COGNITIVE PSYCHOLOGY 8, 553–560 (1976)

Infants' Intermodal Perception of Events

ELIZABETH SPELKE

Cornell University

Four-month-old infants viewed two sound motion picture films of simple, natural events. The films were projected side by side, as one of the two sound tracks was played through a centrally placed speaker. Infants' visual attention to the films was consistently influenced by what they heard: They looked primarily at the event specified by the sound track. The experiment demonstrates that infants are able to perceive relations between sights and sounds in the absence of spatial cues. They respond to a perceived intermodal invariance with increased attention to the event reaching them over two modalities.

Information from one object or event usually reaches us through several modalities. A falling glass is both seen to break into pieces and heard to crash. A fire is seen to glow, heard to crackle, and felt to radiate heat. To an adult, this information specifies unified objects and events. We perceive one breaking, crashing glass and one warm, crackling fire. Furthermore, when information about an object reaches us in one modality, we are likely to seek more information in other modalities. If we see an interesting object, we will reach for it; if we hear a sudden noise, we will look around.

How did we develop the ability to perceive and explore objects over several modalities? We have surely had to learn what kinds of sounds accompany events like a glass breaking or the burning of wood. We may, or may not, have had to learn what kinds of sounds accompany a visibly talking face, and what kinds of tactual sensations are produced by an object that moves rigidly. Developmental research has focused on two related questions: (a) To what intermodal invariants, if any, are we innately sensitive, and (b) How are we able to discover other multimodal properties of events?

One method for studying the young infant's knowledge about the multimodal properties of objects is to "rearrange" the information available to him over two modalities, and to observe his reactions to these artificial, even conflicting, rearrangements. Bower (1971) used

The author is very grateful to Ulric Neisser and to Eleanor J. Gibson for their many helpful comments on the manuscript, and to Richard W. Evans for his assistance in all phases of the study. This research was supported by a Training Grant to Cornell University from the National Institute of Child Health and Human Development (TO1 HD00381-05). The author was a predoctoral fellow on the same grant. Request reprints from Elizabeth Spelke, Department of Psychology, Cornell University, Uris Hall, Ithaca, NY 14853.

polarized lenses and light to create for an infant the visible impression of an object within his reach. The real object producing this effect was farther away and behind a screen. Bower reported that infants as young as 7 days would reach for the "virtual object"; encountering no object, they would begin to cry. Aronson and Rosenbloom (1971) allowed infants to observe their mothers talking through a window, while they listened to her voice over loudspeakers. The speakers were arranged so that the mother's voice seemed to come first from in front of the infant—from the same location as her face—and then from a source 90° to one side. Infants as young as 1 month were reported to show signs of distress when the face and voice appeared to come from different places. From these studies emerged a set of strong conclusions about the perceptual knowledge of the young infant. Infants appeared to expect a seen object to be touchable in its seen location, and a seen and heard event to emanate from the same point in space. When their expectations were violated, they reacted much as would an adult whose beliefs about the world were shaken.

Although these observations have provoked a great deal of interest, they have not been easy to replicate (McGurk & Lewis, 1974; Condry, Haltom, & Neisser, Note 1). It is indeed not clear why a newborn's perceptual expectations should be so definite and so rigid. It would be of considerable genetic cost, and doubtful adaptive significance, for an infant to be born with a set of well-defined beliefs about the perceptual information he will receive. Nature might, with more economy, endow him with a set of flexible strategies for discovering the multimodal properties of objects and events. Thus, infants might tend innately to explore objects and events in several modalities, whenever they can pick up enough information to guide their exploration.

Earlier experiments have found that infants' explorations of an object can be guided by its perceived spatial location. When the location of an object is specified in one modality, infants often seek information, in the same place, over other modalities. Very young infants will sometimes look in the direction of a sound (Wertheimer, 1961; Wolff, 1966). They also appear to reach in the direction of a nearby object that they see (Bower, 1974). Under most circumstances, their multimodal explorations will yield additional visual and tactual information about objects. Infants will thus discover further relations between what they see, hear, and feel.

The present research focuses on a different source of information about multimodal invariance. The experiment asks whether kinetic and temporal patterns can specify an intermodal relation. Will infants explore sights and sounds whose relationship is specified by their internal structure?

Two studies—an initial experiment and a separate replication—were

undertaken. In both studies, two movies were presented side by side to infant subjects, together with a sound track appropriate to one of them. The sound emanated from a source midway between the films. Visual attention to each of the films was recorded, following the "visual preference method" (Fantz, 1961). The research sought to determine if infants would show a consistent preference for the film which was related to the sound track. Such a preference would indicate that infants can perceive an intermodal invariance based on the internal structure of visual and auditory information. It would further demonstrate that infants will respond to such a perceived invariance with increased exploration of the multimodal event.

PROCEDURE

Sixteen infants, aged 3 months, 24 days to 4 months, 19 days (mean age, 4 months, 7 days), participated in the first experiment, for one 15 min session. On two separate trials, each infant viewed two motion picture films, projected side by side. Both films were made in sound. One depicted a woman playing "peekaboo" continuously: hiding her face with her hands, uncovering her face, and saying "hello baby, peekaboo" to the camera. In the other film, a hand holding a wooden baton struck a wood block and a tambourine repeatedly and rhythmically. Both the rhythm of the percussion instruments and the words spoken by the woman varied slightly (and unsystematically) over the course of the films. Both films were shot with Super-8 color film with a magnetic sound track, and each was 2½ min long.

The films were rear-projected onto the left and right halves of a translucent, 50 × 80 cm screen. Each image was about 33 × 36 cm, and they were about 8 cm apart. The sound track of either film was heard through a speaker placed about 1.5 m behind the center of the screen. Its volume approximated that of normal speech. The baby sat in an infant seat, his head about 40 cm from the screen. An observer monitored his visual fixation through a peephole below the screen, recording the duration of each fixation to the left or right by depressing buttons connected to a Harvard event recorder. When fixation was not clearly to one side or the other, neither button was pressed.

One observer (O_1) recorded the fixations of all infants. In addition, either of two other observers (O_2 and O_3) simultaneously recorded the visual fixations of eight of the infants on 13 trials. Reliabilities of O_1 with O_2 and O_3, calculated as the proportion of seconds on which both observers' records agreed, averaged .93 and .88, respectively. O_2 and O_3 were blind to the lateral position of the sound-related film, but O_1 was not. The present analyses are based on the recordings of O_1.

Each baby viewed the movies once with the peekaboo sound track and once with the music sound track. Order of sound tracks (music first vs. peekaboo first), position of the first sound-related movie (left vs. right), and position of the second sound-related movie (same vs. different from the first) were counterbalanced across infants. Two infants received each of the eight possible combinations of sound orders and film positions.

No baby was excluded from the sample of 16 because of crying, sleepiness, or other such problems. (Two babies, not counted in the sample, had participated in the experiment but were rejected because of equipment failures and experimenter errors.) It was never necessary to interrupt or terminate the films on the first run. Infant crying forced an early termination of the second run in two cases, and a brief interruption in a third case. The data to be reported are based on 16 infants for the first session and on 14 infants for the first and second sessions combined.

RESULTS

Looking Preferences: First Run

Looking preferences were expressed, for each infant, as the proportion of his total fixation which was directed at one of the films. There were no significant preferences for either of the films (peekaboo or music) or for either film position (left or right). Proportions of fixation to the peekaboo and to the right-hand film, regardless of the accompanying sound, were .596 and .579, respectively, $t(15) = 1.23$ and .97, $p > .10$. Proportion of fixation to the film which was related to the sound track was analyzed by a 2×2 analysis of variance (ANOVA), with sound track (peekaboo vs. music) and sound-related film position (left vs. right) as between subjects factors. No effects were significant. Accordingly, subsequent analyses collapse over the various conditions and treat the 16 subjects as a single group.

Mean proportion of fixation to the movie projected in sound was .646, significantly above the chance value of .500, $t(15) = 1.97, p < .05$. Infants spent an average of 88.6 sec looking at the film which was congruent with the sound track, and 47.1 sec fixating the film which was unrelated to the sound track. Examination of their individual records revealed that 11 of the babies looked at least 15 sec longer at the sound-related film than at the other film, while four looked 15 or more sec longer at the film which was unrelated to the sound track.

Looking Preferences: Both Runs

Infants' preference for the movie with a concordant sound track becomes clearer when we examine the fixation of those 14 babies who viewed the films twice. Total preference for the sound-related films was expressed, for each infant, as the mean proportion of fixation to the sound-related film on each of the runs. These proportions averaged .643, a significant departure from chance, $t(13) = 3.83, p < .005$. Infants fixated the sound-related films for a mean total of 167.1 sec; the sound-unrelated films were inspected for 95.9 sec. All but one subject fixated the sound-related films more than .500, and nine babies preferred these films by at least 30 sec. Again, there were no reliable film or side preferences, $t(13) = 1.28$ and .28, $p > .10$.

A further set of analyses tested infants' preferences for the sound-related film separately for each sound track. Eight infants heard each sound track on the first run, and seven listened to each sound track on the second run, yielding preference scores for 15 infants in each film condition. Proportion of fixation to the sound-related film was .739 for the peekaboo sound condition and .555 for the music sound condition. Only the former

is significantly greater than .500, $t(14) = 3.21$, $p < .005$; for the music sound track, $t(15) = .67$, $p > .10$.

Temporal Effects

Changes in patterns of looking during the course of the first run showed that preference for the sound-related film was highest at the end of the 2½ min presentation. When the duration of looking at the sound-related and unrelated films is calculated for each 30-sec period, preference for the sound-related film is seen to rise in the last period. The change over periods was significant by a 2 (peekaboo vs. music sound track) × 2 (fixation to sound vs. silent film) × 5 (periods) ANOVA, $F(4,56) = 3.23$, $p < .05$.

The increased preference was produced primarily by the eight infants hearing the music sound track. Among infants hearing the peekaboo sound track, preference for the peekaboo film was uniformly high; among infants hearing the music sound track, preference for the music film emerged toward the end of the session. This group difference was not significant, $F(4,56) = 1.40$, $p > .10$.

Effects of Variations in the Sound Tracks

Although no direct evidence can be given, this experiment suggested one other source of control by the sound track on patterns of visual attention to the films. Both the peekaboo and music sound tracks involved repetitive rhythmic sequences with occasional novelties: a variation in the musical rhythm or in the woman's words. Both also had some periods of silence between beats or vocalizations. The sound tracks therefore varied over time in their informativeness. It was the impression of the principal observer that babies tended to look away from the sound-related film—and toward the other film—during the pauses in the sound track, and to look back when the sound resumed. If an infant were not looking at the sound-related film when a novelty occurred in the sound track, he seemed especially likely to look back at that time. Unfortunately, the method of recording was not sufficiently precise to permit measurement of the timing of infants' visual fixation as a function of events in the sound track. This finer level of observation may prove useful in future studies of infants' exploration over modalities.

A REPLICATION

The first experiment seemed to indicate that infants' visual attention to filmed events was influenced by the relation of those films to a concurrent sound track. However, there are two problems with this research. One concerns the possible effect of observer bias on the recording of

visual preferences. The other concerns the uncertain preference for the music film when it was projected in sound. Infants hearing the music sounds appeared to prefer the music over the peekaboo film, especially toward the end of the session, but the preference was not statistically significant. A replication was undertaken to resolve these uncertainties.

Procedure

Eight infants, aged 4 months, 1 day to 4 months, 11 days (mean age, 4 months, 5 days) participated in this study. No baby was excluded from the sample for any reason, and every baby completed both sessions of the experiment. The design and procedure were identical to those of the first experiment, with two exceptions. First, the lateral position of the peekaboo and music films did not change from the first to the second session, as it had for half the infants in the first experiment. Two infants received each of the four combinations of sound orders (peekaboo first vs. music first) and film positions (peekaboo right vs. music right). Second, the babies' visual attention to the films was recorded by two observers who were unaware of the lateral position of the sound-related film. One observer (O_2 of the previous experiment) recorded the fixations of all the infants. A second observer (O_4) also recorded the fixations of five infants on eight trials. Reliabilities, calculated as in the first experiment, averaged .94.

Results

Visual preferences were calculated and analyzed as in the first experiment. There were no significant preferences for either film (peekaboo or music) or for either film position (left or right), either during the first session alone or during the two sessions combined. Proportion of fixation to the peekaboo film, regardless of sound, was .551 on the first session and .533 on both sessions; proportion of fixation to the right-hand film was .464 on the first session and .451 on both sessions, all ts < 1. Proportion of fixation to the sound-related film on the first session was subjected to a 2×2 ANOVA which revealed no effect of sound track (peekaboo vs. music) or sound-related film position (left vs. right) on the magnitude of the preference for the sound film. The preference scores for infants in all conditions were again combined.

On the first session, the mean proportion of fixation to the sound-related film was .629, $t(7) = 2.29$, $p < .05$. Infants fixated the film projected in sound for an average of 77.9 sec, and the other film for 44.1 sec. Six of the eight babies looked at least 15 sec longer at the sound-related film; the other two preferred the other film. On both sessions combined, the proportion of fixation to the sound-related film averaged .651, $t(7) = 6.76$, $p < .001$. The total duration of fixation, over both sessions, averaged 153.6 sec to the sound-related film and 87.2 sec to the sound-unrelated film. Every infant's fixation to the sound-related film exceeded .500; six of the infants preferred the sound-related film by 30 sec or more.

Mean proportion of fixation to the peekaboo film when its sound track

was played was .685, significantly greater than .500, $t(7) = 2.47, p < .025$. Mean proportion of fixation to the music film, played in sound, was .681, also significantly above .500, $t(7) = 2.26, p < .05$.

The temporal effects observed in the first experiment were not apparent in this replication. The degree of preference for the sound-related film did not change over the course of the first session, $F(4,24) < 1$. Similarly, there was no effect of sound track (peekaboo vs. music) on the time course of visual preferences, $F(4,24) < 1$.

In summary, the second experiment, with blind observers, replicated quite closely the principal results of the main experiment. Unlike the first experiment, however, the replication failed to find any differences between the peekaboo and music sound track conditions. In the presence of either sound track, infants looked primarily to the sound-related film. The preference emerged quickly in both sound conditions.

DISCUSSION

Four-month-old infants' visual preferences between two films were consistently influenced by the structure of the sounds they heard. Even though the sounds came from a source midway between the films, infants preferred the film which depicted the same event as that conveyed by the sound track. They looked primarily at a movie of a woman when they heard her play "peekaboo." At least in the second study, they also looked primarily at the movements of percussion instruments as they listened to the sounds those instruments made. How were the infants able to detect these intermodal relations? Two possibilities emerge: First, 4-month-old infants may already know that a particular visually witnessed event goes with a particular sequence of sounds. When the subjects heard those sounds, they may have begun immediately to search for a particular kind of object. Alternatively, the infants may have gradually come to perceive the relationships between these sights and sounds from the rhythm and the quality of the visual and auditory stimulation.

This research provides tentative support for both possibilities. In both experiments, infants hearing the peekaboo sound track attended to the appropriate film strongly and with little delay; perhaps these infants knew what kind of object to look for. Preferences for the sound music film were weaker and, for some infants, slower to emerge; perhaps infants needed to discover the intermodal relation in the music film, and did so with varying speed and success. Given the nature of the filmed events—one, involving a familiar and important object; the other, involving relatively uncommon objects—these hypotheses are not implausible. Nevertheless, they remain unproven. Further studies, with auditory–visual events whose multimodal structure and whose familiarity are carefully controlled, are needed to specify the stimulus information and

the perceptual knowledge which guide infants' intermodal explorations. The method developed for the present experiments may prove useful in this effort.

The results of this study support several generalizations about infants' perceptual abilities and attentional strategies. First, infants are able to perceive invariant relations between visual and auditory patterns without using any common spatial information: They are evidently able to coordinate sights and sounds on the basis of their internal structure. Second, infants respond to perceived invariances between sights and sounds by attending to those visible events which they also hear. This attentional pattern may be of considerable adaptive significance. An infant who consistently seeks information across modalities will increasingly come to discover multimodal properties of objects and events. Our ability as adults to perceive unified objects and events amidst a flow of multimodal stimulation may be rooted in this exploratory strategy.

REFERENCES

Aronson, E., & Rosenbloom, S. Space perception in early infancy: Perception within a common auditory—visual space. *Science*, 1971, **172,** 1161–1163.

Bower, T. G. R. The object in the world of the infant. *Scientific American*, 1971, **225,** 30–38.

Bower, T. G. R. *Development in infancy*. San Francisco: W. H. Freeman, 1974.

Fantz, R. L. The origin of form perception. *Scientific American*, 1961, **204,** 66–72.

McGurk, H., & Lewis, M. Space perception in early infancy: Perception within a common auditory–visual space? *Science*, 1974, **186,** 649–650.

Wertheimer, M. Psychomotor coordination of auditory–visual space at birth. *Science*, 1961, **134,** 1692.

Wolff, P. H. The causes, controls, and organization of behavior in the neonate. *Psychological Issues*, 1966, **5** (Monograph No. 17).

REFERENCE NOTE

1. Condry, S., Haltom, M., & Neisser, U. *Space perception with conflicting auditory-visual information: A failure to replicate*. Unpublished manuscript, Cornell University, 1975.

(Accepted December 22, 1975)