Focus Article

Development of cross-modal processing

Christopher W. Robinson and Vladimir M. Sloutsky*

The ability to process and integrate cross-modal input is important for many everyday tasks. The current paper reviews theoretical and empirical work examining cross-modal processing with a focus on recent findings examining infants' and children's processing of arbitrary auditory–visual pairings. The current paper puts forward a potential mechanism that may account for modality dominance effects found in a variety of cognitive tasks. The mechanism assumes that although early processing of auditory and visual input is parallel, attention is allocated in a serial manner with the modality that is faster to engage attention dominating later processing. Details of the mechanism, factors influencing processing of arbitrary auditory–visual pairings, and implications for higher-order tasks are discussed. © 2009 John Wiley & Sons, Ltd. *WIREs Cogn Sci* 2010 1 135–141

Most of our experiences are multi-modal in nature. The objects and events that we encounter in the environment can be seen, touched, heard, and smelled. The fact that the brain can integrate this knowledge into a coherent experience is truly amazing given that each modality simultaneously receives qualitatively different types of input (e.g., photons, molecules, pressure, etc.) and this information is processed, at least in the early stages of processing, by dedicated sensory systems. How do people process and integrate multi-modal information into a unitary experience and how do such abilities change in the course of development? The current paper begins addressing these questions by reviewing some of the theoretical and empirical work examining the development of cross-modal processing, with a focus on the development of auditory–visual integration in humans. Due to space limitations, we will have to ignore cross-modal processing in other sensory modalities and cross-species variability (e.g., many nonhuman mammals have an extraordinary sense of smell, whereas some of the birds have extraordinary vision).

THEORETICAL ACCOUNTS AND EMPIRICAL SUPPORT

There are two traditional accounts concerning the development of cross-modal processing (see Refs 1, 2 for reviews). These accounts focus on two critical dimensions: (1) how integrated the sensory systems are in the early stages of development and (2) how the sensory systems change as a function of development and learning. According to the 'early integration' account,[3–5] the young nervous system is surprisingly multisensory, which allows for very young infants to detect aspects of the environment that are redundant across sensory modalities (e.g., rate, tempo, etc.). For example, the rate that a ball is bouncing can be experienced both visually and auditorily. Young infants ably learn this redundant information, especially when it is presented to multiple modalities.[6,7] Throughout development the modalities become more differentiated as the young infant learns how to parse out details that can only be experienced in a single modality (e.g., color of an object, sound an object makes, etc.). According to the 'late integration' account,[8–10] the sensory systems are initially independent of each other. Throughout development the young child must learn how to process and integrate cross-modal input. Therefore, according to the former account, sensory integration is a starting point of development, whereas, according to the latter account, sensory integration is a product of development.

Empirical findings are surprisingly consistent with both accounts. Let us first consider evidence

---

*Correspondence to: sloutsky.1@osu.edu

Cognitive Development Lab, Center for Cognitive Science, The Ohio State University, Columbus, OH 43210, USA

DOI: 10.1002/wcs.12

that supports the 'early integration' account. It is well documented that infants can equate intensity across modalities,[11] orient to one of two video streams that match an auditory stimulus,[12] visually orient to the location of a sound,[13] associate mother's voice with her face,[14] and make cross-modal associations more generally.[15] They also take into account both auditory and visual input when perceiving speech.[16–18] Early cross-modal integration is not limited to auditory and visual modalities—infants can also visually recognize an object they have previously touched but not seen.[19–21] Perhaps most central for the current review is the fact that infants exhibit facilitation and interference effects, with stimuli presented in one modality either facilitating or hindering processing in a second modality.[22–30] It is important to note that neither facilitation nor interference would be possible if the modalities were completely independent of each other.

At the same time, there is evidence supporting the 'late integration' view: under many conditions intersensory integration can be difficult for young infants (see Ref 2 for a review). For example, while 3- and 4-month-old infants have difficulty detecting temporal synchrony across sensory modalities, they can succeed on similar tasks when visual stimuli are dynamic.[31–34] There is also evidence that young infants have difficulty binding arbitrary auditory–visual pairings, while learning these pairings when auditory input is contingent on infants' looking.[15,35] Even infants as old as 14-months of age can have difficulty binding arbitrarily paired words and objects when visual stimuli are static, however, they can form these associations when visual stimuli are dynamic.[36] These findings suggest that early in development intersensory integration may require special conditions.

There is also support for both accounts concerning the developmental trajectory of cross-modal processing, with the direction of these effects interacting with the type of information that infants are required to learn. In particular, when the goal is to learn amodal information (i.e., information that can be expressed in multiple modalities), presenting this information cross-modally often facilitates learning.[6,7] For example, the rhythm of a bouncing ball can be experienced both visually and auditorily. Infants are more likely to learn amodal information when it is presented cross-modally than when the same information is presented to a single modality.[6] Furthermore, it has been argued that learning of amodal information such as rate and tempo may be governed by innate processes.[37] At the same time, learning of modality-specific information in the presence of cross-modal input appears to be more difficult for young infants.[25,26,38]

There is also support for the idea that modalities may become more integrated with development. First, there is neuroscience evidence indicating that cross-modal processing (across a variety of tasks) requires participation of multiple cortical areas (see Ref 39, for a review). The fact that these areas have a different maturational course provides some support for the 'late integration' account. Second, there is behavioral evidence, particularly in the domain of word learning. Word learning is often considered to be a slow and laborious process in the first year of life, however, young children become much more sophisticated word learners during the second and third years of life (see Refs 40–42 for reviews). It is possible that some of the developments in word learning occur because children are becoming more efficient at processing arbitrary auditory–visual pairings.[30]

In summary, the ability to process and integrate across sensory modalities appears to change considerably across development. However, the specific competencies and the developmental course are not determined by a single variable but are likely to be determined by a set of interacting variables. In what follows, we will outline some of the variables underlying cross-modal processing.

PROCESSING OF ARBITRARY AUDITORY–VISUAL PAIRINGS

Some aspects of the environment can be expressed in multiple modalities (e.g., rate, tempo, rhythm, etc.), and young infants are more likely to learn this information when it is presented cross-modally than when the same information is presented unimodally (see Ref 43 for a review). However, there are also many important cognitive tasks that require young infants to process arbitrary cross-modal pairings. For example, to successfully learn the word 'horse', the young word learner must attend to (and bind) the auditorily presented word with visually presented object. While such a task may seem trivial, processing of arbitrary auditory–visual pairings can be quite difficult for infants and young children.

There are at least two theoretical approaches that can account for these difficulties. First, according to the intersensory redundancy hypothesis (IRH), when infants are presented with cross-modal stimuli, such that each modality expresses the same amodal relation (e.g., rhythm), this amodal information is particularly salient.[43] Infants direct their attention to the amodal information and away from modality-specific information. IRH makes two predictions concerning infants' processing of cross-modal input. First, because amodal information is highlighted when presented cross-modally, cross-modal presentation should facilitate learning of amodal information. Second, because modality-specific information is pushed to the background of attention when it is presented cross-modally, cross-modal presentation should attenuate learning of modality-specific information.

Auditory dominance (AD) can also account for poor processing (and binding) of simultaneously presented auditory and visual input. According to the AD account, which will be discussed in more detail in the next section, because of its dynamic nature, auditory input is particularly salient for infants and young children[25,29] (see also Ref 44). When visual stimuli are paired with auditory input, the auditory input quickly engages attention and attenuates processing of visual input. Thus, according to AD, infants and young children should be more likely to process the details of a visual stimulus when it is presented unimodally than when the same visual stimulus is paired with an auditory stimulus.[25–30]

Both accounts predict conditions under which cross-modal input should attenuate processing. According to IRH, cross-modal input which consists of amodal information pulls attention away from processing of modality-specific information. Therefore, interference effects should be particularly evident when cross-modal input consists of amodal information. According to AD, auditory input attenuates processing of corresponding visual input. Therefore, interference effects should be asymmetrical (i.e., auditory stimuli attenuate visual processing more than visual stimuli attenuate auditory processing), and these interference effects should not be limited to cross-modal stimuli consisting of amodal information. The next section will review some of the empirical evidence supporting AD and factors that influence dominance effects found in infants and young children.

EMPIRICAL SUPPORT FOR AUDITORY DOMINANCE EFFECTS

Evidence for AD comes from studies examining infants', children's and adults' processing of arbitrary auditory–visual pairings.[24–26,29,30] In all of these studies, researchers examined how cross-modal presentation affected processing of auditory and/or visual stimuli. In particular, they compared discrimination of auditory and visual stimuli in cross-modal conditions to the respective unimodal baselines. For example, to determine how auditory input affects visual processing, they focused on discrimination of two visual stimuli when these stimuli were presented unimodally (e.g., V_1 and V_2) and discrimination of the same two visual stimuli when presented cross-modally (e.g., A_1V_1 and A_1V_2). To determine how visual input affects auditory processing, they focused on discrimination of two auditory stimuli when these stimuli were presented unimodally (e.g., A_1 and A_2) and discrimination of the same two auditory stimuli when presented cross-modally (e.g., A_1V_1 and A_2V_1). If cross-modal presentation affects processing of auditory and visual stimuli then discrimination in the unimodal condition should differ from that in the cross-modal condition.

Several important conclusions can be drawn using this general methodology. First, while cross-modal input often attenuated processing of visual stimuli in infants and young children,[25,26,29,30] it had little effect on auditory processing.[29,30] This finding is noteworthy given that infants and young children ably discriminated the visual stimuli when presented unimodally, therefore, it was concluded that the auditory input overshadowed (attenuated processing of) the visual input. Second, compared to the unimodal conditions, presenting auditory and visual stimuli cross-modally had no significant effect on adults' discrimination of auditory or visual pairings. Third, the reported effects are not specifically tied to cross-modal stimuli that are presented in synchrony: auditory input can disrupt visual processing even when visual stimuli are presented for a protracted period of time.[26–28] Finally, these effects are not restricted to low-level discrimination tasks—auditory input can also attenuate performance on higher-order tasks such as categorization and individuation.[27,28]

Factors affecting modality dominance

While auditory input can attenuate visual processing, not all cross-modal stimuli lead to AD. Therefore, to obtain a better understanding of the dynamics of cross-modal processing, it is also important to understand factors that attenuate or even reverse AD effects, with such a reversal resulting in visual dominance. First, processing of arbitrary auditory–visual pairings changes considerably in the course of development, with AD effects decreasing with age.[25,29] Second, stimulus familiarity also plays an important role in modality dominance effects. When young children are given relatively little time to process an auditory–visual pairing, the more familiar component of the cross-modal stimulus interferes with processing of the less familiar component.[24] However, under extended presentation time, increasing the familiarity of the auditory stimulus corresponds with able processing of both modalities.[25,26,30] Specifically, in these studies AD effects were observed when infants were familiarized to unfamiliar sounds and unfamiliar pictures. However, infants ably processed the corresponding visual stimuli when given an opportunity to hear the unfamiliar sounds prior to pairing them with visual stimuli.[26,30] Thus, under short stimulus durations, familiar stimuli interfere with processing of less familiar stimuli, whereas, under longer stimulus durations, familiar stimuli are less likely to interfere with processing in a corresponding modality.[30] We believe this complex set of findings highlights the dynamic nature of cross-modal processing and in what follows we propose a mechanism that can account for such effects.

Mechanisms underlying modality dominance

The initial processing of auditory and visual input is likely to be governed by independent (and parallel) sensory systems. For example, there is no reason to believe that the presence of sound can affect how the rods and cones respond to light. However, it seems reasonable to posit that, while initial processing of auditory and visual input may be implemented simultaneously, infants and young children may allocate attention to auditory and visual input in a serial manner. As can be seen below, such an assumption makes many interesting predictions, including the prediction that cross-modal interference effects should be asymmetrical in nature.

In an attempt to explicate the mechanism and the development of auditory–visual processing, we formulated a set of theoretical considerations pertaining to the allocation of attention in the course of cross-modal processing. The overall idea is that cross-modal processing is determined by two critical parameters: (1) the speed of orienting to a modality relative to the competing modality and (2) the dwell time of attention (i.e., speed of processing) to a given modality relative to the overall duration of stimulus presentation. We suggest that both critical parameters undergo developmental change, and the development of these parameters underlies the ability to process and integrate cross-modal information.

We argue that in the beginning of processing, allocation of attention to a given modality is subject to the same choice processes as allocation of attention to objects in visual search tasks and to stimulus dimensions in categorization tasks (see Ref 45 for a discussion). If this is the case, then choice could be instantiated as a race between two modalities (*cf.*, Logan's Instance Theory of Attention and Memory (ITAM) model). It is likely that the outcome of the race is determined automatically by the system (*cf.*, Refs 46–49), with some stimuli being faster to engage attention, thus being more likely to win the race. During the later stages of processing, infants begin processing the details of the stimuli, however, due to the selective nature of sustained attention (see Refs 50–52 for reviews), it is likely that processing of stimuli in the 'winning' modality will be enhanced whereas processing of stimuli in the 'losing' modality will be attenuated. Furthermore, consistent with Jeffrey's[53] serial habituation hypothesis, processing of the 'losing' modality may not start until attention is disengaged from the 'winning' modality. Therefore, according to this account, modality dominance effects should only be found in the early stages of processing when there is not enough time for the 'winning' modality to release attention. Under longer stimulus durations infants and young children should ably process stimuli in both modalities.

Finally, we believe that several factors may give auditory input a 'leg-up' on visual input, thus, making auditory stimuli more likely to win the race. First, auditory stimuli are often transient whereas visual stimuli are often presented for longer durations. Thus, it may be adaptive to first allocate attention to stimuli that are going to quickly disappear. Second, almost all naturally occurring auditory stimuli are dynamic in nature as they change in pitch and amplitude across time. While some visual stimuli can also be dynamic, many visual stimuli are static for extended periods of time. Thus, it is possible that some of the auditory effects stem from the dynamic nature of auditory stimuli. Third, auditory stimuli are processed faster than visual stimuli in adults,[54] and as a result of early maturation of the auditory system, these differences may be even more pronounced early in development.

The current account makes a number of interesting predictions: (1) cross-modal presentation should attenuate processing in the 'losing' modality, while having little or no effect on processing in the 'winning' modality,[29] (2) cross-modal interference effects should change in the course of processing,[28] (3) stimulus properties that affect the speed of engaging attention and the speed of releasing attention (e.g., modality, familiarity, dynamic, etc.) should affect modality dominance,[24,26,30] and (4) modality dominance effects should decrease with age as processing speed increases.[25,29]. In addition, the current account makes predictions regarding higher-order tasks that hinge on cross-modal processing. For example, it is well documented that words and sounds have different effects on categorization and individuation tasks.[55,56] The current account predicts that words and sounds can both interfere with categorization and individuation, unfamiliar sounds should exert stronger cross-modal interference than words, and that these effects should change in the course of processing.[27,28]

UNRESOLVED ISSUES AND FUTURE DIRECTIONS

One critical issue that remains unresolved is how cross-modal information is integrated in the brain. Much of what is known about this issue comes from single cell recordings in nonhumans and Functional Magnetic Resonance Imaging (fMRIs) in adults (see Refs 39, 57, 58 for reviews). At the subcortical level, the superior colliculus receives auditory, visual, and somatosensory input, and this information is represented by overlapping sensory maps. Multisensory cells can be found deep within the superior colliculus, however, these cells are initially very immature and lack the ability to integrate across the senses (see Ref 58 for a review). While the superior colliculus plays an important role in the motor control of orienting responses, multisensory cells in the lateral temporal cortex, ventral temporal cortex, and frontal cortices are likely to play important roles in higher-level tasks such as recognition of arbitrary auditory–visual pairings. As in the superior colliculus, multisensory cells in the cortex are also relatively late to mature and initially show little evidence of synthesizing across modalities.[59]

Localizing the parts of the brain responsible for integrating across sensory modalities is a critical step in understanding cross-modal integration, however, a number of important questions pertaining to cross-modal processing remain to be answered. For example, it is unclear if mechanisms of cross-modal processing differ for binding various features of a single object (e.g., the color and texture of an apple) versus binding of spatiotemporally different features (e.g., a shape of an object and a word). It is also unclear if brain maturation can account for more efficient cross-modal processing in older children. Is it possible that some of these abilities develop as human multisensory cells gradually learn how to integrate across sensory systems (*cf*., Ref 59)? It will also be important to examine how the brain processes arbitrary auditory–visual pairings and amodal relations consisting of auditory and visual input. Amodal relations may be processed earlier in development because they are processed in different (and possibly more primitive) areas of the brain or because multisensory neurons are more likely to fire when cross-modal information is presented in synchrony.

Finally it will be important to further explore the role of attention and memory in cross-modal integration. The current account of cross-modal processing assumes that auditory and visual processing compete for the same pool of attentional resources early in development and that the modality that is faster to engage attention dominates processing. However, research with adults suggests that the story might be more complex. In particular, findings with adults appear to be task dependent—while there are numerous findings consistent with the claim that modalities are competing for the same pool of attentional resources, other findings suggest that modalities have their own attentional resources (see Ref 60 for a review). We believe that examining how infants process arbitrary auditory–visual pairings may shed light on this issue. Future research will also need to examine how cross-modal stimuli are encoded, stored, and retrieved from long-term memory. Are arbitrary auditory–visual pairings such as word–object relations bound and stored in memory as a single unit or are they stored as separate units? If the former is the case, what are the neural structures underlying binding? And do auditory and visual inputs maintain their independent, modality-specific identities in long-term memory? Answers to these questions will help better understand interrelationships between cross-modal processing, attention, and memory.

CONCLUSIONS

In their daily lives, people encounter information that is simultaneously presented to multiple sensory modalities, and the ability to process such information is crucial for many everyday tasks. Historically, two theoretical accounts pertaining to the development of cross-modal processing have been proposed (see Refs 1, 2 for reviews). These accounts differ on the assumed interconnectedness of the sensory modalities in the early stages of development and how the modalities change across development. While previous findings support both accounts, the current review argues that the type of information that young children are required to learn and how this information is presented plays a significant role in learning of cross-modal input (see also Ref 43). When infants and young children are required to learn information that can be expressed in multiple modalities (e.g., rhythm, rate, etc.), presenting this information cross-modally often facilitates learning. However, when the to-be-learned information is specifically tied to a single modality (e.g., color of an object, sound an object makes, etc.), then cross-modal presentation can often hinder learning. The current review briefly considered some of the empirical findings examining processing of arbitrary auditory–visual pairings, examined some of the factors that are known to affect processing of auditory–visual pairings as well as put forward a mechanism that may account for a complex set of empirical findings. While future research is needed, it is likely that such a mechanism can account for many of the reported difficulties found in processing of arbitrary auditory–visual pairings and may ground many sophisticated behaviors in the dynamics of cross-modal processing.

NOTES

This research has been supported by grants from the NSF (BCS-0720135) and from the US Department of Education (R305H050125 and R305B070407) to V. M. S. and from the NIH (RO3HD055527) to C. W. R.

REFERENCES

1. Lewkowicz DJ. Development of intersensory perception in human infants. In: Lewkowicz DJ, Lickliter R, eds. *The Development of Intersensory Perception: Comparative Perspectives*. Hillsdale, NJ: Erlbaum; 1994, 165–203.
2. Lewkowicz DJ. The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol Bull* 2000, 126:281–308.
3. Bower TGR. *Development in Infancy* San Francisco: Freeman; 1974.
4. Gibson JJ. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin; 1966.
5. Werner H. *Comparative Psychology of Mental Development* New York: International Universities Press; 1973.
6. Bahrick LE, Flom R, Lickliter R. Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Dev Psychobiol* 2002, 41:352–363.
7. Bahrick LE, Lickliter R. Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol* 2000, 36:190–201.
8. Birch HG, Lefford A. Intersensory development in children. *Monogr Soc Res Child Dev* 1963, 25:1–47.
9. Birch HG, Lefford A. Visual differentiation, intersensory integration, and voluntary motor control. *Monogr Soc Res Child Dev* 1967, 32:1–87.
10. Piaget J. *The Origins of Intelligence in Children*. New York: International Universities Press; 1952.
11. Lewkowicz DJ, Turkewitz G. Cross-modal equivalence in early infancy: auditory-visual intensity matching. *Dev Psychol* 1980, 16:597–607.
12. Spelke ES. Infants' intermodal perception of events. *Cogn Psychol* 1976, 8:553–560.
13. Muir DW, Clifton R. Infants' orientation to the location of sound sources. In: Gottlieb G, Krasnegor N, eds. *The Measurement of Audition and Vision During the First Year of Life: A Methodological Overview*. Norwood, NJ: Ablex; 1985, 171–194.
14. Spelke ES, Owsley C. Intermodal exploration and knowledge in infancy. *Infant Behav Dev* 1979, 2:13–27.
15. Slater A, Quinn PC, Brown E, Hayes R. Intermodal perception at birth: intersensory redundancy guides newborn infants' learning of arbitrary auditory-visual pairings. *Dev Sci* 1999, 3:333–338.
16. Hollich G, Newman R, Jusczyk P. Infants use of synchronized visual information to separate streams of speech. *Child Dev* 2005, 76:598–613.
17. Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. *Science* 1982, 218:1138–1140.
18. Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. *Percept Psychophys* 1997, 59:347–357.
19. Gottfried AW, Rose SA, Bridger WH. Cross-modal transfer in human infants. *Child Dev* 1977, 48:118–123.
20. Meltzoff AN, Borton RW. Intermodal matching in human neonates. *Nature* 1979, 282:403–404.
21. Streri A. Tactile discrimination of shape and intermodal transfer in 2- to 3-month-old infants. *Br J Dev Psychol* 1987, 5:213–220.
22. Lewkowicz DJ. Sensory dominance in infants: 1. Six-month-old infants' response to auditory-visual compounds. *Dev Psychol* 1988, 24:155–171.
23. Lewkowicz DJ. Sensory dominance in infants: 2. Ten-month-old infants' response to auditory-visual compounds. *Dev Psychol* 1988, 24:172–182.
24. Napolitano AC, Sloutsky VM. Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Dev* 2004, 75:1850–1870.
25. Robinson CW, Sloutsky VM. Auditory dominance and its change in the course of development. *Child Dev* 2004, 75:1387–1401.
26. Robinson CW, Sloutsky VM. Visual processing speed: effects of auditory input on visual processing. *Dev Sci* 2007, 10:734–740.

27. Robinson CW, Sloutsky VM. Linguistic labels and categorization in infancy: do labels facilitate or hinder?. *Infancy* 2007, 11:233–253.
28. Robinson CW, Sloutsky VM. Effects of auditory input in individuation tasks. *Dev Sci* 2008, 11:869–881.
29. Sloutsky VM, Napolitano A. Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Dev* 2003, 74:822–833.
30. Sloutsky VM, Robinson CW. The role of words and sounds in visual processing: from overshadowing to attentional tuning. *Cogn Sci* 2008, 32:354–377.
31. Bahrick LE. Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behav Dev* 1983, 6:429–451.
32. Bahrick LE. Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. *Child Dev* 1988, 59:197–209.
33. Bahrick LE. Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *J Exp Child Psychol* 1992, 53:180–199.
34. Lewkowicz DJ. Infants' response to temporally based intersensory equivalence: the effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behav Dev* 1992, 15:297–324.
35. Gogate LJ, Bahrick LE. Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *J Exp Child Psychol* 1998, 69:133–149.
36. Werker JF, Cohen LB, Lloyd VL, Casasola M, Stager CL. Acquisition of word–object associations by 14-month-old infants. *Dev Psychol* 1998, 34:1289–1309.
37. Slater A, Kirby R. Innate and learned perceptual abilities in the newborn infant. *Exp Brain Res* 1998, 123:90–94.
38. Bahrick LE, Lickliter R. Intersensory redundancy guides early perceptual and cognitive development. In Kail R, ed. *Advances in Child Development and Behavior*. New York: Academic Press; 2002, 30: 153–187.
39. Calvert GA. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 2001, 11:1110–1123.
40. Hollich G, Hirsh-Pasek K, Golinkoff R. Breaking the language barrier: an emergentist coalition model of word learning. *Monogr Soc Res Child Dev* 2000, 65:1–123.
41. Regier T. The emergence of words: attentional learning in form and meaning. *Cogn Sci* 2005, 29:819–865.
42. Woodward AL, Markman EM. Early word learning. In: Damon W, Kuhn D, Siegler R, eds. *Handbook of Child Psychology, Volume 2: Cognition, Perception and Language*. New York: John Wiley & Sons; 1998, 371–420.
43. Bahrick LE, Lickliter R, Flom R. Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Curr Dir Psychol Sci* 2004, 13:99–102.
44. Sloutsky VM. The role of similarity in the development of categorization. *Trends Cogn Sci* 2003, 7:246–251.
45. Logan GD. An instance of theory of attention and memory. *Psychol Review* 2002, 109:376–400.
46. Graham FK. Distinguishing among orienting, defense, and startle reflexes. In: Kimmel HD, van Olst EH, Orlebeke JF, eds. *The Orienting Reflex in Humans*. Hillsdale, NJ: Erlbaum; 1979, 137–167.
47. Graham FK, Clifton RK. Heart-rate change as a component of the orienting response. *Psychol Bull* 1966, 65:305–320.
48. Richards JE, Casey BJ. Heart rate defined phases of infant visual information processing. *Psychophysiology* 1991, 28:43–53.
49. Sokolov EN. *Perception and the Conditioned Reflex* New York: Macmillan; 1963.
50. Berg WK, Richards JE. Attention across time in infant development. In: Lang PJ, Simons RF, Balaban MT, eds. *Attention and Orienting: Sensory and Motivational Processes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 1997, 347–368.
51. Richards JE. Attention in young infants: a developmental psychophysiological perspective. In: Nelson CA, Luciana M, eds. *Handbook of Developmental Cognitive Neuroscience*. Cambridge, MA: MIT Press; 2001, 321–338.
52. Richards JE. Development of covert orienting in young infants. In: Itti L, Rees G, Tsotsos J, eds. *Neurobiology of Attention*. Academic Press/Elsevier; 2004, Chapter 14, 82–88.
53. Jeffrey WE. The orienting reflex and attention in cognitive development. *Psychol Rev* 1968, 75:323–334.
54. Green DM, von Gierke SM. Visual and auditory choice reaction times. *Acta Psychol* 1984, 55:231–247.
55. Balaban MT, Waxman SR. Do words facilitate object categorization in 9-month old infants? *J Exp Child Psychol* 1997, 64:3–26.
56. Xu F. The role of language in acquiring object kind concepts in infancy. *Cognition* 2002, 85:223–250.
57. Amedi A, Von Kriegstein K, Van Atteveldt N, Beauchamp MS, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 2005, 166:559–571.
58. Wallace MT. The development of multisensory processes. *Cogn Process* 2004, 5:69–83.
59. Wallace MT, Carriere BN, Perrault TJ, Vaughan JW, Stein BE. The development of cortical multisensory integration. *J Neurosci* 2006, 26:11844–11849.
60. Burr D, Alais D. Combining visual and auditory information. *Prog Brain Res* 2006, 155:243–258.