A|B|A|I
Association for Behavior Analysis International

Check for updates

Contrasting Accounts of Early Speech Perception and Production

Henry D. Schlinger Jr.[1] ⓘ

Abstract

Language researchers have historically either dismissed or ignored completely behavioral accounts of language acquisition while at the same time acknowledging the important role of experience in language learning. Many language researchers have also moved away from theories based on an innate generative universal grammar and promoted experience-dependent and usage-based theories of language. These theories suggest that hearing and using language in its context is critical for learning language. However, rather than appealing to empirically derived principles to explain the learning, these theories appeal to inferred cognitive mechanisms. In this article, I describe a usage-based theory of language acquisition as a recent example of a more general cognitive linguistic theory and note both logical and methodological problems. I then present a behavior-analytic theory of speech perception and production and contrast it with cognitive theories. Even though some researchers acknowledge the role of social feedback (they rarely call it reinforcement) in vocal learning, they omit the important role played by automatic reinforcement. I conclude by describing automatic reinforcement as the missing link in a parsimonious account of vocal development in human infants and making comparisons to vocal development in songbirds.

Keywords usage-based theories · cognitive linguistics · behavior analysis · language development · automatic reinforcement · infants · songbirds

✉ Henry D. Schlinger Jr.
hschlin@calstatela.edu

1 California State University Los Angeles, Los Angeles, CA, USA

In 2010, I published an article in which I contrasted behavioral and cognitive views of speech perception and production in language-learning children (Schlinger, 2010). In that article, I noted that "speech perception and language acquisition have been studied primarily by cognitively oriented researchers" and that "many of these researchers discount a behavioral account despite" (p. 150) the fact that evidence, even from many cognitively oriented studies, clearly demonstrates that early exposure to speech and interaction with a linguistic community play a significant role in both speech perception and production. Language researchers have also noted that reinforcing consequences from others as well as from infants' own vocalizations contribute to shaping their vocal repertoires; but they do not talk about it in those terms or acknowledge a behavior-analytic account of early speech perception or production. Instead, cognitive theories explain early speech perception and production with mental constructs, the only evidence for which are the very behaviors cognitive researchers observe. In the present article, I make the case that a behavior-analytic theory, based as it is on inductively and experimentally derived laws, provides an adequate and parsimonious account of speech perception and production. To begin, I describe usage-based theories of language acquisition, the most recent in a long line of cognitively based approaches, collectively called "cognitive linguistics" (Lieven, 2016). Many usage-based researchers claim that their theories explain language acquisition by appealing to learning mechanisms. However, these so-called learning mechanisms are vague and ill-defined and not the empirically validated ones discovered by learning scientists.

We should ask whether any "theory" can explain language acquisition without clearly identifying the underlying causal mechanisms. In 1950, B. F. Skinner asked, "Are theories of learning necessary?" (Skinner, 1950). By "theories" he meant explanations that elsewhere (Skinner, 1957) he called explanatory fictions, which have also been called circular explanations (Schlinger, 2018b) or circular statements (Vaughan & Michael, 1982). One notable drawback of most cognitive theories is that they usually consist of circular explanations; that is, the evidence for their explanations is almost always only the very behaviors that are observed. So, we might ask, "Are usage-based or other cognitive theories of language acquisition necessary?" This is a question to which I will return later. But before I contrast cognitive and behavioral approaches to speech perception and production, I briefly describe usage-based theories and evaluate whether or how well they explain language acquisition.

Usage-Based Theories of Language Acquisition

In evaluating usage-based theories of language acquisition, we can ask three questions: (a) What do usage-based theories assume about language? (b) What mechanisms do usage-based theorists propose to account for language acquisition? And (3) What are some problems with explanations based on usage-based theories?

In simple terms, and as their name implies, usage-based theories, unlike theories based on an innate generative universal grammar, stress that all knowledge about language, including meaning, comes primarily from using it in context. Usage-based theories "focus on how meaning-based grammatical constructions emerge from

individual acts of language use" and are based on two assumptions: "meaning is use," and "structure emerges from use" (Tomasello, 2009, p. 69). Usage-based theorists often cite a quote attributed to the philosopher Ludwig Wittgenstein, "Don't ask for the meaning; ask for the use." What Wittgenstein actually said was "the meaning of a word is its use in the language" (Wittgenstein, 1953). Incidentally, behavior analysts might interpret this as "meanings are to be found among the independent variables in a functional account, rather than as properties of the dependent variable" (Skinner, 1957, p. 14).

It is important to point out that usage-based theorists claim their theory is functional and pragmatic, but they use those terms differently than behavior analysts do. For example, Tomasello (2009) states that "in a usage-based view, one must always begin with communicative function" (p. 70) or "using linguistic conventions to achieve social ends" (pp. 69–70). On this view then, usage-based theories, like behavior-analytic theory, assume that the function of language behavior is to have some effect on listeners. The behavior-analytic view is that the behavior of speakers is reinforced through the mediation of others, called listeners (Skinner, 1957), who comprise the environment that determines the speaker's behavior. The usage-based approach, however, puts the determiners inside the individual. Hence, the title of a widely cited article: "From usage to grammar: The mind's response to repetition" (Bybee, 2006). The repetition refers to the frequency of language experience, but in usage-based theories, the mind must transform that experience into the cognitive structure called grammar. And even though usage-based theorists stress the importance of experience with language, the individual is still seen as the agent, making sense of the linguistic input, a position that is the diametrical opposite of behavior theory, which places locus of control (or causation) in the environment. This view is illustrated in such locutions as "by 3 to 4 years of age most children can readily *assimilate* novel verbs to abstract syntactic categories that they bring to the experiment" (Tomasello, 2000, p. 307; emphasis added). Of course, the only evidence that children *assimilate* novel verbs is that they use those verbs correctly.

Usage-based theorists also adhere to the branch of linguistics called pragmatics, which deals with language and the context in which it is used. As Tomasello (2009) put it, usage-based theorists "represent the view that the pragmatics of human communication is primary, both phylogenetically and ontogenetically, and that the nature of conventional languages—and how they are acquired—can only be understood by starting from processes of communication more broadly" (p. 70). If by "communication more broadly" Tomasello means the effects of a speaker's linguistic behavior on the behavior of listeners, then there is probably broad agreement with behavior theorists. For behavior theorists, however, a pragmatic approach is closest to the philosophy of pragmatism, which states that the truth of theories is assessed based on successful practical applications. The practical application of any science can be seen both in the laboratory and in the real world. In both instances, a pragmatic approach boils down to whether we can control and predict behavior. Of course, the control of behavior can only be accomplished using within-subject experimental designs where independent variables can be precisely manipulated to determine their effects on an individual's behavior. When it comes to understanding language, usage-based and behavior-analytic approaches view pragmatism differently. For

usage-based theories, pragmatism only means experiencing and using language in its context. For behavior theorists, it is the epistemological question of how we can know something and the criteria—control and prediction—by which we know it.

Contrary to its name, according to usage-based theories, linguistic structure (i.e., grammar) does not arise solely from using or experiencing language, but rather from the interaction between certain cognitive processes common to all humans and their experience using language. Or, as Bybee (2006) put it, grammar is "the cognitive organization of one's experience with language" (p. 711). So, there are two necessary ingredients: language experience and human cognition. One of the most important features of experience with language that is central to language learning, according to usage-based theorists, is the frequency with which certain linguistic forms are heard or used (e.g., Behrens, 2009; Bybee, 2006; Ibbotson, 2013; Tomasello, 2009).

Before moving to the second question about usage-based theories, I want to say a few words about the concept of frequency. Long before cognitive linguists began stressing frequency of linguistic input as an important determiner of language acquisition, Ernst Moerk, a behavior analyst, noted the importance of frequency and, by extension, of operant learning principles in early language acquisition. Moerk reanalyzed data from Brown's (1973) classic work on language acquisition (e.g., Moerk, 1980, 1983b, 1990, 1992; see also Segal, 1975). Brown was perhaps the first researcher to present exhaustive data on interactions between parents and language-learning children.[1] Brown presented longitudinal data on verbal interactions between a family of three children and their parents and among many other findings, concluded that neither parental frequencies nor perceptual salience influenced the order of development of grammatical forms (Brown, 1970, pp. 343, 362). In his reanalysis of Brown's data, Moerk (1980) came to quite a different conclusion: "that frequency of input was highly related to frequency of production" (p. 105). In addition, Moerk (1983a) found that

> Thirty-nine teaching techniques of the mothers and 37 learning strategies of the children were differentiated. The teaching techniques included conditioned positive reinforcement, obvious linguistic corrections, conditioned punishment, several forms of less obvious corrections, and various forms of modeling. . . . high frequencies of specific teaching techniques and of types of linguistic input were encountered. The interactions between the mothers and the children exhibited not only a considerable degree of structure, that is, the patterns occurred with a frequency that by far surpassed chance co-occurrences, but they also appeared largely to be instructionally highly meaningful. (pp. 129–130)

Thus, even before cognitive theorists abandoned an innate generative grammar approach in favor of usage-based approaches, behavior theorists were stressing the importance of frequency of linguistic input. But more than that, behavior theorists

---

[1] Other researchers have since presented comprehensive longitudinal data on verbal interactions between parents and language-learning children (Hart & Risley, 1995; Roy et al., 2006).

explained language acquisition according to the principles of operant learning (e.g., Moerk, 1983a, 1990; Skinner, 1957).

The second question regarding usage-based theories deals with the mechanisms that underlie language learning. Usage-based theorists assume that learning language is a unique human capacity with a genetic component given that every typically developing child learns language, and an environmental component given that children end up speaking different languages using different specific linguistic forms, so-called "language-specific properties" (Behrens, 2009). It might surprise behavior theorists to learn that usage-based theorists propose that children learn not only the "irregular and peculiar aspects of a language," but also "the more general and predictable patterns of" language by "general learning mechanisms" (Behrens, 2009, p. 384). However, the general learning mechanisms proposed by usage-based theorists are all cognitive, even though some seem to map onto behavior-analytic principles. For example, usage-based theorists assume that "language structure can be learned from language use by means of powerful generalization abilities" (Behrens, 2009, p. 384), which is why usage-based theories are also sometimes called *emergent*. As Behrens (2009) put it, "the concept of emergence tries to explain how relatively small genetic and behavioral differences lead to wide-ranging differences in cognitive abilities, including the competence to use a full-fledged linguistic system" (p. 388). For usage-based theorists, the generalization abilities are not the laws of (stimulus or response) generalization that experimental behavior analysts have discovered. Rather, they are cognitive processes.

According to Tomasello (2009), "children come to the process of language acquisition, at around one year of age, equipped with two sets of cognitive skills, both evolved for other, more general functions before linguistic communication emerged in the human species" (p. 69). Those two cognitive skills are intention-reading and pattern-finding. Intention-reading, which includes the skill of joint attention, refers to how children as listeners recognize the goals and intentions of mature speakers and then learn to use the same grammatical structures to achieve their own ends as speakers. According to Tomasello (2009), intention-reading "is the central cognitive construct in the so-called social-pragmatic approach to language acquisition" (p. 70). Pattern-finding refers to how children go beyond the individual utterances they hear and generalize to more abstract grammatical constructions (Tomasello, 2009). Notice that in both cases, it is the child who must make sense of the linguistic input by *recognizing* speakers' goals and intentions and *generalizing* individual utterances to more abstract constructions. But *recognizing* and *generalizing* are not actions; rather, they are cognitive constructs invented after the fact to explain the very behaviors used to infer them in the first place.

The so-called "general learning mechanisms" (Behrens, 2009, p. 384), or "basic psychological processes" (Behrens, 2009, p. 386), that usage-based theorists claim interact with experience to produce abstract grammatical forms include concepts such as entrenchment, categorization, and schema formation. Entrenchment means that frequently repeated constructions are stored in memory. But it is the mind that must recognize similarities and dissimilarities, weed out features that do not recur, and note commonalities by comparing stored units in memory (from entrenchment) with new units (Behrens, 2009). The mind then categorizes new units according to their similarity to

stored units. Finally, due to the processes of abstraction and generalization, schemas are formed. However, these hypothetical cognitive processes, like all proposed cognitive processes, are not objective, experimentally induced variables that can be studied independently of the behaviors they are said to explain (see Schlinger, 1993). This brings us to our third question.

Our third question about usage-based theories is evaluative. We can assess usage-based theories according to their explanatory schemes and their ability to predict and control the utterances of individual speakers. And this ability is based on the research conducted by usage-based theorists (reviewed by Lieven, 2016), much of which is correlational (e.g., Lieven, 2008; Theakston et al., 2005). The research, which is experimental, is what I have called demonstration research (Schlinger, 2004). In demonstration research, researchers only demonstrate on average the occurrence of certain behaviors under certain circumstances, but they do not carry out the experimental analyses of an individual's behaviors necessary to identify the causal mechanisms. One problem, as already indicated, is that cognitive mechanisms proposed to account for language acquisition are not objectively defined. Nor are these mechanisms observed independently of the linguistic behaviors they are said to explain. Also, talking about cognitive events like schemas and grammatical structures as if they are real entities commits the reification error. Usage-based theorists may not actually believe that these cognitive structures have a concrete existence separate from the behaviors of interest, but they talk like they do. There has also been an emphasis more on linguistic forms than on their function (Lieven, 2016). Finally, as I have argued elsewhere (e.g., Schlinger, 2010), there are more parsimonious explanations of linguistic behavior that rely on objective, independently verifiable variables, which I will describe throughout the remainder of this article.

Let me conclude by returning to the rationale for the discussion of usage-based theories in the first place: whether usage-based cognitive theories can adequately explain language acquisition. Having reviewed several articles by prominent usage-based theorists, I have found that explanations of language acquisition are circular. Moreover, the so-called learning mechanisms they propose are not empirically derived. Finally, the explanations offered by usage-based theorists are less than parsimonious because they make too many assumptions, mostly about unobserved and unobservable cognitive processes. So, to the question of whether usage-based theories are necessary to explain language acquisition, we must answer "no." One of the main points I make in the present article is that a behavior-analytic theory of speech perception and production is parsimonious because it focuses on behavior and the environmental variables of which it is a function—and nothing more. Other researchers have noted that before language researchers invoke de novo learning mechanisms, they should first rule out principles based on operant conditioning (Sturdy & Nicoladis, 2017). That is the approach I take in the present article.

Contrasting Cognitive and Behavior-Analytic Theories

To set the stage for a more detailed comparison, I briefly contrast cognitive with behavior-analytic theories. Behavior-analytic theory, like other theories in the natural sciences, is inductively derived. Simply speaking, behavior-analytic

theory consists of statements (and sometimes equations) that are summaries of thousands of observations of repeatable functional relationships between behavior and environmental variables discovered in the experimental laboratory. These summary statements—the laws of behavior analysis—are then used deductively to explain novel behavioral relations, those that are too complex to study in a controlled setting, or behaviors that cannot be presently observed. Cognitive theories, on the other hand, are proposed absent the discovery of experimentally derived repeatable functional relations between independent and dependent variables. In fact, cognitive scientists cannot discover such relations because their independent and dependent variables (i.e., cognitive structures and processes) have never been and can never be directly observed and measured. Putative cognitive structures, events, and processes (e.g., memories, schemas) are always inferred after the fact only from observed behavior. Behavior analysts simply cut out the middleman and study behavior in its own right, in other words, not as a reflection of some underlying hypothetical cognitive or mentalistic structures or processes. Their efforts have been rewarded with the discovery of numerous laws. Behavior analysis is pragmatic in that its applications, both in the experimental laboratory and in the applied setting, have been extremely successful; that is, they lead to control (through experimentation) and prediction. Thus, behavior-analytic theory is parsimonious because it makes few assumptions, and the ones it does make—those about behavior and environmental variables—can, for the most part, be independently observed and measured and, thus, experimentally tested.

To conclude, I have described usage-based theories of language acquisition and briefly contrasted them with a behavioral approach. I would be remiss, however, if I did not point out a few major areas of agreement. For one, both theories agree that language is learned based on experience with language, both hearing and speaking it. Second, both approaches eschew innateness theories of language acquisition. Third, although both theories do not dismiss the form of language, they emphasize its function, albeit in different ways.

A Behavioral View of Perception

Before specifically addressing cognitive approaches to speech perception and contrasting those approaches with behavioral ones, it might be helpful to look at how behavior theorists view *perception*. To begin with, perception, as a noun, does not denote a person, place, or thing. As I have argued about similar terms (e.g., Schlinger, 2008a), perception is not a real process or structure, even though most psychologists talk about it that way (that is, they reify it). A more scientific approach is to think about "perception" simply as a word we use to describe behaviors in their context. So, if for example, you say that I *perceive* the keyboard in front of me, and I ask you why you said that, in other words, what led you to say that I perceive the keyboard, you would probably say that I look at it (which I have to do because I do not touch type), press the keys, and call it a keyboard, among other possible behaviors. Those are all behaviors of mine that you observe and that lead

you to say that I *perceive* the keyboard. In other words, there is no perception of the keyboard separate from my behaviors of interacting with it. If you didn't observe me interacting at all with the keyboard, you would be hard-pressed to say I perceived it. Incidentally, all my "perceptual" behaviors were acquired through operant learning. Thus, although perception is not a thing, the verbal response "perception" is, if only fleetingly. So, let us look at speech perception and production from both cognitive and behavioral perspectives to understand how a behavior-analytic theory is superior to the task of parsimoniously explaining the observed facts.

Cognitive Views of Speech Perception

Language acquisition researchers refer to hearing and speaking language as speech-perception and speech-production, respectively. Regarding speech perception and production, both behavior-analytic and cognitive theories deal with the same set of observed facts. After all, no matter one's theoretical orientation, there is only one set of observed facts: behavior in its context. The difference, however, is the mechanisms proposed to explain those facts and the methods used to study them, which have important implications both for the control and prediction of behavior and for teaching both typically developing and language-delayed individuals.

> Consider the following description:

> The ease with which a listener perceives speech in his or her native language belies the complexity of the task. A spoken word exists as a fleeting fluctuation of air molecules for a mere fraction of a second, but listeners are usually able to *extract the intended message*. (Holt & Lotto, 2008, p. 42; emphasis added)

This brief description conveys the essence of a cognitive approach to speech perception: that speakers have intentions, that words contain meanings, and that listeners extract meanings. Of course, the only evidence for extracting the intended meaning of an utterance is what the listener specifically does, which is rarely mentioned. In a behavioral account, a speaker's verbal response is itself evoked by the current circumstances (including the presence of a listener) and, in turn, evokes verbal (and nonverbal) behavior in the listener, all because of a history of operant learning. When we contrast this approach to perception with a cognitive account, it is easy to understand why the behavioral approach has not fared well. The cognitive account, despite its logical and scientific problems, is more familiar and accessible, incorporating everyday terms that everyone understands. And it is consistent with the dualistic approach that has defined psychology and philosophy for hundreds of years (see Schlinger, 2018a) in which there are behaviors on the one hand and mental or cognitive events on the other hand.

Researchers often assess auditory perception in infants by using so-called habituation and dishabituation methods (see Thomas & Gilmore, 2004, for a review and critique), for example, by measuring nonnutritive sucking (e.g., Byers-Heinlein, 2014) or changes in heart rate (see Von Bargen, 1983, for a review and critique). One example of a seemingly simple speech perception phenomenon that has

received a considerable amount of attention from cognitive linguistic researchers is categorical perception.

Categorical Perception

Categorical perception refers to the fact that both infants and adults discriminate different categories of phonemes. A phoneme is "the smallest unit of sound that signifies a difference in meaning in a given natural language" (Aslin, 1987, p. 68). Behavior theorists, however, view the phoneme functionally as "the smallest unit of sound that exerts stimulus control over behavior" (Schlinger, 1995, p. 153). For the present purposes, whatever behavior is evoked by the sound of a phoneme is what we mean when we speak of *categorical perception*.

Beginning in the 1970s, researchers began using computers to present synthetic consonant-vowel (CV) sounds that ranged across several consonants (e.g., /bV/-/dV/-/gV/). These synthetic speech sounds varied along a stimulus dimension called voice onset time (VOT). Voice onset time refers to "the point at which vocal cords begin to vibrate before or after we open our lips" (Bates et al., 1987, p. 152). Thus, for sounds we react to as *b* (e.g., *ba*), voicing begins either before or simultaneously with the consonant burst. For sounds that we react to as *p* (e.g., *pa*), voicing begins after the consonant burst. Computers can present stimuli along this VOT continuum from -150 to +150 ms from burst to voice. Studies have shown that English-speaking adults and infants do not respond differentially to VOTs that fall either significantly above or below the boundary between *pa* and *ba*, which is about 25–30 ms. More important, infants only a few months old respond differentially (as measured by differential sucking rates on a nonnutritive nipple) to VOTs within these phonemic categories (Eimas et al., 1971). These findings suggested to some researchers that humans are born with "phonetic feature detectors" that evolved specifically for speech and that respond to phonetic contrasts found in the world's languages (Eimas, 1975).

Other studies, however, have indicated a strong experiential component to categorical perception. For example, nonhuman animals (e.g., chinchillas, monkeys, Japanese quail, and rats) have been trained to respond to phonemic categories just as human adults and children have (e.g., Dooling et al., 1995; Kluender et al., 1987; Kuhl, 1981; Kuhl & Miller, 1975, 1978; Kuhl & Padden, 1982, 1983; Reed et al., 2003; Toro et al., 2005). Further evidence of a significant experiential component is that categorical perception in adults is limited to the phonemes in their respective native languages (Miyawaki et al., 1975).

These and other studies forced researchers to reconsider the prevailing view at the time that infants were born with a discriminative capacity evolved specifically for speech. The alternative view was that infants inherited a general capacity to discriminate auditory stimuli, including speech sounds. Behavior theorists recognize this as the inherited capacity for the behavior of infants to be operantly conditioned and to come under stimulus control. Even cognitive theorists acknowledged that domain-general, rather than species-specific, mechanisms seem to be responsible for infants' tendency to respond to phonetic units (Kuhl, 2000). As Kuhl (1981) stated,

"the evolution of the sound system of language was influenced by the general auditory abilities of mammals" (p. 347). Or, as Bates et al. (1987) put it,

> We assumed that the human auditory system evolved to meet the demands of language; perhaps, instead, language evolved to meet the demands of the mammalian auditory system. This lesson has to be kept in mind when we evaluate other claims about the innate language acquisition device. (p. 154)

Overall, research supported the view that categorical perception results largely from experience and learning. Because researchers have successfully trained discriminative responses to phonemic sounds in nonhuman animals using operant conditioning procedures, behavior theorists would assume that similar contingencies operate naturally for human infants. Many language researchers now agree that the language environment exploits (and can modify) natural boundaries of a general auditory capacity that is common to mammals and some birds (Diel et al., 2004). The question is how that modification occurs, that is, what learning mechanisms are involved. Rather than looking to empirically established principles of learning discovered by operant researchers, however, cognitive linguists have constructed language learning mechanisms de novo (Sturdy & Nicoladis, 2017). To wit, even though many cognitive linguists, including usage-based theorists, propose that certain aspects of language are experience-dependent (vs. experience-independent), and even refer to the experience as "learning," they propose new forms of learning that are mostly human-language-specific instead of appealing to empirically based principles of operant learning (e.g., Saffran et al., 1996). These researchers attribute the similarities across languages to general learning processes that did not evolve solely for language (Saffran, 2003; Tomasello, 2009). But they rarely mention operant conditioning principles (e.g., Kuhl, 2000).

Language researchers have concluded that experience and learning (though not operant learning) play a critical role in language acquisition and speech perception based on the observation that "by simply listening to language, infants acquire sophisticated information about its properties . . ." (Kuhl, 2000, p. 11852), a phenomenon also referred to as "incidental language learning" (Saffran et al., 1996). One example of this incidental language learning is called *statistical learning* (Kuhl, 2000; McMurray & Hollich, 2009).

Statistical Learning

According to Kuhl (2004), "The acquisition of language and speech seems deceptively simple" (p. 831). By that she means that children appear to learn their native language quickly and effortlessly. She wonders, then, how children, but not language theorists, have cracked "the speech code" so easily. This is a little like asking how children cracked the "walking code" or the "throwing the ball" code before physicists understood the laws of motion and gravity. Kuhl states that "children learn rapidly from exposure to language, in ways that are *unique to humans*, combining pattern detection and computational abilities . . . with

special social skills" (p. 831; emphasis added). She is talking about what language researchers refer to as *statistical learning*.

Statistical learning refers to the claim that "infants exploit the statistical properties of the input, enabling them to detect and use distributional and probabilistic information contained in ambient language to identify higher-order units" (Kuhl, 2000, p. 11852). For Kuhl (and many other cognitive linguists), statistical learning is the mechanism "responsible for the developmental change in phonetic perception between the ages of 6 and 12 months" (Kuhl, 2004, p. 833; see also Maye et al., 2002). According to Kuhl (2000),

> Running speech presents a problem for infants because, unlike written speech, there are no breaks between words. New research shows that infants detect and exploit the statistical properties of the language they hear to find word candidates in running speech before they know the meanings of words. (p. 11852)

Thus, researchers have demonstrated that by 6 months of age, infants prefer the phonetic units of their native language (Kuhl et al., 1992). These researchers describe the changes in infant speech perception as a reduction in the ability to discriminate speech sounds that are not found in one's native language. Because "the beginnings and ends of sequences (i.e., the segmentation) of sounds that form words in a particular language are not marked by any consistent acoustic cues" (Aslin et al., 1998, p. 321), such as pauses, and because the acoustic structure of speech across different languages is highly variable, researchers believe that children must use a distributional, rather than an acoustical, analysis to solve the problem of finding the words in a particular language. A distributional analysis refers to the regularities in the relative positions of sounds over a large sample of linguistic input (Aslin et al., 1998). For example, in English, "certain combinations of two consonants are more likely to occur within words whereas others occur at the juncture between words. Thus, the combination 'ft' is more common within words whereas the combination 'vt' is more common between words" (Kuhl, 2000, p. 11853).

According to many language researchers, infants need to discover the phonemes and words in a particular language. Because the speech they are exposed to is so variable and not marked by reliable acoustic cues, researchers believe that "infants use computational strategies to detect the statistical and prosodic patterns in language input" (Kuhl, 2004, p. 831). Some of these researchers are quick to point out that infants are not consciously calculating statistical frequencies, but rather are sensitive to distributional information contained in the linguistic input to which they are exposed (Aslin et al., 1998). Notwithstanding this one disclaimer, most of these researchers still talk about infants, or their brains, extracting statistical information from the linguistic input. Based on the results of certain studies (e.g., Saffran et al., 1996), numerous researchers have concluded that "infants use statistical information to discover word boundaries" (Aslin et al., 1998, p. 321), or they learn "from exposure to the distributional patterns in language input" (Kuhl, 2004, p. 835). However, such conclusions suffer from numerous logical and scientific problems.

Problems with a Cognitive Account of Speech Perception

There are several logical and scientific problems with a statistical learning account of speech perception. The first concerns the issue of agency. Cognitive theorists misplace the agency producing the effects (i.e., the observed behaviors) and put it inside the infant instead of in the linguistic environment, which cognitive linguists clearly believe is critical for such learning. This is illustrated repeatedly in the way researchers talk about language learning infants. For example, Saffran et al. (1996) state that, "One task faced by all language learners is the segmentation of fluent speech into words" (p. 1927). According to Kuhl (2004), "infants use computational strategies to detect the statistical and prosodic patterns in language input, and that this leads to the discovery of phonemes and words" (p. 831). Thus, infants are faced with tasks: they *extract* information, *use* or *exploit* strategies, *abstract* patterns, *discover* rules, and so on. Of course, these are not action verbs and do not specify any behavior. Sometimes it is the (infant's) brain which is assigned the task. For example, the brain is said to be endowed with mechanisms that enable it "to extract the information carried by speech" and to use those mechanisms "to discover abstract grammatical properties" (Mehler et al., 2008, p. 434). But brains do not act, organisms do. Either way, in cognitive accounts of speech perception, the locus of control is placed inside the individual. These ways of talking about what infants presumably do are just redundant descriptions of the observed behaviors in certain contexts. They do not point to specific actions, and when used as explanations of the same behaviors, they are circular.

In general, cognitive linguistic researchers say that it is the job of language learners to make sense of or to detect patterns in vague or complex linguistic information through inferred cognitive processes (e.g., Kuhl, 2000). In usage-based theories, it is called pattern-finding (Tomasello, 2009). This account is at odds with a natural science approach, which looks for physical causes of behavior. Using an evolutionary analogy, it would be akin to saying that the task for individual organisms (or their brains) is to exploit strategies to discover the rules for how to survive in a complex environment. But as Charles Darwin (and Alfred Russell Wallace) correctly theorized, the direction of causation is the other way. The environment selects traits to the extent that on average those traits enable individuals possessing them to live long enough to pass on their genes. A selectionist account of language learning suggests that only some responses of infants to specific stimuli will produce reinforcing consequences and, therefore, continue to occur.

Another problem with some cognitive accounts of language learning is that the questions are the wrong ones to ask. For example, Saffran (2003) asked what infants were learning in a segmentation task: "Are they learning statistics? Or are they using statistics to learn language?" (p. 112). The answer, of course, is neither. To understand what is wrong with questions such as these and with the notion of statistical learning in general, we must distinguish between the researchers' behavior and that of the infants. It is true that a researcher can statistically analyze conditional probabilities of certain sounds or arrangements of sounds within a stream of speech. But neither infants nor their brains are literally carrying out statistical analyses based on the distributional patterns in the speech they hear any more than they are literally

calculating force, resistance, or gravity when they walk. The only evidence of statistical learning in infants is that hearing speech produces changes in their behavior. It is the researchers who are doing the statistical analysis, not the infants. The principle of parsimony suggests that explanations should make the fewest assumptions. Suggesting that infants or their brains are statistically analyzing speech sounds makes too many assumptions and is simply not necessary as an explanation.

In addition to conceptual and logical problems with a cognitive account of speech perception, there are also methodological problems. First, cognitive researchers almost always employ between-subject designs and rarely, if ever, use within-subject, repeated-measures, experimental designs, in which each subject is exposed to the independent variable(s) and serves as his or her own control. This almost exclusive reliance on between-subject designs may be responsible in part for the replication crisis that has plagued psychological research (Normand, 2016). For some experimental questions, a between-subjects design is appropriate and called for. However, for many other research questions, a within-subjects, repeated-measures design would allow researchers to identify the causes of the individual's behavior. This ability to control the behavior would also allow researchers to predict the behavior more accurately.

Another methodological problem is that the infants used in many studies on speech perception are already at least 6 months old, which means that they have had countless interactions with speakers, which researchers acknowledge contributes to speech perception and language learning. Finally, researchers who cite animal studies as evidence against a uniquely human capacity for speech perception claim that statistical learning is uniquely human (Kuhl, 2000, 2004). In fact, studies employing operant conditioning procedures have shown that rats can be taught to perceive (i.e., discriminate) the nuances of human speech (e.g., Reed et al., 2003; Toro et al., 2005), suggesting that operant learning is a plausible explanation for how infants learn to discriminate speech sounds.

A Behavioral View of Speech Perception and Production

Speech Perception

Based on a behavioral theory of perception in general, a behavioral theory of speech perception focuses on what an individual does when we say that he or she *perceives* speech and attributes the acquisition and maintenance of such behavior to operant learning principles. With respect to infants, we can ask, for instance, what behaviors under what circumstances cause us to say that infants *perceive* speech? We can answer that infants *perceive* speech if they turn their head toward the person producing the speech sounds, or if they smile or make their own sounds. In research with infants, *speech perception* refers to responses to speech sounds researchers measure with changes in heart rate and nonnutritive sucking. In the natural environment, the term *speech perception* refers to such behaviors as turning one's head in the direction of the speech, smiling, and making sounds. Saying that infants are "extracting information" from such stimuli adds nothing to the description and muddies the

search for causal variables. In more sophisticated listeners, *speech perception* refers to a much wider range of behaviors, including complying with requests (what behavior analysts call "listener behavior") and listening, that is, subvocally echoing or otherwise talking to oneself about what the speaker is saying (see Schlinger, 2008b). In addition to identifying behaviors that occur when we speak of speech perception, a behavioral theory parsimoniously explains how they come about in the first place by appealing only to empirically established principles, such as reinforcement.

Cognitive developmental linguists have demonstrated the importance of reinforcement in speech perception in infants. Toward that end, researchers have shown that newborn infants prefer to listen to their mother's voice. By *prefer* researchers mean that infants will engage in behaviors that result in hearing (and are reinforced by) (1) the language spoken by their mother during the last trimester of pregnancy; (2) their mother's voice more than another woman's voice; and (3) specific features of their mother's voice over other features (DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Moon et al., 1993; Nazzi et al., 1998). In behavioral terms, at birth or shortly thereafter, certain phonemes in the infant's native language in general and the mother's voice in particular have become potent conditioned reinforcers such that when such stimuli are presented contingently on some infant behavior (e.g., specific sucking patterns), that behavior increases relative to behavior that does not produce those features. These stimuli become conditioned reinforcers (and probably acquire other behavioral functions as well, for example, as conditional stimuli) simply by hearing them. In other words, pairing with other reinforcing stimuli does not appear to be necessary, although that probably regularly happens in the course of daily interactions between parents and infants. Some of these phonetic stimuli, in particular vowel sounds, can apparently acquire their conditioned reinforcing properties even before birth *in utero* (Moon et al., 2013). Thus, reinforcement theory explains "how listeners come to perceive sounds in a manner that is particular to their native language" (Diel et al., 2004, p. 164). Moreover, we do not need to appeal to statistical analyses as explanations. We also do not need to appeal to ad hoc cognitive processes, such as "perceptual representations of speech . . . stored in memory" (Kuhl, 2000, p. 11854) or entrenchment (e.g., Ibbotson, 2013; Tomasello, 2009) because the only evidence for such explanations are the very behaviors to be explained. The behavioral explanation is parsimonious because it makes few assumptions, and it appeals to empirically established laws of behavior. Operant learning principles are much easier to demonstrate in the production of speech, which, ironically, sometimes involves the very same behaviors as perceiving speech.

Speech Production

Infants naturally progress through periods of vocal development. However, even though this progression obviously has a strong maturational component, the sounds are undoubtedly influenced by reinforcement. Early studies demonstrated both social and nonsocial operant conditioning of vocalizations in 3-month-old infants (e.g., Rheingold et al., 1959; Weisberg, 1963). However, Poulson and Nunes (1988) concluded that these early studies did not sufficiently establish the control procedures to demonstrate a clear-cut reinforcement effect. According to Poulson and Nunes,

only two early studies—Sheppard (1969) and Poulson (1983)—did meet such requirements. Since then, other studies have demonstrated the effects of social reinforcement on infant vocalizations (e.g., Goldstein et al., 2009; Pelaez et al., 2011a, 2011b).

Both anecdotal and experimental observations suggest that infants in the first year of life learn to produce not just the intonation and prosody of the language that they hear but the sounds as well (e.g., Levitt & Utman, 1992; Whalen et al., 1991). Research suggests that even the melodic cries of newborns are influenced by hearing the prosodic features of their native language as early as the third trimester of pregnancy (Mampe et al., 2009). Despite the admission that vocal learning depends on hearing the vocalizations of others and of oneself, language researchers admit that "little is known about the processes by which changes in infants' vocalizations are induced" (Kuhl & Meltzoff, 1996, p. 2425). That does not prevent these researchers, however, from inventing ad hoc cognitive explanations. For example, according to Kuhl and Meltzoff (1996), "infants listening to ambient language store perceptually derived representations of the speech sounds they hear which in turn serve as targets for the production of speech utterances," and both adults and infants "have an internalized auditory-articulatory 'map' that specifies the relations between mouth movements and sound" (p. 2426). These explanations are like those proposed by usage-based theorists mentioned previously that appeal to such cognitive "learning" mechanisms as entrenchment, categorization, and schema formation. The problem with such explanations is that they (a) require many untestable, unfalsifiable assumptions about unobserved, inferred events; (b) are often just redundant descriptions of the observed behaviors; and (c) are not necessary to explain those behaviors. A behavioral account, by contrast, looks at the direct relationship between infants' vocalizations and their possible reinforcing consequences. The role that reinforcement can have on infants' speech can be clarified by comparing vocal development in human infants with that of songbirds.

The Reinforcement of Vocal Sounds in Infants and Songbirds

Some researchers have noted many parallels between vocal development in humans and the development of songs in certain species of birds (Brainard & Doupe, 2002; Doupe & Kuhl, 1999; Kuhl, 2000, 2004). For one, social contingencies of reinforcement play an important role in vocal learning in both human infants and songbirds (Goldstein et al., 2003, 2009). But it is not only songbirds who benefit from postnatal experience. Lickliter and colleagues (e.g., Harshaw & Lickliter, 2007) have demonstrated that "postnatal presentation of an individual maternal call contingent on quail neonates' own vocalizations dramatically modifies the acquisition and maintenance of their species-typical auditory preferences in the first days following hatching" (Lickliter & Bahrick, 2016, p. 10). The postnatal presentation of the maternal call functions as reinforcement. In addition, researchers agree that hearing the vocalizations of others and of oneself is necessary for vocal development in infants (Kuhl & Meltzoff, 1996) and in songbirds (Brainard & Doupe, 2002; Doupe & Kuhl, 1999). In infants and in many songbirds, immature vocal sounds

are shaped into more mature sounds in large part by the feedback produced by making sounds. However, as Sturdy and Nicoladis (2017) have noted, these researchers do not mention reinforcement or operant learning, even though many describe how "infants' *successive approximations* of vowels would become more accurate" due to the "acoustic consequences of their own articulatory acts" (Kuhl & Meltzoff, 1996, p. 2426; emphasis added), or how "during sensorimotor song learning, motor circuitry is gradually *shaped* by performance-based feedback to produce an adaptively modified behaviour" (Brainerd & Doupe, 2002, p. 355; emphasis added).[2] Successive approximations and shaping are operant learning concepts. Other researchers describe how sounds emitted by infants and songbirds "are then gradually molded to resemble adult vocalizations" (Doupe & Kuhl, 1999, p. 574). Finally, some researchers explicitly acknowledge a selection process involved in early vocal production. For example, de Boysson-Bardies (1999) writes,

> The vocal productions of children are thus modeled by selection processes. The phonetic forms and intonation patterns specific to the language of the child's environment are progressively retained at the expense of forms that are not pertinent to the phonological system of this language. The process begins at birth, if not before. However, the first effects on vocal performance are delayed, particularly by the slow course of motor development. (p. 56)

Even though these scholars are describing operant conditioning processes (i.e., reinforcement) as a form of selection by consequences (see Skinner, 1981), they do not acknowledge it. Perhaps it is because they hold a limited view of reinforcement as something tangible deliberately given to one individual by another. In fact, even Sturdy and Nicoladis (2017), who have encouraged language researchers to exhaust operant explanations before inventing de novo learning mechanisms, appear to hold this limited view of reinforcement. In their article, they cite the work by Goldstein and colleagues (Goldstein et al., 2003; see also Goldstein & Schwade, 2008, and Goldstein et al., 2009), which has shown that contingent social attention by mothers increased the vocalizations of 8-month-old infants. Such results, however, would not have been a surprise to operant researchers who for decades have produced research showing the effects of social reinforcement on infants' vocalizations (e.g., Dunst et al., 2010; Poulson, 1983, 1988; Rheingold et al., 1959; Sheppard, 1969; Todd & Palmer, 1968; Weisberg, 1963).

The Missing Link: Automatic Reinforcement of Vocal Learning in Infants and Songbirds

Perhaps a clearer understanding of just what the law of reinforcement is would help language researchers in their quest for general learning mechanisms to explain speech perception and production. Reinforcement, as a consequence produced by

---

[2] Numerous studies have demonstrated the operant conditioning of vocalizations in a variety of species, including male zebra finches in which hearing their own song functioned as a reinforcer for song learning before sexual maturity (see Adret, 1993a, 1993b).

behavior, is not defined based on any formal characteristics, such as where the reinforcement originates or what it looks or feels like, but by its effect on behavior, which is to increase similar responses under similar circumstances. Much of what reinforces behavior consists of the stimuli produced by the behavior itself and, thus, is not deliberately mediated by other individuals. Skinner (1957) called this automatic reinforcement and wrote,

> Automatic reinforcement may shape the speaker's behavior. When, as a listener, a man acquires discriminative responses to verbal forms, he may reinforce himself for standard forms and extinguish deviant behavior. Reinforcing sounds in the child's environment provide for the automatic reinforcement of vocal forms. (p. 164)

Of course, Skinner did not mean that the speaker literally reinforces himself. He meant that if the sounds a speaker produces resemble the sounds that speaker has heard there will be an immediate strengthening effect, increasing the probability of similar sounds under similar circumstances. Automatic reinforcement plays a crucial role in the development of the vocal repertoire of human infants, parrots, and some songbirds.

Like infants, parrots and songbirds start out with a repertoire of immature or unrefined sounds. When they hear themselves making sounds that match what they have heard from others, those sounds are automatically strengthened (i.e., reinforced), in the sense that they occur with a greater frequency relative to sounds that do not match what they have heard from others (e.g., Konishi, 1965, 1985; Watanabe & Aoki, 1998). In other words, the parity achieved when produced sounds are closest to heard sounds automatically strengthens the produced sounds (Palmer, 1996). The concept of automatic reinforcement answers the claim by some researchers that vocal learning occurs without much in the way of external reinforcement (Doupe & Kuhl, 1999, but see Goldstein et al., 2003, 2009). The key word is *external*. Such a position represents two possible misunderstandings of the concept of reinforcement by language researchers. The first, as was previously mentioned, is that reinforcement must be deliberately and consciously delivered by one individual to another. Of course, adult birds do not deliberately reinforce vocal sounds of their young if by *deliberately* we mean consciously. The second and related misunderstanding is that, although not conscious and deliberate, reinforcement must still originate from another individual. This is illustrated in the studies by Goldstein and his colleagues who compared the development of infant vocalizations to those of songbirds and concluded that vocal development in human infants, as in songbirds, is influenced by the actions of social partners (Goldstein et al., 2003). What Goldstein and his colleagues have demonstrated is that social reinforcement plays a role in the vocal development of birds and human infants. But it does not play the only role.

Some aspects of the vocal learning of infants and certain species of birds occur independently of any reactions from others. Such shaping takes place as a function of automatic reinforcement, that is, reinforcement not deliberately mediated by other individuals (Vaughan & Michael, 1982). Automatic reinforcement (though not by that name) in the form of parity between self-produced auditory feedback and the sounds heard from others has been recognized in

song learning in birds (e.g., Konishi, 1965, 1985; Watanabe & Aoki, 1998). For example, according to Konishi (1985), "a bird's use of auditory feedback in song development resembles learning by trial and error; the bird corrects errors in vocal output until it matches the intended pattern" (p. 134). Of course, the bird does not correct the errors; reinforcement in the form of auditory feedback simply selects specific aspects of the song, usually those that most resemble a parent. Also, by "trial and error," Konishi meant operant learning, which is more accurately described as trial and success.

Automatic reinforcement also plays an important role in early language acquisition (Schlinger, 1995, pp. 158–160; Smith et al., 1996; Sundberg et al., 1996). For example, several studies suggest that use of the passive voice can be acquired simply as a function of hearing someone else use it in context (e.g., Dal Ben & Goyos, 2019; Whitehurst & Ironsmith, 1974; Wright, 2006). Automatic reinforcement is the learning mechanism responsible for vocal learning that many language researchers have alluded to when they talk about how sounds emitted by infants and songbirds "are then gradually molded to resemble adult vocalizations" (Doupe & Kuhl, 1999, p. 574). There is also a lot of circumstantial evidence of the role of automatic reinforcement in vocal behavior. As just one example, infants who are congenitally deaf produce the same vocal sounds as hearing infants until about 6 months of age when the range of babbling sounds becomes more restricted for deaf infants (Lenneberg, 1964). In such cases, babbling decreases due to the lack of immediate auditory feedback (i.e., automatic reinforcement) from babbling (Schlinger, 1995, p. 160). Moreover, because an automatic reinforcement hypothesis requires very few assumptions and is consistent with known scientific principles, it is a more parsimonious explanation than appealing to such concepts as statistical learning, entrenchment, and schema formation.

One final point needs to be made. Automatic reinforcement is reinforcement. Calling it automatic was only meant to counter claims that reinforcement must be deliberately mediated by another individual (Vaughan & Michael, 1982). The definition of reinforcement, however, includes no such constraints. Reinforcement is any consequence of behavior that increases the probability of similar behavior under similar circumstances. Thus, all behavior that continues to occur does so in large part because of the natural consequences of the behavior. Vocal verbal behavior is no exception, as it constantly produces auditory feedback for the speaker and is sometimes accompanied by social reinforcement. Thus, reinforcement—both automatic and social—as an explanatory principle can parsimoniously explain vocal development not only in songbirds, but, perhaps more important, in human infants.

Summary and Conclusion

In the present article, I have contrasted cognitive and behavioral accounts of speech perception and production, with a special emphasis on usage-based theories of language acquisition as an example of a modern cognitive linguistic

theory. I suggested that cognitive theories in general and usage-based theories in particular fare less well than a behavior-analytic theory of language learning in explaining language acquisition in human infants. I concluded this despite the fact that cognitive linguists and developmental psychologists have provided valuable research that is revealing about early language acquisition. Such research, however, lacks a strong, empirically based, unifying theoretical framework that also fails to impart practical knowledge that can enable practitioners to reliably teach individuals with language delays. As I have previously written, however, "we needn't throw the baby out with the bathwater. Instead, we ought to ask whether it is possible to make sense of apparently unrelated data according to a single unifying theory" (Schlinger, 1995, p. vii). The goal of the present article is to show how some of the facts of early speech perception and production can be explained within an operant learning framework. Thus, the take-home point is that even though both cognitive and behavior-analytic researchers have shown that early and constant exposure to a linguistic environment has a direct and enormous impact on child speech perception and production, a behavior-analytic theory, based as it is on empirically derived laws of learning, can most parsimoniously explain the observations.

The most consistent behavioral mechanism that accounts for both speech perception and production is reinforcement. One form of reinforcement that is alluded to by cognitive researchers, but never identified as such, is automatic reinforcement. In the present context, this refers to the products of one's own vocalizations reinforcing those very vocalizations. A behavior-analytic theory of speech perception and production with its focus on reinforcement, both social and automatic, is consonant with a sizeable amount of research on human infants as well as that on songbirds and other avian species over the past 60 years. The focus on reinforcement as an observable, measurable, and manipulable feature of the environment has an added benefit: it affords a way to teach language effectively and successfully to individuals with language delays.

Compliance with Ethical Standards

Conflicts of Interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Adret, P. (1993). Vocal learning induced with operant techniques: An overview. *Netherlands Journal of Zoology, 43*, 125.

Adret, P. (1993b). Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Animal Behaviour, 46*(1), 149–159.

Aslin, R. N. (1987). Visual and auditory development in infancy. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 5–97). Wiley.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Bates, E., O'Connell, B., & Shore, C. (1987). Language and communication in infancy. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 149–203). Wiley.

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47*, 383–411. https://doi.org/10.1515/LING.2009.014

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature, 417*(6886), 351–358.

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

Byers-Heinlein, K. (2014). High amplitude sucking procedure. In P. J. Brooks & V. Kempe (Eds.), *Encyclopaedia of language development* (pp. 263–264). Sage.

Dal Ben, R., & Goyos, C. (2019). Further evidence of automatic reinforcement effects on verbal form. *Analysis of verbal behavior, 35*(1), 74–84.

de Boysson-Bardies, B. (1999). *How language comes to children: From birth to two years*. MIT Press.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mother's voices. *Science, 208*, 1174–1176.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development, 9*, 133–150.

Diel, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55*, 149–179.

Dooling, R. J., Best, C. T., & Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave/ra-la/continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *Journal of the Acoustical Society of America, 97*, 1839–1846.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*(1), 567–631.

Dunst, C. J., Gorman, E., & Hamby, D. W. (2010). Effects of adult verbal and vocal contingent responsiveness on increases in infant vocalizations. *Center for Early Literacy Learning, 3*(1), 1–11.

Eimas, P. D. (1975). Developmental studies in speech perception. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: Vol. 2. From sensation to cognition (pp. 193–231)*. Academic.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*, 515–522.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, 100*(13), 8030–8035.

Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from adults. *Child Development, 80*(3), 636–644.

Harshaw, C., & Lickliter, R. (2007). Interactive and vicarious acquisition of auditory preferences in Northern bobwhite (*Colinus virginianus*) chicks. *Journal of Comparative Psychology, 121*(3), 320.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17*, 42–46.

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology, 4*, 255.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can form phonetic categories. *Science, 237*, 1195–1197.

Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow 1. *Zeitschrift für Tierpsychologie, 22*(7), 770–783.

Konishi, M. (1985). Birdsong: From behavior to neuron. *Annual Review of Neuroscience, 8*, 125–170.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70*, 340–349.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academies of Sciences, 97*, 11850–11857.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience, 5*, 831–843.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100*, 2425–2438.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69–72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America, 63*, 905–917.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics, 32*, 542–550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73*, 1003–1010.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindbloom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Lenneberg, E. H. (1964). Speech as a motor skill with special reference to nonaphasic disorders. *Monographs of the Society for Research in Child Development*, 115–127.

Levitt, A. G., & Utman, J. G. A. (1992). From babbling towards the sound systems of English and French: A longitudinal two-case study. *Journal of Child Language, 19*(1), 19–49.

Lickliter, R., & Bahrick, L. E. (2016). Using an animal model to explore the prenatal origins of social development. In *Fetal development* (pp. 3–14). Springer.

Lieven, E. (2008). Learning the English auxiliary: A usage-based approach. In H. Behrens (Ed.), *Corpora in language acquisition research: Finding structure in data* (pp. 60–98). John Benjamins.

Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language & Cognition, 8*(3), 346–368.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994–1997.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McMurray, B., & Hollich, G. (2009). Core computational principles of language acquisition: Can statistical learning do the job? Introduction to Special Section. *Developmental Science, 12*, 365–368.

Mehler, J., Nespor, M., & Peña, M. (2008). What infants know and what they have to learn about language. *European Review, 16*, 429–444.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., & Jenkins, J. J. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18*, 331–340.

Moerk, E. L. (1980). Relationships between parental input frequencies and children's language acquisition: A reanalysis of Brown's data. *Journal of Child Language, 7*(1), 105–118.

Moerk, E. L. (1983a). A behavioral analysis of controversial topics in first language acquisition: Reinforcements, corrections, modeling, input frequencies, and the three-term contingency pattern. *Journal of Psycholinguistic Research, 12*(2), 129–155.

Moerk, E. L. (1983b). The mother of Eve—as a first language teacher. *Monographs on Infancy, Mo, 3*, 158.

Moerk, E. L. (1990). Three-term contingency patterns in mother-child verbal interactions during first-language acquisition. *Journal of the Experimental Analysis of Behavior, 54*(3), 293–305.

Moerk, E. L. (1992). *A first language taught and learned*. Paul H. Brookes.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development, 16*, 495–500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica, 102*(2), 156–160.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 756–766.

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people. *Frontiers in Psychology, 7*, 934.

Palmer, D. C. (1996). Achieving parity: The role of automatic reinforcement. *Journal of the Experimental Analysis of Behavior, 65*(1), 289.

Pelaez, M., Virues-Ortega, J., & Gewirtz, J. L. (2011a). Reinforcement of vocalizations through contingent vocal imitation. *Journal of Applied Behavior Analysis, 44*(1), 33–40.

Pelaez, M., Virués-Ortega, J., & Gewirtz, J. L. (2011b). Contingent and noncontingent reinforcement with maternal vocal imitation and motherese speech: Effects on infant vocalizations. *European Journal of Behavior Analysis, 12*, 277–287. https://doi.org/10.1080/15021149.2011.11434370

Poulson, C. L. (1983). Differential reinforcement of other-than-vocalization as a control procedure in the conditioning of infant vocalization rate. *Journal of Experimental Child Psychology, 36*(3), 471–489.

Poulson, C. L. (1988). Operant conditioning of vocalization rate of infants with Down syndrome. *American Journal of Mental Retardation, AJMR, 93*(1), 57–63.

Poulson, C. L., & Nunes, L. R. (1988). The infant vocal-conditioning literature: A theoretical and methodological critique. *Journal of Experimental Child Psychology, 46*(3), 438–450.

Reed, P., Howell, P., Sackin, S., Pizzimenti, L., & Rosen, S. (2003). Speech perception in rats: Use of duration and rise time cues in labeling of affricate/fricative sounds. *Journal of the Experimental Analysis of Behavior, 80*, 205–215.

Rheingold, H. L., Gewirtz, J. L., & Ross, H. W. (1959). Social conditioning of vocalizations in the infant. *Journal of Comparative & Physiological Psychology, 52*(1), 68–73. https://doi.org/10.1037/h0040067

Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., Mavridis, N., Tellex, S., Salata, A., Guinness, J., Levit, M., & Gorniak, P. (2006). The human speechome project. In *International workshop on Emergence & Evolution of Linguistic Communication* (pp. 192–196). Springer.

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

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Schlinger, H. D. (1993). Learned expectancies are not adequate scientific explanations. *American Psychologist, 48*, 1155–1156.

Schlinger, H. D. (1995). *A behavior-analytic view of child development*. *Plenum Press.* https://doi.org/10.1007/978-1-4757-8976-8

Schlinger, H. D. (2004). Why psychology hasn't kept its promises. *Journal of Mind & Behavior, 25*(2), 123–144.

Schlinger, H. D. (2008a). Consciousness is nothing but a word. *Skeptic, 13*, 58–63.

Schlinger, H. D. (2008b). Listening is behaving verbally. *The Behavior Analyst, 31*(2), 145–161. https://doi.org/10.1007/BF03392168

Schlinger, H. D. (2010). Behavioral vs. cognitive views of speech perception and production. *Journal of Speech Language Pathology—Applied Behavior. Analysis, 5*, 150–165.

Schlinger, H. D. (2018a). The heterodoxy of behavior analysis. *Archives of Scientific Psychology, 6*, 159–168. https://doi.org/10.1037/arc0000051

Schlinger, H. D. (2018b). A behavior-analytic perspective on development. *Revista Brasileirade Terapia Comportamental e Cognitiva (Brazilian Journal of Behavior & Cognitive Therapy), 20*(4), 116–131.

Segal, E. F. (1975). Psycholinguistics discovers the operant: A review of Roger Brown's A First Language: The early stages 1. *Journal of the Experimental Analysis of Behavior, 23*(1), 149–158.

Sheppard, W. C. (1969). Operant control of infant vocal and motor behavior. *Journal of Experimental Child Psychology, 7*, 36–51.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review, 57*, 193–216. https://doi.org/10.1037/h0054367

Skinner, B. F. (1957). *Verbal behavior*. *Prentice-Hall.* https://doi.org/10.1037/11256-000

Skinner, B. F. (1981). Selection by consequences. *Science, 213*, 501–504.

Smith, R., Michael, J., & Sundberg, M. (1996). Automatic reinforcement and automatic punishment in infant vocal behavior. *Analysis of Verbal Behavior, 13*, 39–48.

Sturdy, C. B., & Nicoladis, E. (2017). How much of language acquisition does operant conditioning explain? *Frontiers in Psychology*, *8. Article, 1918.* https://doi.org/10.3389/fpsyg.2017.01918

Sundberg, M. L., Michael, J., Partington, J. W., & Sundberg, C. A. (1996). The role of automatic reinforcement in early language acquisition. *Analysis of Verbal Behavior, 13*, 21–37.

Todd, G. A., & Palmer, B. (1968). Social reinforcement of infant babbling. *Child Development, 39*(2), 591–596. https://doi.org/10.2307/1126969

Tomasello, M. (2000, September). A usage-based approach to child language acquisition. *Annual Meeting of the Berkeley Linguistics Society, 26*(1), 305–319.

Tomasello, M. (2009). The usage-based theory of language acquisition. In E. L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 69–87). Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511576164.005

Theakston, A., Lieven, E., Pine, J., & Rowland, C. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics, 16*, 247–277.

Thomas, H., & Gilmore, R. O. (2004). Habituation assessment in infancy. *Psychological Methods, 9*(1), 70.

Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2005). Effects of backward speech and speaker variability in language discrimination by rats. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 95–100.

Vaughan, M. E., & Michael, J. L. (1982). Automatic reinforcement: An important but ignored concept. *Behaviorism, 10*(2), 217–227.

Von Bargen, D. M. (1983). Infant heart rate: A review of research and methodology. *Merrill-Palmer Quarterly, 29*(2), 115–149.

Watanabe, A., & Aoki, K. (1998). The role of auditory feedback in the maintenance of song in adult male Bengalese finches Lonchura striata var. domestica. *Zoological Science, 15*(6), 837–841.

Weisberg, P. (1963). Social and nonsocial conditioning of infant vocalizations. *Child Development, 34*(2), 377–388. https://doi.org/10.2307/1126734

Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French-and English-learning infants. *Journal of Child Language, 18*(3), 501–516.

Whitehurst, G. J., & Ironsmith, M. (1974). Selective imitation of the passive construction through modeling. *Journal of Experimental Child Psychology, 17*(2), 288–302.

Wittgenstein, L. (1953). *Philosophical investigations*. Macmillan.

Wright, A. N. (2006). The role of modeling and automatic reinforcement in the construction of the passive voice. *Analysis of Verbal Behavior, 22*(1), 153–169.