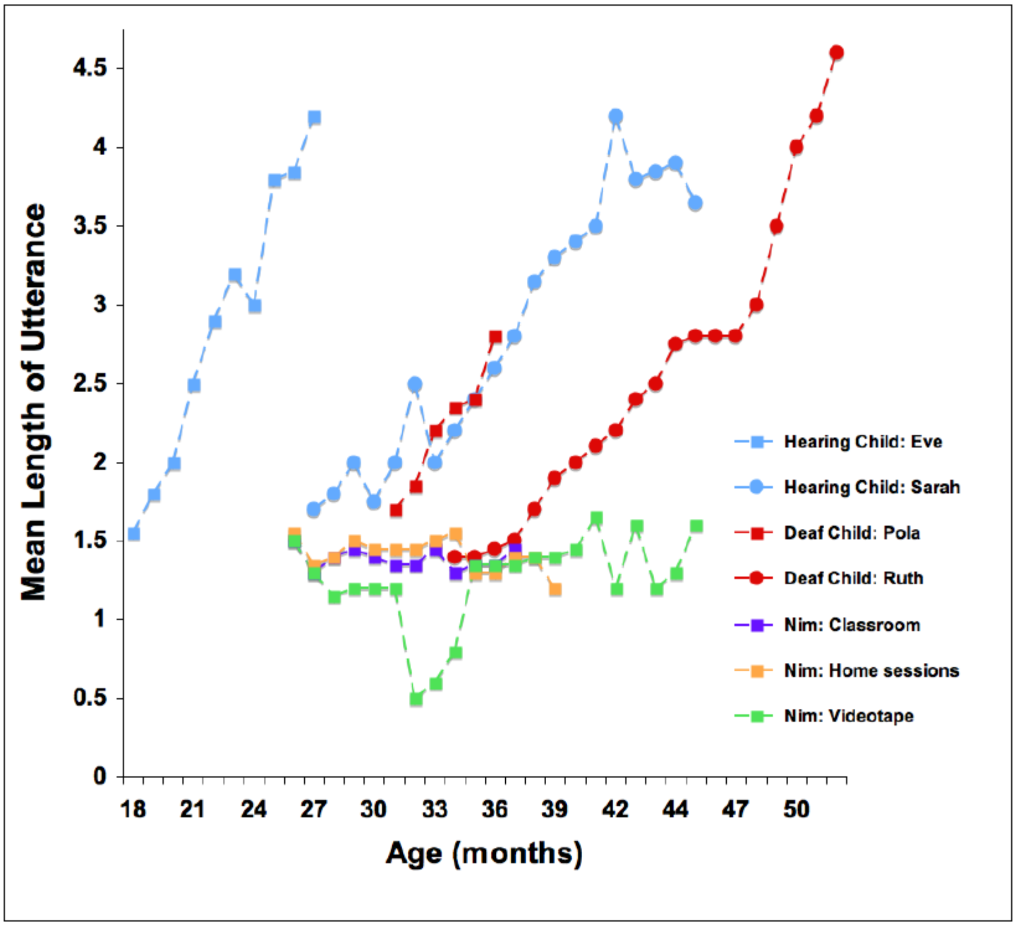


Figure 5.

Mean length of signed utterances of Nim and three deaf children and mean length of spoken utterances of two hearing children. The functions showing Nim's MLU are based on data obtained from teachers' reports between January 1976 and February 1977 (age, 26 to 39 months) are based on data obtained from teachers' reports; the function showing Nim's MLU between February 1976 and August 1977 (age, 27 to 45 months) is based upon video transcript data.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

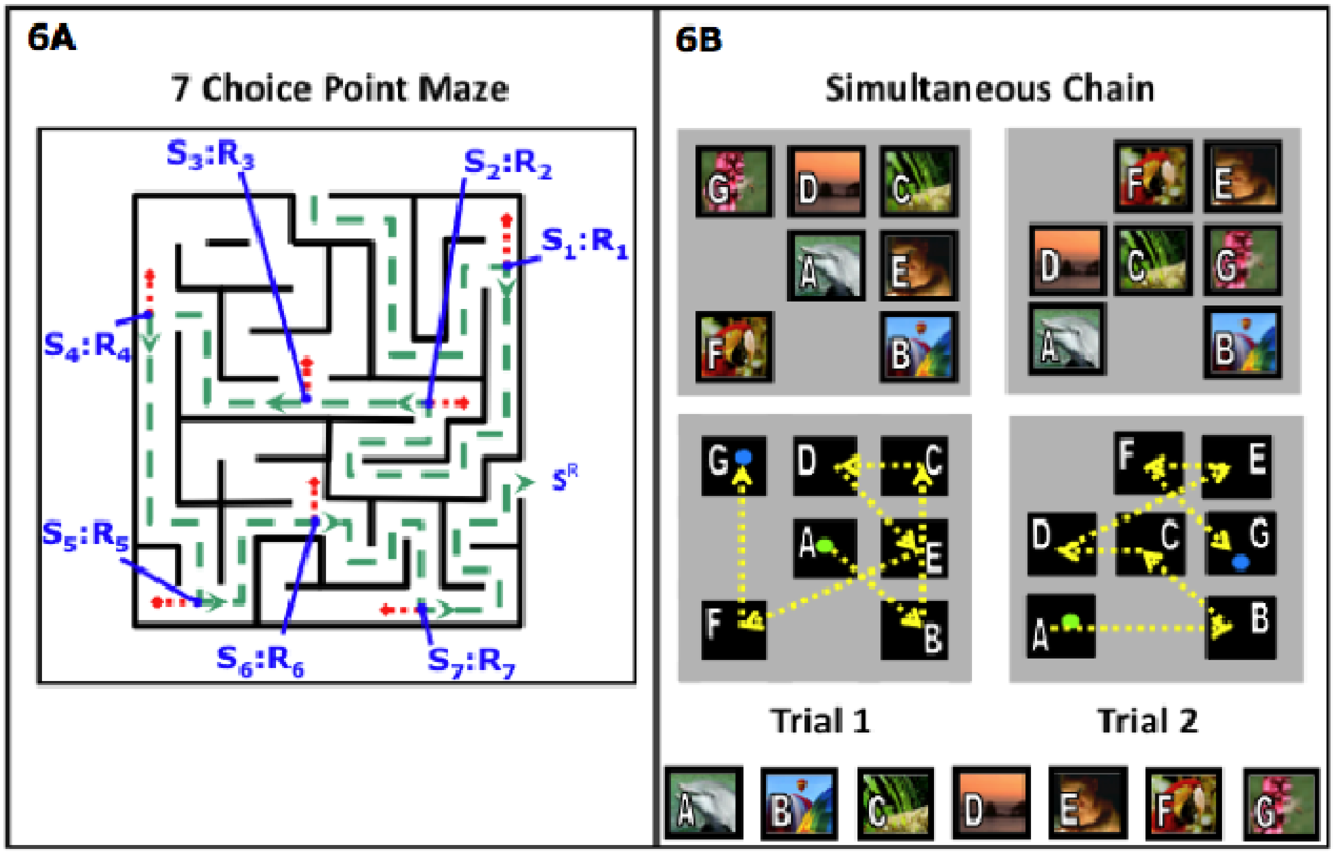


Figure 6.

Figure 6A. Basic plan of a traditional maze. Each choice point is separated, both spatially and temporally from the others. Accordingly, a run through the maze can be characterized as a *successive* chain that is composed of a series of S-R associations.

Figure 6B. *Simultaneous chaining* paradigm. Each of the upper panels depicts a trial during training on a 7-item simultaneous chain. The order in which the subject must respond to these photographs are shown on the bottom of the figure. The configuration of the items, typically travel photographs, varies randomly from trial to trial. The bottom portion of each panel depicts the route that a subject must follow with respect to the items shown in the top panel in order to receive a reward. Barring a 1 in 5014 guess, a simultaneous chain must be learned by trial and error. A reward is presented only after the subject has responded, to each item in the correct order. An error at any point of the sequence ends the trial. Another configuration of the same stimuli appears on the next trial.

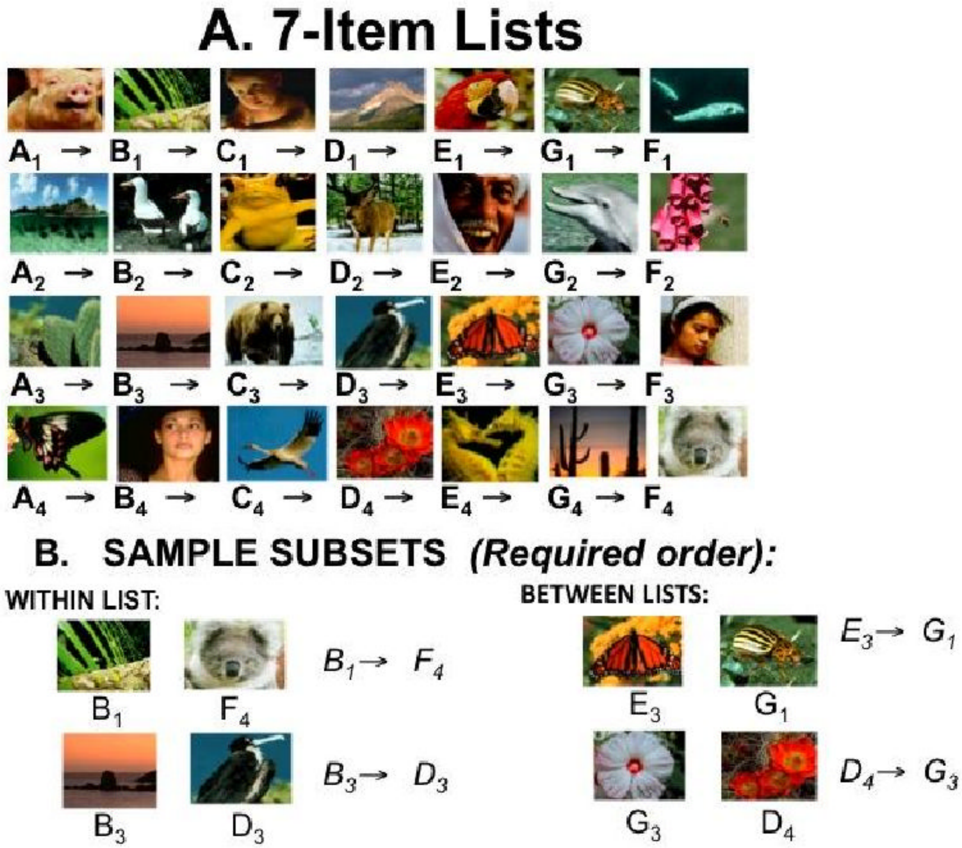
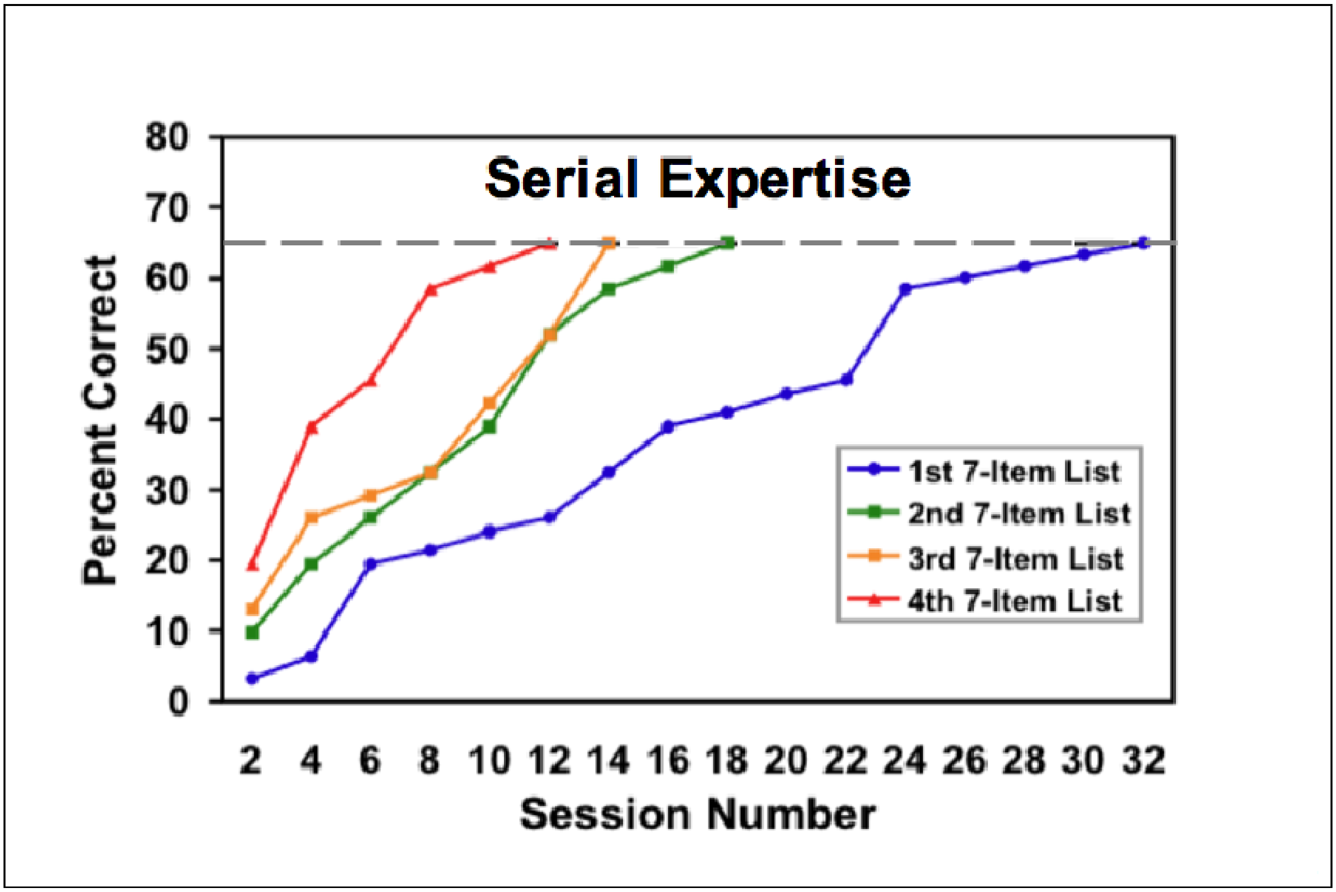


Figure 7.

Figure 7A. The four 7-item lists used in the expertise experiment (Terrace, Son, & Brannon, 2003).

Figure 7B. Examples of within- and between-list 2-item subsets.


Serial Expertise
Percent Correct
0
10
20
30
40
50
60
70
80
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32
Session Number
1st 7-Item List
2nd 7-Item List
3rd 7-Item List
4th 7-Item List

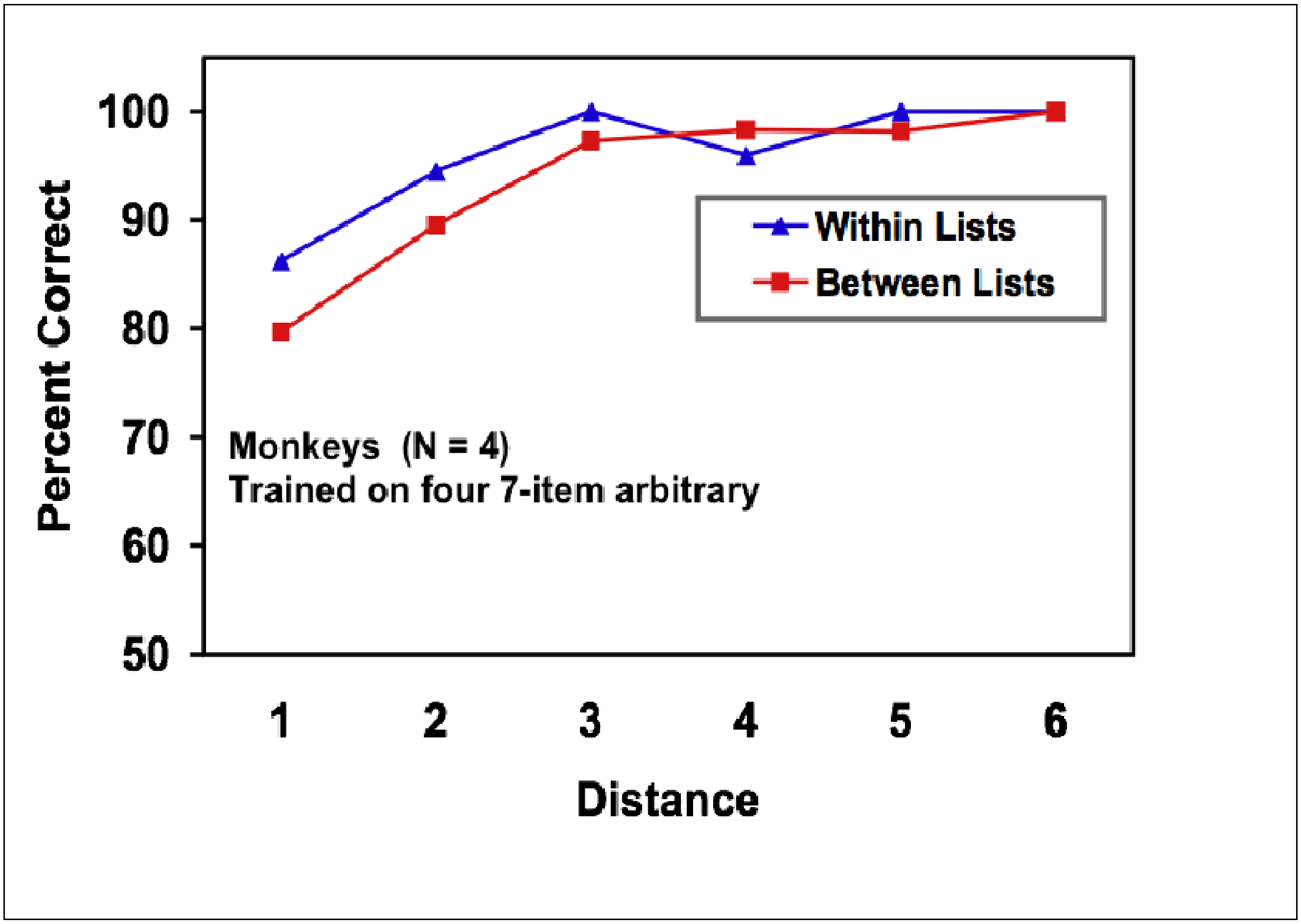


Figure 9.
Mean accuracy of responding to between- and within-list subsets as a function of the distance between items on original lists (Terrace, Son, and Brannon. 2003).

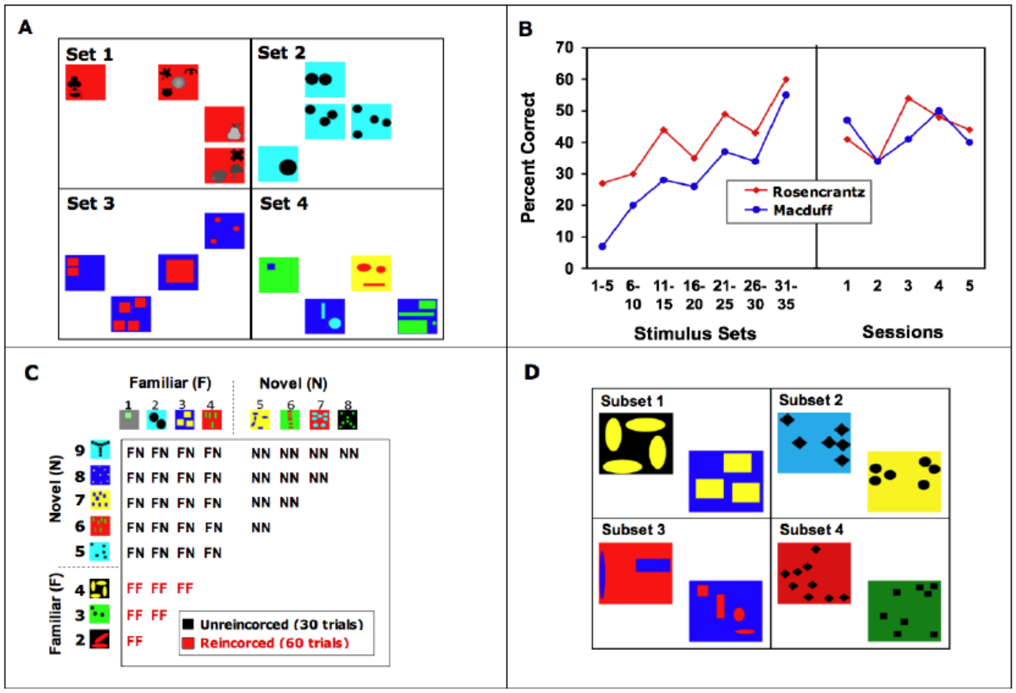


Figure 10.

Figure 10A. Examples of 4-item simultaneous chains, each composed of 4 numerical stimuli (Brannon & Terrace, 1998).

Figure 10 B. *Left panel:* Percentage of correctly completed trials by monkeys during the first session of each of 35 training stimulus sets (blocks of five sessions). *Right panel:* Percentage of correctly completed trials on the 150 test sets, each composed of novel stimulus (Brannon & Terrace, 1998).

Figure 10 C. The 36 pairs of the numerosities 1 through 9 used in subset test (Brannon and Terrace, 1998). These were defined with respect to the subjects' prior experience with the constituent numerosities: familiar-familiar (FF), familiar-novel (FN), novel-novel (NN). Only FF subsets were reinforced.

Figure 10 D. Exemplars of subsets composed of novel numerical stimuli (Brannon & Terrace, 1998).

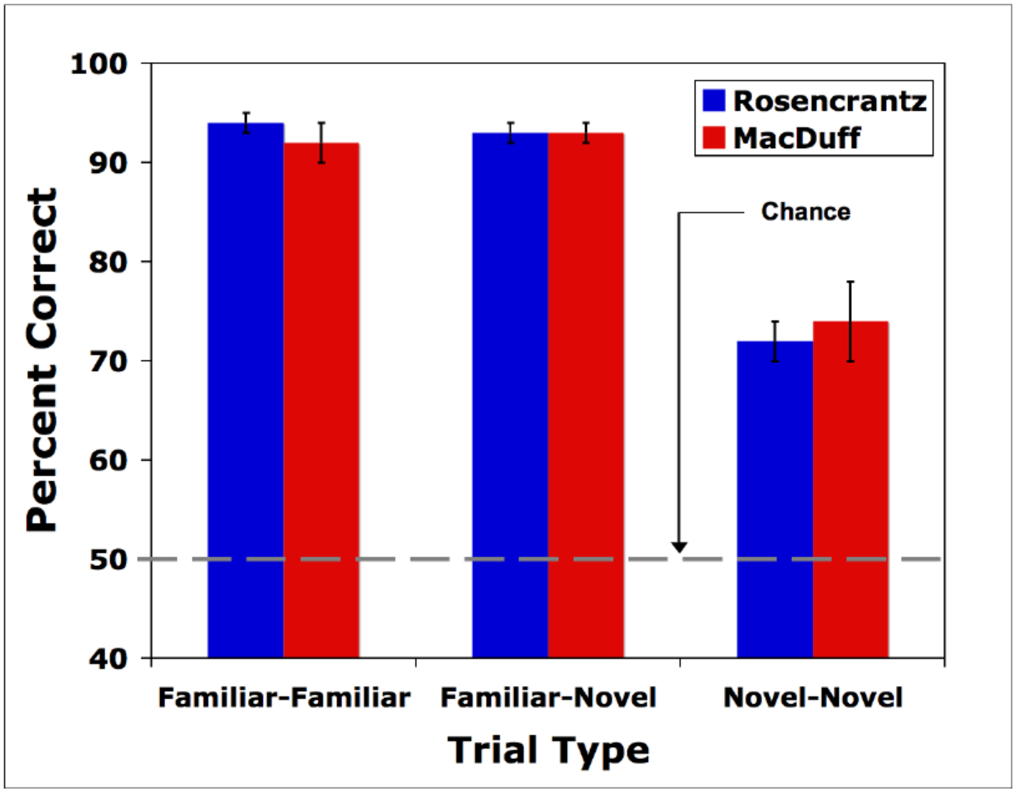


Figure 11.
Accuracy of responding by monkeys on familiar-familiar, familiar-novel, and novel-novel subsets (Brannon and Terrace, 1998).

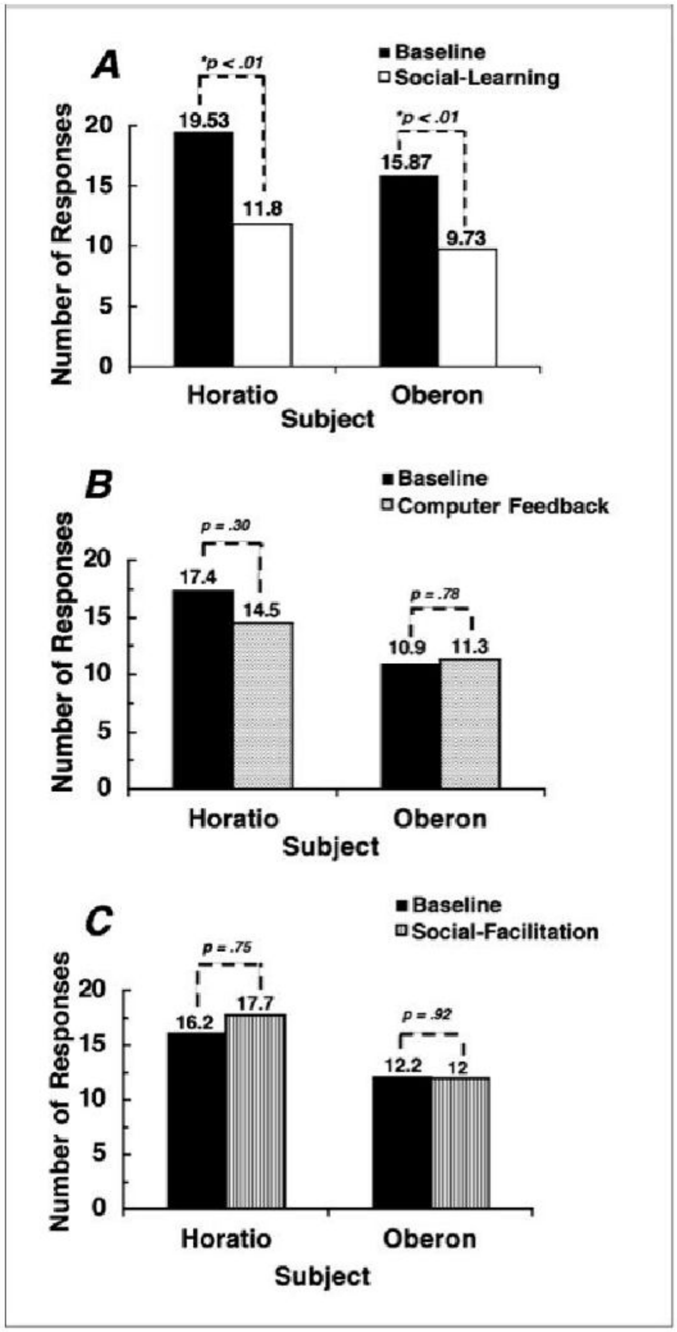


Figure 12.
Figure 12A. Summary data for cognitive imitation
Figure 12 B. Summary data for computer-generated feedback
Figure 12 C. Summary data for social facilitation

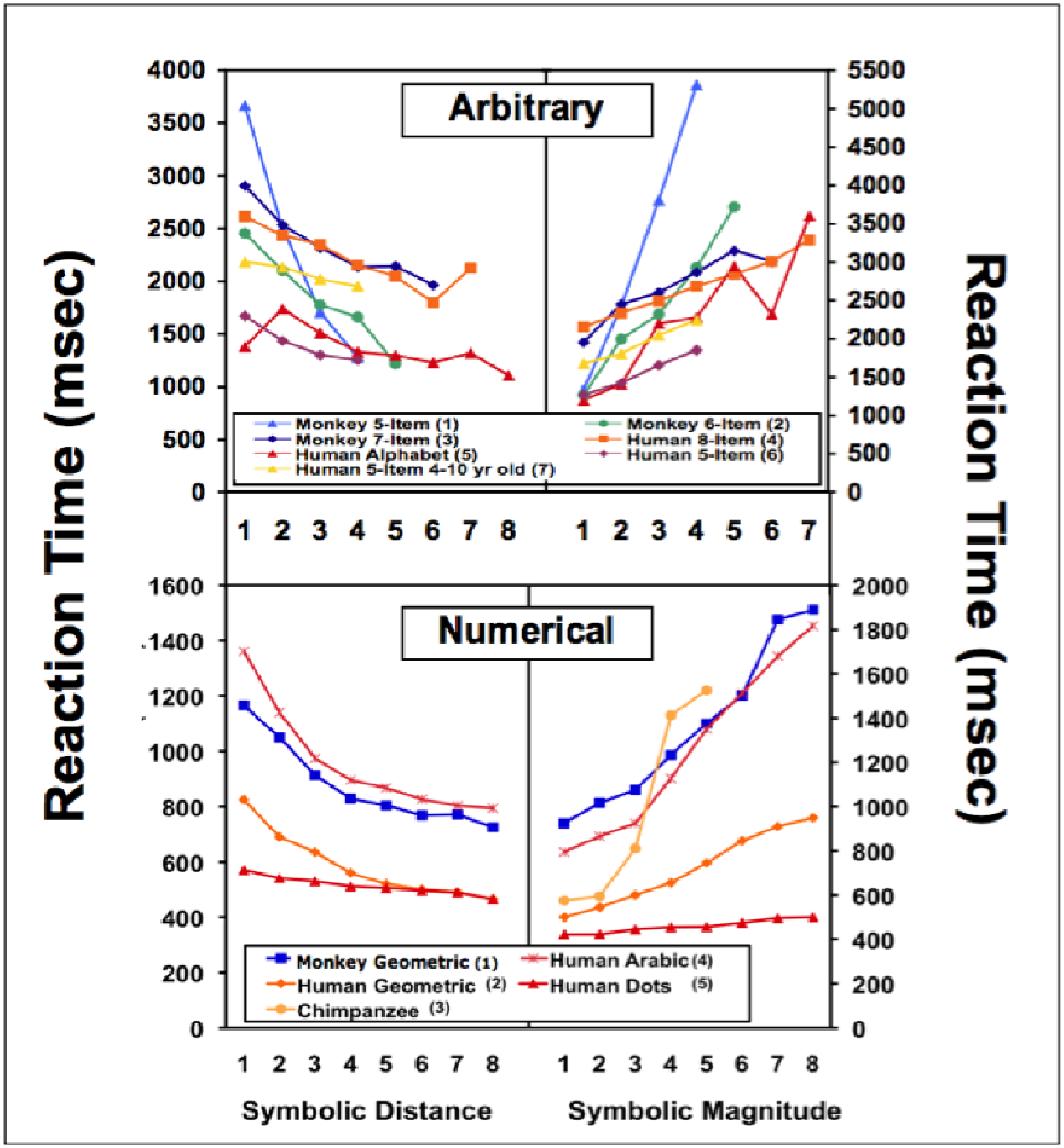


Figure 13.

Distance and magnitude functions obtained from human adults and children and monkeys and chimpanzees on subtest tests composed of numerical and arbitrary items (1) Brannon and Terrace, 1998 (2) Brannon and Terrace, 2002 (3) Murofushi, 1997 (4) D'Amato and Colombo, 1988 (5) Terrace, 2001b (6) Terrace et al., 2003 (7) Terrace, 2001, (8) Hamilton and Sanford, 1978, Brannon and Terrace, 2002 (9) Colombo and Frost, 2001, Guyla and Colombo, 2004 (10) Guyla and Colombo, 2004 (11) Brannon and Terrace, 2002 (12) Buckley, 1974 (13) Moyer and Landauer, 1967, Buckley, 1974. All of the nonhuman primates were first trained by the simultaneous chaining paradigm to learn lists of arbitrary or numerical stimuli. They were then tested with 2-item subsets composed of items from a given list. The data shown in this figure are the median reaction times of correct responses to the first item of each subset. Also shown are distance and magnitude functions obtained from human subjects that were tested on their ordinal knowledge of Arabic numbers and letters of the alphabet. These are similar to functions obtained from non-human primates.

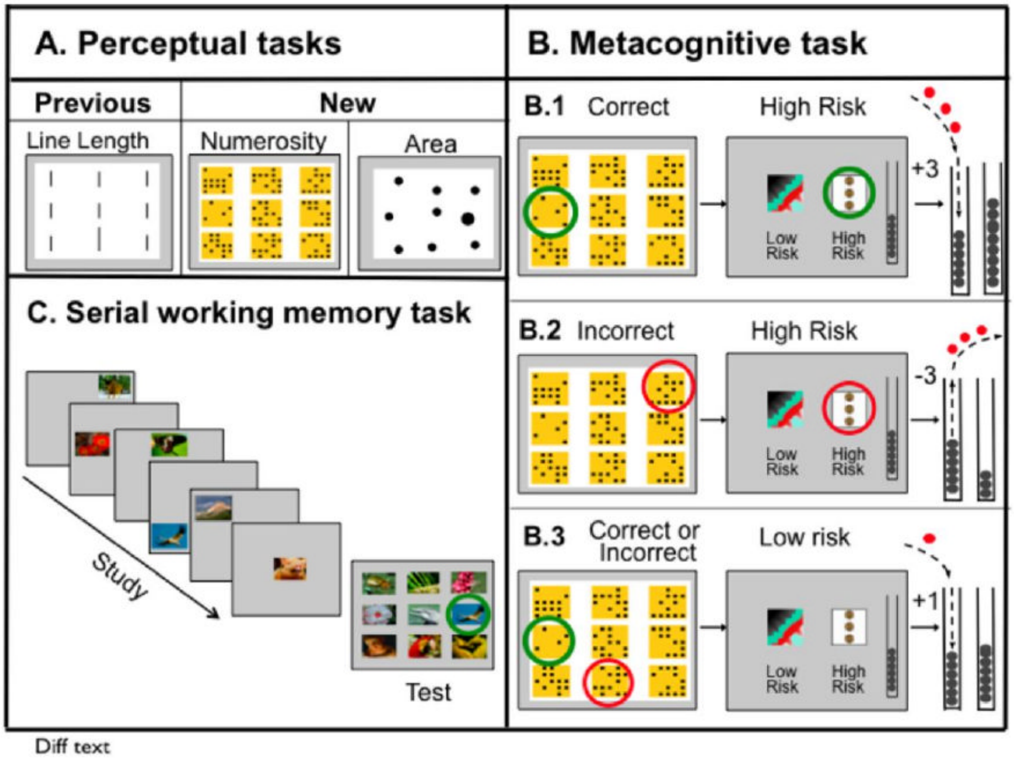


Figure 14.

Figure 14A. In Task 1 (trained prior to this study) subjects had to select the longest line. Task 2 was to select the item with the largest (or, for one subject, the smallest) number of items. Task 3 was to select the smallest (or, for one subject, the largest) circle.

Figure 14 B. In the metacognitive task, the high- and low-confidence icons were presented immediately after the subject made its selection on the perceptual task. A response to the high confidence icon resulted in a gain of 3 tokens after a correct response (b.1). Tokens were accumulated in a “bank” located to the right of the high- and low-confidence icons. There were 9 tokens in the bank at the start of every trial. When 12 or more tokens accumulated in the bank, two pellets were delivered and the number of tokens reset to 9. A response to the high confidence icon after an incorrect response resulted in the loss of 3 tokens (b.2). A response to the low-confidence icon always resulted in a gain of one token, whether or not the subject responded correctly (b.3).

Figure 14 C. Trial structure of the serial working memory task: six sample pictures were displayed successively, each separated by a 2 s interval. Following the last sample and a .5 s pause, the test phase of the trial began. One of the six sample photographs was selected at random. Either that stimulus or an unfamiliar stimulus was shown and the monkey had to identify it as familiar or unfamiliar. The same two confidence icons that were displayed during the perceptual tasks were presented immediately after the subject responded to the probe.