The Development of Infants' Sensitivity to Arbitrary Intermodal Relations

Lorraine E. Bahrick
Florida International University

The development of infants' ability to detect the arbitrary relation between the color/shape of an object and the pitch of its impact sound was investigated using an infant-control habituation procedure. Ninety-six infants of 3, 5, or 7 months were habituated to films of two objects differing in color and shape, striking a surface in an erratic pattern. One object produced an impact sound of a high pitch, and the other object produced a low pitch. During test trials, infants in the experimental conditions received a change in the pairing of pitch with color/shape, whereas controls received no change. Results indicate that visual recovery to the change in pitch-color/shape relations was significantly greater than that of age-matched controls at 7 months, but not at 3 or 5 months. A prior study demonstrated that by 3 months, infants were able to discriminate the color/shape and pitch changes of these events. However, it is not until 7 months that they show evidence of detecting the arbitrary relation between these attributes.

Studies of intermodal perception have revealed that infants are remarkably sensitive to amodal invariant relations. From a very early age, they are attuned to information that is redundant across vision and audition. They detect temporal synchrony, changing distance, rhythm, vowel sounds, tempo of action, affective expression, and object substance and composition across vision and audition (Bahrick, 1983, 1987, 1988; Kuhl & Meltzoff, 1984; Pickens, 1990; Spelke, 1979; Walker, 1982; Walker-Andrews & Lennon, 1985). Detection of these relations is present by 3–5 months of age and requires little or no learning (see Bahrick & Pickens, in press, for a review).

Requests for reprints should be sent to Lorraine E. Bahrick, Department of Psychology, Florida International University, Miami, FL 33199.
Some intersensory relations, however, are not amodally specified or redundant across sense modalities. Rather, they are arbitrary and modality-specific. For example, the relation between the shape or color of a container and the smell or taste of its contents, or the visual appearance of a moving object and the pitch of its impact sound, or the location of an object and its identity, are arbitrary and unpredictable. So is that between the particular sound of a woman's voice and the appearance of her face, her hairstyle, and coloring as contrasted with the appearance of another woman. These relations are not permanent and may even change from one context to another. An object's impact sound may change depending on the surface it strikes but its visual appearance will not; a particular food may come in several different types of containers, and its location in the store may change; one woman's voice could just as easily belong to another woman, or a given woman's hair color or style may change but her voice will not. These relations must be learned, and they are important for developing expectations about objects and events in our environment. Nevertheless, adults appear to easily detect and remember numerous arbitrary relations. One need only imagine a typical trip to a grocery store, where we are familiar with the products and their locations, to realize how we rely on intermodal knowledge of arbitrary relations. In contrast, a trip to a grocery store in a foreign country, where we are unfamiliar with the product packaging and their locations, can take a great deal longer than at home.

Only a few studies, however, have investigated intermodal perception of arbitrary relations by young infants. By 3½ months, infants recognize the relation between the sight of their mother's face and the particular sound of her voice (Spelke & Owsley, 1979). Female infants at 4 months learned the relation between the visual appearance of a toy and a distinctive odor (Fernandez & Bahrick, in press). Seven-month-old infants were able to learn the relation between the color of a cup and the taste of the food it contained (Reardon & Bushnell, 1988), but they were unable to learn the relation between color and temperature (Bushnell, 1986).

How might these arbitrary intermodal relations be learned? One possibility is that they are learned by association, on the basis of co-occurrence. Similarly, arbitrary intermodal relations can be learned through training or conditioning in early infancy. For example, conditioning effects are demonstrated by the nonnutritive sucking technique in which infants modulate their sucking rate for a visual reinforcement (Siqueland & DeLucia, 1969), and the mobile conjugate reinforcement method in which infants learn to move their legs to cause a crib mobile to move (Rovee-Collier & Fagen, 1981). This type of conditioning may also play a role in the infant's learning about social contingencies. However, research has demonstrated that in the first months of life, infants show a unique sensitivity to amodal information and need no special training or conditioning to detect amodal relations. Bahrick & Pickens (in press) suggest that young infants are selectively tuned to abstract amodal information prior to abstracting
arbitrary relations, and this progression may operate even within a given learning session. Consistent with this notion, Bahrick (1992) proposed that intermodal knowledge about arbitrary relations often develops by first detecting amodal relations. For example, Spelke (1979, 1981) found that through detecting amodal synchrony and tempo uniting the sounds and sights of a bouncing stuffed animal, infants were then able to remember the relation between the appearance of the animal and the particular sound it produced. Bahrick (1988) found that only after detecting two types of amodal invariant relations were infants able to learn to pair a soundtrack with one of two distinctive objects. Thus, by detecting amodal information, infants' attention is directed toward appropriate audio-visual relations and, in turn, may lead to intermodal knowledge about arbitrary relations. For example, by attending to a synchronous sight and sound, the infant may eventually differentiate the arbitrary relation between its pitch and its color and shape. In this way, detection of amodal invariants can precede and guide learning about arbitrary relations. This progression is consistent with Gibson's (1969) view that perceptual learning proceeds in order of increasing specificity.

Consistent with this view, Bahrick (1992) found that 3-month-old infants could detect two amodal relations but not an arbitrary relation between the visual and acoustic stimulation produced by objects striking a surface. Infants detected a change in synchrony between the sights and sounds of an object's impacts, and a change in information specifying object composition. The audible and visible temporal microstructure at each impact specified object composition, that is, whether the object was composed of a single, unitary element, or a group of smaller elements. However, infants failed to detect a change in the arbitrary relation between the pitch of a sound and the color/shape of the object, even though the visual and acoustic contrasts were discriminable to them. It was thus proposed that detection of amodal relations developmentally precedes detection of arbitrary relations within a given domain. This pattern was seen as adaptive in that it could promote the development of veridical perception. Because amodal relations detected in one context can generalize meaningfully to new contexts and events, they are important as a basis for perceptual development. On the other hand, arbitrary relations often cannot be generalized meaningfully across contexts and events. Thus, by first attending to amodal relations, perceptual learning will be based on more permanent, context-independent relations, and it will not be disrupted by learning unpredictable relations that vary from one context or event to another.

The present research assesses the development of infants' detection of an arbitrary audio-visual relation. It extends the work of Bahrick (1992) by asking when, if at all, do infants detect the pitch-color/shape relations in the events used in that study. All objects can be characterized as having a distinctive color and shape, and when they hit a surface, they have an impact of a particular sound. The impact sound may be characterized as having a distinctive pitch.
Given that infants of 3 months are able to discriminate the objects on the basis of color and shape and the sounds on the basis of pitch, at what age do infants attend to and remember the arbitrary relation between these attributes?

**METHOD**

Participants

Ninety-six infants of 3, 5, and 7 months participated in the study. The 3-month-old infants \( n = 36; 18 \) males, 18 females) had a mean age of 112.7 days \( (SD = 4.1) \), and their data were reported in a prior study (see procedure for details); the 5-month-old infants \( n = 30; 15 \) males, 15 females) had a mean age of 156.8 days \( (SD = 6.0) \); the 7-month-old infants \( n = 30; 15 \) males, 15 females) had a mean age of 216.2 days \( (SD = 6.0) \). The data of 31 additional infants were rejected: 14 because of excessive fussiness \( n = 9, 2, \) and 3 at 3, 5, and 7 months, respectively), 5 for equipment failure or experimenter error, 9 for failure to meet habituation or fatigue criteria (see procedure section for details), 2 for falling asleep, and 1 for failure to attend to task. The babies were primarily middle class, and their parents met a minimum criterion of 12 years of education. They were recruited through the local birth records, and all were healthy with no major complications at delivery.

Audio–Visual Displays

The audio–visual events were the same as those used by Bahrick (1992). They consisted of video films of 12 events, each portraying objects with distinctive colors and shapes striking a surface repeatedly and producing natural impact sounds of a high or low pitch. The 12 events comprised three event sets, each with two event pairs. Each set was constructed of a different material—metal, plastic, or wood—and depicted a different type of impact motion. The members of each event pair differed visually from one another only in terms of color and shape; they were of the same substance and composition, and they underwent the same types of motion. One pair within each set depicted single, large objects (single objects), and the other pair depicted clusters of smaller objects (compound objects). The compound object, as a unit, was about the same size as the single object. The event sets consisted of the following:

1. **Metal set.** In the single-object pair, a single, large, metal, object, suspended from a string, was raised and abruptly dropped against a wooden surface. One object was a yellow, donut-shaped washer about 8.5 cm in diameter and the other was an orange, hexagonal nut about 6 cm in diameter. In the compound-
object pair, a cluster of small, yellow washers and a cluster of small, orange nuts were also raised and abruptly dropped against a wooden surface.

2. Plastic set. In the single-object pair, a single, large, plastic fruit (a yellow pear, 7 cm in diameter or a red tomato, 8.5 cm), was held from behind by a small stick (with an unseen hand) and abruptly struck against two surfaces of different heights in an erratic back and forth motion. In the compound-object pair, a cluster of small, red, tomatoes and a cluster of small, yellow, pears were moved as just described.

3. Wooden set. In the single-object pair, a single, large, colorful board (approximately 20 cm in width) was hinged to a horizontal surface and slammed shut with a string from above. One board was semicircular, with pink and purple stripes and the other was rectangular, with yellow and blue stripes. In the compound-object pair, a group of six small (approximately 6.5 cm in width) semicircular objects, with pink and purple stripes and a group of six small, rectangular objects, with yellow and blue stripes, each hinged to a wooden surface (arranged in two rows of three), were opened and slammed shut with a common string from above.

The natural impact sounds of the objects were modified by an auditory signal processor. The pitch was raised or lowered by 3.5–4 whole steps such that the high- and low-pitched sounds for each event were at least one octave apart. Twenty adult observers judged both the high and low sounds to be plausibly produced by the visual events. The single objects produced single, abrupt, banging sounds at impact, whereas the compound objects produced discrete, but slightly more prolonged and complex, impact sounds. The objects struck the surface in an erratic, unpredictable pattern, and each event averaged approximately 40 impacts per min.

In addition, a video film was also created to serve as a control display. It portrayed a green and white, plastic toy turtle that had front legs that spun around, creating a whirring sound.

Apparatus

Infants sat in a standard infant seat, facing a 19-in. (Panasonic® BT–S1900N) video monitor, approximately 55 cm away. A mechanical toy dog and a set of colored Christmas tree lights were positioned above and to the right of the video display. The video monitor was surrounded by black posterboard. Two apertures, one to the upper-right and the other to the upper-left side of the monitor, allowed observers to view the infants’ visual fixations.

The audio–visual events were videotaped using a Panasonic® (WV 3170) color video camera and a Sony® (EMC–150T) remote microphone. They were edited and presented with a Panasonic® (VHS NV–A500) edit controller that was connected to two Panasonic® video decks (NV–8500 and AG–6300). By using
two video decks, we were able to switch from one video display to another without the extra time or noise that results from changing cassettes. All soundtracks were presented at approximately 65–70 db. through an auditory signal processor (Yamaha® SPX 90) connected to a speaker positioned just below the video monitor.

A trained observer, unaware of the infant's condition (experimental vs. control), monitored visual fixations by depressing a button while the infant fixated the video image. The button box was connected to an Apple® IIe personal computer that was programmed to record visual fixations on line, signal when the infant had looked away for 1.5 s, and signal when the habituation criterion had been reached. The signal was transmitted to the experimenter through a small speaker and headphone. A permanent record of the infant's visual fixations was printed during the experiment using a Qume® (Letterpro 20–S) printer. The observations of the primary observer controlled the audio–visual presentation, whereas those of the secondary observer were simply recorded for later calculation of interobserver reliability.

Procedure

Infants were tested using an infant–control habituation procedure (Horowitz, Paden, Bhana, & Self, 1972) to determine whether they could discriminate a change in the pairing of object pitch with color/shape. Most of the data of infants in the 3-month-old group were collected first, and they were described as part of a prior study (Bahrick, 1992, Experiment 1) where it was found that they could not detect changes in the pairing of pitch with color/shape. Thus, the 5- and 7-month-old infants were tested concurrently, using methods and procedures identical to those of the 3-month-old infants, to determine at what age this ability emerged. Because all three age groups were tested within a close time frame, were under identical procedures, and will be compared in analyses, the data of the 3-month-old group will be described here along with those of the two older groups.

Infants of 3, 5, and 7 months were randomly assigned to either an experimental or a control group (n = 24, 18, and 18, respectively, in the experimental groups; n = 12 in each control group). Each baby was habituated to two events that were members of the same event pair and differed only in color and shape from one another. One member of the pair was presented with an impact sound of a high pitch, whereas the other was presented with an impact sound of a low pitch. Event set and event pair were counterbalanced within each experimental and control condition. One third of the infants in each condition received an event pair from the set of metal objects, one third from the wooden objects, and one third from the plastic objects. Within each of these cells, half received the pair of single objects and half received the pair of compound objects. It was randomly determined which of the two events would be presented first. Presen-
tation order of the two events then followed an alternating sequence. The assignment of the high-versus low-pitched sound to the member of each event pair was also randomly determined.

Following the habituation sequence, those in the experimental conditions received a change in the pairing of pitch with color/shape for the test trials, whereas those in the control conditions received no change from habituation to test. For example, for habituation, a baby might receive the large yellow pear with the high-pitched sound and the large red tomato with the low sound; for test, the pear was then played with the low sound and the tomato with the high sound for experimental infants, whereas controls received no change. Because, originally, these objects had essentially the same sound when they struck the surface, the assignment of a high- or low-pitched sound was arbitrary. From habituation to test, only the pairing of pitch with the object's color and shape changed for experimental infants.

The habituation sequence consisted of an initial control trial, four mandatory habituation trials, and was terminated after the infant reached the criterion for habituation and received two posthabituation trials. Each trial began when the infant fixated the visual display and terminated after the infant looked away for 1.5 s. In addition, a ceiling of 60 s was set for the maximum trial length, and 20 trials was set as the maximum number of trials. Two babies were rejected for failure to habituate within this number of trials. The habituation criterion was defined as a fixation decrement of 50% or greater on two consecutive trials, relative to the infant's initial fixation level (the average number of seconds of fixation during the first two habituation trials). After the two criteria trials, two (no-change) posthabituation trials were presented. These trials served to establish a more conservative criterion for habituation, reducing chance habituation and taking into account spontaneous regression effects by assessing visual recovery in relation to these trials (see Bertenthal, Haith, & Campos, 1983, for a discussion of regression effects). Following the habitation sequence, infants received two test trials—one of each event—to assess visual recovery and then a final trial with the control display (turtle). The two test trials were presented in the same order as the first two events seen for habituation. The intertrial intervals were approximately 6 s long and the colored lights were illuminated during this time. The mechanical toy dog was only infrequently activated, always at some point prior to the posthabituation trials, to encourage babies to attend to task.

The infant's data were examined to determine whether two criteria (set a priori, used previously in Bahrick, 1992) had been met. To ensure that they had in fact habituated to the events, data from infants whose mean posthabituation fixation level exceeded that of their mean initial fixation level (baseline) were excluded from the study ($n = 2$ at 3 months; $n = 1$ at 7 months). Second, to exclude the data of babies who were overly fatigued and unable to show visual recovery, the fixation duration to the moving turtle during the final control trial
was compared to that of the initial control trial. If the final level was less than
20% of the initial level, the infant's data were rejected \( n = 3 \) at 3 months; \( n = 1 \) at 7 months. The remaining babies showed a large visual recovery on the final
control trial (median = 100% of initial fixation level).

A secondary observer monitored visual fixations for 21 of the infants for the
purpose of assessing interobserver reliability. Total fixation time on each trial
was calculated independently on the basis of observations made by the primary
and the secondary observer. For each infant, a Pearson product–moment
correlation between these observations was calculated and the mean correlation
across subjects was .99.

RESULTS

Primary Analyses

Visual recovery to the two test displays was calculated by subtracting the mean
fixation time across the two (no-change) posthabituation trials from that of the
two test trials for each infant. Mean visual recovery for experimental versus
control infants at each age is depicted in Figure 1. To assess any differences

![Figure 1: Mean visual recovery to a change in pitch–color/shape relations.](image-url)
across age, a two-way analysis of variance (ANOVA) was performed with age (3, 5, and 7 months) and condition (experimental vs. control) as main factors. Results indicated no significant main effects or interactions (all ps > .1). In addition, visual recovery scores showed unequal variance across age (p < .001, according to a Cochran C test), where younger babies showed greater variance than older babies. Thus, the visual recovery scores were not well suited to ANOVAs that include age as a factor.

To address the main research question—did infants show evidence of detecting the change in pitch–color/shape relations—planned t tests were conducted at each age, comparing the visual recovery scores of the experimental infants with those of the age-matched controls. Results indicated no significant difference at 3 or 5 months; t(34) = -.16, p > .1, t(28) = .33, p > .1, respectively; however, by 7 months, visual recovery of the experimental infants was significantly greater than that of the age-matched controls t(28) = 2.31, p < .05. For individual infants, 15 of the 18 7-month-old infants showed visual recovery scores greater than the mean of the control group (p < .005, binomial test) and 13 of the 18 7-month-old infants in the experimental condition showed visual recovery scores that were greater than 0 (p < .05, binomial test). Further, when the visual recovery of the 7-month-old infants was tested for significance against the chance value of 0, it also showed significance according to a single sample t test, t(17) = 2.15, p < .05, whereas that of the 3- and 5-month-old infants did not (ps > .1). Thus, by the age of 7 months, but not at 3 or 5 months, infants show evidence of detecting the relation between the pitch of an impact sound and the color and shape of an object.

In order to determine whether visual recovery differed as a function of event set or pair, three-way ANOVAs were conducted for each age group separately with condition (experimental, control), event set (metal, plastic, wood), and event pair (single, compound) as main factors. Results indicated no significant main effects or interactions (all ps > .1).

Secondary Analyses

Analyses were also conducted to evaluate the performance of infants during the habituation phase, prior to the introduction of the test trials. Habituation performance was examined with respect to five measures: (a) baseline, defined as the mean length of fixation across the first two habituation trials, (b) mean number of trials needed to reach the habituation criterion, (c) mean number of seconds to habituation, summed across trials, (d) mean length of fixation across the two criteria trials, and (e) mean length of fixation across the two (no-change) posthabituation trials. Because three of these measures (baseline, seconds to habituation, and posthabituation) showed unequal variance across age, overall analyses were not performed using age as a factor. Thus, one-way ANOVAs were conducted separately for each of these measures at each age to determine
whether experimental and control conditions differed a priori as a function of any of these variables. Results indicated only one significant main effect: Experimental infants at 7 months took less time to reach the habituation criterion than did controls, $F(1, 28) = 5.51, p < .05$. However, they showed no difference in their mean length of fixation across the two criteria trials or across the two posthabituation trials ($p > .1$ for these and all other main effects). Thus, infants in the experimental and control groups at each age did not differ a priori in their initial interest in the events or their final interest level prior to the introduction of the test trials.

**DISCUSSION**

By the age of 7 months, but not at 3 or 5 months, infants show evidence of detecting the arbitrary relation between the pitch of an impact sound and the color and shape of an object striking a surface. Although infants in the experimental and control conditions did not differ in their initial or final levels of attention to the habituation displays, they did differ in their attention to the test displays by the age of 7 months. Infants in the experimental condition showed significant visual recovery to a change in the pairing of pitch with object color/shape as compared with controls, who received no change. In contrast, at 3 and 5 months, babies showed no evidence of visual recovery.

An alternative hypothesis, that infants at 3 and 5 months failed to detect the changes in pitch–color/shape relations because they were unable to discriminate the visual or acoustic changes, was disconfirmed by results of Bahrick (1992, Experiments 2 and 3). In these studies, infants of 3 months detected changes in both the pitch and color/shape of all the events used here. Thus, although 3-month-old infants can discriminate the changes in the pitch of the impact sounds and the distinctive color/shapes of these objects, it is not until 7 months that they attend to the relation between these visual and acoustic attributes under these conditions.

Results of Bahrick (1992) also cast doubt on another alternative hypothesis, that the development of infants' general attentional capabilities accounts for the development in detection of the pitch–color/shape relations. In Experiment 1, 3-month-old infants showed significant visual recovery following habituation to a shift in the pairing of objects and sounds when the shift violated an amodal audio–visual relation. That is, when the impact sounds of a single, large object and a compound object were synchronized but mismatched, infants showed visual recovery. They were sensitive to the audible and visible temporal microstructure specifying object composition. In this study, the same events were presented under the same habituation procedures and exposure conditions as in the Bahrick (1992) study, but the shift violated an arbitrary audio–visual relation. Thus, it appears that the infant's selective attention, or sensitivity to
certain object–sound relations changes with age. Given similar exposure conditions, sensitivity to amodal synchrony and object composition relations is greater than that to arbitrary pitch–color/shape relations at 3 months. Sensitivity to pitch–color/shape relations increases with age, and by 7 months infants attend to these relations and detect changes in this arbitrary pairing.

Learning about arbitrary relations, such as that between the pitch of an impact sound and the color and shape of an object, typically does not generalize appropriately to new objects or contexts. A yellow, hexagonal object will not always make an impact sound of a high pitch. In contrast, learning about amodal relations can generalize meaningfully to new objects and contexts. A single object will typically make a single impact sound. Results of this study along with those of Bahrick (1992) are consistent with the view that detection of amodal relations developmentally precedes detection of modality-specific arbitrary relations for a given event. This pattern is adaptive because it can promote the development of knowledge about persistent relations prior to the acquisition of knowledge about more variable, idiosyncratic relations. This developmental pattern is consistent with an increasing specificity view of perceptual development (Gibson, 1969). That is, more global, context-independent relations are learned prior to more specific, context-dependent relations. Further research will assess whether this developmental pattern is evident across other domains. However, it is expected that the ages at which the amodal and arbitrary relations are detected will vary depending on familiarity with the events, their complexity, and other factors, whereas the developmental pattern will remain constant.

These findings also add to our knowledge of the arbitrary intermodal relations that infants are able to detect. All moving objects in the natural environment produce an impact sound of a characteristic pitch and have a characteristic color and shape. Apparently, by the age of 7 months infants become attuned to the relation between these attributes.

Why do infants at 3 months perceive and discriminate color/shape and pitch in their respective modalities, whereas not until 7 months do they abstract the intermodal relation between them? No similar developmental lag has been found to characterize infants’ detection of audible and visible information that is amodal or redundant across sense modalities. Detection of relational information that is amodal appears to be well established by the age of 4 months or earlier (see Bahrick & Pickens, in press). This developmental lag may be specific to the perception of arbitrary relations. If so, this lag could facilitate the development of veridical intermodal perception by, in effect, buffering against premature learning of idiosyncratic relations that do not generalize appropriately. The notion of such a buffer is supported by the findings of Bahrick (1988). Three-month-old infants were able to learn to pair single and compound objects with their natural impact sounds when the sounds and motions were synchronous. However, under identical training conditions, they were unable to learn to pair the objects with the wrong sounds or with their natural sounds when
they were asynchronous. The important role of amodal relations in constraining early intermodal learning may thus allow time for infants to first gain knowledge about more global, context-free relations—that certain types of objects and certain kinds of sounds belong together. For example, information about object substance, composition, number, or pattern of motion is both visually and acoustically given. Then learning about more specific, context-dependent or arbitrary relations may develop within the appropriate constraints.

ACKNOWLEDGMENTS

This research was supported by National Institute of Child Health and Human Development Grant 1–RO1–HD25669 to the author.

Special thanks are extended to Linda Alberga, Maricel Cigales, Christine Fadil, Maria Hernandez, Lisa Moss, Gloria Peruyera, and Dianelys Soutullo for their assistance in the collection and analysis of data.

REFERENCES


