Psychological Review
2010, Vol. 117, No. 2, 496–516

0033-295X/10/$12.00 DOI: 10.1037/a0019049

Invariance Detection Within an Interactive System: A Perceptual Gateway to Language Development

Lakshmi J. Gogate
Florida Gulf Coast University

George Hollich
Purdue University

In this article, we hypothesize that *invariance detection*, a general perceptual phenomenon whereby organisms attend to relatively stable patterns or regularities, is an important means by which infants tune in to various aspects of spoken language. In so doing, we synthesize a substantial body of research on detection of regularities across the domains of speech perception, word segmentation, word–referent mapping, and grammar learning. In addition, we outline our framework for how invariance detection might serve as a perceptual gateway to more sophisticated communication by providing a foundation for subsequent emergent capacities. We test our hypothesis using the domain of word mapping as a case in point, emphasizing its epigenetic nature: Word mapping is rooted in the real-time interactions between the infant and the physical world. The present account offers an alternative to prior theories of early language development and helps to link the field of early language development with more general perceptual processes.

Keywords: invariance detection, infancy, language development, dynamic organism-environment interaction, computational modeling

> Cognitive economy makes sense, and discovery of invariance is the essence of economy, finding order in change. (E. J. Gibson, 1991, p. 354)

To learn a language, infants must discover how sounds (phonemes) combine to form meaningful units (morphemes); how those morphemes map onto objects, actions, and events in the real world (semantics); and how words combine to form new meanings (grammar). Our central hypothesis is that perceiving relatively stable regularities or consistencies, called *invariance detection*, is the fundamental starting point for infants to tune in to their ambient communication systems at these component levels: phonetic, morphemic, semantic, and grammatical. In essence, by discovering contextually consistent patterns in caregivers' communication, infants develop the foundation to progress to successively more sophisticated levels of communication. Because of the epigenetic nature of this process, the invariants detected (and offered by caregivers) will necessarily change over time. Our main goal is to provide a framework that suggests that invariance detection by the infant during ongoing real-time interactions between the infant and the physical world provides a perceptual gateway to language development. This framework offers a perceptual solution to problems that have preoccupied researchers concerned with the origins of language and its development—psychologists, philosophers, linguists, speech scientists, anthropologists, cognitive scientists, and neuroscientists.

The mechanism of invariance detection has traditionally been used in perception to describe phonetic feature detection (Jakobson & Halle, 1956) or increasing differentiation (e.g., perception of different types of wines; J. J. Gibson & Gibson, 1955). However, given the importance of invariance detection as a mechanism for perception, we see no reason why it cannot explain the perceptual beginnings of language, even those aspects that are traditionally considered to be the domain of complex cognition, including rule learning and word mapping. Like E. J. Gibson (1966), by invariance detection we mean selective attention to *relatively stable patterns* or *structural regularities* in the changing stimulus array. This is in contrast to the colloquial definition of invariance, which links the term to stimuli that do not change or that remain constant. Invariant patterns are thus not always obvious in the changing stimulus array, nor do they occur 100% of the time. Nonetheless, as long as these patterns are consistent within a context, they can be detected and used by the child. Invariants are detected at multiple levels and times. Some stimuli that are perceived as random at first may in fact have a comparatively stable pattern that can be discovered with experience (e.g., reading an ultrasound). Likewise, invariance detection at a more basic level serves as a building block for invariance detection at more complex levels as language development progresses. Our broad definition accommodates all types of regularity detection—from the detection of perceptual invariants to more abstract invariants. Although this phenomenon has many specific names within each domain (e.g., phonological perception, normalization, categorization, generalization, statistical learning, and rule learning), the idea that infants come to detect commonalities seems relatively incontrovertible.

Lakshmi J. Gogate, Department of Psychology, Florida Gulf Coast University; George Hollich, Department of Psychological Sciences, Purdue University.

We are grateful to Christopher Prince for his comments on an earlier version of the multisensory underpinnings of lexical comprehension hypothesis. The writing of this article was supported by funding from the Thrasher Research Fund (02819-1) and the March of Dimes Birth Defects Foundation to Lakshmi J. Gogate.

Correspondence concerning this article should be addressed to Lakshmi J. Gogate, Room 29, Modular 1, Psychology, Florida Gulf Coast University, 10501 FGCU Boulevard South, Fort Myers, FL 33965. E-mail: lgogate@fgcu.edu

We suggest that, fundamental to this viewpoint, the process of invariance detection takes place within a developing interactive system. By this we mean that infant perception and knowledge are situated or rooted in the ever-changing real-time interactions between the infant and its physical world. Furthermore, certain phenomena exist only in these interactions. As a case in point, consider infant-directed (ID) speech. It is difficult to get an infant's attention. So difficult, in fact, that during communication, caregivers provide large frequency modulations that adults would find odd to hear but infants find salient. Caregivers do not use ID speech in the absence of infants, and infants have difficulty attending to speech if these modulations are not present (Pegg & Werker, 1992; Thiessen, Hill, & Saffran, 2005). Thus, ID speech exists only in the interaction between infants and their caregivers, and changes in one may have a ripple effect and cause changes in the other. For example, more mature phonemic perception by the infant will increase language sensitivity. This, in turn, might obviate the need for ID speech by the caregiver.

Thus, we suggest that perception of invariants is not built in or represented solely in the brain of the infant but rather is distributed across brain, body, and world (Smith, 2005; Wilson, 2002). Pfiefer and Scheir (1999, p. 21) suggested that *embodiment*, the existence of a body interacting with an environment, is one of the key factors in understanding intelligence. A. Clark and Chalmers (1998) elaborated on this notion by saying that "once we recognize the crucial role of the environment in constraining the evolution and development of cognition, we see that extended cognition is a core cognitive process, not an add-on extra." They suggested that a sea of words "envelopes us from birth. Under such conditions, the plastic human brain will surely come to treat such structures as a reliable resource to be factored into the shaping of on-board cognitive routines" (p. 17).

By acknowledging the joint contribution of the infant and the immediate environment, our invariance detection framework is compatible in spirit with other general systems theories (Bronfenbrenner, 1979; Lerner, 2001; Rogoff, 1990; Sameroff, 1975) and with many prior accounts of language development (e.g., L. Bloom, 1998; E. Clark, 1995, 2003; Nelson, 1988; Snow, 1972; Tomasello, 2006). Yet, it differs from these accounts in one important respect. These accounts have established that the input to the infant contains a laundry list of regularities (what we call invariant properties or stable patterns) and that infants are quite adept at picking up these regularities from the input. These accounts have not, however, established how the regularities present in the input are tightly coupled with, or even driven by, infants' changing perception of these regularities.

In this article, we espouse a novel approach to address, in part, how infants' language develops, forcing us to find *matches* or *points of stability* between the changing properties of the communicative environment and perception of those very properties by the infant. These matches have also been referred to as a *relational invariant* (proposed by E. J. Gibson, 1991, for perceptual development in general). The rationale for seeking these matches is as follows. Just because a specific type of regularity (e.g., the past tense morpheme –*ed* in English) is present in the input, it does not automatically imply that infants will pick up this regularity. At any given time, some regularities in caregivers' communication will be perceived more readily than others, possibly because these regularities are more functionally relevant to the developing infant or are more perceptually salient (L. Bloom, 1998). Reciprocally, just because an infant is capable of perceiving certain regularities presented in a controlled experiment, one cannot assume that these regularities are naturally present in the infant's communicative environment during the same developmental period.

Consideration of the progressive matches between caregiver and infant could reveal important insights about the ongoing environment–organism interaction and about the invariant properties that recruit infants' attention to ambient communication at different points in development. For example, it is possible that infants' natural preference for ID speech over adult-directed (AD) speech (Cooper, Abraham, Berman, & Statska, 1997; Cooper & Aslin, 1990) stems from the *match* between infants' developing propensity to perceive invariance in the speech signal early on and the heightened regularities in caregivers' ID speech. This match might serve as a catalyst for further invariance detection and language learning (also see Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989). Thus, detailed analyses of the invariant properties of ID speech combined with tests of infants' perception of these very properties could reveal important insights into ongoing organism–environment interaction. For example, McRoberts and Best (1997) demonstrated parents' fundamental frequency adaptations in speech that match with their infants' propensity for speech perception. The matches represent points of stability or preferred states in ongoing organism–environment interaction. They have also been referred to as *equilibrium states* in epigenetic theory (e.g., Thelen & Smith, 1998).

Equally important as it is to describe matches is to describe how the system then moves from one point of stability (match) to instability to the next point of stability. The mechanism for change within a dynamic system is self-organization. In a multicausal system, local-level perturbation and resultant deviation from these points of stability by either infants' ability to perceive a specific type of invariance or in the caregivers' provision of a specific invariant could result in the system reorganizing itself around that perturbation as it seeks a new point of stability (also see Thelen & Smith, 1998). Instabilities likely occur when key variables (e.g., infants' ability to detect a specific type of invariance) reach *critical mass,* the point prior to peak instability (see Section III; also see Smith, 1999).

One such inflection point occurs when the infant becomes more mobile. There are often hidden dangers that are not perceptually obvious. For instance, a mother talking to an infant who is more mobile might use statements such as "It's hot. If you touch it, it will hurt" to highlight hypothetical danger. This might drive instability in the system until the infant understands these statements and learns to pause even when encountering perceptually innocuous things when a warning statement is made. Other potential sources of instability include organismic processes such as developing visual acuity or growing stronger. Thus, infants are able to perceive invariants that they were unable to before and organize their movements in ways that they were unable to before (e.g., transition from crawling to walking). In such cases, the organism or the environment or both must reorganize or work around such instabilities to bring about change and stability. Thus, within a fluid or dynamic system, self-organization results not only in stability but also in instability that leads to further self-organization.

Consider babbling as another example. The quality and timing of the caregiver's response directly affects subsequent infant vocalization. Thus, contingent maternal speech is causally related to the quality of infant babble (Goldstein & Schwade, 2008). If there is a match between infant vocalizations and maternal responses, infants tend to babble more; if not, babbling might be inconsistent until either infant or caregiver or both can tune in to one another. In this manner, advances in babbling are a direct result of ongoing interaction between infant and caregiver. Babbling differences depend on the caregiver's input; French infants babble sounds that are more French (iambic), and Arabic infants babble sounds that are more Arabic (trochaic; De Boysson-Bardies, Sagart, & Durand, 1984). If the caregiver's input is not utilized, as is the case with hearing-impaired infants, the onset of canonical babbling never occurs or is delayed (Oller & Eilers, 1988) and babbling in general diminishes (Volterra, Iverson, & Castrataro, 2005). Again, invariance detection is an emergent product of the context, past experience, and the match between what the infant can perceive and what the environment provides.

In summary, the main thesis of this article is that invariance detection drives language development, which is an interactive multilevel, multicausal process that involves caregiver provision of stable patterns during communication and infant perceptual pickup of these patterns. Thus, we seek connections in invariance detection at multiple levels across perception and cognition, assuming no clear-cut demarcations between perceptual and cognitive levels (also see Smith & Thelen, 2003). Although there are multiple levels and causes for an outcome, no level is assigned special status.

The remainder of this article is organized as follows. In Section I, first we outline our theory of invariance detection and language development in comparison with that of several others. We provide examples to illustrate how invariance detection could be a *perceptual gateway* to language development. We then review the literature to show how the interactive process works. Reviewing infant experiments, we illustrate that infants detect all types of invariance—*suprasegmental* (prosodic) and *segmental* (phonetic, lexical, and grammatical)—and invariance across auditory and visual senses when caregivers communicate to them.

Section II is dedicated to the narrower domain of word learning, a fundamentally interactive, multimodal activity involving at least auditory and visual modalities. In this section we illustrate how co-occurring, invariant *multimodal* and *unimodal* properties of caregivers' multisensory naming combined with infant perception of these properties can explain the origins of word mapping and provide an example of our theoretical approach. A central goal is to highlight invariance detection as a perceptual mechanism underlying word mapping development. To elucidate how the perceptual framework might play out for early word mapping, and to generate further research in this domain, we propose *the multisensory underpinnings of lexical comprehension hypothesis*. According to this hypothesis, the interactive process of word mapping begins with infants' learning of *amodal relations* (involving temporally and spatially coordinated sounds and objects) and proceeds to the learning of auditory–visual relations involving *increasing degrees of arbitrariness* (sound-symbolic and non-sound-symbolic word–referent relations) in caregivers' multisensory naming, eventually leading to referential abilities by the second year.

In Section III, we chart the direction for future cross-disciplinary investigations of infants' invariance detection during word mapping, including developmental experiments, computational modeling of invariance detection by infants and of invariance provision by caregivers, and neurophysiological studies of caregiver–infant interaction. The final section, Section IV, summarizes how invariance detection paves the way for infants' initiation into the world of spoken language. We conclude with a discussion of the interactive system within which invariance detection and language development take place.

I. Invariance Detection: Its Relevance to Language Development

A Theoretical Overview of Detection of Regularities in Language

The study of language development has historically been fragmented into several subdomains. Whereas some have focused on how infants learn word meaning (semantics), others have focused on the sounds of language (phonology), and still others have focused on grammatical development. Although this approach, focusing on individual subdomains, worked optimally to chart the details of a phenomenon as complex as language development, we are now in a position, owing to this vast body of research, to study similarities across these domains. One such similarity is that researchers within many of these domains discuss the learning of regularities in language structure, or what we call *invariance detection*.

In grammatical development, learning about the structural regularities, or patterns of one's language, has traditionally been attributed to the induction of rules (e.g., adding –*ed* to make the past tense) by a preprogrammed language acquisition device (Chomsky, 1965; Pinker, 1984, 1987). Similarly, within the domain of semantic development, partially influenced by the Chomskian revolution, examinations of learning about semantic regularities have focused on lexical constraints or preformed biases as a means for infants to gain insight about word meaning (e.g., children assume one word goes with one referent; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989). Likewise, early research on invariance in phonological development emphasized the infants' inborn ability to categorize native-language phonemes (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Common across these accounts is the notion that the "mind has [preformed] mechanisms and representational formats that allow it to represent, extract, and generalize relationships—known as rules" (Marcus, 2001, p. 5).

More recent accounts, in contrast, highlight organism–environment interaction—rather than built in rules, constraints, or biases—as an underlying mechanism for language development (e.g., Smith, 2005; see further Thelen & Smith, 1994, for a general view). In these more recent accounts and in ours, the organism and its environment are tightly intertwined and constantly adapting to one another. In line with this view, organism and environment are softly assembled instead of being hardwired to learn language. It is the soft assembly that enables the ongoing (fluid) interaction between the organism and its environment—evident in the stability or matches as well as instability during ontogenetic and phylogenetic development. In a softly assembled system, environment, behavior, and genes can shape each other's destiny. Dedicated neuronal structures (e.g., Broca's area for speech–motor function) eventually form as a result of such development during the life span; they do not exist from the start.

A pioneering advocate of this bidirectional view, E. J. Gibson (1969) suggested that salient aspects of the input are highlighted for infants, by educating infants' attention to *invariants* (relatively stable patterns) in the input. We propose, therefore, that invariance in language structure at various component levels is made available by caregivers during everyday adaptive communication with infants and that this might be one means by which caregivers scaffold infants' language development (also see Zukow-Goldring, 1990). Thus, we learn about the behavior of an organism by examining its adaptations to the immediate environment, and we can learn about environmental adaptations to an organism by studying the organisms that live in it (Lickliter, 2006; see L. Bloom, 1998, for a similar ecological view of language development).

Several other accounts also suggest a bidirectional relation between developing infants and caregivers' communication (Adamson & Bakeman, 2006; Deak, 2000; Gogate, Bolzani, & Betancourt, 2006; Gogate, Walker-Andrews, & Bahrick, 2001; Matatyaho & Gogate, 2008; Matatyaho, Mason, & Gogate, 2007; Sullivan & Horowitz, 1983) and that language development takes place within a multilevel system as a result of ongoing interaction between the developing infant and the immediate environment (Smith, 2005). Language takes two—it seems to develop whenever there are two or more humans trying to communicate (Senghas, 1995). Thus, it is necessary to elucidate the dynamic interaction between the language novice (the infant) and the expert (the caregiver) to help explain the developmental process. Deak (2000, p. 47) suggested, consistent with this view, that "word learning tendencies can only be explained in terms of complex systems of interactive contingencies among properties of the child and of the environment." Hockema and Smith (2009, pp. 460) suggested, also consistent with this view, that during the process of language learning, a *synthesis* of basic "learning mechanisms with the environmental regularities triggers the formation of new perceptual and conceptual 'units' [phonemes, words] that can then be hierarchically applied to further speed the process of learning language." General agreement with this hypothesis can be found in E. J. Gibson's (1991) thesis that perceptual development takes place as a result of the infant's ability to "detect the information specifying both, the required environmental supports and its own capabilities and, above all, the relation between them, a kind of abstract *relational invariant*" (p. 613). Thus, invariance or structural regularity detection from caregiver input by the infant is an important basic learning mechanism for language development.

Invariance Detection: A Gibsonian View

According to E. J. Gibson (1966), at any given time, there is too much information in the total stimulus flux. The infant must, therefore, develop some strategies for information pickup—of *selection*, *adaptation*, and *inhibition*. Invariance detection by infants is part and parcel of these developing perceptual strategies. It is the process by which the perceptual system seeks to reduce uncertainty in the stimulus flux and seeks order amid change. Thus, invariance detection might involve pickup of a single relatively stable pattern (feature) in the stimulus array or pickup of a set of relatively stable features from several stimuli in the array. In the latter case, invariance detection underlies categorization of objects and events, in which members that share the same or a similar set of stable features are grouped together or are judged to be of the same class.

Contrary to the notion that invariance detection involves finding constants, in our view it is the process by which infants find what is *relatively stable* in the stimulus flux. Invariance versus variance must then be perceived along a continuum. Perceptual learning involves *active exploration* and increased differentiation of the stimulus array through the process of discovering distinguishing or salient features of objects and events and discovering invariance in stimulation (E. J. Gibson, 1966). The end product of this educated perception and attention is a filtered stimulus—information that is relevant to the infant that is also an *affordance* of the environment. Other information that has no utility for differentiation is tuned out by this educated perception and attention (E. J. Gibson, 1969). Thus, invariance in the stimulus array recruits infants' attention to salient and relevant properties of the array.

E. J. Gibson (1969) noted two important characteristics of invariance detection. First, she noted that invariant properties of a stimulus array are perceived or are perceived better only amid variation. Thus, if a greater number of varying medial elements are presented in synthesized syllabic strings, 18-month-olds and adults detect the invariant nonadjacent dependencies or correlations between nonneighboring segments such as *pel* and *jic* in the strings better than if only a small number of variant elements are presented (e.g., pel *wadim* jic; Gomez, 2002). Thus, Gomez (2002) suggested that the detection of invariant probabilistic cues (high probabilistic transitions) is a means for learning about other types of invariant language structure (e.g., serial word order). Enhanced invariance detection amid variation is not limited to the language domain. Even 5.5-month-olds show long-term memory for everyday activities (e.g., brushing teeth) if during initial familiarization they are shown dynamic but not static video displays consisting of different orientations of the actor (Bahrick, Gogate, & Ruiz, 2002).

Second, E. J. Gibson (1966) noted that infants' detection of relevant invariant information can change over time. As the infant develops, information that was not previously relevant can become relevant and vice versa, making way for new affordances to be perceived whereas old ones take on a less important status. In this manner, perception of different types of invariance can become fine-tuned to the infant's developing cognitive capacities. For example, information about grammatical frames may not be relevant to young infants, who may be more focused on finding the invariant properties that signal where words begin and end (e.g., stress cues, and transitional probabilities; see additional examples in Section III). Only later might infants focus on invariant properties signaling grammatical frames. In this manner, segmentation is a perceptual gateway to learning grammatical frames. One type of invariance detection could provide a perceptual gateway for detecting other invariants, leading to language development.

Invariance Detection and Infant Perception of Phonemes, Words, and Grammar

The speech perception literature abounds in studies that suggest that, prior to 1 year, infants tune in to their native language by detecting invariance in language structure at various component levels, including native-language phonetic categories (Kuhl, 2000). For example, by 1 month, infants discriminate /ba/ from /pa/ by perceiving invariance in the first and second formants, which are relatively stable within phonetic categories (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Studies suggest that with development, infants' perception becomes canalized, or narrowed, to native-language-specific phonemes. Whereas the canalization of language-specific vowel categories takes place by 3 to 6 months (perceptual magnet effect; Bosch & Sebastian-Galles, 2003; Kuhl et al., 1992), the canalization of language-specific consonants takes place later, around 10 months (Werker & Tees, 1984). By this time, infants also detect invariant phonotactic patterns in native-language syllables (Jusczyk, Luce, & Charles-Luce, 1994).

Invariance detection also facilitates word recognition. At 4.5 months, infants recognize their own names earlier than those of others (Mandel, Jusczyk, & Pisoni, 1995), on the basis of their familiarity with or *relative invariance* in the syllabic constituents, and syllable-initial stress patterns of these names. By 6 months, infants use their own names as anchors to segment speech (Bortfeld, Morgan, Golinkoff, & Rathburn, 2005). Infants detect word boundaries in passages by 7.5 months (Jusczyk & Aslin, 1995). They do this by detecting stable phonetic properties and stress patterns (Jusczyk, Cutler, & Redanz, 1993) in words across multiple tokens of the same talker (Houston & Jusczyk, 2000) and across multiple talkers (Houston & Jusczyk, 2003). Thus, infants use *segmental invariance* in the syllables and *suprasegmental invariance* in stress patterns to segment words. Further, 9-month-olds detect stable stress patterns across several words strung together, indicating sensitivity to the metrics of the language (Gerken, 2004).

Invariance detection also underlies native-language grammar learning. There is no shortage of grammatical invariants to which infants and children could be sensitive. Statistical regularities in the distribution of words could lead infants to ascertain parts of speech (Elman, 1993). Furthermore, toddlers are able to use the fact that words of a particular part of speech in both English and French are surrounded by the same sets of words (or frames) to learn the parts of speech (Mintz, 2003; also Chemla, Mintz, Bernal, & Christophe, 2009). Likewise, transitional probabilities can be used to detect phrases as well (Thompson & Newport, 2007). In addition, many functional approaches to grammatical acquisition (Jackendoff, 2002; Tomasello, 2006) posit language regularities that could enable children to discover the "rules" of language.

Although there are many possible invariants that infants could learn, thus far the research supporting infants' sensitivity to grammar is limited to a few studies of transitional probabilities and a few simple algebraic rules. Between 7 and 8 months, infants use pause duration and final word lengthening in the input to detect stable grammatical units (e.g., phrases and clauses) in the speech stream (Hirsh-Pasek et al., 1987; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). By 18 months, infants detect invariant nonadjacent dependencies in the speech stream as do adults (Santelmann & Jusczyk, 1998). Although these studies show that infants can detect some grammatical regularities in the input at an early age, there remains a glaring paucity of studies that show that infants and children learn the grammatical regularities that their caregivers provide. One notable exception is a study by Cameron-Faulkner, Lieven, and Tomasello (2003), which demonstrates that 2- to 3-year-olds learn and produce the same item-based frames as used by their mothers (e.g., Can you . . ., Here's . . ., or Let's . . .) and that their usage of the frames correlates with their mothers' frequency of use. Such findings suggest that specific matches exist between children's detection of invariants and the invariants that their caregivers provide.

As a further case in point, infants learn invariant relations in artificially created syllabic strings using powerful computational abilities that underlie word segmentation and grammatical development. For example, 8-month-olds utilize transitional probabilities (the probability of co-occurrence of two neighboring syllables corrected by the base rate of the first syllable in the pair) within segments (e.g., *pa* and *bi* in *pabiku*) and between segments (e.g., *ku* and *da*, across *pabiku* and *daropi*) of a synthetic syllabic string (e.g., . . . *pabikudaropitibudopabikugolatu* . . .) to detect invariant segmental boundaries (Saffran, Aslin, & Newport, 1996). Thus, learning of transitional probabilities is one type of invariance detection. Computational abilities also enable the learning of more complex invariant relations such as adjacent and nonadjacent dependencies in syllabic strings (Gomez, 2002; Gomez & Gerken, 1999), as well as the learning of abstract algebraic rules (e.g., ABA vs. ABB; Marcus, Fernandez, & Johnson, 2008; Marcus, Vijayan, Bandi Rao, & Vishton, 1999), which may be useful in native-language grammar learning.

Further evidence for invariance detection during language development can be seen in children's ability to pick up relatively stable patterns in the semantic and grammatical structure of the ambient language. The stable patterns or regularities become generalized and internalized (some call these "rules") upon repeated encounters with the ambient language. Exceptions to the rule (i.e., irregular forms) are learned one at a time from the environment, by first making the irregular form fit the invariant pattern (the general rule), often evidenced in overgeneralization or overextension of the rule. For example, Deak (2000, p. 52; J. J. Gibson & Gibson, 1955) suggested that semantic overextensions (e.g., naming a lion as *cat*) are the result of a child's success in perceiving invariant properties (between a cat and a lion) but a failure in as yet making fine perceptual distinctions between basic-level categories. This phase of learning gradually gives way to learning of the correct form, again with repeated exposure to it.

As another classic example, consider English-speaking children learning to produce grammatical morphemes and using them reliably over time. Consider children first learning the regular plural morpheme *–s* or the past tense morpheme *–ed* and then overextending this regular pattern to the exceptions, resulting in ungrammatical forms (e.g., **sheeps* and **runned*; Brown, 1973; Fletcher & Garman, 1987; Hoff-Ginsberg, 1997). Although these are often cited as examples of rule learning, from our perspective, they are outcomes of invariance detection and developing attention to specificity (E. J. Gibson, 1969). If children learn the regular forms first and overgeneralize the now familiar and regular pattern to exceptions, then they must perceive and abstract invariance readily available in the regular forms. Therefore, invariance detection must be part and parcel of children's perceptual learning of the ambient language. The correct forms for the exceptions (e.g., *sheep* and *ran*) appear gradually in children's repertoire. The learning of exceptions too likely results from gradual differentiation of the exceptions from the regular or invariant morphemic pattern (rule). Consequently, we see the learning of structural regularities proceed from attention to individual forms to general rules to fully differentiated understanding of both rules and exceptions (Pinker, 1999). Initially, children might operate on a case-by-case basis. Later, they begin to notice the invariant properties and generalize or regularize across cases. During this phase, children fail to perceive that the general rule does not apply to the exceptions; the exceptions are not differentiated from the regular pattern. Gradually, the exceptions and regular pattern are differentiated, internalized, and correctly utilized. In fact, the tendency to detect invariance is so strong that when the ambient language is grammatically inconsistent, children invent invariant forms (see Hudson-Kam & Newport, 2005, for regularization of spoken language in the lab; see Singleton & Newport, 2004, and Senghas & Coppola, 2001, for regularization in American and Nicaraguan Sign Language, respectively). These examples taken together speak to invariance detection as a perceptual gateway for children's learning of grammatical morphemes from caregivers' language.

Invariance Detection in ID Unimodal and Bimodal Communication

In the previous subsection, we demonstrated that infants perceive invariance in phonetic, lexical, and grammatical domains. In the present subsection, we consider invariants that occur across these domains and that arise through the interaction between infants, caregivers, and their linguistic environment. This includes unimodal ID speech, or motherese, and bimodal auditory–visual (voice–face) communication.

Detection of invariance in unimodal speech. The pickup of invariant information is evident right from the start when infants encounter their caregivers or other adults speaking to them. For example, detection of *prosodic or suprasegmental invariance* in intonation, stress, and rhythmical patterns in maternal speech seems critical to infants' recognition of their mother, which, in turn, is critical to infants' survival and the survival of the species (Fernald, 1992; Fernald & Simon, 1984; see Figure 1). On the basis of talker-specific invariant pitch and timbre heard in utero, even newborns distinguish and prefer their own mother's voice to that of another woman (DeCasper & Fifer, 1980). In addition, in ID speech, *segmental invariance* in the form of simplified phrases and clauses, repetitive utterances, and placement of novel words more often in phrase- or clause-final position co-occurs with the intonation peaks of these segments (Fernald & Mazzie, 1991; also see Figure 1). Placement of novel words in phrase- or clause-final position occurred frequently in Caucasian and Hispanic maternal speech to 6- to 8-month-old infants (76%, Gogate et al., 2006). Additionally, 7.5-month-old infants find it easy to segment words in clause-initial or -final position (Seidl & Johnson, 2006). Mothers present novel words in varied contexts that might highlight the invariant transitional probabilities of these words amid contextual variation (Goodsitt, Morgan, & Kuhl, 1993). Thus, invariance in maternal speech likely matches with and facilitates preverbal infants' detection of invariance in the speech signal. In addition to modifying the speech signal, caregivers modify their actions in specialized ways—called *motionese*—during interactions with their infants (e.g., exaggerated synchronized movements of the hand or body while talking) relative to when speaking with other adults (Brand, Baldwin, & Ashburn, 2002; Brand, Shalcross, Sabatos, & Massie, 2007). These modifications when combined with ID speech might serve to direct infants' attention in two modalities and aid infants in segmenting speech. Infants and 2.5-year-olds, respectively, prefer ID to AD speech when detecting statistical regularities during word segmentation (Thiessen et al., 2005) and when learning grammatical categories such as nouns and verbs using invariant structures or frames that surround these words (Mintz, 2003).

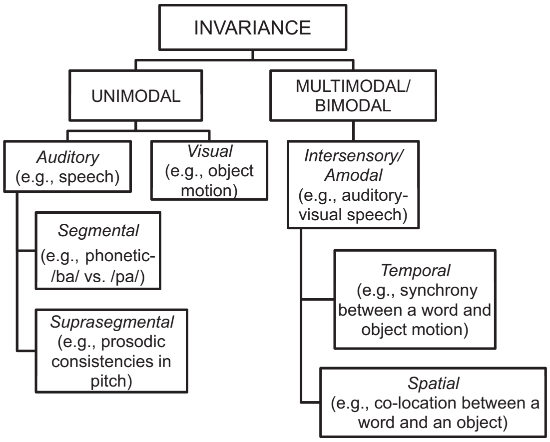

Figure 1. A taxonomy of invariants provided during caregivers' communication and invariants perceived by infants.

Detection of invariance in bimodal (face and voice) speech. Infants not only hear their mother's or an adult's voice during speech directed to them but they also see the person's face. Vocalizations and mouth movements are naturally spatially co-located; closely related in time; and share common tempo, rhythm, and intensity shifts. When adults speak, their simultaneous vocalizations and mouth movements provide *redundant,* or *amodal,* information; some of the same information is presented across modalities (see Figure 1). Because amodal invariance is perceived across at least two sense modalities, it is a type of bimodal invariance.

A wealth of evidence suggests that young infants detect various invariant amodal properties of auditory–visual speech (e.g., temporal synchrony, a common amplitude shift, spatial co-location). For example, infants as young as 7 weeks prefer the face and voice of a person using ID over AD speech regardless of the person's gender (Pegg & Werker, 1992). By 2 months, infants perceive amodal information in bimodal speech. They match particular vowel sounds with mouth shapes on the basis of invariant temporal synchrony and spatial co-location between the voice and lip movements (Patterson & Werker, 2003; also see evidence at 3–4 months in Kuhl & Meltzoff, 1982, 1984, 1988). By 3–4 months, infants imitate mouth movements only if the audio–visual components of vowels are temporally coordinated and invariant (Legerstee, 1990). Infants of this age also visually prefer the temporal invariance between a synchronized face and voice during continuous speech (Dodd, 1979; Pickens et al., 1994). By 3 months, infants detect affect (e.g., happy vs. sad) better if amodal invariance is presented between their mother's face and voice than between their father's face and voice or a male or female stranger (Montague & Walker-Andrews, 2002). In addition, English- and Cantonese-learning infants of 4.5 and 9 months prefer ID over AD

Cantonese only when the voice of the female speaker is presented in synchrony with the face (Werker, Pegg, & McLeod, 1994). By 5 months, infants can also detect a change in language membership (English and Spanish) if the face and voice of the speaker are synchronous during habituation (Bahrick & Pickens, 1988) and can match bimodally presented bisyllables on the basis of invariant amplitude shifts in their vowel nuclei (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). By 7 months, infants match faces and voices on the basis of age and gender, again if the faces and voices are synchronous during habituation (Bahrick, Soutullo-Netto, & Hernandez-Reif, 1998; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991). And finally, by 7.5 months, infants use audiovisual synchrony to tune in to one continuously speaking voice over another. In summary, these studies suggest that preverbal infants perceive several types of invariant properties—synchrony, spatial co-location, and amplitude shifts—across auditory–visual modalities during ongoing communication.

II. Invariance Detection and Word-Mapping Development

To fully illustrate the nature of invariance detection within an interactive system and show how invariance detection is a perceptual gateway to language development, we use word mapping as a case in point. In several theories, researchers have attempted to address the origins of word mapping and lexical comprehension in infants. Whereas in traditional views some have advocated the existence of preformed biases or constraints (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989; Quine, 1960; Waxman & Lidz, 2006), others have adopted a sociopragmatic view (Baldwin, 1993; L. Bloom, 1998; Callanan & Sabbagh, 2004; E. Clark, 1995; Tomasello, 1995). Still others have embraced a systems approach to the origins of lexical comprehension (Gogate, Walker-Andrews, & Bahrick, 2001; Smith, 2005), whereas others who do not necessarily adopt a systems approach have also favored a more inclusive view that underscores the coalition of environmental and organismic cues for lexical comprehension to develop (e.g., Hollich, Hirsh-Pasek, & Golinkoff, 2000). Many of these approaches have yielded important clues to the origins of word mapping. For example, recent experimental and quasi-experimental studies have shown that preverbal infants map *novel* words (Gogate et al., 2006; Matatyaho et al., 2007; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Schafer, 2005) and *familiar* words such as *mommy* and *daddy* (Tincoff & Jusczyk, 1999) onto specific referents in the second half of the first year. What, then, is the process by which preverbal infants begin to map words onto objects in the first year?

For the typically developing preverbal infant, mapping a word onto an object or an action is a perceptually driven, *multisensory activity*. We emphasize that preverbal infants begin to learn novel word-referent mappings by attending to *unimodal and multimodal invariance* in caregivers' naming. For example, young infants perceive *auditory invariance* such as the acoustic–phonetic properties of words, *visual invariance* such as the shape of an object or the specific path of its motion, and *temporal and spatial invariances* that unify words and objects. Detection of intersensory invariance in particular facilitates infants' early perception of words and referents as unified events. By delineating the co-occurring *multimodal and unimodal invariant properties* of caregivers' communication, and by elucidating how infants perceive these invariant properties in naming contexts, we can explain, in part, how word mapping develops. Therefore, we seek *relational invariants,* or matches between invariants in maternal communication and infants' perception of these invariants, to explain how word mapping develops in the second half of the first year. The specific interactive systems approach to word mapping that we espouse here offers a novel perceptual solution to an age-old question—how do infants map a word onto an appropriate referent (Quine, 1960)? We suggest that in a world of competing referents, by detecting invariants, within and across modalities, infants find correct word–referent mappings while tuning out incorrect ones.

Invariance in Naming Contexts

Caregivers name tangible or concrete objects and actions for their infants. Typically, referents are either visually present or are presented by the caregiver at the same time as the label is spoken (e.g., Masur, 1997; Nelson, 1978; Snow, 1972). Thus, at first, mothers communicate to their young infants by naming objects and actions in the immediate context. It is only after toddlers have had sufficient experience with referents and communication with adults (18–30 months) that mothers' conversations begin to consist of speech that is displaced in time and space from the referential context (Adamson & Bakeman, 2006). This *here and now* quality of caregivers' initial communication to preverbal infants has profound developmental and ecological consequence for infants' learning to map words. It provides ample opportunities for caregivers to tailor their language use in a manner well suited to the perceptual and cognitive limitations of their infants and provides developmentally appropriate opportunities for infants to learn word–referent mappings. Specifically, it enables caregivers to provide invariance within and across multiple sensory modalities during naming, especially early on, when the infant is a novice at word mapping. Reciprocally, infants perceive various types of invariants in word-mapping contexts. Because unimodal invariants (e.g., phonetic invariance and invariant motion) often occur in conjunction with bimodal (amodal) invariants (e.g., synchrony) during caregivers' communication, infants likely perceive these invariants as a unified event. We discuss infants' perception of bimodal and unimodal invariance in word-mapping contexts in separate units solely for clarity of explication purposes. Below we provide examples of invariants in caregivers' naming and examples of infant perception of these same invariants.

Bimodal invariance. Cross-cultural research on maternal communication to infants has revealed that caregivers often name objects or actions for their preverbal infants (5–11 months) by uttering a name and simultaneously holding and moving a tangible object in the infant's line of sight (Gogate, Bahrick, & Watson, 2000; Gogate et al., 2006; Gogate, Friedman, & Bewley, 2001; Zukow-Goldring, 1997). Messer (1978) reported that 73% to 95% of British English mothers' references to toys were simultaneous with toy manipulations during play with their 11- to 24-month-olds. Similar findings were reported from a field study of Mexican and Caucasian American mothers (Zukow-Goldring, 1997) and a sample of Asian Indian mothers from a pediatric clinic (Gogate et al., 2001). In Gogate et al. (2000), when Caucasian and Hispanic American mothers of 5- to 30-month-olds were asked to teach two object names and two action names to their offspring during semistructured play, mothers of 5- to 8-month-old preverbal infants used synchrony more often between words and referents than did mothers of 21- to 30-month-old lexically advanced toddlers. In addition, mothers of the middle group of infants and toddlers (9–17 months) used synchrony more often between verbs and object motions (72%) than between nouns and object motions (43%), perhaps reciprocating their offsprings' need for greater assistance with verbs than nouns. A recent study of Japanese and American mothers also reported naming in synchrony with actions to 2-year-olds (Yoshida, 2004).

Infants, in turn, are highly adept at perceiving temporal invariance between an adult's utterance and hand-held object motion. When adults utter a novel word and simultaneously move or shake an object in 6- to 8-month-olds' line of sight, the infants pair what they hear with what they see using the shared onset, offset, and duration of the otherwise unrelated word and moving object. Preverbal infants detect such *temporal invariance* and learn two syllable–object or word–object relations. This learning is demonstrated in habituation experiments. Infants of 7–8 months were presented with two alternating video stimuli in which a woman uttered one of two vocalic syllables—/a/ and /i/ (Gogate & Bahrick, 1998, 2001)—or more complex syllables—/tah/ and /gah/ (Gogate, 2010) or /wem/ and /bæf/ (Matatyaho & Gogate, 2007)—while holding and moving a toy object. Following habituation to two syllable–object pairings, infants detected a switch in the syllable–object pairings in the synchronous but not in the asynchronous condition. Under less tightly controlled procedures as well, 6- to 8-month-old infants learned the novel relations between two words—*gow* and *chi*—and two toy objects when their mothers named each object for 1.5 min during a semistructured interactive play episode (Gogate et al., 2006). In these studies, infants learned only when the spoken word or syllable and object motion contained temporal invariance (i.e., a common onset, offset, and duration), which helped infants to perceptually unify the otherwise unrelated word and moving object across auditory–visual modalities. Even 2-month-olds are sensitive to such temporal invariance. Following habituation to a single syllable–object pairing, in a synchronous or an asynchronous condition, upon testing, infants in the synchronous condition alone detected a change in the syllable and the object (Gogate, Prince, & Matatyaho, 2009).

In the caregiver studies, during naming, mothers provided *temporal invariance*—shared onset, offset, and duration between the utterance and the object's motion. The temporal properties shared by spoken word and object motion are an example of *amodal invariance* (E. J. Gibson, 1969), akin to the shared temporal properties of synchronous voice and lip movements during speech. Several other types of amodal invariance—such as a common tempo, rhythm, and intensity shifts between the spoken word and the object motion, as well as spatial co-location (name and referent are typically proximal)—might also be communicated by caregivers and perceived by infants. Temporal invariance is a type of *relational invariant*—it is information specifying the invariant relation between the required environmental supports and infants' perceptual ability (E. J. Gibson, 1991). It must be noted that the temporal alignment between spoken word and object motion need not be perfect. Preverbal infants' developing perceptual system will accept a relatively moderate temporal alignment relative to adults (Lewkowicz, 1996).

Thus far, the examples of bimodal invariance we have provided in word-mapping contexts have been *suprasegmental* (superimposed onto the segment utterance and object; e.g., synchrony). A recent study showed that Japanese- and English-speaking mothers' action naming to 2-year-olds frequently consisted of *mimetics* (Yoshida, 2004), in which the segmental (phonetic) properties of the word resemble the action to which it refers (e.g., words such as *bang*, *crush*, or *swoosh* co-occurred with a miming action). Reciprocally, the 2-year-olds learned these words more easily than regular words that did not resemble the actions to which they referred, once again suggesting a match between environment and organism.

Anecdotal reports indicate that infants' first words are often disproportionately onomatopoetic or iconic. A thing or action is named with a vocal imitation of the sound associated with it (e.g., "ruff-ruff" for *dog*, "tick-tock" for *clock*; Nelson, 1978; Werner & Kaplan, 1967). These universally present *sound-symbolic* words (Kita, 1997; Nuckolls, 1999; also see the bouba-kiki effect, Ramachandran & Hubbard, 2001; also Maurer, Pathman, & Mondloch, 2006), which are also frequent in infants' early receptive and productive vocabulary (see MacArthur Communicative Development Inventory [infant version]; Fenson et al., 1994), could be of critical importance to word-mapping development. We suggest that preverbal infants easily glean relations between onomatopoetic words and their referents early on because they share a greater degree of invariance across word and referent relative to regular words and referents (see hypothesis later in the Multisensory Underpinnings of Lexical Comprehension Hypothesis section; also Gogate, Friedman, & Bewley, 2001).

Unimodal invariance. When a mother names an object or an action for her infant and moves or performs an action with the object, she provides, in addition to invariant information across the senses, invariant information within modalities, or *unimodal invariance*. For example, when a caregiver repeats a novel word, multiple tokens of the same utterance provide opportunities for the infant to detect *invariant acoustic–phonetic properties* of that word among other words. Similarly, if the caregiver picks up and repeatedly moves the object during naming, invariant properties or features of the object such as its shape, texture, color, and size become available to the infant. Furthermore, if the caregiver happens to pick up and name several similar objects of a category, then the caregiver conveys invariant physical properties shared by all of these objects. And finally, if the caregiver repeatedly moves an object in a specific direction in tandem with a specific word, then *invariant path of motion* is conveyed to the infant.

Consistent with our hypothesis on caregivers' provision of invariant motion during object naming, a field study of Mexican and Caucasian American mothers reported that they often used *showing* gestures to their preverbal infants (5–11 months) during object naming (Zukow-Goldring & Ferko, 1994). The gestures consisted primarily of shaking or looming an object in the infant's line of sight. More recently, a quantitative study showed that when asked to teach the names of two novel objects, Hispanic and Caucasian American mothers of 6- to 8-month-olds used shaking or looming (forward/downward) motions more often than backward or upward motions during naming (Matatyaho & Gogate, 2008). In turn, preverbal infants perceive many types of within-modality, or unimodal, invariance. For instance, they perceive *motion invariance* during caregivers' object naming. Six- to 8-month-olds learned two word–object mappings better if their mothers used showing gestures consisting of shaking looming motions during synchronous object naming relative to upward or backward motions (Matatyaho & Gogate, 2008). In infant-controlled habituation experiments as well, 8-month-old infants attended to and learned word–object mappings such as /wem/-fish and /bæf/-dragonfly only if the words were spoken simultaneously with an adult's shaking or looming object motions but not sideways or upward motions, or if the four object motions were concatenated together in equal distribution (Matatyaho et al., 2007). Attention to motion invariance also appears to be important for understanding the concepts underlying certain motion verbs (e.g., *jump*). Following habituation to a moving object, on test trials, 7-month-olds detected changes in the manner and path of object motion (Pulverman & Golinkoff, 2004).

Furthermore, as early as 6 months, infants perceive *invariant features of an object category* when a single label is given to members of the category (Fulkerson & Waxman, 2007). As discussed earlier, following habituation to several exemplars of an object category and a label, infants detected a change in the label on test trials when a novel label was presented. In addition, when learning novel word–referent mappings, 10-month-olds attend to the interesting object and ignore the boring object even if an adult names it (Pruden et al., 2006), suggesting that object salience is key to early learning of novel word mappings (Hollich et al., 2000, provide similar evidence from 12-month-olds).

As another example, infants perceive *phonetic invariance* across multiple tokens of words in word-mapping tasks. Thus, 8-month-olds mapped two *similar* syllables—/tah/ and /gah/—onto two objects only when synchrony was provided between the syllables and moving objects. In contrast, 7-month-olds failed to map the same syllables onto objects in a synchronous condition, although they discriminated the syllables and objects in a nonmapping task (Gogate, 2010). However, 7-month-olds successfully mapped two *distinct* syllables—/tah/ and /gih/ or /gah/ and /tih/—onto the same two moving objects, again in a synchronous condition (Gogate, 2010; also see Gogate et al. 2006; Matatyaho et al., 2007). These studies suggest that preverbal infants detect unimodal invariance in words, objects, and object motions in caregivers' naming.

Detection of invariance in objects and phonetic properties is also implicit in some word-mapping studies of infants in their second year. For example, 14-month-old infants learn syllable–object pairings if the objects move but are not static and if the syllables are highly distinct (e.g., /nim/ and /If/) but not similar (e.g., /bIh/ and /dIh/; Stager & Werker, 1997; Werker, Cohen, Lloyd, Casasola, & Stager, 1998). By 20 months, infants map the highly similar syllables onto objects, once again facilitated by object motion (Werker, Fennell, Corcoran, & Stager, 2002). Objects in motion enable infants to abstract invariant shape across many different transformations of the object (E. J. Gibson, 1969). Thus, even 2-year-olds rely heavily on invariant object shape to assign a familiar name to novel exemplars of an object (Smith, Landau, & Jones, 1998).

Word Mapping in the Absence of Ostensive Naming and Bimodal Invariance

Some field studies suggest that ostensive naming to young infants is not typical of all mothers. In some cultures, mothers do not directly address their infants during conversations (e.g., the Kaluli of Papua–New Guinea; Schieffelin, 1979; and the Kwara'ae of Malatia in the Solomon Islands; Watson-Gegeo, & Gegeo, 1986). In these cultures, however, it is conceivable that infants learn from ostensive naming and multimodal communication provided to them by secondary family members (e.g., older siblings) if not from their mothers. It is also conceivable that these infants learn word–object mappings by detecting invariants when they eavesdrop on others' communication (similar to reports by Akhtar, Jipson, & Callanan, 2001; Floor & Akhtar, 2006).

When and how infants from such cultures begin to map words onto referents remain open to debate in the absence of developmental studies. The existing literature on word mapping raises a larger set of questions. What might account for preverbal infants' word mapping onto referents in the absence of synchrony? Might object perception be enhanced when greater variance is provided in the visual stimuli (see Section I, the Invariance Detection: A Gibsonian View subsection) during word mapping? Could the enriched visual stimuli override infants' need for temporal invariance? Infant studies have demonstrated that, in some contexts, invariants other than synchrony play an important role in facilitating word mapping. For example, if words have been repeated many times during everyday communication (e.g., *mommy* and *daddy*), then on a test, 6-month-olds look longer at the correct referent (e.g., their mother and father) than the incorrect one (Tincoff & Jusczyk, 1999). In addition, presenting 9-month-olds with multiple exemplars of an object category (e.g., different types of birds or dinosaurs) likely obviates the need for temporal invariance between the words and referents and highlights the categorical invariants (Balaban & Waxman, 1997). Similarly, if caregivers' naming to their 9-month-olds is contingent upon infant attention (e.g., the mothers name an object that the infants are looking at), then infants' vocabulary is larger at 12 and 18 months (Rollins, 2003). Even newborns relate auditory–visual pairings that are less tightly coupled when the auditory information is made to appear contingent upon infants' attention to the visual information (Slater, Brown, & Badenoch, 1997; Slater, Quinn, Brown, & Hayes, 1999). These findings suggest changes in the organism (infant) in response to a changing environment, where various stimulus properties (e.g., synchrony, word or object familiarity, or contingency) carry different weights in differing contexts.

The Multisensory Underpinnings of Lexical Comprehension Hypothesis

Throughout this article we have suggested that infants perceive different invariants provided by caregivers. Caregivers perceptually highlight salient aspects of communication for infants by educating infants' attention to *unimodal and bimodal invariance* (E. J. Gibson, 1969). Infants in turn are highly adept at detecting all types of invariance—*suprasegmental* (e.g., synchrony) and *segmental* (e.g., sound-symbolic words and their referents), bimodal (across sensory modalities) and unimodal (within sensory modalities)—to successfully map words onto referents. Thus, early domain-general sensitivities, such as intersensory perception, fall on a developmental continuum with and lead to language-specific abilities, such as word comprehension.

To further illustrate how the process of invariance detection could act as a perceptual gateway to language development, we focus now on the specific case of word learning. In 2001, Gogate, Walker-Andrews, and Bahrick outlined how this might work (also see Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Sullivan & Horowitz, 1983; Zukow-Goldring, 1990). A modified version of this outline is presented here in the form of the *multisensory underpinnings of lexical comprehension hypothesis* (MULCH). Using the gateway metaphor, we hypothesize primarily that the developmental trajectory for infants' lexical mapping proceeds along a continuous path in a sequence (see Figure 2). Beginning with the learning of auditory–visual relations that share invariance across modalities (*amodal* relations; e.g., sound and sight of a barking dog), infants proceed to the learning of relations that share less invariance across auditory–visual modalities (sound-symbolic words and referents, e.g., "ruff-ruff" and a dog, Nelson, 1978; or gestures such as panting to depict a dog, Goodwyn, Acredolo, & Brown, 2000). Later, they learn non-sound-symbolic word–object relations in the presence of minimal invariance (e.g., synchrony between a word and simultaneous motion of a handheld object), then proceed to learning arbitrary word–object relations (regular adultlike nouns or verbs and their concrete referents, e.g., the word *doggie* and a dog) in the absence of invariance, and eventually develop their referential abilities (relating a spoken word with an absent referent). Thus, word-mapping development begins with infants' learning of amodal relations (involving temporally and spatially coordinated sounds and objects) and proceeds to the learning of auditory–visual relations involving *increasing degrees of arbitrariness* (sound-symbolic and non-sound-symbolic word–referent relations; Gogate, Walker-Andrews, & Bahrick, 2001), eventually leading to referential abilities by the second year.

Our model predicts referential ability with increased memory for the disparate elements, the word and the referent, and their relations, despite the temporal and spatial displacements between the elements. Infants learn symbol–referent relations with greater temporal and spatial displacements from 18 to 30 months (Adamson & Bakeman, 2006). Also, 24-month-olds learn word–referent relations that are spatially aligned but temporally displaced (Baldwin, 1993). Furthermore, our theoretical model predicts that over time, repeated encounters with the same invariant word–referent relations will lead infants to expect words rather than gestures or nonwords as appropriate symbols for referents (Hollich et al., 2000; also Namy & Waxman, 1998).

Recall that *amodal* auditory–visual relations refers to relations that have a great deal of invariance present across modalities (e.g., mouth movements and corresponding vocalizations; Kuhl & Meltzoff, 1982). Such invariance includes synchrony, a common tempo, rhythm, intensity shifts, and spatial co-location in auditory–visual events. *Arbitrary* auditory–visual relations refer to relations in which information available to the auditory modality is relatively distinct from that available to the visual modality (e.g., a spoken word and an object's shape or color). *Sound-symbolic* relations refer to word–object relations in which the word resembles salient properties of the object (e.g., an adult's use of the term *tick-tock* to refer to a clock with moving hands that makes a ticking sound). We theorize that infants learn these relations easily because phonetic properties of the word resemble salient properties of the referent (Gogate, Walker-Andrews, & Bahrick, 2001). *Arbitrary* (non-sound-symbolic) word–object relations, in contrast, refer to relations in which the phonetic properties of the word do not resemble salient physical properties of the referent (e.g., an adult's use of the term *clock* to refer to a clock that has moving hands and makes a ticking sound).

A major assumption of MULCH is *ongoing emergence,* in which later abilities emerge from earlier abilities (see Prince, Helder, & Hollich, 2005, and Smith, 2005, for reviews). For example, infants' perception of arbitrary word–object relations emerges from the earlier perception of amodal (suprasegmental) invariance (e.g., synchrony in auditory–visual relations) and segmental invariance perceived across sound-symbolic words and referents. In Figure 2, the notion of ongoing emergence is depicted

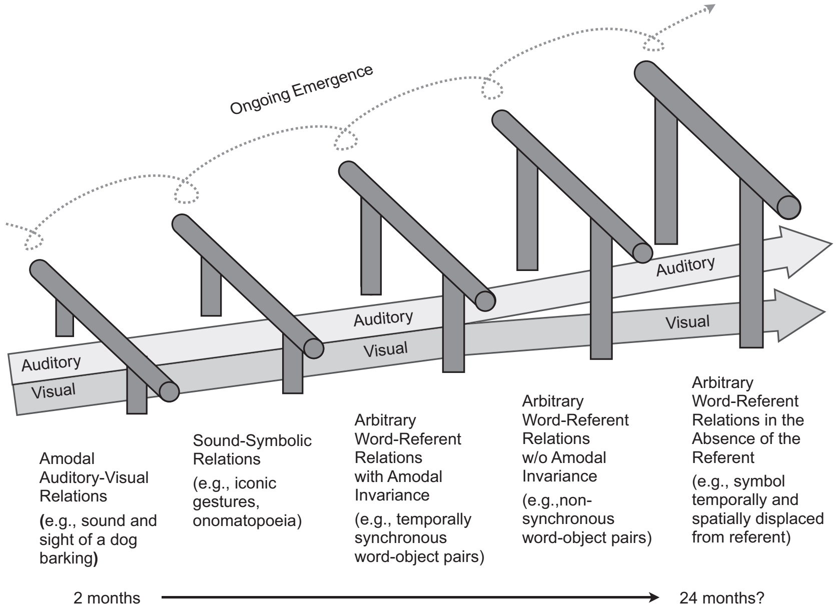

Figure 2. A schematic diagram of the multisensory underpinnings of lexical comprehension hypothesis.

by a series of gateways, each representing a type of auditory–visual relation with progressively increasing degrees of arbitrariness across sensory modalities. The first gateway depicts amodal auditory–visual relations with a great degree of redundancy across modalities. This redundancy scaffolds early learning of auditory–visual relations and provides a foundation for the learning of sound-symbolic relations, which themselves have a great deal of perceptual redundancy. Thus, even 2.5-year-olds easily detect sound-symbolic relations such as between a rounded object and the word *bouba* and an angular object and the word *kiki* (Maurer et al., 2006) because of the perceived relatedness between auditory and visual elements. In contrast, more arbitrary word–referent relations are relatively difficult to learn. Only over time, as a critical mass of sound-symbolic relations accrue, is it possible to learn more arbitrary relations. Each successive gateway becomes increasingly arbitrary and more removed from perceptual support. The arrows below the gateways depict this general progression along a continuum—from learning auditory–visual relations in the presence of intersensory invariance to learning these relations independent of intersensory invariance.

How might prior abilities serve as a building block for the emergence of later abilities? Specifically, what causes infants to attend first to sound-symbolic relations and then to increasingly arbitrary relations? One factor that might cause this change is the increased ability to hold perceptually unrelated auditory and visual elements long enough in memory to map one element to another (Gogate, 2010). Easing of memory load also occurs when 17-month-olds map isolated words far better than words embedded in continuous speech onto target versus distractor objects (Plunkett, 2006).

Within the present framework, developmental change is caused by local-level perturbations (or mismatches) during caregiver–infant interaction, resulting in instability and subsequent reorganization of either maternal communication or infant perception or both to get to a new point of stability. Mothers might use a great deal of invariant intersensory information early on (e.g., sound-action connections such as "coo-chi-coo-chi-coo") to scaffold their infants' internally heightened attention to this information until this type of word mapping reaches a critical mass. At this point, instability in the interaction and reorganization coupled with increasing perceptual differentiation should facilitate increased attention to unimodal information and the learning of increasing levels of arbitrary word–referent relations in ambient communication.

Another major assumption of MULCH is that during lexical development, closely coupled infant–caregiver adaptive behaviors give rise to *matches between infants' internal capabilities and salient properties of caregivers' unimodal and bimodal communication*. Thus, at times, during development, we should find points of stability during caregiver–infant interaction when there is congruence between a specific type of invariance that caregivers provide and the invariance infants perceive. An example of such congruence would be the predominant use of synchrony in maternal object naming for preverbal infants coupled with infants' learning of word–referent relations when synchrony is provided. We suggest that, in line with the dynamic systems framework (Thelen & Smith, 1998), the point in time when infants most utilize synchrony to learn word–object relations and caregivers use synchrony most often during naming is a *preferred,* or *equilibrium* (stable), *state* in the ongoing organism–environment interaction. At other times during development, points of instability should be found as the system shifts from one stable state to another. Such instability might be observed when, for example, a mother predominantly names static objects at a point in time when her infant prefers synchronous naming. The instability may be triggered by either a change in caregiver practice or infant perception and will continue until one, the other, or both reorganize around the instability. These points of stability and instability are an example of self-organization during mother–infant interaction. Points of stability and instability are depicted by spirals versus linear trends in the single line above the gateways (see Figure 2). The next subsection illustrates how infant–caregiver matches lead to word mapping.

Infants' Invariance Perception and Word-Mapping Development

Caregiver repetition and the reduction of uncertainty. When an infant's perceptual–cognitive system is sufficiently developed to pick up word–referent relations (e.g., with a basic working memory for words and referents, and expectations about the world of entities: object permanence, animacy, agency, and causality), caregivers' repetition likely foregrounds and highlights the invariance, enabling the infant to focus upon the most stable patterns of the communicated signal, which over time become salient and relevant. Repetition of invariance across modalities in the communicated signal, therefore, should educate infants' attention to the co-occurrence between a specific word and its referent whenever the specific information becomes an *affordance* and receives infants' attention (E. J. Gibson, 1969). Preverbal infants, who are novices at mapping words onto referents, find intersensory invariance (e.g., synchrony, rhythm, tempo, and intensity shifts) salient, which, if repeatedly presented across modalities, unifies and facilitates learning of each word–referent relation. With increasing encounters and infants' developing memory for each relation, the initially perceived intersensory invariance might help infants to later predict the referent upon hearing its name in the absence of such invariance.

Word–referent unification via the use of intersensory invariance reduces the degree of uncertainty between the word and its referent for the word-mapping novice. When adults name an object or an action, if the auditory–visual information presented to infants is temporally invariant across sense modalities, the degree of arbitrariness of the word–referent relation is reduced, thereby reducing uncertainty. The end result is that intersensory invariance reduces the perceptual and cognitive demands on preverbal infants. In the absence of intersensory invariance, the separation between the word and referent would make it more difficult for infants to discern a relation between them. Thus, no sensitivity to or learning of syllable–object pairings early on can be seen in the absence of synchrony in habituation experiments (Gogate, 2010; Gogate & Bahrick, 1998, 2001; Gogate, Prince, & Matatyaho, 2009; also Stager & Werker, 1997; Werker et al., 1998).

Invariance detection as a guide to correct word–referent mappings. The key question that we ask here is: How does the infant attend to caregivers' intersensory invariance in naming contexts and learn novel word–referent mappings? When caregivers name and simultaneously move an object, the temporal invariance between the name and the object's motion triggers infants' head and/or eye turn from their caregiver toward the moving object and enables learning of the word–object relation. Thus, infants attend to the salient invariant properties afforded in word-mapping contexts. Six- to 8-month-olds mapped two novel words onto two objects during play, if their mother often used synchrony between the words and objects' motion during naming and if infants switched eye gaze from their mother to the named object most often during her synchronous naming (Gogate et al., 2006). During this emerging phase of joint attention and word mapping, a greater number of infants switched eye gaze from their mother to an object during her synchronous (17/24) than during asynchronous (7/24) naming, suggesting an intrinsic relation between infants' perception of temporal invariance and developing joint attention. The utility value of this intrinsic relation is manifold. In a world of competing referents, being able to turn successfully in the direction of an object that is named, aided by caregivers' temporal invariance, by default enables infants to tune out all other potential referents and attend to the correct referent for the name. The outcome of this interactive process is that infants detect the mapping that the caregiver intended.

What might be the underlying mechanism linking infants' perception of synchrony with their gaze-switching ability? We speculate, using the Hebbian learning principle (Hebb, 1988), that neurons for perceiving temporal invariance between words and moving objects, which fire together with neurons for gaze-switching behavior, are wired together and share neuronal space. Our speculation is supported to an extent by cross-species evidence from nonhuman primates. These reports suggest that the *superior colliculus*, where multisensory neurons show response enhancement to temporally and spatially coordinated auditory–visual stimuli, is also the control center for gaze-shifting behavior (see review by Stein, Jiang, & Stanford, 2004). Thus, as discussed earlier, when animals hear a sound that is temporally aligned with a visual stimulus, they shift their gaze in the direction of the sound to locate the visual stimulus. Early multisensory experience is critical for shared communication and resultant neuronal projections from two cortical regions (the anterior ectosylvian sulcus and the lateral suprasylvian sulcus) to the superior colliculus (Stein, 2005; Wallace, Carriere, Perrault, Vaughan, & Stein, 2007). We speculate that for human infants as well, multisensory neurons that perceive and compute invariance in auditory–visual stimuli fire together with neurons that enable gaze switching and are wired together. In the case of animals, the original eye-gaze position is in the direction of prowling gait and the shift in eye gaze is to the location of the prey (see Figure 3). In the case of human infants, the original eye-gaze position is in the direction of the mother and the shift in eye gaze is to the labeled object. In human infants, the behavioral correlates of neurons for invariance detection and gaze switching from mother to object could be acting together in the service of word mapping during maternal synchronous naming (Gogate et al., 2006). In support of our speculation, researchers have suggested a positive correlation between infants' joint attention (gaze following) in the first year and word comprehension in the second year (Morales, Mundy, & Rojas, 1998; Silven, 2001). In sum, word mapping emerges in the first year as a result of infants' perception of temporal invariance between caregivers' utterances and object motion, which triggers early joint attention and enables understanding of caregivers' intended reference. Thus, for word mapping to take place, at a minimum, both abilities—perception of temporal invariance and the capacity for gaze shifting—need to be in place.

MULCH and the Quinean quandary revisited. In this subsection, we consider how the perception of temporal invariance can address Quine's (1960) question concerning how an infant can arrive at the correct referent when an adult names it amid referential uncertainty. In addition, we show how within an interactive system, infants' detection of invariance in caregivers' multisensory naming provides an alternative explanation for some lexical hypotheses or preformed biases (constraints). These biases have been shown to play a key role in children's word learning after they have had considerable experience with word mapping, but their origins have remained unclear to the present day.

For instance, according to the novel name–nameless object (category), or N3C, principle (Golinkoff et al., 1994; Markman, 1989), infants assume that a novel name refers to an unfamiliar object rather than to a familiar object or that labels are mutually exclusive. Alternatively, we suggest that infants' assumptions about word–referent mappings originate from their ability to pick up invariance across caregivers' naming and gestures. By perceiving temporal invariance between the caregivers' spoken word and simultaneous hand movements when they shake or loom the intended object, and by shifting gaze toward that object and narrowing the number of plausible options as to referents, infants successfully disambiguate the correct referent from the incorrect ones. Yu, Ballard, and Aslin (2005) provided a similar explanation for infants' word mapping using adults' body movements to glean the intended referent, called *embodied intention* (adopted from *embodied cognition*; A. Clark, 1997). Within such a system, attention to caregivers' intersensory invariance eliminates referential uncer-

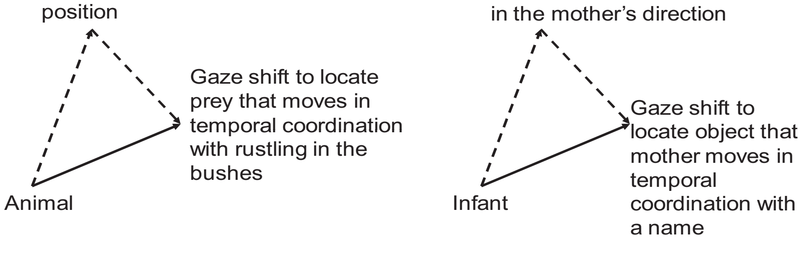

Figure 3. Gaze shifting by animals to locate prey and human infants to map words onto objects in the presence of temporally coordinated auditory–visual stimuli.

tainty for infants. Thus, the mutual exclusivity hypothesis likely originates in infants' perception of temporal invariance. Further, during conversations with infants who have less than 50 words in their productive vocabulary, caregivers also tend to use single mutually exclusive labels rather than multiple labels for basic-level object categories (Callanan & Sabbagh, 2004), supporting our view that a match between caregiver input and infant word learning within an interactive system reduces referential ambiguity.

To further address how caregivers make salient the relation between a word and its correct referent, consider the possible perceptual origins of another preformed bias, *the whole object hypothesis*, which presumably constrains novel word mappings so that infants attend first to whole rather than to parts of objects. The origins of infants' attention to a whole object as a potential referent can be found, once again, in the interactive process of lexical learning and in the match between the invariant properties that the caregivers provide and those that the infants perceive during word mapping. Because early word learning focuses on global amodal relations, caregivers will necessarily tend to move whole objects while simultaneously naming these objects. During the vast majority of naming, caregivers pick up whole objects and move them along with their parts (Gogate, Laing, Brangwin, & Perreira, 2009). Reciprocally, infants likely attend to the amodal invariants between whole objects and their names. Further supporting our claim about how caregivers constrain naming contexts during interactions, when mothers name objects for their preverbal infants, they always name a whole object when they first introduce a novel word–object relation (Gogate, Laing, et al., 2009; Masur, 1997). Only later during the play episode do mothers name salient parts of the object for their infants (see similar evidence from caregiver input to 3- to 4-year-olds in Saylor, Sabbagh, & Baldwin, 2002). Reciprocally, even when the parts of an object are given a name more often than is the whole during training, 12- and 19-month-olds attend longer to the whole object than to its parts when the same label is given during testing (Hollich, Golinkoff, & Hirsh-Pasek, 2007). These findings once again suggest a match between caregiver provision of names for whole versus part objects and infants' ability to map labels onto these referents. If the caregiver moves a whole object while simultaneously naming it, the preverbal infant likely picks up the temporal invariance between the whole object's motion and the spoken word (for whole vs. part perception during infants' categorization of moving objects, see Rakison & Poulin-Dubois, 2002). Preverbal infants will likely pick up the temporal invariance between the whole object and the spoken word first. Only later might they glean the temporal invariance between the moving part and the spoken word. Thus, by attending to caregivers' hand movement and utterance, infants can successfully map a word to its whole object referent or its parts. Infants' hypothesis about names for whole objects versus parts also originates in their perception of invariance provided by caregivers.

We predict that, within an interactive system, local-level changes in either caregivers' naming or infants' ability to perceive invariants at increasing levels of specificity, from whole objects to parts, and mapping these onto labels result in developing reference for whole versus part objects. The process proceeds as follows: Infants notice global amodal relations. Parents adapt to this focus and highlight these global amodal relations. This process tends to highlight wholes over parts. Infants tend to adapt to this and develop the default assumption that caregivers are labeling wholes (the so-called whole-object bias). Thus, if caregivers want to label a part, they must do something special. They might name the whole object before naming the part or move that part separate from the whole object while naming it. In this manner, caregivers' multisensory naming is coupled with infants' attending to amodal relations in word–referent relations. Caregivers' multisensory naming enables infants to "know" which invariants to selectively attend to and hold in memory and which invariants to ignore, at a point during development when infants are ready to pick up a specific type of invariant information. Thus, temporal invariance detection serves as a perceptual gateway for word-mapping development in preverbal infants.

In summary, in light of the previous examples of perception and learning of word–object relations, we suggest, as others have before us (e.g., Deak, 2000), that infants' word-mapping hypotheses are not preformed or built in but instead develop over time as a result of infants' interactive experience with naming contexts provided by caregivers. Moreover, as we have seen, the hypotheses do not emanate from infants or caregivers in isolation but emanate during ongoing caregiver–infant interaction. Preverbal infants' environment is initially enriched from the infants' point of view (Nelson, 1988) by caregivers' use of temporal invariance, between words and gestures (e.g., shaking or looming an object), which in turn triggers joint attention and word-mapping behavior that at the outset might resemble a preformed bias.

III. Future Directions for Invariance Detection Research in Language Development

In the preceding sections, we have tried to provide some insights into the unifying power of invariance detection to explain a wide range of phenomenon in language development. Specifically, we have shown that considering invariance within the dynamic relation between infant and caregiver communication provides a perceptual gateway to language understanding, including learning of transitional probabilities (Saffran et al., 1996) and abstract patterns (Gomez, 2002; Marcus, 2001), as well as the learning of word–referent relations.

Despite the clear advantages of such an interactive approach, much work remains. We recommend three converging lines of research to advance our understanding of the mechanisms of invariance detection and validate the theory. Researchers must (a) conduct developmental experiments in which they examine infants' detection of and caregivers' provision of specific types of invariance and the reciprocal changes that occur over time in each, (b) explore sensory-oriented computational models that capture how those dynamic interactions work in the real world, and (c) validate a neural theory that moves us beyond the domain general/specific debate and defines the specifics of neural implementation.

Developmental Experiments

The study of invariance detection by infants that match with the invariance in caregivers' input is in its infancy. Researchers have just begun to tap the range of invariant properties that infants may be sensitive to and are only just beginning to understand how perception of invariance may change over time, driven by infants' interaction with the environment. Infant–environment interactions are bidirectional. For example, with respect to word mapping, it is clear that parents' tendency to move objects in synchrony with a word changes over time (Gogate et al., 2000; also see Adamson & Bakeman, 2006). It is also clear that infants' weighting of synchrony for learning word–referent relations decreases in the second year of life, as detection of other social and linguistic invariants comes to the fore (Akhtar & Tomasello, 1996; Baldwin, 1993; Baldwin et al., 1996; also see Hollich et al., 2000, for change in weighting of perceptual salience from 12 to 24 months; also Moore, Angeloupolus, & Bennet, 1999). As discussed previously, this is likely due to infants' changing responses to parents' use of invariance and/or parents' changing responses. Such dynamics are also applicable to other invariant properties, which could likewise be detected through changing patterns of parent–child interaction. We argue that close examination of the dynamics of parent–child interaction could provide useful insights concerning the invariant properties of the environment and infants' changing sensitivities to and weightings for these properties. After all, if mothers use an invariant property consistently, it is likely because infants at that age are particularly sensitive to it and respond. Conversely, if mothers stop using an invariant property or begin to emphasize others, it is likely because infants are no longer responding in the same way. It is also likely that evidence of these dynamics from meta-analyses of prior studies of caregiver input and infant experiments could be found (e.g., for developmental changes in maternal directive vs. follow-in labeling and infant word learning supporting these dynamics, see Gogate et al., 2006, and Akhtar, Dunham, & Dunham, 1991).

An additional area of empirical research in invariance involves longitudinal studies that connect early audiovisual and perceptual abilities with mature language-development and word-learning principles. An oft-cited criticism of the work on infants' invariance detection at 6 months is that it has no more to do with the development of mature-language principles than does any other perceptual mechanism: It is necessary but not integral to the process (Werker & Patterson, 2001). After all, chimps (Savage-Rumbaugh et al., 1993), parrots (Pepperberg, 1999), and chinchillas (Kuhl & Miller, 1978) can detect phonetic, lexical, and grammatical invariance, yet they fall far short of the *sophisticated* language-learning behavior of a human at 24 months. In short, studies that directly explore the connection between early perception and advanced language are needed. An excellent example of how such connections might be made can be found in the work of Smith and her colleague (Smith, 1999, 2003; Yoshida & Smith, 2005). In studying the shape bias, the tendency in older children to spontaneously extend a new word to other similarly shaped objects, Smith (2003) found that toddlers, from 17 to 25 months, with less than 100 object names in their productive vocabulary do not recognize abstract depictions of objects (e.g., a caricaturized model of a pizza). In contrast, toddlers with more than 100 object names in their vocabulary connect the label to the abstraction. From such evidence, Smith (2003) suggested that the tendency to extend new words on the basis of shape similarity appears only in infants laboriously learning a critical mass of words (producing approximately 100 object names) and then making generalizations from these about how new words may extend. Thus, an apparent mature principle of word learning, the shape bias, results from the working of "dumb attentional mechanisms" (Smith, Jones, & Landau, 1996, pp. 143; cf. Booth & Waxman, 2003; Cimpian & Markman, 2005; Diesendruck & Bloom, 2003; Smith, 2002). We would suggest that this is true for other apparently sophisticated mechanisms. Thus, other presumed preformed biases (see Section I) might also have their origins in infants' ability to perceive and attend to invariance and caregivers' provision of these invariant properties.

In a further line of empirical research, the origins of intersensory invariance detection mechanisms—whether the senses are unified at birth or become integrated via associative learning mechanisms—would be examined. Some researchers have theorized that detection of intersensory invariance, or redundancy across the senses, is possible right from birth, owing to an integrated perceptual system. With development, via the process of increasing differentiation, infants come to perceive information that is specific to each sensory modality (E. J. Gibson, 1969; the *increasing specificity hypothesis*, Bahrick & Lickliter, 2000; Bahrick & Pickens, 1994; Walker-Andrews, 1994). Furthermore, some researchers have suggested that intersensory invariance recruits infants' attention and enables infants to unify the auditory and visual components of multisensory patterns of stimulation (see *intersensory redundancy hypothesis*, Bahrick & Lickliter, 2000; also see Gogate, Walker-Andrews, & Bahrick, 2001; Walker-Andrews, 1994). By detecting intersensory invariance and perceptually unifying auditory–visual patterns of stimulation, infants treat them as single bimodal events rather than as separate events in individual modalities. In contrast, others have theorized that the senses are separate at birth and become integrated only with experience (the *associationist view*, Birch & Lefford, 1967). Future studies need to elucidate the nature of invariance detection mechanisms and specify whether they are present at birth or develop during the first few weeks of life to provide a perceptual gateway to language development.

Another line of research concerns the degree to which invariance detection is domain-general. Do infants use similar perceptual processes and computational mechanisms to detect invariance in nonspeech stimuli as they do in speech? For example, 8-month-olds detect transitional probabilities in synthesized syllabic strings (Saffran et al., 1996) and tonal strings (Saffran, Johnson, Aslin, & Newport, 1999). In addition, infants at 2, 5, and 8 months detect co-occurrence frequencies in visual sequences containing discrete, looming shapes (Kirkham, Slemmer, & Johnson, 2002). These studies suggest that domain-general processes (mechanisms)—perception and attention to structural regularities or invariance in stimuli—enable probabilistic learning. Furthermore, in studies of speech and nonspeech, temporal invariance assists infants in unifying otherwise disparate auditory and visual information and facilitates infants' learning of auditory–visual relations. Infants at 7 months learn the auditory–visual relations between two vowel sounds—/a/ and /i/—and two objects if the sounds are vocalized in synchrony with the motions of the objects (Gogate & Bahrick, 1998). Similarly, in the nonspeech domain, infants at 7 months, but not 3 or 5 months, learn the auditory–visual relations between two different pitches (high and low) of impact sounds and two objects, again if synchrony is provided between the sounds and the objects' motions (Bahrick, 1994).

A related issue—whether the detection of specific types of invariance starts out as domain-general and becomes more language-specific with development—also remains an empirical question. For example, in their 2nd year, infants respond to gestures, object sounds, and mouth noises when embedded in naming frames just as they do with words (Hollich et al., 2000; Namy & Waxman, 1998; Woodward & Hoyne, 1999; cf. Balaban & Waxman, 1997). However, by the end of the 2nd year they will not respond to these nonwords or gestures as they do with words. And finally, if presented with auditory–visual patterns of stimulation, infants at 7.5 months detect invariance in continuous speech when the amplitude changes of an oscilloscope as well as a talking female face are synchronized with the speech stream (Hollich, Newman, & Jusczyk, 2005). It is possible that with development, these infants might not accept such stimuli as candidate visual complements.

Sensory-Oriented Computational Models

In future work researchers must also explore algorithmic implementations of specific mechanisms of invariance detection and test the performance of these algorithms in developing interactive systems. Demonstrating empirically that infants are sensitive to some range of social, cognitive, and linguistic invariants (e.g., Hollich et al., 2000) is only the first step in validating an ecological theory; one must also specify the algorithm (in the form of a computational model) and the precise set of parameters and variables necessary for this detection (Marr, 1982).

One way to do this is through the use of sensory-oriented models (Hollich & Prince, 2009). Sensory-oriented models are computational models that accept raw perceptual inputs and enable us to (a) study specific mechanisms; (b) discover parameters; and (c) explore, at an unprecedented level of detail, the nature of infants' multisensory perception and learning (Prince & Hollich, 2005). Until recently, computational models have relied on rather caricatured versions of the environmental input. Indeed, many connectionist models of word learning abstractly encode their inputs (e.g., Colunga & Smith, 2005; Li, Farkas, & MacWhinney, 2004; Plunkett, Sinha, Moller, & Strandsby, 1992). Although this is an excellent beginning, it is through the direct interaction with real-world input that one can test the validity and scalability of the simplifications (Yu et al., 2005). In other words, it is not enough to hypothesize that a particular invariant property exists and demonstrate that infants are sensitive to it. One must explore the interactions and fine-grained behavior that result in the real world from this sensitivity.

For example, recent work on modeling infants' temporal invariance detection (Gogate, Prince, & Matatyaho, 2009; Prince & Hollich, 2005) reveals that infant performance closely matches that of a computational model of synchrony detection. Prince and Hollich (2005) found that a model of visual salience based on sudden visual change or movement can account for 43% of the looking behavior of infants in an ostentatiously audiovisual task. Similarly, work by Ogata, Hattori, Kozima, Komatani, and Hiroshi (2006) on robot learning of arbitrary sound–action relations based on amodal information provides an algorithmic *proof of concept* for one way in which word learning might proceed: by creating a robot that learns to associate particular visual events that co-occur with particular sounds (e.g., link a bouncing ball to its impact sounds) and to reproduce those sounds when the event is presented later. In these studies, the models operated on raw sensory input (a camera and microphone feed), not abstract encodings of it. Thus, by exhibiting a tighter coupling between model and reality, these models represent the next step in computational modeling. Future models of invariance detection of rhythm, intensity shifts, tempo, and other invariant properties would also help ground the theorizing about invariance detection mechanisms of infants.

More sophisticated models should include caregivers as part of the system to predict fine-grained interactions between infants and their caregivers and examine the changes in these interactions at different points during language development. By way of example, work by Poulin-Dubois (2006) on contingent naming by a robot caregiver and word learning by human infants could help show intentionality and the types of invariance that facilitate infants' learning at different phases of lexical mapping. Kozima, Nakagawa, and Yasuda (2005) demonstrated that the response of an autistic child changes depending on the degree of contingency in an interactive robot. Finally, Nagai and Rohlfing (2007) demonstrated that a saliency-based attention model shows greater attention to a parent's ID talking face during real-time parent–infant interaction with objects than to the same parent's face during parent–adult interaction.

Finally, through these models researchers should also be able to explicate the developmental process by showing how invariance detection at one level might facilitate invariance detection at the next, more advanced level. For example, Monaghan, Chater, and Christiansen (2005) have shown that the learning of phonological invariants can facilitate the detection of invariant grammatical categories by a neural network. Also, Christiansen, Onnis, and Hockema (2009) have shown that invariance detection at the phonological level can facilitate invariance detection of lexical categories (see also Jolly & Plunkett, 2008). Finally, child-directed speech models, which tend to have a higher frequency of occurrence of certain words, seem to set off an earlier vocabulary growth spurt (McMurray, 2007).

Neurophysiological Studies

The final piece of the puzzle involves specification of the precise neurophysiological mechanisms involved. Whereas empirical studies can show that infants *are* sensitive to invariance and computational models can show *how* they might detect such invariance, cognitive neuroscience can help determine the neurophysiology of invariance detection. The growing knowledge of this neurophysiology can, in turn, help constrain theorizing in empirical studies and computational models.

Our current hypothesis about the neurophysiology of invariance detection is that children possess a collection of different invariance detectors, each operating at the confluence of the different sensory modalities. For example, at the confluence of audio and visual streams of processing lie neurons that are sensitive to information from both modalities, so-called superadditive neurons, and these indeed respond more when the information is redundant across the senses (Stein et al., 2004). Neurophysiological studies of cats, rhesus monkeys, and rodents reveal such *response enhancement* in the *superior colliculus* (mediated by projections from cortical regions, the *anterior ectosylvian sulcus and the rostral lateral suprasylvian sulcus,* which also contain multisensory neurons) to species-specific temporally and spatially aligned auditory–visual stimuli (e.g., calls of adult conspecifics) compared with unimodal visual or auditory stimulation (see review by Stein et al., 2004).

This midbrain region is also a major control center for gaze-shifting behaviors. Owing to shared neuronal *space*, upon hearing a sound occurring simultaneously with a visual stimulus, animals shift their eye gaze in the direction of the sound to locate the visual stimulus. Such cross-species evidence is congruent with the view that in humans, intersensory invariance detection is an interactive process involving infants' perceptual, neural, and motor responses to caregivers' multisensory communication.

Furthermore, although multisensory neurons may be located in separate parts of the brain, they are unified by the fact that they all detect some consistent patterns within a changing stimulus array. Indeed, such *pattern associator* networks are one of three general types of networks found throughout the brain (Rolls & Treves, 1998). Thus, it is not surprising that invariance detection is fundamental to a range of behaviors including language. For example, in the visual domain, the ability to recognize an object across multiple views appears to be calculated by a hierarchical network of neurons organized from the lateral geniculate nucleus to area TE of the inferior temporal visual cortex (Tovee, Rolls, & Azzopardi, 1994). Each successive neural area appears to be responsible for detection of ever more abstract representations, until position-independent recognition is achieved. It is likely that similar network architecture allows one to recognize the same word spoken by two different speakers in successive areas of the auditory cortex and temporal lobe. In this manner, we posit that hierarchically layered pattern–associator networks distributed throughout the cortex abstract consistent patterns across a changing stimulus array within and across modalities. Thus, invariance detection is not domain-specific in the sense of a specific localized portion of the cortex. Instead, it is domain-general in that it is implemented by similar types of network architectures wherever there is a confluence of information.

Although this theory is, as yet, empirically untested, there are established means to tell whether there is a *localized* general mechanism or many *distributed* mechanisms for invariance detection (Conway & Christiansen, 2005). The first is the existence of a double dissociation between detection of two different invariant properties. This could be tested in clinical populations of patients who, following focal lesions, might detect some kinds of invariance but not others. Dissociations can also be tested in the impaired behavior of normal adults under conditions of distraction, with, for example, the Garner (1974) paradigm. In this paradigm, reaction time is used to examine the relative independence of two skills. If infants can detect one invariant property as quickly whether or not another distracting invariant property is presented, one can surmise that the mechanisms for detection are relatively independent. Finally, a more direct way of studying localized versus distributed cortical mechanisms would be to examine the neurophysiology of invariance detection in event-related potential or functional magnetic resonance imaging studies.

In addition to studying the neurophysiology of infants' invariance detection, one might also study the changing neural signals that occur during interactions between parents and children. In all prior studies of which we are aware, only the signal from one individual is measured. From behavioral studies presented here earlier, it is clear that mothers and infants are constantly adjusting their behavior from moment to moment in intersubjective responsiveness. It would be informative to gain some understanding of the concomitant neurological changes that underlie these dynamic adjustments in behavior. For example, does intersubjectivity during mother–child interaction reflect in neurological intersubjectivity?

IV. Conclusions

In this article, we theorized that infants learn about speech and language using rudimentary but powerful domain-general invariance detection abilities that are well developed in the first year of life (see the introduction). Caregivers complement these invariance detection abilities by providing communication that is well matched to their infants' developing sensitivities. In support of this thesis, we showed that infants attend to invariance in unimodal and bimodal patterns of caregivers' multisensory communication (see Sections I and II). They learn language by perceiving invariant structure and economically utilizing it to glean what is predictable in the ambient language. Furthermore, we illustrated how caregivers provide invariant patterns in their communication to infants across several domains of language. Thus, infants' invariance detection abilities coupled with caregivers' provision of invariance serve as a perceptual gateway for learning the ambient language. Thus, language development takes place within a complex, multicausal, multilevel, interactive system comprising infants and their immediate environment. Invariance is present in abundance in the ambient language. Perception of invariance, aided by neural structures for computing invariance, leads to overt actions upon the environment (e.g., gaze-shifting behavior).

To evaluate our interactive account of invariance detection with respect to word learning, in Section II we considered at length the real-time match between caregivers' synchronous naming and infants' ability to perceive synchrony and shift their gaze from their mother to an object. Specifically, we showed that invariant temporal properties of caregivers' naming contexts are perceived and utilized by infants in the service of word mapping, at least by age 6–8 months. Infants perceive that which is made salient by their caregiver—*the temporally invariant relation between a novel name and a novel object* —during ongoing caregiver–infant interaction. By providing invariance between word and gesture, caregivers perceptually highlight or foreground a specific word–referent relation so as to make the picking up of that relation simpler for their word-mapping novices. Other potential referents, not highlighted in this manner, fade away into the background and are tuned out of infants' immediate perceptual fields. In this manner, in a world of competing referents, detection of invariance between caregivers' utterances and handheld object motions helps infants find correct word–referent mappings and tune out incorrect ones. Furthermore, we showed that within an interactive system, caregivers' provision of temporal invariance between utterance and specific types of handheld object motions triggered infants' gaze (attention) shift from caregiver to object and perception of the relation between the word and caregivers' intended referent. Thus, temporal invariance and invariant object motion might also serve as perceptual gateways to infants' understanding of caregivers' *intentionality*, of importance to infants' learning of word meaning (P. Bloom, 2000).

In our view, the present work expands upon previous theories by providing a single perceptual framework from which to understand the origins of language. It also explicitly forces us to consider the dynamic real-time interaction between infants and caregivers during language learning. Although this approach is advocated by many for the study of development in general (E. J. Gibson, 1969, 1991; Lickliter, 2006; Smith & Thelen, 2003; Thelen & Smith, 1994, 1998), and language development in particular (L. Bloom, 1998; Tomasello, 1995), prior approaches have not (with the exception of Cameron-Faulkner et al., 2003; Gogate et al., 2006; Matatyaho & Gogate, 2008; also see Schafer, 2005) explicitly examined the real-time matches between children's receptive language learning and caregiver input (also see Bates, Bretherton, & Snyder, 1988, for integrative analyses of parents' input and children's language learning and of continuities in the production of gestures, words, and syntax).

Our thesis, which focuses mainly on prelinguistic infants' capabilities, is consistent with the work of a few other researchers (e.g., Smith, 2005) on older infants' language development. It provides a perceptual basis for the hypotheses or rules that infants arrive at in several language domains. In one aspect, our work is also consistent with traditional theories of language acquisition (Marcus, 2001; Pinker, 1999). It proposes that infants or children detect invariant patterns (or rules) that are salient in the ambient language, but it provides a perceptual basis for the rules that they detect within an interactive system. Thus, it is also comparable with more recent theories of language acquisition that underscore the role of input or perceptual support (Jackendoff, 2002; Tomasello, 2006). In addition, this theory provides a concrete perceptual framework from which phonemes, words, and grammar emerge within an interactive system.

In conclusion, language development in infants is a truly interactive process, involving caregivers' communication and provision of invariance and infant detection of invariant patterns in caregivers' communication. Future studies of language development (see Section III) should describe particular invariant properties and their changing role over time, provide increasingly detailed sensory-oriented computational models, and specify the neural correlates of invariance detection during caregiver–child interactions.

References

Adamson, L., & Bakeman, R. (2006). The development of displaced speech in early mother-child conversations. *Child Development, 77,* 186–200.

Akhtar, N., Dunham, F., & Dunham, P. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language, 18,* 41–49.

Akhtar, N., Jipson, J., & Callanan, M. (2001). Learning words through overhearing. *Child Development, 72,* 416–430.

Akhtar, N., & Tomasello, M. (1996). Twenty-four-month-old children learn words for absent objects and actions. *British Journal of Developmental Psychology, 14,* 79–93.

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology, 6,* 111–123.

Bahrick, L. E., Gogate, L. J., & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: The salience of actions over faces in dynamic events. *Child Development, 73,* 1629–1643.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36,* 190–201.

Bahrick, L. E., & Pickens, J. N. (1988). Classification of bimodal English and Spanish passages by infants. *Infant Behavior and Development, 11,* 277–296.

Bahrick, L. E., & Pickens, J. N. (1994). Amodal relations: The basis for intermodal perception and learning in infancy. In D. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 205–233). Hillsdale, NJ: Erlbaum.

Bahrick, L. E., Soutullo-Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development, 69,* 1263–1275.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology, 64,* 3–26.

Baldwin, D. (1993). Early referential understanding: Infants' ability to understand referential acts for what they are. *Developmental Psychology, 29,* 832–843.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67,* 3135–3153.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York, NY: Academic Press.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* New York, NY: Cambridge University Press.

Birch, H., & Lefford, A. (1967). Visual differentiation, intersensory integration and voluntary motor control. *Monographs of the Society for Research in Child Development, 32*(1–2, Serial No. 110), 1–83.

Bloom, L. (1998). Language acquisition in its developmental context. In W. Damon, D. Kuhn, & R. S. Siegler (Eds.), *Handbook of Child Psychology: Vol. 2. Cognition, perception, and language* (pp. 309–370). New York, NY: Wiley.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Booth, A. E., & Waxman, S. R. (2003). Mapping words to the world in infancy: Infants' expectation for count nouns and adjectives. *Journal of Cognition and Development, 4,* 357–381.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathburn, K. (2005). Mommy and me: Familiar words help launch babies into speech stream segmentation. *Psychological Science, 16,* 298–303.

Bosch, L., & Sebastian-Galles, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language & Speech, 46,* 217–243.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science, 5,* 72–83.

Brand, R. J., Shallcross, W., Sabatos, M., & Massie, K. P. (2007). Fine-grained analysis of motionese: Eye-gaze, object exchanges, and action units in infant- versus adult-directed action. *Infancy, 11,* 203–214.

Bronfenbrenner, U. (1979). *The ecology of human development.* Cambridge, MA: Harvard University Press.

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

Callanan, M., & Sabbagh, M. (2004). Multiple labels for objects in conversations with young children: Parents' language and children's developing expectations about word meaning, *Developmental Psychology, 40,* 746–763.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech, *Cognitive Science, 27,* 843–873.

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using frequent frames: What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science, 12,* 396–406.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Christiansen, M., Onnis, L., & Hockema, S. (2009). The secret is in the sound: From unsegmented speech to lexical categories. *Psychological Science, 12,* 388–395.

Cimpian, A., & Markman, E. (2005). The absence of a shape bias in children's word learning. *Developmental Psychology, 41,* 1003–1019.

Clark, A. (1997). *Being there: Putting brain, body, and world together again.* Cambridge, MA: MIT Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58,* 7–19.

Clark, E. (1995). *The lexicon in acquisition* (Cambridge Studies in Linguistics Series, No. 65). Cambridge, England: Cambridge University Press.

Clark, E. (2003). *First language acquisition.* New York, NY: Cambridge University Press.

Colunga, E., & Smith, L. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112,* 347–382.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 24–39.

Cooper, R. P., Abraham, S., Berman, S., & Statska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development, 20,* 477–488.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584–1595.

Deak, G. O. (2000). Hunting the fox of word learning: Why constraints fail to capture it. *Developmental Review, 20,* 29–80.

De Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language, 11,* 1–15.

DeCasper, A. J., & Fifer, W. P. (1980, June 6). Of human bonding: Newborns prefer their mothers' voices. *Science, 208,* 1174–1176.

Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development, 74,* 168–178.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11,* 478–484.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971, January 22). Speech perception in infants. *Science, 171,* 303–306.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71–99.

Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242).

Fernald, A. (1992). Human vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford, England: Oxford University Press.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27,* 209–221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mother's speech to newborns. *Developmental Psychology, 20,* 104–113.

Fletcher, P., & Garman, M. (1987). *Language acquisition* (2nd ed.). Cambridge, MA: MIT Press.

Floor, P., & Akhtar, N. (2006). Can 18-month-olds learn words by listening in on conversations? *Infancy, 9,* 327–339.

Fulkerson, A. L., & Waxman, S. R. (2007). Linking object names and object categories: Words (but not tones) facilitate object categorization in 6- and 12-month-olds. *Cognition, 105,* 218–228.

Garner, W. R. (1974). *The processing of information and structure.* Mahwah, NJ: Erlbaum.

Gerken, L. A. (2004). Nine-month-olds extract structural principles required for natural language. *Cognition, 93,* 89–96.

Gibson, E. J. (1966). Perceptual development and the reduction of uncertainty. In H. Tajfel, G. Jahoda, C. Nemeth, J. Campbell, & N. Johnson, *Proceedings of the 18th International Congress of Psychology* (pp. 7–17). Moscow, Russia: Nauka.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York, NY: Appleton–Century–Crofts.

Gibson, E. J. (1991). *An odyssey in learning and perception.* Cambridge, MA: MIT Press.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32–41.

Gogate, L. J. (2010). Learning of syllable-object relations by preverbal infants: The role of temporal synchrony and syllable distinctiveness. *Journal of Experimental Child Psychology, 103,* 178–197.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in 7-month-old infants. *Journal of Experimental Child Psychology, 69,* 133–149.

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for syllable-object relations. *Infancy, 2,* 219–231.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71,* 876–892.

Gogate, L. J., Bewley, S., Wan, P., Kremen, I., & Chen, L.-F. (2003, May). *Preverbal infants' learning of arbitrary relations between complex syllables and objects.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Tampa, FL.

Gogate, L. J., Bolzani, L. H., & Betancourt, E. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy, 9,* 259–288.

Gogate, L. J., Friedman, A., & Bewley, S. (2001, April). *Cross-cultural evidence for multimodal motherese: The role of temporal synchrony between verbal labels and gestures.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Minneapolis, MN.

Gogate, L. J., Laing, K., Brangwin, H., & Pereira, V. (2009, April). *Revisiting the whole object hypothesis: Built-in or environmental constraints?* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Denver, Colorado.

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable-object pairings: The role of temporal synchrony. *Journal of Experiment Psychology: Human Perception and Performance, 35,* 508–519.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science, 4,* 1–18.

Goldstein, M., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19,* 515–523.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21,* 125–155.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20,* 229–252.

Goodwyn, S., Acredolo, L., & Brown, S. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior, 24,* 81–103.

Hebb, D. O. (1988). *The organization of behavior.* Cambridge, MA: MIT Press.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269–286.

Hockema, S. A., & Smith, L. B. (2009). Learning your language, outside-in and inside-out. *Linguistics, 47,* 453–479.

Hoff-Ginsberg, E. (1997). The development of syntax and morphology: Learning the structure of language. In E. Hoff-Ginsberg (Ed.), *Language development* (pp. 129–183). Pacific Grove, CA: Brooks/Cole.

Hollich, G., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Young children associate novel words with complex objects rather than salient parts. *Developmental Psychology, 5,* 1051–1061.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development, 65*(3, Serial No. 262).

Hollich, G. J., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development, 76,* 598–613.

Hollich, G., & Prince, C. G. (2009). Do you see what I hear? Comparing infants' preference for synchronized audiovisual speech with signal-level computational models. *Developmental Science, 12,* 379–387.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1570–1582.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 1143–1154.

Hudson-Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adults and child learners in language formation and change. *Language Learning and Development, 1,* 151–195.

Jackendoff, R. (2002). *Foundations of language.* Oxford, England: Oxford University Press.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language.* The Hague, The Netherlands: Mouton.

Jolly, H., & Plunkett, K. (2008). Inflectional bootstrapping in 2-year-olds. *Language and Speech, 51,* 45–59.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1–23.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for predominant stress patterns of English words. *Child Development, 64,* 675–687.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to the phonotactic patterns of the native language. *Journal of Memory and Language, 33,* 530–545.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist in language learning. *Journal of Child Language, 16,* 55–68.

Kirkham, N., Slemmer, J., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism, *Cognition, 83,* 35–42.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics, 35,* 379–415.

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication-care: A case-study in autism therapy. *IEEE International Workshop on Robot and Human Interactive Communication,* (ROMAN-2005), 341–346.

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

Kuhl, P. K., & Meltzoff, A. N. (1982, December 10). The bimodal perception of speech in infancy. *Science, 218,* 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development, 7,* 361–381.

Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy: The Minnesota Symposia on Child Psychology* (Vol. 20, pp. 235–266). Mahwah, NJ: Erlbaum.

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., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992, January 31). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608.

Legerstee, M. (1990). Infants' use of multimodal information to imitate speech sounds. *Infant Behavior and Development, 13,* 343–354.

Lerner, R. M. (2001). *Concepts and theories of human development* (3rd ed.). Mahwah, NJ: Erlbaum.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants, *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1094–1106.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17,* 1345–1362.

Lickliter, R. (2006). Dynamical systems and psychological science. In B. D. Midgley & E. K. Morris (Eds.), *Modern perspectives on J. R. Kantor and interbehaviorism* (pp. 173–194). Greenwood, CT: Context Press.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983, March 18). Infant intermodal speech perception is a left hemisphere function. *Science, 219,* 1347–1349.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6,* 314–317.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism with cognitive science.* Cambridge, MA: MIT Press.

Marcus, G. F., Fernandez, K., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18,* 387–391.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999, January 1). Rule-learning by 7-month-old infants. *Science, 283,* 77–80.

Markman, E. (1989). *Categorization and naming in children.* Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision: A computational approach.* New York, NY: Freeman.

Masur, E. F. (1997). Maternal labeling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language, 24,* 427–439.

Matatyaho, D., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word-object relations. *Infancy, 13,* 172–184.

Matatyaho, D., Mason, Z., & Gogate, L. J. (2007). Word learning by 8-month-old infants: The role of object motion and synchrony. *Proceedings of the 7th International Conference on Epigenetic Robotics: Lund University Cognitive Studies, 135,* 201–202.

Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9,* 316–322.

McMurray, B. (2007, August 3). Defusing the childhood vocabulary explosion, *Science, 317,* 631.

McRoberts, G. W., & Best, C. T. (1997). Accommodation in mean f_o during mother-infant and father-infant vocal interactions: A longitudinal case study. *Journal of Child Language, 24,* 719–736.

Messer, D. (1978). The integration of mothers' referential speech with joint play. *Child Development, 49,* 781–787.

Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child-directed speech. *Cognition, 90,* 91–117.

Monaghan, P., Chater, N., & Christiansen, M. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition, 96,* 143–182.

Montague, D. P. F., & Walker-Andrews, A. S. (2002). Mothers, fathers and infants: The role of person familiarity and parental involvement in infants' perception of emotion expressions. *Child Development, 73,* 1339–1352.

Moore, C., Angeloupolus, M., & Bennet, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35,* 60–68.

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, 21,* 373–377.

Nagai, Y., & Rohlfing, K. J. (2007, April). *Can motionese tell infants and robots "what to imitate"?* Paper presented at the meeting of the Society for Artificial Intelligence and the Simulation of Behaviour, Bristol, England.

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretation of different forms of symbolic reference. *Child Development, 69,* 295–308.

Nelson, K. (1978). Early speech in its communicative context. In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities: Early behavioral assessment* (pp. 443–473). Baltimore, MD: University Park Press.

Nelson, K. (1988). Constraints on word learning? *Cognitive Development, 3,* 221–246.

Nuckolls, J. B. (1999). The case of sound symbolism. *Annual Review of Anthropology, 28,* 225–252.

Ogata, T., Hattori, Y., Kozima, H., Komatani, K., & Hiroshi, O. (2006, September). *Generation of robot motions from environmental sounds using intermodality mapping by RNNPB.* Paper presented at the Sixth International Workshop on Epigenetic Robotics, Paris, France.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development, 59,* 441–449.

Patterson, M., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6,* 191–196.

Pegg, J. E., & Werker, J. F. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15,* 325–345.

Pepperberg, I. (1999). *The Alex studies: Cognitive and communicative abilities of grey parrots.* Cambridge, MA: Harvard University Press.

Pfiefer, R., & Scheier, C. (1999). *Understanding intelligence.* Cambridge, MA: MIT Press.

Pickens, J., Field, T., Nawrocki, T., Martinez, A., Soutullo, D., & Gonzalez, J. (1994). Full-term and preterm infants' perception of face-voice synchrony. *Infant Behavior & Development, 17,* 447–455.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: MIT Press.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 399–441). Mahwah, NJ: Erlbaum.

Pinker, S. (1999). *Words and rules: The ingredients of language.* London, England: Weidenfield & Nicolson.

Plunkett, K. (2006). Learning how to be flexible with words. In Y. Munakata & M. Johnson (Eds.), *Processes of change in brain and cognitive development* (pp. 234–250). New York, NY: Oxford University Press.

Plunkett, K., Sinha, C., Moller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science, 4,* 293–312.

Poulin-Dubois, D. (2006, June). Infants' concept of mentalistic agent: Who can have intentions and goals? In S. Itakura, K. Hiraki, A. Arita, K. Kamewari, R. Baillargeon, & D. Poulin-Dubois, *Infants' Conceptual Representation of Intentional Agent.* Symposium conducted at the International Conference on Infant Studies, Kyoto, Japan.

Prince, C. G., Helder, N. A., & Hollich, G. J. (2005). Ongoing emergence: A core concept in epigenetic robotics. *Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems: Lund University Cognitive Studies, 123,* 63–70.

Prince, C. G., & Hollich, G. J. (2005). Synching infants with models: A perceptual-level model of infant synchrony detection. *Journal of Cognitive Systems Research, 6,* 205–228.

Pruden, S., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. (2006). The birth of words: 10-month-olds learn words through perceptual salience. *Child Development, 77,* 266–280.

Pulverman, R., & Golinkoff, R. M. (2004). Seven-month-olds' attention to potential verb referents in nonlinguistic events. In A. Brugos, L. Micciulla, & C.E. Smith (Eds.), *Proceedings of the 28th Annual Boston University Conference on Language Development* (Vol. 2, pp. 473–480). Somerville, MA: Cascadilla Press.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Rakison, D., & Poulin-Dubois, D. (2002). You go this way and I'll go that way: Developmental changes in infants' detection of correlations among static and dynamic features in motion events. *Child Development, 73,* 682–699.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies, 8,* 3–34.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in a social context.* New York, NY: Oxford University Press.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships to later language. *Applied Psycholinguistics, 24,* 221–234.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function.* Oxford, England: Oxford University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tonal structure by infants and adults. *Cognition, 70,* 27–52.

Sameroff, A. (1975). Transactional models in early social relations. *Human Development, 18,* 65–79.

Santelmann, L., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations of processing space. *Cognition, 69,* 105–134.

Savage-Rumbaugh, S. E., Murphy, J., Sevcik, R., Brakke, K., Williams, S., & Rumbaugh, D. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, 58*(3–4): 1–256.

Saylor, M., Sabbagh, M., & Baldwin, D. (2002). Children use whole-part juxtaposition as a pragmatic cue to word meaning. *Developmental Psychology, 38,* 993–1003.

Schafer, G. (2005). Infants can learn decontextualized words before their first birthday. *Child Development, 76,* 87–96.

Schieffelin, B. (1979). Getting it together: An ethnographic approach to the study of the development of communicative competence. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics.* New York, NY: Academic Press.

Seidl, A., & Johnson, E. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9,* 565–573.

Senghas, A. (1995). The development of Nicaraguan Sign Language via the language acquisition process. In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of the 19th Boston University Conference on Language Development* (pp. 543–552). Somerville, MA: Cascadilla Press.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science, 12,* 323–328.

Silven, M. (2001). Attention in very young infants predicts learning of first words. *Infant Behavior and Development, 24,* 229–237.

Singleton, J., & Newport, E. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology, 49,* 370–409.

Slater, A., Brown, E., & Badenoch, M. (1997). Intermodal perception at birth: Newborn infants' memory for arbitrary auditory-visual pairings. *Early Development and Parenting, 6,* 99–104.

Slater, A., Quinn, P., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of auditory-visual pairings. *Developmental Science, 2,* 333–338.

Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language* (pp. 277–304). Mahwah, NJ: Erlbaum.

Smith, L. B. (2002). Learning how to learn words: An associative crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, A. L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich. (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). Oxford, England: Oxford University Press.

Smith, L. B. (2003). Learning to recognize objects. *Psychological Science, 14,* 244–250.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review, 25,* 278–298.

Smith, L. B., Jones, S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60,* 143–171.

Smith, L. B., Landau, B., & Jones, S. (1998). Object shape, object function, object name. *Journal of Memory and Language, 38,* 1–27.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7,* 243–348.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development, 43,* 549–565.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49,* 249–267.

Stager, C. L., & Werker, J. F. (1997, July 24). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388,* 381–382.

Stein, B. E. (2005). The development of a dialogue between cortex and midbrain to integrate multisensory information. *Experimental Brain Research, 166,* 305–315.

Stein, B. E., Jiang, W., & Stanford, T. (2004). Multisensory integration in single neurons of the midbrain. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (Vol. 15, pp. 243–264). Cambridge, MA: MIT Press.

Sullivan, J., & Horowitz, F. D. (1983). Infant intermodal perception and maternal multimodal stimulation: Implications for language development. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 2, pp. 183–239). Norwood, NJ: Ablex.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Thelen, E., & Smith, L. B. (1998). Dynamic systems theories. In W. Damon & R. M. Lerner (Eds.), *Handbook of Child Psychology: Vol. 1. Theoretical models of human development* (pp. 147–198). New York, NY: Wiley.

Thiessen, E., Hill, E., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7,* 53–71.

Thompson, S., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3,* 1–42.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10,* 172–175.

Tomasello, M. (1995). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology: Vol. 2. Cognition, Perception, and Language* (pp. 255–298). New York, NY: Wiley.

Tovee, M. J., Rolls, E. T., & Azzopardi, P. (1994). Translation invariance and the responses of neurons in the temporal visual cortical areas of primates. *Journal of Neurophysiology, 70,* 1049–1060.

Volterra, V., Iverson, J. M., & Castrataro, M. (2005). The development of gesture in hearing and deaf children. In B. Schick, M. Marschark, P. E. Spencer (Eds.), *Advances in sign language* (pp. 46–70). New York, NY: Oxford University Press.

Walker-Andrews, A. S. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz & R. Lickliter (Eds.*), The development of intersensory perception: Comparative perspectives* (pp. 39–56). Mahwah, NJ: Erlbaum.

Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology, 3,* 55–75.

Wallace, M. T., Carriere, B. N., Perrault, T. J., Vaughan, W., & Stein, B. E. (2007). The development of cortical multisensory integration. *Journal of Neuroscience, 26,* 1184–1189.

Watson-Gegeo, K., & Gegeo, D. (1986). Calling out and repeating routines in Kwara'ae children's language socialization. In B. B. Schieffelin & E. Ochs (Eds.), *Language socialization across cultures* (pp. 17–50). New York, NY: Cambridge University Press.

Waxman, S. R., & Lidz, J. (2006). Early word learning. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 299–335). New York, NY: Wiley.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34,* 1289–1309.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3,* 1–30.

Werker, J. F., & Patterson, M. (2001). Does perceptual learning really contribute to word learning? *Developmental Science, 4,* 26–28.

Werker, J. F., Pegg, J., & McLeod, P. J. (1994). A cross-linguistic investigation of infant preference for infant-directed communication. *Infant Behavior and Development, 17,* 323–333.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.

Werner, H., & Kaplan, B. (1967). *Symbol formation: An organismic-developmental approach to language and the expression of thought.* New York, NY: Wiley.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9,* 625–636.

Woodward, A., & Hoyne, K. L. (1999). Infants learn about words and sounds in relation to objects. *Child Development, 70,* 65–77.

Yoshida, H. (2004). Iconicity in language learning: The role of mimetics in word learning tasks. *Dissertation Abstracts International: Section B. Sciences and Engineering, 65*(2-B), 1055.

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16,* 90–95.

Yu, C., Ballard, D., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science, 29,* 961–1005.

Zukow-Goldring, P. (1990). Socio-perceptual bases for the emergence of language: An alternative to innatist approaches. *Developmental Psychobiology, 23,* 705–726.

Zukow-Goldring, P. (1997). A social ecological realist approach to the emergence of the lexicon: Educating attention to amodal invariants in gesture and speech. In C. Dent-Read & P. Zukow-Goldring (Eds.), *Evolving explanations of development: Ecological approaches to organism-environment systems* (pp. 199–252). Washington DC: American Psychological Association Press.

Zukow-Goldring, P., & Ferko, K. R. (1994). An ecological approach to the emergence of the lexicon: Socializing attention. In V. John-Steiner, C. Panofsky, & L. Smith (Eds.), *Sociocultural approaches to language and literacy: Interactionist perspective* (pp. 170–190). New York, NY: Cambridge University Press.

Received July 20, 2007
Revision received December 21, 2009
Accepted December 22, 2009 ■