The intersensory origins of word comprehension:
an ecological–dynamic systems view

Lakshmi J. Gogate, Arlene S. Walker-Andrews and Lorraine E. Bahrick

1. The State University of New York at Brooklyn, USA
2. Rutgers University, USA
3. Florida International University, USA

The learning of language by the child is not simply the associative naming or labeling of impressions from the world. It is also, and more importantly, an expression of distinctions, abstractions, and recognitions that the child is coming to achieve in perceiving.

(J.J. Gibson, 1966, p. 281)

Abstract

How do infants begin to understand spoken words? Recent research suggests that word comprehension develops from the early detection of intersensory relations between conventionally paired auditory speech patterns (words) and visible objects or actions. More importantly, in keeping with dynamic systems principles, the findings suggest that word comprehension develops from a dynamic and complementary relationship between the organism (the infant) and the environment (language addressed to the infant). In addition, parallel findings from speech and non-speech studies of intersensory perception provide evidence for domain general processes in the development of word comprehension. These research findings contrast with the view that a lexical acquisition device with specific lexical principles and innate constraints is required for early word comprehension. Furthermore, they suggest that learning of word–object relations is not merely an associative process. The data support an alternative view of the developmental process that emphasizes the dynamic and reciprocal interactions between general intersensory perception, selective attention and learning in infants, and the specific characteristics of maternal communication.

Introduction

A first step in word learning may involve the discovery that the continuous speech stream contains smaller units or words (Jusczyk, Cutler & Redanz, 1993; Jusczyk & Aslin, 1995; Saffran, Aslin & Newport, 1996). Furthermore, the process of understanding words may entail the realization that unique spoken words (nouns or verbs) are conventionally paired with and refer to perceivable objects or actions in the environment (for a review see Sullivan & Horowitz, 1983). Infants and young children in a language community must recognize that (a) heard speech (words) and seen objects or actions are related, (b) specific words function as symbols for objects or actions, and (c) such words denote specific objects or actions. We propose that the development of the first ability, in particular, has its origins in the interaction between infants' intersensory perception and the specific qualities of adult multimodal communication to infants during the first year (Sullivan & Horowitz, 1983; Zukow-Goldring, 1997; Gogate, Bahrick & Watson, 2000). Focusing primarily on the infant prior to symbolic–lexical development, therefore, we suggest that the development of lexical comprehension is the result of a complex ongoing reciprocal interaction between the organism (the infant) and the environment (the language addressed to the infant). This thesis is consistent with the ecological (J.J. Gibson, 1986) and...
dynamic systems approaches to development (Thelen & Smith, 1994; Lickliter, in press).

How could the infant's intersensory perception, a general capacity, facilitate the development of lexical comprehension during infancy? The current proposal contrasts sharply with the more popular view of innate language-specific processes (e.g. Chomsky, 1980; Fodor, 1983) and challenges an early statement by Chomsky (1980): 'multi-purpose learning strategies are no more likely to exist than general principles of "growth of organs" that account for the shape, structure, and growth of the kidney' (pp. 138–139). We argue that infants detect invariant, amodal or redundant information in bimodal non-speech and speech events (e.g. the temporally synchronous sight and sound of a bouncing ball, or the mouth movements and vocalizations of a person) very early in life (E.J. Gibson, 1969). According to E.J. Gibson, perception is initially global and becomes increasingly differentiated over time. Infants' general sensitivity to redundant information in bimodal events enables them to detect such relations in arbitrarily linked patterns of audible and visible stimulation, including words and objects or actions: The central role of perception in early lexical comprehension, in particular the perception of temporal contiguity between words and referents, is acknowledged in past research (e.g. Sullivan & Horowitz, 1983; Harris, 1992; Budwig, 1995; Bloom, 1998). However, the studies reported here are the first to show that, around 7 to 8 months, preverbal infants tune in to more precise redundant information (temporal synchrony), to learn the relation between a word and its referent.

We also propose that maternal communication to infants is abundantly endowed with the very information that infants initially detect. Mothers tend to embed words in multimodal events tailored to the perceptual skills of their infants. For example, mothers temporally synchronize their naming and showing of objects to infants when infants most require temporal synchrony to detect arbitrary speech–object relations (Gogate & Bahrick, 1998; Gogate et al., 2000; Gogate, under review; also see Zukow-Goldring, 1997). Perceptual abilities in infancy dovetail with environmental stimulation to promote the development of lexical comprehension. As will be reviewed, experimental research on the development of intersensory perception offers data to explain how infants learn to pair words and objects. Descriptive studies of maternal communication to infants point to ways in which the environment is organized to facilitate infants' word comprehension. The infant's perceptual and environmental (maternal language) systems change with time, and the co-action between these changing systems enables progress in infants' detection of intersensory information and, eventually, comprehension of words in the language. Because the two systems are fluid, the changes in one system are probably driven by a combination of constantly changing dynamics (instabilities) within and external to the system. Furthermore, the systems are self-organizing (i.e. instabilities give rise to temporary points of stability during development). Therefore, the developmental changes of these systems are constrained locally (not by a pre-program) by the properties of each (e.g. developmental factors in the child) and the real-time interactions between them (see Thelen & Smith, 1994, p. 36). The salience of any given parameter over others is a function of its relative weight (e.g. the salience of temporal synchrony may be a function of infants' attention). The environment in a complementary fashion affords different parameters interacting with the infant's developmental abilities. Because of the close fit between the two systems, the changing nature of one fits the changing nature of the other.

In general, the focus of this paper is on the process rather than the end-state of development. Many earlier cognitive accounts of early word learning have suggested that infants simply associate or 'map' words with perceivable objects and events (Ellis & Wells, 1980; Bates & MacWhinney, 1987; Stember, 1989). These accounts do not explain exactly how mapping takes place. Other accounts of word learning emphasize the 'association between words and the functions of words in the situation in which the child hears them' (Bloom, 1998, p. 322; Budwig, 1995). These accounts do not consider organismic development (see Bloom, 1998). In contrast, adhering to the ecological and dynamic systems views, we describe and emphasize the interaction between the organism and its environment to show how infants learn to relate words and referents. Furthermore, we attribute much more to the process by which infants detect word–object relations than mere association. In our view, infants seek out and attend to the affordances of the environment (the very environmental properties that facilitate the discovery of word–object relations; see J.J. Gibson, 1966). They do not simply 'associate' idiosyncratic patterns of stimulation available to different modalities whenever temporal contiguity or spatial co-location exists. Rather, they detect intermodal invariants, redundant information in patterns of stimulation, especially as their attention becomes educated by both experience and maternal scaffolding. Finally, unlike some modular approaches to language (e.g. Chomsky, 1980; Fodor, 1983), we suggest that language-specific systems are not required to account for the detection of arbitrary relations between words and referents (see Dent, 1990; Bates & Elman.
E.J. Gibson, in particular, perceptual development. The developmental process is manifested behaviorally in the early detection of redundant information and the later detection of information that is specific to a given modality. Perceptual information such as temporal synchrony, intensity shifts, rhythm and tempo are redundant or 'amodal' because the same information can be perceived across two or more modalities. Recently, the dynamic systems view of development has underscored the importance of intersensory perception for the development of general cognition including language (Thelen & Smith, 1994). In this section, and in the next, we will illustrate how the infant's initial undifferentiated perception, particularly in the auditory and visual domains, enables the perception of arbitrary relations between heard words and seen referents.

A wealth of findings on early intersensory organization across avian and mammalian species provide insights into its phylogenetic and ontogenetic development in humans. Newborns across many avian species detect redundant information across modalities, suggesting a possible innate basis for this ability. Changes in the prenatal environment applied to one sensory modality can influence the postnatal development of a different modality, indicating that the senses are integrated and undifferentiated early in development (Lickliter, 1990a, 1990b; Lickliter & Stoumbos, 1991). These effects of early experience can be seen across tactile, vestibular, auditory and visual modalities in bobwhite quail chicks (Lickliter & Banker, 1994) and in ducklings (Gottlieb, Tomlinson & Radell, 1989; Gottlieb, 1993). Furthermore, the fetal and newborn brains of some animals (e.g. owls) exhibit more plasticity in response to intersensory stimulation than do the brains of adult con-specifics, again presenting evidence for early intersensory integration (Knudsen, 1983; Stein, Meredith & Wallace, 1994).

Similar to their avian counterparts, human neonates show evidence for an integrated multisensory system. For example, human neonates can detect the common shape of a pacifier, and recognize an object's substance across visual and haptic modalities (Meltzoff & Borton, 1979; E.J. Gibson & Walker, 1984; Kaye & Bower, 1994). They can also imitate another person's facial motions (Field et al., 1983; Meltzoff & Moore, 1989). With respect to the visual and auditory modes, neonates detect changes in synchrony relations between the sound and sight of objects in motion as early as 4 weeks (Bahrick, 1996). These findings support the view that for humans the senses are integrated early during development and probably lay the foundation for continued intersensory perception across auditory and visual domains, including bimodal speech.

A majority of findings on infants' sensitivity to redundant information in bimodal speech has come from older infants and often cross-sectional approaches. Nonetheless, some developmental patterns can be identified. For example, between 2½ and 4 months, infants attended more to the synchronous than the asynchronous visible lip movements and audible speech patterns of a person reciting a nursery rhyme (Dodd, 1979). Other studies have established that infants detect redundant information in acoustic or phonemic segments, suggesting that infants attend to more specific aspects of bimodal speech. Four-month-old infants matched the visual mouth shape of a person with a corresponding vowel sound across several vowel pairs, in the presence of temporal synchrony between the lip movements and the vowel sounds (Kuhl & Meltzoff, 1982, 1984, 1988). These data suggest that young infants might detect correspondences between the visual mouth movements and vocalizations of adults at a relatively global level, progressing later on to abstract finer correspondences such as specific speech sounds and mouth shapes. In addition, infants detect intensity shifts in bimodal speech across auditory and visual modalities. At 5 months, infants detected the common intensity shift between a gradually opening mouth and an increase in the vowel sound amplitude of spoken disyllables consisting of consonants and vowels. They also detected the common intensity shift between a gradually closing mouth and a decrease in the vowel sound amplitude of these disyllables (MacKain, Studdert-Kennedy, Spieker & Stern, 1983).

Infants perceive redundant temporal synchrony and intensity shifts across auditory–visual non-speech events as well. Around 3 to 4 months, infants selectively attended to a specific object hitting a surface, based on...
the temporal synchrony between the object's motion and the timing of its sound of impact (Bahrick, 1983, 1987, 1992). Infants also matched approaching and receding objects with an increase and a decrease, respectively, in their sound amplitude (Walker-Andrews & Lennon, 1985; Schiff, Benasich & Bornstein, 1989; Pickens, 1994). Furthermore, at this age infants matched auditory–visual events based on the common (redundant) rhythm and tempo (Spelke, 1979; Bahrick & Lickliter, 2000). These parallel findings across speech and non-speech domains suggest that infants may use general perceptual processes for detecting redundant information in speech and non-speech events.

Intersensory redundancy and the detection of arbitrary speech–object relations

When observers look at and listen to events, they perceive redundant information across the senses as well as the information specific to each (see Figure 1). Observers may find some of these redundancies to be more informative than others. For example, the sight and sound of a person humming a tune (closed mouth) provides spatial co-location and temporal contiguity between the face and the voice, but not temporal synchrony, or a common rhythm, tempo and intensity shifts. In contrast, the sight and sound of a person singing or speaking with matching mouth movements provides more redundant information, co-location, temporal synchrony, a common rhythm, tempo and intensity shifts. In addition, the types of relations may vary. Given that articulation of a speech sound requires an appropriate mouth shape, the perception of these auditory–visual relations entails abstracting the available identical information across two or more senses. However, perception of the relations between a person’s face and his/her specific voice qualities (e.g. fundamental frequency or accent) requires finding similarities between two distinct but co-occurring events, because, taken separately, one event does not convey the same information as the other. Walker-Andrews (1994) asserts that arbitrary relations such as these can be categorized into two types (Figure 1).

<table>
<thead>
<tr>
<th>Intersensory Relations in (Auditory-Visual) Bimodal Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic Associations</strong></td>
</tr>
<tr>
<td>Temporal Contiguity</td>
</tr>
<tr>
<td>e.g., a person humming a tune with mouth closed</td>
</tr>
<tr>
<td><strong>Irredundant or Amodal Relations</strong></td>
</tr>
<tr>
<td>Temporal Synchrony, Shared Intensity Shifts, Rhythm and Tempo</td>
</tr>
<tr>
<td>e.g., sight and sound of a person talking or singing</td>
</tr>
<tr>
<td><strong>Arbitrary Natural Relations</strong></td>
</tr>
<tr>
<td>Predictable Relations</td>
</tr>
<tr>
<td>e.g., face–voice pairing, onomatopoeia</td>
</tr>
<tr>
<td><strong>Arbitrary Artificial Relations</strong></td>
</tr>
<tr>
<td>Unpredictable Relations</td>
</tr>
<tr>
<td>e.g., word–object or word–action pairing</td>
</tr>
</tbody>
</table>

Figure 1 A taxonomy for intersensory relations in human bimodal communication.

© Blackwell Publishers Ltd. 2001
Some arbitrary relations, such as that between a person's face and the sound qualities of his/her voice, predictably occur together in nature and may be referred to as 'arbitrary natural relations'. The detection of this arbitrary natural relation may be facilitated because a person's face and voice are spatially co-located. 'Arbitrary artificial relations' such as between a spoken word and an object, the color (or shape) of an object and the pitch of its sound on impact, or the color of a container and the taste or smell of the food it contains are idiosyncratic (Reardon & Bushnell, 1988).

In the case of word-object relations in particular, detecting the arbitrary relation can be complicated further because the spoken word does not emanate from the object but emanates from a person using the word to refer to the object. There is no spatial co-location. Notwithstanding this complication, the ability to detect these arbitrary relations is necessary for the development of lexical comprehension in the view proposed here.

How do infants begin to detect the arbitrary relations between words and their referents, given their ability to abstract redundant information in auditory-visual events? Some studies of intersensory perception of non-speech events in infancy demonstrate that the detection of redundant relations precedes the detection of arbitrary relations within the same context (Bahrick, 1994; Hernandez-Reif & Bahrick, 1995). In these experiments, in addition to redundant information, specific properties of an object for which information is available only to one modality were paired with those available to another, to determine whether infants could perceive redundant as well as arbitrary (natural or artificial) relations. In one set of experiments, infants were habituated to two events in which visible properties such as color and shape were paired with an audible property such as pitch. At test, the arbitrary relation between the two events was mismatched. The 7-month-old infants, but not 3- and 5-month-olds, detected the mismatch in the arbitrary relations between the color and shape of an object and the pitch of its impact sound (Bahrick, 1994). The 7-month-olds detected the changes in arbitrary relations when they were also provided redundant information such as temporal synchrony between the impact sounds and visible motions of the objects. The 3- and 5-month-olds, however, only detected changes in the redundant relations. In a more recent set of experiments, 6-month-olds but not 4-month-olds detected the arbitrary artificial visual-tactual relations set up between the color-pattern and the shape of an object. The 4-month-olds, however, detected the redundant shape information across visual and tactual modalities (Hernandez-Reif & Bahrick, 1995). In these studies, infants detected redundant relations across modalities developmentally prior to detecting arbitrary relations, and when they detected the arbitrary relations it was only in the presence of intersensory redundancy.

Given that young infants detect redundant information in speech and non-speech events, one might expect findings in the domain of speech to parallel those from non-speech research (Bahrick, 1994; Hernandez-Reif & Bahrick, 1995). Specifically, one might expect infants to detect arbitrary relations between speech patterns and objects around 6 to 7 months, but not before. Kuhl, Williams and Meltzoff (1991) reported such findings. Four- to 5-month-old infants did not match two pure tones with the visible mouth shapes of a person articulating the vowels /a/ and /i/. Adults, however, had no difficulty, even though they did not identify the tones as speech sounds. Therefore, the arbitrary relation imposed by the experimenter was the only information uniting the tones and the visible mouth shapes. Given that 3- and 5-month-old infants did not detect the arbitrary relations between the color of an object and the pitch of its impact sound (Bahrick, 1994), it is not surprising that 4- to 5-month-olds (Kuhl et al., 1991) did not detect the arbitrary relations between a mouth shape and a pure tone. Infants at this age did, however, match the mouth shape with a corresponding vowel sound in prior studies (Kuhl & Meltzoff, 1982, 1984, 1988). Findings such as these, from speech and non-speech research, suggest that the sequence in the development of intersensory perception is movement from the detection of redundant relations to the detection of arbitrary relations in both domains (see Bahrick & Pickens, 1994, for a review).

The ability to detect redundant information across bimodal events not only precedes the detection of arbitrary relations but also guides the detection of arbitrary relations within a given context (Sullivan & Horowitz, 1983; Bahrick, 1994). With respect to arbitrary natural relations, in a recent set of experiments, infants detected the relations between the faces of two persons and their matching voices given temporal synchrony between the visible mouth movements and audible vocalizations (Bahrick, Hernandez-Reif & Cigales, 1988). Following habituation training to video-films of alternating same-sex faces of two persons reciting a nursery rhyme, each paired with a distinctive synchronous voice, 4- and 6-month-olds detected the arbitrary relations between the faces and voices. When the face-voice pairings were switched (but synchrony between the faces and voices was maintained), infants looked longer to these novel pairings. These results indicate that the presence of redundant information such as temporal synchrony might facilitate the detection of arbitrary natural relations.
With respect to arbitrary artificial relations, prior results have shown that 6- to 7-month-old infants also detect these relations in the presence of temporal synchrony that unifies the auditory–visual non-speech events (Spelke, 1979; Lawson, 1980; Bahrick, 1994). However, only recently was the salience of redundant temporal synchrony for the detection of arbitrary artificial relations in speech established by testing infants under synchronous and asynchronous conditions. In one set of experiments, intersensory redundancy (temporal synchrony) resulted in 7-month-old infants learning the arbitrary relations set up between the speech sounds /a/ and /i/ and moving objects (Gogate & Bahrick, 1998). Infants learned the arbitrary relation only when the timing of the vocalizations coincided with that of object motions. Infants failed to learn these relations when temporal contiguity between a vowel sound and a static object or temporal contingency was presented (i.e. when the sounds occurred during the pauses between object motions; Gogate & Bahrick, 1998). Using identical methods, a second set of experiments showed that 8-month-olds (but not 7-month-olds) detected the arbitrary relations between complex monosyllables /tah/ and /gah/ and moving objects, again only in the presence of temporal synchrony (Gogate, under review). Likewise, 8-month-olds did not detect the arbitrary relations between syllables such as /nim/ and /lil/ and two objects in the absence of temporal synchrony (Werker, Cohen, Lloyd, Casasola & Stager, 1998). These studies underscore the importance of intersensory redundancy for infants' detection of arbitrary word–object relations. They also suggest that intrinsic developmental differences in infants' perception between 7 and 8 months interact with properties of the communicative environment such as syllabic complexity and intersensory redundancy to influence infants' detection of arbitrary relations. One further study, discussed earlier, has shown that the presence of redundant information facilitates the detection of arbitrary relations in other modalities as well. In a study of non-speech events, 6-month-old infants learned the arbitrary relations between the visible shape and color-pattern of objects only if they were given the opportunity to feel (the shape of) the object (Hernandez-Reif & Bahrick, 1995).

The requirement for precise temporal synchrony in concurrent patterns of stimulation across sensory modalities appears to develop gradually during the first 6 months, probably as infants' sensitivity improves (Kellman & Arterberry, 1998). Thus, newborns can detect arbitrary relations in speech events when temporal contiguity (but not temporal synchrony) and spatial co-location are provided, and when the stimuli are quite distinct (Slater, Brown & Badenoch, 1997). In this study, after newborns were familiarized with arbitrary pairings of a red line and the word 'mum' in a male voice, and a green line and the word 'teat' in a female voice, they looked longer at test to alterations in the word–voice–line combinations. In this instance, temporal contiguity was provided by presenting the words only when the infants looked at the visible lines, and spatial co-location was provided when the words were played via a speaker located directly above the visual screen. For newborns, spatial co-location between paired stimuli is necessary for learning sound–object relations (Morrow, Fenwick & Chance, 1998). Furthermore, for newborns, temporal contiguity alone may suffice for learning the arbitrary auditory–visual relations (Figure 1). In addition, the task was simplified by using a female voice for one word and a male voice for the other. The newborns could have detected the arbitrary relations at a global level, between the voice quality (male or female) and the red or green line, rather than the specific word and the object. In the studies with 7- and 8-month-old infants (Gogate & Bahrick, 1998; Gogate, under review) intersensory redundancy (temporal synchrony) was provided. Further, all syllables, /a/, /i/, /tah/ and /gah/, were presented by a single female speaker. Thus, detection of the arbitrary relation hinged solely upon learning the pairing between the specific syllable and the object.

Conclusions

In general, these findings support the conclusion that, by the second half of the first year, preverbal infants use intersensory redundancy (temporal synchrony given via visual and auditory experience, or shape given via visual and haptic experience) to detect arbitrary relations. These studies show how intermodal perception might contribute to infants' detection of word–referent pairings. As infants attend to temporal synchrony in speech events, the arbitrary relations are highlighted, allowing infants to discover the links between specific words and referents. In this view, temporal synchrony serves as a bootstrap (and is a control parameter) for infants to tune in to arbitrary word–object relations. The redundant information serves to unify two otherwise arbitrarily co-occurring stimulus patterns. Because temporal synchrony specifies that the vocalization and the object belong together, it probably assists the infant in narrowing down the possibilities to find the most likely referent to pair with the word. An earlier report suggested that symbolic–lexical development may entail knowing that a word 'stands for' a tangible referent rather than knowing that a word 'goes with' a referent.
which is indicative of pre-symbolic development (Golinkoff, Mervis & Hirsh-Pasek, 1994). We suggest that the infant's knowledge of how a word and object 'go together' may well be an important developmental precursor to knowing that the word 'stands for' the object. Although we have emphasized the role of temporal synchrony here, other types of redundant information such as intensity shifts, tempo and rhythm between spoken syllables and moving objects were also provided in Gogate's studies (Gogate & Bahrick, 1998; Gogate, under review). These types of information also may facilitate the detection of arbitrary relations. Controlled experiments are needed to examine whether they, too, can facilitate the learning of word-object relations.

The findings suggest that within the auditory–visual domain infants might detect arbitrary natural relations, such as those between faces and voices of the same sex, developmentally prior to detecting arbitrary artificial relations such as those between specific speech patterns and objects. Detecting the arbitrary relations between co-occurring events that are predictable in nature may be easier than detecting the arbitrary relations between less predictably co-occurring events. However, given that this conjecture derives from the results of independent studies (Bahrick et al., 1988; Gogate & Bahrick, 1998; Gogate, under review), future controlled experiments are required. Nonetheless, if predictability facilitates detection of word–referent relations, it casts an interesting light on the perception and production of onomatopoeia in infants' early words. For example, infants use terms such as bow wow to refer to 'things of action' such as a 'dog' prior to using words such as doggie (see Werner & Kaplan, 1963; Nelson, 1973; Fenson et al., 1994). In this case, infants may detect the relation between the object and the term easily, because the term resembles the sound that the object makes. Infants may have many opportunities to detect the intersensory relations between the barking animal, its visible opening and closing mouth, and the term bow wow. Only later might infants detect the arbitrary artificial relation between the term doggie and the same visible animal (Figure 1). Thus, infants' word perception and production may be closely related. In addition, infants' ability to detect more predictable arbitrary relations and their use of onomatopoeia may be reciprocated by adults' frequent use of these forms in their communication to infants. (See Fernald and Morikawa (1993) for mothers' greater use of onomatopoeia to 6-month-olds.) These findings would show, in keeping with the theme of this paper, that maternal language and infants' intersensory perceptual abilities are reciprocally related.

Developmental changes in the prerequisites for detecting spoken label–object relations

Changes in the requisite conditions under which infants learn speech pattern–object relations in experimental settings reflect the dynamic nature of infants' developing perceptual–lexical system. At first, although newborns detect changes in redundant information (Bahrick, 1996), they do not appear to require temporal synchrony to detect arbitrary auditory–visual relations (Slater et al., 1997). During this early period, perhaps owing to low visual acuity, temporal contiguity may be sufficient for detecting arbitrary relations between auditory–visual events at a fairly global level (Figure 1). Later, around 7 to 8 months, infants require temporal synchrony to detect speech pattern–object relations (Gogate & Bahrick, 1998; Gogate, under review). Temporal contiguity alone in the absence of temporal synchrony, such as a word uttered in the presence of a static object, is no longer sufficient. Similarly, temporal contingency, when vocalizations occur during the pauses between the motions of an object, does not result in learning. Still later, this requirement seems to diminish with the lexical development that it enables. With advances in perception and memory, coupled with the onset of the implicit knowledge that words stand for objects, and with the growing ability to use words consistently to refer to objects, infants may no longer require temporal synchrony (or other redundant information) to detect arbitrary relations between words and objects. Thus, although 7-month-olds (Gogate & Bahrick, 1998) and 8-month-olds (Werker et al., 1998; Gogate, under review) require temporal synchrony, 14-month-olds detected a switch in syllable–object pairs even though vocalizations were presented out-of-synchrony with the moving objects (Werker et al., 1998). But, these experiments have also shown that infants require object motion to detect speech–object relations up to the age of 14 months. Neither 7-month-olds (Gogate & Bahrick, 1998) nor 14-month-olds (Werker et al., 1998) detected a switch in the speech–object pairs when static objects were presented concurrently with speech patterns. Why did these infants require object motion? Earlier research suggests that seeing objects in motion enables infants to abstract invariant properties of objects across different transformations (see review

---

1 For example, infants may detect the arbitrary natural relation between the term tick–tack and the object 'clock' by first detecting the redundant relation between the object (the moving hands of the clock) and its temporally synchronous ticking sound (see Nelson, 1973).
in E.J. Gibson, 1991; E.J. Gibson, Owsley, Walker & Megaw-Nyce, 1979). Thus, movement itself highlights the invariant properties of the object and may permit easier detection of arbitrary relations between the objects and vocalizations. However, older infants who have presumably developed a repertoire of receptive and expressive vocabulary can detect word–object relations in the absence of object motion, temporal synchrony or perhaps even temporal contiguity (Baldwin, 1993a).

The strong role of multisensory information for detecting novel word–object relations during the first year and its relative loss of importance during the second year is evident from a comparison of Gogate’s findings (Gogate & Bahrick, 1998; Gogate, under review) with those of Baldwin and her coworkers (Baldwin et al., 1996). In Gogate’s studies, 7- and 8-month-old infants detected the syllable–object relations if an object in a speaker’s hand moved synchronously with the speaker’s vocalizations, even though her face was not visible (on videotape). In contrast, Baldwin et al. (1996) found that older infants required a view of the speaker’s face (during live stimulus presentations). For infants of 18 to 20 months, multisensory relations alone between spoken words and objects, in the absence of the speaker’s face and referential intent, did not ensure learning of word–object relations. These older infants required a social criterion (a visible speaker naming objects while displaying visual interest in them) for establishing stable links between words and objects (Baldwin et al., 1996). Baldwin et al.’s finding was substantiated in a recent study where even 13-month-olds detected word–object relations during live stimulus presentations when an object was named while the infant and the speaker attended to it (Woodward & Hoyne, 1999). Thus, during later infancy, joint attention becomes a more critical factor for detecting word–object relations than multisensory information. The increasing importance of joint attention is especially relevant given that even 6-month-olds are capable of following the direction of mother’s eye gaze, which is positively correlated with their lexical comprehension at 12 months (Morales, Mundy & Rojas, 1998). Consistent with the dynamic systems view, the pattern of findings suggests that the non-linguistic, contextual requirements for learning word–object relations change with development. In general, a dynamic systems view is unique in that it emphasizes the constant fluctuations in and across parameters rather than a ‘genetic blueprint’ or a reliance on the most basic associations. Changes over time may look, on the surface, like regressions or U-shaped functions because of the co-action of systems and relative weight of parameters.

Thus far, we have seen that multiple non-linguistic factors interact in a dynamic fashion to facilitate word comprehension during infancy. How do linguistic factors such as syllabic complexity interact with non-linguistic factors? Seven-month-olds (Gogate & Bahrick, 1998) detected the relations between the vowels /a/ and /i/ objects moving in temporal synchrony but did not detect the relations between more complex syllables, /lah/ and /gah/, and the same objects. Although infants at this age could distinguish the two complex syllables and the two objects from one another, they did not link them with the objects. However, 8-month-olds successfully detected these relations under identical conditions (Gogate, under review). In contrast, two prior studies (Stager & Werker, 1997; Werker & Stager, in press) reported that 8-month-olds were unable to ‘map sound onto meaning’ when two spoken syllables /blh/ and /dlh/ or /Ilif/ and /nim/ were each paired with a moving object. This was the case even though infants perceived the voicing distinction between /blh/ and /dlh/ and the oral–nasal distinction between /Ilif/ and /nim/. However, in these experiments, temporal synchrony was not provided. Such findings show how linguistic and non-linguistic factors can interact and affect infants’ detection of word–referent relations.

The perceptual–lexical system: general or language-specific processes?

The findings reviewed above show that young infants are adept perceivers of redundant information across bimodal (auditory–visual) speech and non-speech events, and suggest that the development of general intersensory perceptual abilities may foster lexical comprehension. Specifically, because some of the same types of redundant information guide infants’ detection of arbitrary relations in speech and non-speech events, the findings suggest that general or multipurpose perceptual processes are at work. Similar results with older infants (Namy & Waxman, 1998; Woodward & Hoyne, 1999) and young children (Markson & Bloom, 1997) also point to general processes for word comprehension. These data offer an alternative to the view that a unique acquisition device with innate constraints is necessary for the development of early lexical comprehension. Further, these findings suggest that development of intermodal speech does not proceed along paths unique and different from those found for other domains of bimodal perception (cf. Chomsky, 1980; Fodor, 1983).

The current data also call into question the notion that autonomous modules are present at birth for processing language-like and non-language-like patterns of bimodal stimulation. Preverbal infants as young as 7 months detected both arbitrary non-speech (Bahrick, 1994) and speech–object (Gogate & Bahrick, 1998; Gogate, under review) relations following approximately 2–3 min of...
exposure. We hypothesize, therefore, that infants learn such relations using powerful general perceptual–cognitive processes that may be sensitive to the probability of co-occurrence between bimodal patterns of stimulation. Such a thesis, though speculative, is in keeping with other recent findings. For example, 8-month-olds can detect word boundaries from an artificial language speech stream based solely on the statistical probability of their co-occurrence with adjacent speech sounds (Saffran, Aslin & Newport, 1996; see Bates & Elman, 1996; Seidenberg, 1997). Identical findings have been reported in 8-month-olds' detection of boundaries when tones replaced the syllables in a sequence (Johnson, Saffran, Aslin & Newport, 1998). Thus, infants rely on some process that enables grouping of auditory stimuli based on their transitional probabilities for both syllables and tones. Further studies must establish whether infants access powerful general learning processes and use probabilistic optimization as a means for early word learning and comprehension. Consistent with the dynamic systems view of development, such findings suggest that computation of probabilities is one among many learning strategies available to the infant for word learning (see Kelso, 1997).

The nature of infants' learning

What strategies do young infants use to relate words and their referents? Several recent studies suggest that the detection of word–object relations is the result of associative learning of words and concepts or referents (see discussion in Jusczyk & Hohns, 1997; Stager & Werker, 1997; Werker et al., 1998). In our view, there is much more to the process by which infants detect word–object relations. As discussed earlier, infants seek out and attend to specific affordances of the environment very early in life depending on their perceptual sensitivities (J.J. Gibson, 1966). They are unlikely to simply associate idiosyncratic information available to different modalities whenever temporal contiguity or spatial co-location is provided (see Slater et al., 1997, for data on newborns). Thus, between 3 and 5 months, infants selectively attend to intersensory redundancies only when natural structural correspondences, such as an opening mouth with a gradual increase in vocal amplitude (MacKain et al., 1983) or a single object with a single impact sound (Bahrick, 1988), are also provided. Selective attention to (and detection of) intersensory redundancies facilitates the discovery of arbitrary word–object relations. Thus, in our experiments, 7- and 8-month-old infants did not form idiosyncratic associations between a novel word and a moving hand, or part of the word and part of the object, despite the temporal synchrony between these events. Rather, they selectively attended to the (hand-held) whole object and syllable by tuning in to the intersensory redundancy. They did not learn the arbitrary relations when temporal contiguity between a word and a static object was provided, but required a more precise temporal synchrony between a word and an object's motion. They also did not learn these relations when the object motion consistently preceded the vocalization (delayed conditioning; cf. Markham, Schwartz & Gordon, in press). Therefore, these findings are inconsistent with the association view of early word learning (Bates & MacWhinney, 1987).

Furthermore, the methods we have used to study infants' learning of speech–object relations reflect infants' active participation. Infants set the pace for learning the relations during the habituation sequence (Horowitz, 1974). Although infants learn the arbitrary relations in one context (habituation to two auditory–visual displays, one at a time), they actively search for and select one of the two object displays in a second context (two choice intermodal preference reflecting memory) (Gogate & Bahrick, in press). Schafer and Plunkett (1998) raise a similar point with regard to the difference between their familiarization and testing procedures used to study word–object mapping. These behaviors reflect more than a passive forming of associations that can be induced for two randomly co-occurring (temporally contiguous or contingent) stimuli during conditioning under identical or similar contexts (cf. Rovee-Collier, 1986; see E.J. Gibson, 1969, in press).

Infants might also compute the probability of co-occurrence between words and referents to learn these relations. Some studies of word recognition suggest that the higher probability of occurrence of specific words in the environment may favor infants' recognition of those words. Thus, 4.5-month-olds are familiar with their own name (Mandell, Jusczyk & Pisoni, 1995) and 7.5-month-olds recognize words from a passage or a list after exposure to it (Jusczyk & Aslin, 1995). Similarly, if detection of arbitrary relations between words and objects is influenced by the probability of their co-occurrence, then infants should detect word–object pairings better as the probability of co-occurrence increases. Such findings would explain parsimoniously how infants might learn the most obvious meaning of a word prior to learning secondary meanings of the word. Furthermore, infants should detect arbitrary word–referent relations at a young age if the words and referents occur relatively frequently in their

Seidenberg argues, against modular theories, that language acquisition is based on the frequency of linguistic and non-linguistic sources (see Seidenberg, 1997, footnote 29; and Bates & Elman, 1996).
environment. Tincoff and Jusczyk (1999) conducted experiments that speak to this prediction. Six-month-old infants detected the arbitrary relations between familiar words such as *mommy* and *daddy* and the faces of their own mother and father, but not those of another infant's mother and father. In another study, when a familiar word such as *ball* was presented, older infants (24-month-olds) looked to the named object in a pair even before the word was completely uttered (Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998). Fifteen-month-olds, however, looked at the named object only after the word was completely uttered. Consistent with the dynamic systems view, these findings underscore an interaction between environmental factors such as the probability of co-occurrence and organismic factors such as infants' speed of processing and perceptual acuity in detecting word–referent relations.

Memory for arbitrary speech–object relations in infancy

To understand spoken words on a regular basis, it is not sufficient that infants merely recognize speech patterns and detect the arbitrary relations between the patterns and objects. Infants must also remember the arbitrary relations in order to reference them upon later encounters. The development of lexical comprehension is therefore contingent upon developing memory and can be affected by memory limitations. What factors might facilitate infants' memory for arbitrary relations? Temporal synchrony not only facilitates infants' learning of arbitrary relations but may also enable memory for these relations. In a study discussed earlier, 6-month-olds, but not 4-month-olds, detected a switch in arbitrary face–voice pairings, and remembered them for 10 min following habituation (Bahrick et al., 1998). In another study, 7-month-old infants indicated memory for vowel–object pairings (Gogate & Bahrick, in press). Infants who were habituated to two alternately presented objects whose movements were temporally synchronized with the vocalizations /a/ or /i/ remembered the vowel–object pairs after 10 min or 4 days. Infants who were habituated and tested with temporally asynchronous or static displays of objects did not remember the pairings. The conditions that facilitate young infants' initial learning of arbitrary relations between spoken labels and objects appear to operate during delayed learning or memory for these relations as well. For older infants, other factors such as adult labeling of (static) objects might suffice. Thus, 10- and 14-month-old infants looked longer during a subsequent play period to objects that had been both pointed at and labeled but not to those to which adults had merely pointed (Baldwin & Markman, 1989).

Further, at 13 months, infants regulated their own attention to objects, and the act of adult labeling when infants' attention was focused on the objects facilitated infants' memory for the word–object relation (Woodward, Markman & Fitzsimmons, 1994).

Taken together, the findings suggest progression in infants' memory for word–object relations. Seven-month-old infants require redundant information such as temporal synchrony to remember syllable–object relations. Older infants, between 10 and 13 months, remember word–object relations under different conditions, such as when adults label a static object during play episodes, or when adults label an object to which the infant's attention is directed. Future studies might focus on this progression. For example, at what age might infants detect and remember the relation between objects and words that are presented together but asynchronously?

The dynamic nature of infants' perceptual–lexical system: a summary

An integrated multisensory system, present from birth, allows the detection of redundant information in intersensory patterns of stimulation. Thus, very young infants are sensitive to changes in redundant information provided in auditory–visual patterns of stimulation and the ability to detect such changes is well established early in life (Figure 2). For example, infants of 2.5 to 4 months are visually attracted to the temporal synchrony between audible speech sounds and visible lip movements. The ability to detect redundancies, in turn, guides infants to detect arbitrary relations between auditory and visual events by the second half of the first year. The detection of arbitrary relations in temporally synchronous events is possible because temporal synchrony highlights the relation between the bimodal patterns of stimulation. Infants progress thereafter, detecting and remembering first arbitrary relations between predictable events such as talking persons and their specific faces, and later the relations between unpredictable events such as words (spoken syllables) and temporally synchronous moving objects. As discussed earlier, the latter ability is an important precursor to lexical comprehension, because the lexicon consists of words that are arbitrarily assigned to specific objects by convention. Once this ability is well developed, infants no longer require temporal synchrony for the detection of and memory for arbitrary relations. The present findings underscore the dynamic nature of infants' developing perceptual–lexical system, and give rise to questions about what environmental factors might contribute to infants' detection of and memory for word–object relations.
Maternal communication and the development of infants' lexical comprehension

In the previous section, we discussed the dynamic characteristics of the infant's developing perceptual system that aid in the detection of and memory for the arbitrary relations between audible speech patterns and visible objects. Given our main hypothesis that lexical comprehension develops from the dynamic interaction between organismic and environmental systems, and given that infants are sensitive to intersensory redundancy in bimodal communication, in this section we discuss specifically how the environment contributes to infants' lexical comprehension. We ask to what extent the information that facilitates detection of and memory for word-referent relations is available to infants in their everyday lexical environment. Next, we ask whether the information provided in the lexical environment is compatible with infants' developing lexical comprehension.

Intersensory redundancy in maternal³ communication

Soon after birth, maternal interaction with infants is multimodal in nature, providing auditory, visual, tactile

³In most research the participant caregiver has been the infant's natural or adopted mother. Therefore, we use the terms 'mother', 'maternal' and 'motherese' throughout.

and olfactory information about the mother to the infant. Infants show a visual preference and positive affect for simultaneous auditory and visual presentations of both female and male adults' infant-directed speech during the first 6 months (Werker & McLeod, 1989). Of primary importance to the thesis of this paper is that mothers provide a variety of multimodal naming contexts during everyday interactions with their infants. Prior descriptive studies characterize the lexical environment of infants as consisting of numerous instances in which mothers label objects and provide visual stimulation by showing or moving the objects (Messer, 1978; Masur, 1997; see Sullivan & Horowitz, 1983). Mothers also use gestures to highlight meaning by pointing to or touching an object when referring to it (Zukow-Goldring, 1997). They acoustically highlight specific words in the speech stream through the use of exaggerated prosody (greater stress and pitch variation), use shorter sentences and place a word in sentence final position (see review by Fernald, 1992). We suggest that the multimodal aspects of maternal communication may aid in the development of infants' communication.

In particular, mothers provide redundant information such as temporal synchrony when they simultaneously name and show objects or actions to their infants, creating a multimodal event (Zukow-Goldring, 1997). We call this intermodal infant-directed communication
'multimodal motherese' (Gogate et al., 2000). Temporal synchrony, as noted earlier, facilitates infants' early learning of novel word–object relations (Gogate & Bahrick, 1998; Gogate, under review). Synchronous naming and showing of objects, in particular, occurs in maternal communication to infants as old as 11 months (Messer, 1978). Recently a descriptive study of maternal multimodal communication to 24 American (Caucasian and Hispanic) infants, ranging from the age of 5 months to 30 months, and 12 Asian Indian infants ranging from 5 to 17 months, was conducted. This study showed that mothers exploit temporal synchrony, but not temporal contiguity or contingency, in their communication to infants of all ages. The mothers’ utterances of novel words coincided with the objects they moved or actions they performed, and sometimes, even with touch. Second, this study showed that American mothers used multimodal communication styles that were well suited to their infant’s age, which is typically correlated with the level of lexical comprehension (Gogate et al., 2000). When naming objects and actions, the mothers of 5- to 8-month-olds used temporally synchronous object motions and words more often than those of older infants (9–17 months and 21–30 months). These results, together with the experimental data discussed earlier (Gogate & Bahrick, 1998, in press; Werker et al., 1998; Gogate, under review), suggest that mothers’ use of temporal synchrony actually decreases over time, corresponding with infants’ increasing ability to detect word–object relations in the absence of synchrony (Figure 2).

The findings, specifically, point to the significant role of integrated multimodal information for promoting the development of lexical comprehension. Bates (1993) raised a similar point. She suggested that infants require integration of information from multiple modalities to comprehend words because they are still in the process of ‘cracking the speech code’ (p. 237). Even older infants with a working vocabulary predominantly understand words that refer to the immediate context rather than those that refer to previously experienced events. This necessitates that adults use relatively context-bound speech (Nelson, 1973; Messer, 1978; Bates, 1993) in a manner that highlights the arbitrary word–referent relations for young infants. The temporal synchrony provided by mothers when they name objects may facilitate integration of multisensory information perhaps because synchrony highlights the intermodal relations for the infant. The temporal synchrony need only be fairly precise. Infants allow some ‘slippage between auditory and visual inputs’ (within 350 ms) before they will perceive them as asynchronous (Lewkowicz, 1992, p. 35). Thus, even when mothers loosely synchronize words with moving objects, it may unify the two events and highlight the arbitrary relations, facilitating infants’ attention, detection and eventual memory for word–referent relations. Additionally, the loose synchrony may be compensated for by other types of redundant information such as matching intensity shifts, rhythm and tempo between a word and a moving object.

Pointing to static objects in maternal communication

Other types of multimodal communication, including maternal pointing to a static object or touching a static object while naming it, may become useful in directing attention to the arbitrary relations between labels and objects later in infancy, when infants become sensitive to these gestures (Murphy & Messer, 1977). Gogate et al. (2000) found that pointing was a relatively minor component of mothers’ multimodal communication to young infants. The frequency of pointing was low, perhaps because the mothers named target objects that were in close proximity to themselves and their infant. However, even under these conditions where pointing was sporadic, mothers pointed to objects and labeled them slightly more frequently during interactions with 9- to 17-month-olds than with 5- to 8-month-olds or 21- to 30-month-olds. This finding complements that of studies showing that, between the ages of 9 and 17 months, infants are particularly receptive to the pointing gesture (Butterworth, 1991) and that attention to pointing is a robust predictor of lexical comprehension (Mundy, Kasari, Sigman & Ruskim, 1995). The findings, once again, speak to mothers’ sensitivity to infants’ capabilities (Figure 2).

Naming of static objects in the presence of joint attention between infant and mother

During interactions with older infants, adults may utter words in the presence of static objects without pointing to the objects or touching the infants (Baldwin et al., 1996). Mothers’ and infants’ joint attention to objects alone may suffice for the learning of word–object relations, once infants begin to detect these relations on

4In two cultures reported thus far, the Kaluli of New Guinea (Schieffelin, 1979) and the Kwara’ae of Malaita in the Solomon Islands (Watson-Gego & Gego, 1976), mothers may not provide the direct multimodal communication to young infants underscored in this paper. It is our hypothesis that, in these cultures, infants receive redundant multimodal information either indirectly from mothers or directly from other members of the community. These types of secondary input may be important for word learning. Recent studies suggest that infants are just as capable of learning ‘object labels when they eavesdrop on others’ interactions as when they are directly addressed’ (Akhtar, Jipson & Callanan, 1998).
their own and demonstrate a working vocabulary. In accordance with this thesis, Gogate et al. (2000) found that mothers named static objects, which they simply held for the infant, significantly more often during interactions with 21- to 30-month-olds relative to the two younger age groups (5-8 months and 9-17 months). In these instances, both mother and infant attended to the object or event being named. Given that the experimental data showed that 14-month-olds did not learn the arbitrary relations between syllables and static objects in the absence of joint attention (Werker et al., 1998), this variation by age is appropriate (Figure 2).

Infant-regulated multimodal communication

In the previous sections, we suggested that mothers regulate the naming context by structuring the environment according to the infants' level of perceptual-lexical competence. Consonant with infants' requirement for integrated auditory and visual information to detect speech pattern-object relations, the descriptive data showed that mothers provide integrated information to younger infants, perhaps highlighting the arbitrary relations between words and referents. Older infants may have other means for gleaning word-referent relations and may not need explicit maternal guidance for learning the relations. Once again, mothers of older infants seem to be attuned to their infants' requirements, and are more likely to simply name objects without also moving them.

During the infants' second year, however, this multimodal communication seems to undergo a qualitative shift, away from predominant adult regulation and towards increased infant regulation. For example, mothers of the oldest group of infants observed by Gogate et al. (2000) were likely to attend to and name an object already being explored and manipulated by infants (also see Masur, 1997). Mothers of the younger infants in this study (5-8 months) rarely showed this pattern of behavior (Figure 2). These results support the conclusion that, with experience, older infants regulate the lexical input they receive. This pattern is paralleled by data on infants' word learning. In the second year, infants learned words rapidly when adults named or verbally encouraged infants to act upon the objects on which infants' attention was already focused (Akhtar, Dunham & Dunham, 1991; Woodward et al., 1994). In comparison, their rate of vocabulary learning was slower when mothers redirected infants' attention to objects outside their focus of attention (Akhtar et al., 1991). These patterns speak to the dynamic nature of mother-infant interaction during the development of lexical comprehension. Further studies are required to examine the dynamic nature of maternal communication and to examine the extent of its adaptation to infants' perceptual and communicative requirement.

Conclusion

A diverse set of experimental findings suggests that early lexical comprehension owes much to infants' developing ability to perceive intersensory relations in auditory-visual events. Initially infants attend to and detect global information across the senses, progressing to the detection of more specific relations. In this respect, the changes in the perceptual-linguistic system embody Werner's orthogenetic principle. They show that development proceeds from a state of 'relative globality and lack of differentiation to a state of increasing differentiation, articulation, and hierarchic integration' (Werner, 1957, p. 126; also see Bower, 1974; Walton & Bower, 1993; Hernandez-Reif & Bahrick, 1995; Bahrick, 1996). With respect to bimodal speech, improvement in intersensory perception culminates in infants' detection of and memory for arbitrary relations between different speech patterns and their object or action referents. The development can be directly observed in infants' attention to and selection of salient properties of the environment when learning word-referent relations. Detection of arbitrary word-referent relations begins long prior to an infant's first word production, as illustrated by the precedence of language comprehension skills relative to word production (Benedict, 1979; see Huttenlocher, 1974; Reznick, 1990; Golinkoff & Hirsh-Pasek, 1995). Although infants do not produce words until the second year, in part because of motor constraints, some perceptual abilities necessary for word comprehension are well in place by the end of the first year.

The development of infants' perceptual-lexical system is only part of the story, however. We are not the first to emphasize the maternal communication environment in language acquisition (Garnica, 1977; Cooper & Paccia-Cooper, 1980; Sullivan & Horowitz, 1983; Gleitman, Newport & Gleitman, 1984; Hirsh-Pasek et al., 1987; Kelly, 1992; Lewkowicz, 1996; Bloom, 1998). In addition, however, we wish to call attention to its multimodal nature and the dynamic properties of this particular system (see Tucker & Hirsh-Pasek, 1993). Mothers provide a variety of developmentally compatible multisensory contexts (see Figure 2). When they name objects or actions for infants, they often provide temporal synchrony, they point to objects and actions, or they provide no
motion at all. Moreover, they tailor these contextual conditions to the developmental status of their infant. Mothers use temporal synchrony primarily in communication with younger infants, which may scaffold infants' developing ability to detect word–referent relations. They phase out some aspects or provide a different set of conditions for older infants, such as naming an object in motion without temporal synchrony, or naming an object that is held immobile. In addition, mothers may provide socially contingent contexts that facilitate infants' detection of word–object relations. Such a context is provided when infants encounter a speaker who names and visually attends to the object being named, indicating referential intent. As they develop, infants begin to increasingly regulate these contexts. As discussed in the previous section, infants' attention to objects or events in themselves can initiate maternal-naming. These findings illustrate that the maternal multimodal communication system ('multimodal motherese') is dynamic and well suited to the infant's developmental status.

We do not suppose that the participants or 'owners' of each system are necessarily cognizant of the dynamic interaction between the two systems. In fact, we would emphasize that mothers and infants are not conscious of the entrainment of the two systems. For example, the infant probably does not view the temporal synchrony provided by the mother as a resource for detecting word–object relations, nor is the infant aware of his/her own ability to detect these relations. In the process of discovering the meanings of words, the infant, motivated to learn, may stumble upon the intersensory redundancy between words and referents. Similarly, the mother is unlikely to provide temporal synchrony consciously nor is she aware necessarily that the infant is relying on such information for detecting a relationship between a spoken word and a moving object. Furthermore, infants' developing lexical competence is not the result of a master plan held by either party, but the 'workings of a myriad of interactions at a local level' (van Geert, 1993, p. 266). The observed interactions between the environmental system (maternal multimodal communication) and the infant system (perceptual–lexical) is the product of acting together of numerous diverse factors, each with multiple degrees of freedom. We have identified some of these factors for developing lexical comprehension – temporal synchrony, motion, and pointing by the mother, and the developing sensitivity to intersensory relations and syllabic complexity on the part of the infant – but there are bound to be many others, each contributing jointly to lexical development. During the ebb and flow of interactions, neither party need be cognizant of these factors, although one provides the environment and the opportunity for the adaptations of the other.

Furthermore, we suggest that the two systems are self-organizing or order seeking (Thelen & Smith, 1994; Reed, 1996; Kelso, 1997). During the course of development, one system provides the affordances encountered by the other. Affordances are properties of the environment that are salient to the perceiver (J.J. Gibson, 1966, 1979; see Reed, 1996). As a result of self-organization, the infant's perceptual–lexical system may regulate itself in response to changes in the affordances of the maternal communication system. Analogously, the maternal communication system may regulate itself in response to systemic pressures exerted by the affordances of the infant's perceptual–lexical system. An outcome of self-organization in the two dynamic systems is that the maternal (adult) communication system strives to adapt itself to the infant as much as the infant's system to maternal communication (see Studdert-Kennedy, 1991).

In conclusion, early lexical development can be seen as a process of ongoing reciprocal interactions between the organism and the environment. During the process of adapting to one another, the infant's intersensory perceptual and maternal multimodal communication systems momentarily exhibit stable patterns revealing, at the outset, a consonance between the two systems (preferred states), where predominant factors or control parameters influencing each system are observable. The ongoing changes that give rise to these momentary points of stability may well be driven by a combination of (as yet) unobservable instabilities in one or the other of the two systems (see Kelso, 1997). The instabilities of one system and the self-organization of the second result in changes in the latter that seem consonant with those of the former. In this manner, the changes of one system dovetail with the changes of the other. Environment–organism interactions, therefore, are an important unit for study because everyday social interactions between the mother (or other communicating individuals) and the infant provide rich opportunities for lexical development and ensure the successful transmission of the lexicon from one generation to the next.

Acknowledgements

Writing of this paper was supported by a General University Research Grant to the first author from the University of Delaware. We thank Robert Lickliter and
References


Gogate, L.J. (under review). Intersensory redundancy facilitates learning of arbitrary minimal pair-object relations by 8- but not 7-month-old infants: evidence for a dynamic system.


Received: 28 January 1999
Accepted: 9 December 1999
There is little doubt that language learning in infancy is a thoroughly multimodal affair. Babies rarely, if ever, only hear speech. As they interact with caregivers and social partners, they also see and sometimes even feel it. Based on the Gibsonian ecological view of perception, Gogate, Walker-Andrews and Bahrick assert that the intersensory redundancy inherent in multimodal speech is directly available to the perceptual system and that it facilitates infant learning of word–object associations. They argue that infants selectively attend to the spatial and temporal contiguity, temporal synchrony and amodal relations inherent in multimodal speech. There is much to recommend in this view because of its well-placed emphasis on the multimodal character of speech, its use of dynamical systems principles to conceptualize language acquisition as a product of the bidirectional interplay between organismic and environmental factors through developmental time, and its rejection of nativistic/modularistic explanations of language development.

Despite its general appeal, however, when one delves into the specifics of the proposed view, at least four areas appear to be problematic. First, although the principal advantage of the ecological view of perception is that it constrains the number of places that we need to look for the ‘objects’ of perception, it also shackles us needlessly. Second, Gogate et al.’s representation of the current state of knowledge regarding infant intersensory abilities is sufficiently selective that it glosses over some of the inconsistencies in a rather complex body of findings. Third, the perception of non-redundant multimodal events is ignored despite the fact that they are part-and-parcel of intersensory perception. Finally, recent neurophysiological findings have cast doubt on the theoretical assumptions underlying the concept of direct perception of intersensory relations. Thus, I believe that our current state of knowledge about intersensory perception and its development makes the application of Gogate et al.’s theoretical perspective to a new domain premature – a case of putting the cart before the horse.

According to the Gibsonian ecological view of perception, the world is a unified place because attributes that specify it usually go together in space and time and thus provide us with a great deal of redundant and amodal structure. For example, when we watch and hear someone speak we simultaneously perceive ambient auditory and visual arrays that, according to J.J. Gibson, specify the underlying physical reality and thus are sufficient to provide us with a direct and unified percept of the talking person; no binding of features nor internal, inferential processes are necessary. The ecological view makes the key assumption that intermodal redundancies and amodal invariants are directly available in the ambient array and all the observer has to do is ‘pick them up’ in order to perceive meaningful affordances for action. E.J. Gibson (1969) posited that infants come into the world ready to pick up intermodal structure, and Gogate et al. argue that since then empirical studies have provided ample evidence of infants’ ability to do just that. Based on this evidence, they propose that infants use this ability to form their first word–object associations and that they thus begin their long journey into lexical acquisition.

The problem is that the assumption that intersensory relations are ready-made and directly available makes the question of how they are made moot. The fact is, however, that intersensory perception is not a unitary, homogenous process and that it is far from clear whether or not the perception of intersensory relations in infancy can be simply reduced to a question of whether they can be picked up or not.

Furthermore, the Gibsonian idea of intermodal perception appears simple and elegant, making it seem as if the world is ready-made for the infant to step in and partake of the panoply of intersensory structure that is
underlying mechanisms of integration per se it does not emergence (Lewkowicz, 2000). Finally, not all the data on emergence nor the underlying processes mediating their emergence support for the perception of intersensory relations. In other words, once differentiation of the array structure occurs, the perception of intermodal invariance occurs automatically. The problem is that the concept of differentiation is too general and for those of us who are interested in the underlying mechanisms of integration per se it does not offer sufficient explanatory power. In addition, E.J. Gibson proposes that differentiation of an increasingly greater number of amodal invariants occurs over development, but she does not specify the order of their emergence nor the underlying processes mediating their emergence (Lewkowicz, 2000). Finally, not all the data on infant intersensory perception provide clear, unambiguous support for the perception of intersensory relations. For example, infants do not perceive certain temporal intersensory relations regardless of whether they are specified by simple, inanimate displays (Lewkowicz, 1985, 1992) or by faces and voices (Lewkowicz, 1999b).

Gogate et al.'s use of the concept of redundancy implies that more is better, but that is not always the case. When heteromodal inputs specify the same property, they do make perception more economical, but when they specify unique, modality-specific properties, often these can conflict with one another. In the latter case, the perceptual system must decide between them and this can produce sensory dominance effects and/or such well-known illusions as the McGurk or the ventriloquism effects. These effects cannot be ignored for they are as important to intersensory perception as are the cases of amodal perception. Indeed, the fact that intersensory illusions occur raises important questions precisely because they lead to radical reorganization of the global percept. For example, the perception of an audible syllable can change dramatically in the presence of a conflicting visible syllable (McGurk effect), or the perception of the spatial localization of an auditory source can change in the presence of a spatially disparate visual stimulus (ventriloquism effect). Might these types of effects play a role in the detection of word–object relations? This is quite likely given evidence that infants experience the McGurk effect and given recent findings from our laboratory that infants' perception of ambiguous visual motion becomes unambiguous in the presence of sound.

Indeed, the ecological view of perception does not deal well with the case of intersensory conflict (Stoffregen & Bardy, in press) and Gogate et al. recognize this fact by noting that there are times when a spoken word may not emanate from the object directly but from a person who is spatially discrepant. Curiously, they dismiss this problem and simply state that the ability to detect these arbitrary relations is necessary for the development of lexical comprehension. I would argue, however, that this is a key problem for their view because the auditory referent is almost never spatially or temporally concordant when an adult labels an object. Fortunately, infants are able to ignore relatively large temporal asynchronies (Lewkowicz, 2000) but we still do not know how large an intersensory spatial discrepancy infants can tolerate nor how the interaction of specific spatio/temporal discrepancies affects perception and how that might change in development.

Finally, contrary to Gibsonian theory, recent neuro-physiological findings show that, at least in the case of spatial multimodal cues, the nervous system does not automatically pick up the relations but instead actively synthesizes them. Stein and his colleagues (Stein, 1998) have shown that the deep layers of the mammalian superior colliculus contain mostly multimodal cells. These cells exhibit a marked enhancement in activity when two different sensory stimuli are spatially concordant but a profound response depression when they are spatially discordant. Moreover, these cells are not present at birth but appear later in development, and when they do they still do not exhibit mature functional properties. The most troubling finding for the ecological direct perception view is that when the association cortex is made inactive the collicular cells cease to integrate the heteromodal sensory cues converging on them which, in turn, results in a failure to produce appropriate behavioral responses. In other words, perception is in the head, not out there!

In conclusion, although the ecological view of perception correctly focuses researchers' attention on the multisensory richness of language, the issues and problems raised in this article make its application to the study of word–object relations no less problematic than simple associationism.

References


Lewkowicz, D.J. (1992). Infants' response to temporally based intersensory equivalence: the effect of synchronous sounds

© Blackwell Publishers Ltd. 2001
The dynamics of language development: from perception to comprehension

Robert Lickliter

Virginia Tech, USA

One of the most important insights achieved by psychological sciences is that behavior is not represented in a genetic or neurological template prior to its emergence. This hard-won insight of probabilistic epigenesis replaced the longstanding notion of predeterminism, in which behavioral development was seen to progress in an orderly and preordained sequence under the control of genetic substrates (see Gottlieb, 1992, for a historical overview). From this view, genes were seen as deterministic, guiding the nervous system to mature in a predetermined fashion, and giving rise to ‘innate’ or ‘instinctive’ behavior. Thanks in large part to nearly a half century of comparative and developmental research (see Michel & Moore, 1995), most psychologists now appreciate that behavior does not unfold from some predetermined template or blueprint. Behavior is increasingly viewed and studied as a ‘system’ property, an emergent phenomenon resulting from transactions between the organism and its specific environments over the course of individual ontogeny (Oyama, 1985; Thelen & Smith, 1994). Conventional assumptions of innate or hard-wired behavior have gradually given way to the realization that genes cannot, in and of themselves, directly produce behavioral traits or characteristics (Gottlieb, Wahlsten & Lickliter, 1998). As a result, psychological sciences now generally recognize the value of grounding the study of behavioral development in a system of multiple influences, rather than continuing to rely on simple dichotomies such as genes and environment, instinct and learning, or maturation and experience. The study of language development, however, has proven particularly slow to yield to this multidetermined, systems perspective. Nativist accounts of the emergence and developmental course of language comprehension continue to populate the developmental literature.

In recent years, however, a growing number of investigators have shifted their research focus to the discovery and mapping of the relations between levels of organization in the organism–environment system that can contribute to the course of speech perception and production. The target article by Gogate, Walker-Andrews and Bahrick is an important contribution to this shift in research focus, providing a useful review of recent findings concerned with the dynamic relationship between internal and external factors involved in scaffolding the infant’s path to word comprehension and offering a working model of how these factors co-act and change over the course of early development. Conceptually, the authors’ view of the determinants of language development is based on an appreciation that there is no top-down or bottom-up explanation of language development that will be adequate to the task. A strict emphasis on any single domain or level in the developmental system, be it genes, physiology, neuroanatomy or neurochemistry, social interactions, or culture is seen to be too limited to successfully address the emergence and development of speech perception.

Address for correspondence: Department of Psychology, 5081 K Derring Hall, Virginia Tech, Blacksburg, VA 24061, USA; e-mail: duckling@vt.edu

© Blackwell Publishers Ltd. 2001
and production skills. Gogate, Walker-Andrews and Bahrick provide a compelling argument for broadening the scope of research regarding language development to include an empirical concern with the relationship between the emerging perceptual and cognitive capabilities of the infant and the specific social and communicative behaviors of the caretaker.

Their review of a number of recent studies of infant intersensory perception and maternal communication styles serves to support their general conclusion that it is the reciprocity between the perceptual-cognitive responsiveness of the infant and the structure of its physical and social environment that is likely to provide the initial basis for bootstrapping the remarkable process of word comprehension. According to the model proposed by Gogate, Walker-Andrews and Bahrick, the infant's sensitivity to specific types of information and the caretaker's emphasis on and coordination of this information are viewed as equal partners in the path to language acquisition. This dynamic, multidetermined view of the development of word comprehension rejects the more simplistic (and popular) modularity view of language development for a more general-process view, in which the infant's selective attention and increasing differentiation, the salience of intersensory redundancy, and maternal multimodal communication all co-act to support the emergence of initial word comprehension capacities.

This emphasis on the nature of the relationship between the structure of the organism and the structure of its environment is an ecologically grounded alternative to more traditional language-specific views, in which causal factors are typically attributed to abstract internal mechanisms such as 'language acquisition devices' and the like. From this older view, development is seen to proceed based on internal programs and constraints and the child's activities and experiences are typically viewed as 'supportive' rather than fundamental to the process. This article embraces the insight that development is fundamentally experience or activity dependent, a perspective that has only recently become more commonplace in psychological sciences. Many psychologists and linguists continue to assume that language development is determined by more 'primary' or 'hard-wired' processes that occur at the genetic or neurological level. However, this type of linear, unidirectional, bottom-up view overlooks a central tenet of the target article, namely that no single element or factor in the developmental system necessarily has ontological primacy or causal privilege. The functional significance of genes or neural structures on language development are to be understood only in relation to the larger developmental system of which they are a part. The authors emphasize this point throughout the target article, providing a useful working model of how the structures and activities of the organism and the specific (and changing) set of physical, biological and social factors with which it interacts over ontogeny serve as dynamic partners in both guiding and constraining the achievement of early word comprehension skills.

In this light, the article is effective in pointing out that one of the most challenging tasks facing a developmental systems approach to language is to interpret the concept of the environment in such a way that it incorporates an appropriately dynamic view of the changing relations between the developing infant and its context over time. Empirical findings from developmental science have repeatedly highlighted that, as an organism develops, its relation to the external world changes, such that its effective environment, the actual physical, biological and social factors with which it interacts, also changes (Bronfenbrenner & Morris, 1998). One of the major strengths of this paper is its attempt to characterize and map the course and direction of the nature of this change and its consequences for the process of language acquisition. For example, their review of how caretakers adapt and adjust their interaction styles to the changing perceptual and cognitive abilities of the child highlights the dividends of expanding our research focus to include how processes change over time as a result of individuals' histories and activities. The authors' call for an increased empirical appreciation of the 'differences' that make a difference to young infants, such as unimodal versus multimodal stimulation or static versus dynamic stimulation, is an important message and can provide a useful framework for pursuing a more developmental approach to the process of word comprehension.

References


Intersensory origins of word comprehension: but what's the mechanism?

Graham Schafer

University of Reading

Gogate, Walker-Andrews and Bahrick have given us an excellent survey of the dynamic development of intersensory perception in the first 2 years. Gogate et al. argue that early word learning is underpinned by the development of intersensory perception; they are surely right. Nonetheless, their theoretical account of these data misses an important opportunity. How do infants, or infant-caregiver dyads, progress along the trajectory so carefully mapped by the authors?

Infants' changing sensitivity to cross-modal events is well documented (see for example the edited volume by Lewkowicz & Lickliter, 1994). In particular, it appears that the infant's ability to link events in two modalities becomes increasingly robust over development. That is, the infant is more and more able to cope with noisy correlation between stimuli in (say) the visual and auditory modalities. The authors argue that this ability is intimately connected with the learning of words. Three features of their Gibsonian approach are as follows:

1. Innate affordances specify cross-modal links that the infant is likely to encounter in early language experience.
2. ‘Process’ is emphasized over mechanism.
3. Comprehension in the ‘presymbolic’ period is the result of a complex interaction between two loosely coupled systems: the infant, and the multimodal environment provided by the caregiver. It is argued that the co-action of these systems enables a succession of stages in which infants are differentially sensitive to temporal relations between events in different modalities. Caregivers tailor their behavior to the infant's current ability to detect multimodal events.

Let us now briefly consider these issues from the standpoint of simple associative learning. Complex emergent behaviors can be explained by domain general, low-level, associative mechanisms (e.g. Plunkett, Sinha, Möller & Strandsby, 1992; Bates & Elman, 1996; Elman et al., 1996; Schafer & Mareschal, in press). Thinking about mechanism can help us understand the developmental process.

Are innate representations of naturally occurring contingencies really necessary?

The issue of whether innate representations of naturally occurring contingencies are really necessary has been dealt with extensively elsewhere (e.g. Elman et al., 1996). In the context of word learning, labels for objects will in general co-occur with their referents. Such contingency is noisy: sometimes temporal synchrony or contingency, contiguity etc. will be manifest between label and object, and sometimes not. Furthermore, the referent will often be present without the label, and vice versa. Nonetheless, on average, the most frequent co-occurrence between any given label and any given referent will be that in which the label matches the referent. This environmental regularity is no doubt manipulated by the caregiver, as elegantly demonstrated in the article; it is less certain that this manipulation is an innately specified sine qua non of infant lexical-semantic development. We should acknowledge that all the rich affordances of objects, language and other cognitive domains are present in the input, as a priori statistical regularities. In the framework of nativists such as the Gibsons, and Spelke, these regularities are represented explicitly in the infant
mind (e.g. E.J. Gibson, 1969; J.J. Gibson, 1979; Spelke, Breinlinger, Macomber & Jacobson, 1992; Spelke, 1994). But why reify these facts about the world when they are there to be discovered by association?

A mechanism? Infants' use of correlated and contextual information

A central contribution of Gogate et al.'s paper is to chart the development of the infant's ability to detect correspondences between information in two (or more) modalities. The authors suggest no causal mechanism for the pattern of evidence in the article. In fact, they are dismissive of 'mechanism', on the grounds both that it appeals to teleology and that 'process' is more important. This emphasis on process (description?) over mechanism fails to give insight into how different states might be achieved. The simplest candidate mechanism is learning by association. Gogate et al. argue that infants 'do not simply "associate" idiosyncratic patterns of stimulation available to different modalities' (p. 2). Of course they don't. Rather, infants reject truly idiosyncratic stimulus pairings precisely because these do not co-occur at all that often, whereas conventional (i.e. 'correct') mappings are much more likely. And infants are notoriously slow learners, a feature that makes them well suited to the rejection of spurious mappings. 'Stages' emerge during this process as infants learn the reliability of successively more arbitrary cues (e.g. the parental pointing gesture).

The early word learning task can be characterized as one in which it is ultimately important to decouple representations of label and referent. Eventually, representations of word and referent are capable of existences separate from one another. Infants' associative learning is well suited to the task of finding out about simple contingencies (e.g. synchrony) at the outset, and ultimately about decoupled representations of label and referent, in just the fashion described in the article. Consider the role of contextual information. Contextual information may be thought of as any information present during the learning episode which is not strictly part of the mapping to be learned. Such information is a cue to the veracity of the intermodal mapping the infant is currently experiencing. This cue might be a pointing gesture by the parent, or a particular intonational contour (e.g. the near-universal falling contour used in ostension: Papousek & Papousek, 1992). At a detailed level, it might be the color of a tablecloth upon which new toys are presented in a word learning experiment (Samuelson & Smith, 1998). Most generally, it is that part of an infant's brain state not directly encoding the attributes of an array of stimuli. Such 'irrelevant' information enables infants to form meaningful representations from the noisy but lawful environment.

The potentiation of learning by co-occurring stimuli is a general learning principle, and furthermore it is one with a developmental profile. The developmental trajectory of human infants' responsiveness to intersensory temporal attributes of stimuli is well documented (e.g. Lewkowicz, 1994; Bahrick & Pickens, 1994; Gogate, Walker-Andrews & Bahrick, target article). In fact, the developmental change in such responsiveness is not specifically human. This is important because it suggests that such mechanisms are not peculiar to such human domains as the word learning task. For example, the infant-rat differs profoundly from the adult rat in its treatment of compound stimuli. Learning in infant rats is potentiated by 'irrelevant' or extra contextual information, whereas in adults such irrelevant information causes reduction in the learning of mappings between new stimuli. And interestingly, like humans, the infant rat shows an increasing ability to map temporal disparity between stimuli presented in auditory and visual modalities (Spear & McKinzie, 1994).

Gogate et al. class informative contextual arrays of stimulation as intermodal invariants. But need we assume that infants are innately endowed with knowledge of these affordances? Might they not simply be honoring the statistical cues in their environment? Evidence for statistical learning is strong (e.g. Saffran, Aslin & Newport, 1996; Gomez & Gerken, 1999; Marcus, Vijayan, Rao & Vishton, 1999). And attention to contextual information can further provide a handle for infants to learn arbitrary mappings. An elegant example of such learning is supplied in the report of Samuelson and Smith (1998), albeit with somewhat older (2-year-old) children than those focused on in the target paper. In Samuelson and Smith's experiment, it was shown that learning a new word for a new object could arise simply from contextual conditions, and not necessarily from the pragmatics of the situation as had previously been argued by Akhtar, Carpenter and Tomasello (1996). At a lower level, it can be shown that attention to consistently labeled objects, explained on pragmatic grounds by Baldwin and Markman (1989), can be resolved on a purely statistical basis: stimuli that co-occur with repeated irrelevant information are more readily learned than those which do not. This can be demonstrated in simple associative networks (see http://www.reading.ac.uk/~sxs97gws/baldwin.html). Thus, the implication of findings from associative learning in young rats, to experimentation such as that of Samuel-
Developmental change in the detection of increasingly decorrelated contingency: systems or cues?

The evidence for a developmental trajectory in infants’ ability to pair events from more than one modality in increasingly decorrelated and abstract situations is compelling. But whether much is gained at this point by the description of the caregiver and infant as two loosely coupled and interrelated systems is moot. The evidence for developmental shifts in caregivers’ utterances is well presented by Gogate et al., but their conclusion that such shifts are necessary – and indeed causally related to infants’ subsequent ability to learn words – is premature. Little in the way of direct evidence for this claim is offered. Contextual information of the kind discussed in the previous section is likely to be provided as a matter of course by caregivers. The suggestion of the authors that this correlative information is specifically tailored to the infant’s ability to detect it is attractive. However, evidence that such tailoring is an essential aspect of lexical–semantic development is not yet available. And, as already indicated, the notion of a process which is not a mechanism is problematic. The authors make no real effort to explain why the infant–caregiver dyad might switch from one mode to another. The associationist approach offers a potential way forward here also.

Simple recurrent networks (SRNs: Jordan, 1986; Elman, 1990) form representations of information correlated in time and space, or a combination of the two. Such networks demonstrate interesting developmental behavior. One compelling example comes from Elman, who has shown that such SRNs can learn long distance dependencies but only if they are constrained to learn first about more local contingencies (Elman, 1993). Though these simulations learn about embedded relative clauses, the logic of the task is similar to infants’ learning in the second half of the first year, in which they are increasingly able to detect lawful correspondences at a (temporal) distance. Gogate et al. suggest that infants’ early use of synchrony relations might be used as a ‘bootstrap’ or ‘control parameter’ for subsequent learning of arbitrary relations. The mechanism for this useful behavior may in fact turn out to be nothing more complex than associative learning.

References


The mystery of how children come to understand that arbitrary sound sequences stand for concepts is still not fully understood. Few theorists today would characterize word learning as either a strictly associative or a parameter setting process. Current approaches emphasize built-in biases, social-contextual factors, general-purpose learning mechanisms, conceptual development, or the interaction of multiple factors. Gogate and colleagues propose a developmental account of word learning within a Gibsonian and dynamic systems framework.

Over the past decade, we and a number of other infant speech perception researchers have directed our attention to word learning (Aslin, Woodward, LaMendola & Bever, 1996; Jusczyk, 1997, 1999; Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998) and have asked how developing perceptual capabilities might contribute to word learning. In recent chapters (Stager & Werker, 1998; Werker & Tees, 1999), we suggest that the perceptual processes of syllable discrimination, preference for native phonotactic and rhythmical properties, segmentation of words, and recognition of acoustic forms (some of which may be words) reveal how infants learn "about" words, but do not inform us about actual word learning (see also Jusczyk, 1997). The ability to associate a label with an object is a step towards word learning but is not equivalent to referential understanding. Moreover, within the realm of associative word–object learning, there may be several levels of complexity from recognizing a mismatch in a habituation procedure to looking correctly to one of two visual displays in response to a heard word to the even more complex task of reaching for a named object. We, like others before us, restrict full 'word learning' to (1) the ability to use a word to stand for an object even when that object is absent, (2) an arbitrary relation between word and referent, and, when appropriate, (3) the child's understanding that some words extend to other members of the same category. Within this framework, our work on associative word–object learning (e.g. Werker, Cohen, Lloyd, Casasola & Stager, 1998) requires minimal referential understanding. The associative task does, however, require a definitive element of word learning: the infant must form an arbitrary link between word and object (see Stager & Werker, 1998; Werker & Tees, 1999; Werker & Stager, in press).

We believe that the requirements for word learning outlined above emerge sequentially. However, recent work showing a 'discontinuity' between the perceptual information available in syllable discrimination tasks compared to word–object association tasks (Stager & Werker, 1997, 1998) has made us reconsider the link between earlier developing perceptual abilities and the emergence of true word learning. In particular, we are no longer confident that perceptual learning during infancy is linked in a causal way to the conceptual advances required for word learning. With this skepticism, which developed from 10 years of unexpected findings (see Stager & Werker, 1997, 1998), we turn to a discussion of the model proposed by Gogate, Walker-Andrews and Bahrick.

Gogate et al. argue that mapping sound sequences to seen objects or actions originates in the interaction...
between intersensory perception and the specific qualities of adult multimodal communication to infants. They review research from the past 20 years showing that young infants detect intermodal information in both speech and non-speech events, but with certain kinds of intermodal information being easier for the child to detect than others. This includes what Gibsonians refer to as ‘amodal’ relations in which the properties of the stimulus in one domain (e.g. sound) specify the same generating source in another domain (e.g. sight). The intermodal matching of vowel information in face and voice may be an instance of such amodal integration (see Patterson & Werker, 1999a). Gogate and colleagues discuss how redundancies in the stimulus array such as temporal synchrony, temporal contiguity and co-location enable infants to show intermodal matching at a very early age. The detection of amodal relations precedes that of ‘typical but natural’ arbitrary relations which in turn precedes detection of ‘arbitrary artificial’ relations (Walker-Andrews, 1994). Our work is consistent with this sequential model. Infants aged 4.5 months detect amodal vowel information across face and voice with both male and female models, but it is not until 8 months of age that infants match the arbitrary but natural cues specifying gender in face and voice using the same isolated vowels (Patterson & Werker, 1999b, under review). Mapping words to objects is an instance of the more advanced ‘arbitrary artificial’ kind of intermodal matching. Furthermore, ‘the spoken word does not emanate from the object’ (Gogate et al., p. 5); thus spatial contiguity cannot act to facilitate detection of the link. We would argue that the intermodal learning of word–object relations at 7–8 months, when acoustic stimuli are presented in temporal synchrony with a change in object motion (Gogate & Bahrick, 1998, under review), is not comparable to the word learning seen at 14 months when temporal synchrony is absent (Werker et al., 1998).

We are convinced by Gogate and colleagues’ argument that there is a developmental sequence underlying the kinds of intermodal relations infants can detect and that the earlier emerging abilities ‘guide’ achievement of each subsequent milestone in perceptual learning. Our skepticism stems from the idea that earlier emerging developmental achievements necessarily guide – or contribute in any causal way – to the qualitatively distinct achievement required to link the ‘arbitrary and artificial’ relation of a word and object when support from spatial contiguity and temporal synchrony are removed. Instead, we would suggest that an advance in conceptual thinking that allows the infant to form an association in the absence of perceptual cues is required. Indeed, after this advance in conceptual thinking, word–object associative learning may be possible with or without previous intermodal experience.

The same issues are relevant to Gogate, Walker-Andrews and Bahrick’s review of parent–infant communicative interactions. Gogate et al. convincingly argue that caregivers’ interactions with their infants highlight amodal cues such as temporal and spatial contiguity which facilitate infants’ learning of intermodal relations. Undoubtedly, these caregiver behaviours not only improve the quality of parent–child interaction but also help the child understand and predict events in their world. However, detecting intermodal perceptual relations does not necessarily lead to the insight that words can stand in an entirely arbitrary relationship with objects. Perhaps the understanding of intentionality is the critical step here. With this insight, the child can use cues such as line of regard and pointing (Baldwin & Tomasello, 1998) as qualitatively distinct substitutes for the earlier available intermodal cues of temporal and spatial contiguity.

In summary, Gogate, Walker-Andrews and Bahrick present a model of word learning based on dynamic systems and Gibsonian theory. Although we agree that there is a step-wise progression in intermodal perceptual development, we remain skeptical that word learning can be understood simply as another step in this process.

References


Gogate, L.J. (under review). Intersensory redundancy facilitates learning of arbitrary relations by 8- but not 7-month-old infants: evidence for a dynamic system.


**Perceiving referring actions**

**Patricia Zukow-Goldring¹ and Nancy de Villiers Rader²**

1. University of California, Los Angeles, USA
2. Ithaca College, USA

A mother invites her son, Zeke, to feel the surface of a broom while saying the word 'sticky'; The saying of 'st' accompanies Ann's fingers as they rather smoothly traverse some of the bristles .... Her voice quality changes to a more raspy, vibrating 'i::' in synchrony with her fingers repetitively touching and interrupting the flow of her hand across the broom bristles .... With a slight pause on the 'ck' as her hand ends its horizontal movement ..., Ann lifts away as she says, 'y' .... Then, ... Zeke touches the bristles, saying, '[k]'.

(ZR, 12 months, 9.29.88; Zukow-Goldring, 1997, pp. 232–233)

Gogate, Walker-Andrews and Bahrick refer to a chapter entitled 'A social ecological realist approach to the emergence of the lexicon: educating attention to amodal invariants in gesture and speech' in which Zukow-Goldring (1997) asked,

Given infants are immersed in a continuous perceptual flow, how do these cultural novices begin to detect and participate in assembling the structure and organization of everyday events? In particular, how do infants come to perceive the relation between the auditory structure in speech and a specific subsegment of the ... perceptual structure available at any one moment ....? (p. 201)

Drawing on a Gibsonian theory of perception, she posed the following answer: 'Amodal invariants detectable in gesture and speech may make the routine relation between word and world perceivable by providing an equivalence across dissimilar kinds' (p. 227). Using cross-cultural samples of infants, Zukow-Goldring presented microanalyses of mother–infant interactions that support this possibility.

While Gogate *et al.* provide welcome, confirming experimental evidence relative to Zukow-Goldring's

Address for correspondence: Patricia Zukow-Goldring, Center for the Study of Women, UCLA, Kinsey 288, 405 Hilgard Avenue, Los Angeles, CA 90095-1504. USA; e-mail: zukow@ucla.edu

© Blackwell Publishers Ltd. 2001
theoretical stance regarding the perceptual origins of the lexicon, they ignore its substantial history and rich observational support (Zukow, 1989, 1990; Zukow-Goldring & Ferko, 1994; Zukow-Goldring, 1996, 1997, 2001). In this commentary, we document the sources of four important hypotheses developed by Zukow-Goldring over the past decade and restated by Gogate et al. These hypotheses were derived from longitudinal investigations conducted by Zukow-Goldring in situ in Central Mexico among rural and urban Spanish-speaking families as well as in the western USA among working-class, Spanish-speaking, Latino families and among those from the English-speaking middle-class.

Hypothesis 1: A general perceptual process (namely, detecting amodal invariants) allows infants to perceive the correspondence between word and referent

The vignette opening this commentary illustrates the social ecological realist approach to the emergence of the lexicon used by Zukow-Goldring. 'The delicate timing of this caregiver coordinating the audio-visibility of saying-and-doing “stickiness” gives 12-month-old Zeke an opportunity to detect amodal invariants specifying a particular routine relation between word and world' (Zukow-Goldring, 1997, p. 233). Loudly, if not so clearly, Zeke repeats her speech as he leans forward to touch the broom bristles.

Hypothesis 2: Infants can and do detect amodal regularities prior to the comprehension of speech

In presenting this hypothesis, Zukow-Goldring (1997) summarized the laboratory research of Bahrick, Walker-Andrews, Lewkowicz, Spelke and others showing that infants 'from three months on ... can detect amodal invariant relations uniting the auditory and visual flow that infants perceive in dynamic events' (pp. 217–218). Thus, she argued, this general perceptual ability would be available at least by the second half of the first year, a time of initial word learning. Clearly, without this ability, infants would be unable to take advantage of such information should it occur in caregiver messages.

Hypothesis 3: Caregivers present amodal regularities as they gesture and speak to infants from 6 months of age

Gogate et al. do not acknowledge the fact that Zukow-Goldring has already presented the idea of ‘multimodal motherese’, if not the name, in her studies documenting caregivers’ communicative practices (Zukow, 1989, 1990; Zukow-Goldring & Ferko, 1994; Zukow-Goldring, 1996, 1997, 2001). Zukow (1990, p. 718) noted that caregiver gestures have characteristics that parallel those found in Fernald’s studies of infant-directed speech (motherese), suggesting that they might serve as ‘a means for detecting amodal specification ... available to infants. An equivalence in gestural and prosodic contour suggests a means by which infants might detect an invariant relation between object and word.’ She stated specifically that ‘caregivers make amodal invariant relations (timing, tempo, intensity, and rhythmicity) prominent as they gather and direct attention to animate beings, objects, and actions. Infants pick up these regularities in their speech and action’ (Zukow-Goldring, 1997, p. 240).

For instance, when a caregiver shakes an open bottle of sweet-smelling vitamins ... [as] she says and does ‘shakey = sha::key = SHA:::KEY?’, she coordinates in action, vision, odor, and sound the ... perceptual pick-up of amodal invariant relations through tempo, rhythmicity, synchronous initiation-termination, accelerating intensity and so on .... Thus, we have a tangible way of seeing how infants detect invariance ... and ... equivalence across dissimilars in action and speech .... (Zukow-Goldring, 1997, p. 221)

Hypothesis 4: Temporal synchrony of action and word, not temporal contiguity, is of crucial importance to lexical development

'Simple spatio-temporal contiguity will not serve as a satisfactory explanation of how children detect the relations between words and world, although this assumption seems to be a hidden default in perspectives appealing to innate biases' (Zukow, 1990, p. 712). 'Temporal contiguity' refers to events occurring closely in time — perhaps when someone says a word with one tempo or rhythm while moving an object in another. Another vignette vividly demonstrates what can go

1 In the Oxford English Dictionary contiguity is defined as follows: 'Psychology. Proximity of impressions or ideas in place or time'.
‘wrong’. While saying ‘frog’, the mother’s ‘speech does not match the action in tempo, rhythm, or initiation’ (Zukow-Goldring, 1997, p. 237) as she quickly re-arranges and shows a finger puppet. Then the mother says ‘ribbit’ synchronously with looming the toy towards the child. To the mother’s consternation, the child subsequently calls it a ‘ribbit’ (KJ, 17 months, 2.1.89). Zukow-Goldring (1997, p. 239) comments, ‘This segment illustrates the speculation that natural laws manifested through perceptual structure (amodal invariant relations of tempo, rhythmicity, intensity, and timing) detected across modalities may underlie the emergence of the early lexicon.’

Gogate et al. do make a significant contribution in bringing together a number of recent infant language studies whose findings support the claim that temporal synchrony is an important component of multimodal motherese. While one could question the degree to which habituation studies tell us anything more than that the infant has learned to link the word as a sound and the image/object presented, these findings are exciting. Particularly strong are the findings that 7- and 8-month-old infants required a precise temporal synchrony, and not just contiguity, when learning the relation between a word and an object’s motion, tuning in to the joint intersensory regulation. Equally important is the fact that the role of multisensory information is key during the first year, waning in importance in the second. Alternative methodologies should be developed to ascertain whether or not the habituation findings relate to perceiving referring actions. Ultimately, we all want to know how infants come to know what words mean (Zukow-Goldring, 1996, 2001).

As we have summarized here, careful observation of caregiver practices in ecologically valid environments is, in actual fact, anything but a ‘trivial pursuit’: it is often the unacknowledged but indispensable step in determining hypotheses worthy of testing. Other than Zukow-Goldring, few researchers have explored the close coupling of caregiver speech and actions prior to and after infants’ first words; these include Harris, Jones and Grant (1983) and Messer (1983). Yet the study of the mundane world of everyday interaction, finding out what caregivers actually do day-in and day-out, should lead the way in the design of experiments (Pick, 1989; Kessen, 1993). Johnston (1985) and Miller (1997) have stressed the complementary move of taking findings from normal environments back to the laboratory. Gogate et al. have shown how using well-defined, artificially altered circumstances in relation to the complex social system of infant and caregiver can enrich our understanding of development.

References

RESPONSE

More on developmental dynamics in lexical learning

Lakshmi J. Gogate¹ and Arlene Walker-Andrews²

¹. State University of New York at Brooklyn, USA
². Rutgers University, New Brunswick, USA

We appreciate the commentaries on our paper. Each commentator has addressed a different set of issues. Therefore, we address each commentary separately in alphabetical order.

The rewards of putting the cart before the horse (Lewkowicz)

Often times in a developing field such as developmental psychology, using a new framework (systems theory) can be challenging yet a worthwhile pursuit with bountiful rewards. Researchers in their pursuit of developmental explanations formulate hypotheses (e.g. that infant perceptual and maternal multimodal communication systems are dynamic and interact) in advance of extensive data in the hope that others will continue this endeavor allowing for a test of the hypotheses. We agree with Lewkowicz that we need more evidence. We continue to find empirical evidence for our hypotheses (see Gogate, Betancourt, Bolzani & Watson, under review; Gogate & Bahrick, in preparation). Our data continue to help us formulate further hypotheses. Such is the nature of any scientific process. Theory is often ahead of empirical support and application. In our case, putting the cart before the horse has yielded much. We have allowed our insights from developmental psychology in general to guide us in our language data gathering efforts, to combine controlled experiments with semi-structured observations, and to scrutinize our data before interpreting it (if we fail our reviewers do not). So why reject a theory before it is proven wrong? The popularity of nativism in the study of child language to date, despite evidence to the contrary, bears testament to this democratic scientific process.

Lewkowicz’s treatment of the current paper as a mere extension of the Gibsonian view of development is unfortunate. We adopted the ecological and the dynamic systems approaches to early word–object mapping because they complement one another. Nonetheless, Lewkowicz points to four reasons that he claims make the basis for this paper premature. First, he asserts that in our view ‘intersensory perception is a unitary, homogeneous process, and simply a matter of whether (or not) intersensory redundancies can be picked up in their ready-made final form’. Although we provide extensive evidence to show that infants detect intersensory redundancy early in infancy, we certainly do not claim that this detection is immune to development. Rather, we discuss prior evidence to show that infants’ detection of intersensory redundancies develops gradually from global to more specific invariants in auditory–visual events (p. 3), and eventually facilitates the detection of arbitrary relations. Thus, infants are highly sensitive to the intersensory redundancy present in otherwise arbitrary syllable–object relations by about 7 months.

Second, Lewkowicz claims that young infants do not always detect intersensory redundancies (or map syllables onto objects). We agree. For example, when syllables are presented with a checkerboard pattern, where the visual display covers the entire screen (the same stimuli used by Lewkowicz, 1985, 1992), Stager and Werker (1997) have shown that 14-month-olds act as if participating in a non-mapping task. Similarly, from another domain, 7-month-old infants do not discriminate vocal expressions of emotion when they accompany a checkerboard, yet do when the expressions are accompanied by faces (Walker-Andrews & Lennon, 1991). We see no reason why these null findings should diminish the strength of our findings when whole objects moving in temporal synchrony with vocalizations were given to infants.

Third, Lewkowicz asserts that we have ignored well-known intersensory illusions such as the McGurk and
ventriloquism effects. We refrained from discussing these effects because infants are unlikely to encounter them in caregivers’ multimodal communication. For example, mothers will probably never say ‘baba’ and present a visual mouth shape which looks like ‘gaga’ requiring the infant to neutralize the conflicting stimuli into ‘dada’ (the McGurk effect). Nor is the mother likely to throw her voice to make it appear as if it were coming from the object being named. In fact, in our multimodal motherese video records (Gogate, Bahrick & Watson, 2000; Gogate et al., under review), mothers who pretended during play that the hand puppets were talking to their infants never used ventriloquism. Thus, these illusions are irrelevant to our discussion. Although we admitted that spoken words, in general, do not emanate from referents (p. 5), our observational studies (Gogate et al., 2000, under review) showed that, for preverbal infants, typically the mother, the word and the object (she holds) are proximal and therefore spatially concordant (also see Zukow-Goldring, 1997). Contrary to Lewkowicz’s view about temporal discordance, our data show that mothers’ multimodal naming to preverbal infants is most often temporally concordant.

Finally, Lewkowicz misrepresents the Gibsonian view when he claims that in this view perception is wholly ‘out there’ and not ‘in the head’. The emphasis has been on the spontaneous and self-initiated exploratory activity of the perceiver rather than on the neural underpinnings and a description of the information for perception available to the perceiver. But, Lewkowicz uses Stein’s (1998) neurophysiological reports of intersensory perception in cats and monkeys as counter-evidence for the Gibsonian view of increasing specificity. In this view, perceptual development is mediated by an integrated sensory system that enables abstraction of intersensory invariants by young human infants and later increasing differentiation that enables perception of modality-specific information. Lewkowicz claims that the senses are not likely to be integrated early on in human infants because Stein (1998) reports that multisensory neurons in the superior colliculus (SC) are not present at birth in kittens. Unfortunately, Lewkowicz does not consider Stein’s (1998) cautious statement that (only) some structures may be common across species (p. 134). Further, Stein (1998) reports that deactivation of the auditory association cortex (in adult cats) inhibits response enhancement to spatially and temporally concordant stimuli in the multisensory neurons of the SC, suggesting to Lewkowicz that ‘perception is in the head, not out there’. However, Lewkowicz does not consider Stein’s report that multisensory neurons are present not only in the midbrain (SC) but also in the cortex (auditory ectosylvian sulcus, AES) and the thalamus (p. 125). Stein does not report response inhibition in the multisensory neurons of these areas as in the SC. The study of neurophysiological correlates of multisensory perception is intriguing but also in its infancy: Little is known about whether, as a result of evolution, human infants have more multisensory neurons in the AES than in the SC that are relatively resistant to response inhibition and are involved in functional roles such as word–object mapping and other language functions. Testing this hypothesis could be a fruitful endeavor, but another case of putting the cart before the horse!

Emphasis on bidirectionality in developmental systems (Lickliter)

Lickliter and Berry (1990) explain the ‘concept of a developmental system’ as that ‘which holds that what is inherited in reproduction is more than just the genes; rather, what is inherited are the genes and the developmentally relevant aspects of the organism’s stimulative environment’ (p. 355). Robert Lickliter primarily conducts research on how sensory systems and their respective stimulation histories influence one another during late prenatal and early postnatal development. He finds that such sensory experiences serve to maintain, facilitate or interfere with the usual course of intersensory development. Therefore, it is not surprising that Lickliter has taken our attempts to describe how human infants might learn label–object pairings and expanded the discussion to incorporate development more broadly. He argues here that it is crucial that we come to understand what specific developmental contexts offer in the form of stimulation to developing individuals and how new structural and functional properties and competencies emerge at all levels of analysis as a consequence of co-actions of all parts. We can but agree.

Self-organization: the ‘mechanism’ for change in dynamic systems (Schafer)

Schafer laments that our emphasis on process over mechanism fails to give insight into how different states (or stages) might be achieved. He evokes an association mechanism in the infant, and the higher frequency of correct word–referent pairings relative to idiosyncratic pairings (e.g. the word and the actor’s hand in our experiments) in the environment, to account for the development of word–referent mapping in infants. Stages are said to emerge as infants (simply) learn
about the reliability of successively more arbitrary cues (e.g. first infants require temporal synchrony to learn word-object relations (stage 1), and later they learn these relations in the absence of temporal synchrony (stage 2)). Early word learning is 'characterized as one in which it is ultimately important to decouple representations of label and referent' – a linear developmental process. No explanation is provided, however, for how infants come to decouple these. Schafer fails to acknowledge our explanation for nonlinear developmental change in both infants' perceptual–lexical and maternal multimodal communication systems. We find it necessary, therefore, to reiterate that changes from one (temporary) preferred state to the next result from self-organization in one system in response to systemic pressures of the other. Self-organization occurs due to local instabilities within and outside a system when multiple factors constantly fluctuate and interact in real time. We do not think that 'facts about the world ... are ... discovered by association' alone. Similarly, Schafer and Plunkett (1998) pointed out that infants' responses on an interval word-mapping task were 'unlikely to result from simple classical conditioning' (p. 318). We propose here that the ability to associate is simply one of many intrinsic factors available during word-mapping development. Infants' ability to attend to, relate and remember arbitrary word–object pairings, their speed of processing and the ability to compute probabilities (detect statistical regularities) too are intrinsic factors that interact to facilitate lexical mapping. The frequency of co-occurring words and referents is an extrinsic factor. Treating any one of these factors as the mechanism for change results in a predicament where one investigator's mechanism becomes another's extraneous factor (see a parallel problem with the nativistic distinction between competence and performance limitation; Telen & Smith, 1994, p. 26). By monitoring organismic changes in intrinsic variables and the organism's responses to extrinsic variables in carefully controlled experiments we can chart developmental change (Gogate, under review). Data-driven descriptions of interactions between multiple variables at a given time, and their dips and peaks over time, are alone sufficient to explain developmental dynamics without evoking mechanisms. Our review of the word-mapping data suggested a decline in the importance of association (learning based on temporal contingency and contiguity) and an increasing requirement of a more precise temporal synchrony (object motion in temporal alignment with vocalization) in the first year, and social cues such as adults' pointing and/or eye-gaze in the direction of objects being named during the second year.

Several comments about the present article are confusing to us. For example, Schafer uses the term 'innate affordances', and categorizes E.J. and J.J. Gibson as 'nativists'. In contrast, we reiterate that affordances are the very properties of the environment that are salient to (and perceived by) the organism. E.J. Gibson (1969) and J.J. Gibson (1979) recommended that 'what is out there' and 'what the infant perceives' be studied as a unit, rather than studying the environment or the organism in isolation. (This does not make them nativists.) After all, even though properties of objects and spoken words may be 'present in the input as a priori statistical regularities' they would fail to hold any relevance to the organism (the infant) unless they were also salient to the organism at a given point in time during its development. Affordances are not 'manipulated' by the caregiver as Schafer asserts (no awareness is assumed). Neither do we think that they are the 'innately specified sine qua non of infant lexical-semantic development'. Knowledge in the form of statistical regularities is not represented explicitly in the infant's mind. Schafer confuses the ecological view with that expressed by Spelke (Spelke, Breinlinger, Macomber & Jacobson, 1992). The ecological approach in particular speaks of affordances and regularities in the environment waiting to be discovered by the organism. Perception and attention to salient stimuli can lead to learning and memory, not only in neonates and young infants, but also in utero by the fetus (e.g. Decasper & Fifer, 1980; Fifer & Moon, 1989). We examined infants' attention to and perception of salient properties of the communication stimulus array, and abstraction of invariants when these stimuli were made available to auditory–visual modalities.

In the present paper, we acknowledged that basic association is one type of lower level learning (Figure 1). However, all learning cannot be chalked up to associations based on statistical regularity or frequency of co-occurrence. Our studies (Gogate & Bahrick, 1998; Gogate, under review) showed that infants learned the syllable–object pairings, but not idiosyncratic pairings such as a hand and the syllable, from bimodal patterns of stimulation when the statistical regularities were held constant for these and the idiosyncratic pairings. Next, Schafer stated that maternal multimodal communication and infants' intersensory perceptual systems are 'loosely coupled'. To the contrary, we argued – 'Because of the close fit between the two systems, the changing nature of one fits the changing nature of the other' (p. 2). Finally, Schafer places a unidirectional emphasis on shifts in maternal multimodal communication that causally relate to infants' subsequent ability to learn words. This puts the onus on the caregiver to regulate
changes in the infant, but does not shed light on the systemic changes in the infant that might regulate caregivers' communication. In contrast, we suggest a bidirectional relation between dynamic infant perceptual and maternal multimodal communication systems. A more recent study (Gogate et al., under review) has revealed an inverse correlation between mothers' use of 'multimodal motherese' (temporal synchrony) and infants' age (6 to 8 months), and between infants' attention to 'multimodal motherese' and infants' age, suggesting a bidirectional relation between infant and maternal systems and multiple co-acting factors typical of dynamic systems. Further, infant attention to 'multimodal motherese' and mothers' use of 'multimodal motherese' were significantly correlated with each other and with infants' learning of word-object relations on a subsequent intermodal word-mapping test. These data question the usefulness of single-factorial unidirectional explanations for lexical development.

Can lexical comprehension develop without perception (Werker & Patterson)?

Werker and Patterson acknowledge that we present a developmental account of word learning. However, they remain skeptical of our view that lexical comprehension originates from infants' intersensory perception of auditory-visual events. In our view, this skepticism stems from their focus on three main sources. First, they focus on unimodal data that 'reveal how infants learn about words but do not [we agree] inform us about actual word learning'. For young infants, the natural word-learning environment requires that words and referents be present in the immediate context. How else could infants learn the conventional relations between words and their referents? We should not be surprised, therefore, if well-known experimental studies of speech perception, where infants are provided with only words or syllables but no referent objects or actions (Werker lists numerous references), can teach us nothing about actual word learning, and more precisely word-referent mapping.

Second, Werker and Patterson focus on the end-state of development of full word learning (like others before them). This focus has kept researchers away from trying to explain how infants begin to attend to auditory-visual information and end up knowing that specific words stand for specific objects or actions (see Gogate & Bahrick, in press, for a review). In contrast, we have attempted to address this issue by focusing on a description of the process of early lexical development. Our paper focuses on how infants come to know that a word and an object go together (pre-symbolic development; see p. 1). We agree with Werker and Patterson that the ability to put together a word and an object is not reference, where a word 'stands for a referent'. But we are interested in charting the road infants take in their long journey to reference. Werker and Patterson assert that, for one to grant referential ability, 'the infant must form an arbitrary link between word and object'. Again, we agree, but how do infants develop this ability? We assert that development cannot take place in a vacuum (Thelen & Smith, 1994), and researchers must account for the early milestones leading to reference sooner or later. Focusing on the end-state will not explain development.

Third, Werker and Patterson are not convinced that 'perceptual learning' during infancy is linked in a causal way to 'the conceptual advances required for word learning'. Their skepticism stems from their own findings where 8-month-olds did not put together two syllables ([imir] and [ilir]) and two objects when object motion alone was provided in the absence of temporal synchrony (Werker, Cohen, Lloyd, Casasola & Stager, 1998). But consider the evidence from a set of studies that suggest that the perceptual requirement for word-referent mapping changes during development. Seven- and 8-month-olds learned syllable-object relations (/a/ and /i/, and /tah/ and /gah/, respectively) only in the presence of temporal synchrony in two studies (Gogate & Bahrick, 1998; Gogate, under review). Werker et al. (1998) showed that only 14-month-olds put together distinct syllables and objects in the absence of temporal synchrony, and only in the presence of object motion (a perceptual property). Their 14-month-olds and our 7-month-olds failed to learn syllable-object relations when two static objects accompanied two syllables (Werker et al., 1998; Gogate & Bahrick, 1998; cf. Moore, Angelopoulos & Bennett, 1999, with older infants). Thus, we concluded that perceptual properties of the syllable-object stimuli such as intersensory redundancy and object motion facilitate the detection of syllable-object relations. Furthermore, syllable distinctiveness influences word mapping. Stager and Werker (1997) found that 14-month-olds did not detect two syllable-object relations when the syllable onsets differed only by a single feature ([bIh] and [dIh]). We have recently found similar evidence. Fourteen-month-olds did not put together two syllables and objects when the syllables differed only in their onset phonemes, [tab] and [gah] (even though 8-month-olds did in a prior experiment in the presence of temporal synchrony; Gogate, under review), but succeeded when the syllables differed in their onset and nucleus, such as [tah] and [gih] (Gogate & Bahrick, in preparation). The objects'
motions in our videos were out of synchrony with the syllables, similar to Stager and Werker's (1997) and Werker et al.'s (1998) studies. Rather than conclude that infants' perceptual abilities are unrelated to the conceptual advances leading to word learning, we maintain that these findings taken together suggest changing perceptual requirements for word-referent mapping (in keeping with the dynamic systems view). Infants require temporal synchrony, then object motion, and phonetically distinct stimuli in the absence of synchrony at different points during lexical development.

Werker and Patterson fail to acknowledge that our 'multimodal motherese' data (Gogate et al., 2000) complement their valuable experimental findings and ours. Mothers of preverbal infants (5- to 8-month-olds) predominantly named objects and actions in temporal synchrony with objects' motions, and mothers of lexically advanced infants (21-30 months) used this strategy less often (also see Zukow-Goldring, 1997). These findings were replicated in a more recent study (Gogate et al., under review). In this study we showed, further, that 6- to 8-month-olds learned two word-object relations on an intermodal word-mapping test if they attended to their mother's use of temporal synchrony between target words and moving toy objects during an earlier play episode. In addition, we have found that these infants' attention to toy objects when mothers named them during play was positively correlated with the mothers' reports of their infants' (receptive) knowledge of names for toys on the MacArthur Communicative Development Inventory (Fenson et al., 1994). These findings suggest that early perceptual learning of word-object relations is closely related to word learning in the real world (also see Schafer's concern). By this, we do not suggest that learning of word-object relations at 7–8 months in the presence of temporal synchrony is comparable to learning of word-object relations in the absence of temporal synchrony at 14 months (Werker & Patterson's question). Fourteen-month-olds may have a productive vocabulary, whereas 7–8-month-olds do not. But we also know from earlier studies that infants' receptive vocabulary is far in advance of their productive vocabulary (Benedict, 1979). Therefore, we have every reason to believe that Werker's (and our) findings reflect infants' actual word-mapping capabilities. Questions comparing younger and older infants remind us of an old developmental question that has now been answered – is early stepping comparable to later walking? Studies have shown that early stepping movements resemble and are a precursor to later walking (Thelen & Cooke, 1987; see Thelen, 1988, for a review).

Finally, Werker and Patterson maintain that an advance in conceptual thinking (e.g. understanding a speaker's intentionality), rather than intersensory perceptual abilities, is the key to 'true word learning in the absence of perceptual cues'. However, they provide no explanation for how the understanding of another's intentions (one cognitive ability) might lead to reference (another cognitive ability). We suspect that the infants' ability to understand adult speakers' intentions too has its humble origins in the early joint attentional (perceptual) abilities seen emerging at about 6 months (see Bakeman & Adamson, 1984) which are highly correlated with receptive vocabulary in the second year (Morales, Mundy & Rojas, 1998). We conclude that many perceptual abilities are intertwined with cognitive abilities such as reference and the understanding of others' intentions. Distinctions between perception and cognition are more clear-cut in the minds of researchers and analysts than in the minds of babies (Thelen & Smith, 1994).

Infants' intersensory perceptual–lexical and maternal multimodal communication systems: a history of the bidirectionality hypothesis (Zukow-Goldring and de Villiers Rader)

Zukow's invaluable observations of mothers' communication with infants have inspired our studies, as have others' research. The present paper is an example of how researchers come to write a timely collaborative, theoretical piece after coming to similar ideas from their own research, exposure to various literatures, and communication of these ideas to one another. In 1981 one of us (AW-A) reviewed a paper by Sullivan and Horowitz (published in 1983) which proposed that 'The infant's experiences with intermodal stimulation over the first year of life are very likely to provide the abilities necessary for learning the arbitrary associations between words and objects. ... The learning of word-object associations, will be accomplished in the context of the infant's multimodal interactions with its mother' (p. 210). This exciting idea gained currency as Zukow conducted exquisite observational studies that support the initial proposal and document the rich communicative environment provided by caregivers (see Zukow, 1990; Zukow-Goldring, 1997, in press; see papers listed in her commentary). In 1995, another of us (LJG) completed her doctoral training in developmental psychology with an emphasis on phonological and symbolic development in infancy and early childhood (see McCune, Vihman, Roug-Hellichius, Delery & Gogate, 1996; Yavas & Gogate, 1999), yet regretted not having found a method to test preverbal infants' developing lexical comprehension. A method became

© Blackwell Publishers Ltd. 2001
available from the work of Bahrick (1994), who showed that 7-month-olds detected arbitrary color–pitch relations following habituation to two distinct objects (yellow and red) each making a distinct sound (of high or low pitch) on impact against a surface. From then on began a fruitful exchange of concepts and methodology between intermodal perception and word–referent mapping development, resulting in the current paper. Gogate has modified Bahrick’s (1994) method to conduct laboratory studies of infants’ perception and memory for syllable–object relations (Gogate & Bahrick, 1998; in press; in preparation) and simultaneously Gogate and Bahrick have conducted quasi-naturalistic studies of maternal multimodal communication to infants. The two strands of research indicate that infants’ perceptual requirement for syllable–object mapping complements their experiences with multimodal maternal communication (see response to Werker & Patterson). Gogate and her students have found further that mothers’ communication to preverbal infants using temporal synchrony is correlated with these infants’ learning of word–object relations from their mothers as seen on an intermodal word-mapping test (Gogate et al., under review). The converging evidence on the importance of temporal synchrony to infants from experimental procedures and infants’ learning of word–object relations from their mothers dispels any reservations about the relevance of the experimental procedures to infants’ word learning in the real world. Continued research by us, by Zukow-Goldring, and others who focus more directly on infants’ learning of word–object relations (e.g. Werker, Schafer and their colleagues) will lead to further tests of the bidirectionality hypothesis.

References


Gogate, L.J. (under review). Intersensory redundancy facilitates learning of arbitrary minimal pair–object relations by 8- but not 7-month-old infants: evidence for a dynamic system.


