

## REVIEW ARTICLE

# Semantic confusion regarding the development of multisensory integration: a practical solution

Barry E. Stein,<sup>1</sup> David Burr,<sup>2</sup> Christos Constantinidis,<sup>1</sup> Paul J. Laurienti,<sup>3</sup> M. Alex Meredith,<sup>4</sup> Thomas J. Perrault Jr,<sup>1</sup> Ramnarayan Ramachandran,<sup>1</sup> Brigitte Röder,<sup>5</sup> Benjamin A. Rowland,<sup>1</sup> K. Sathian,<sup>6</sup> Charles E. Schroeder,<sup>7</sup> Ladan Shams,<sup>8</sup> Terrence R. Stanford,<sup>1</sup> Mark T. Wallace,<sup>9</sup> Liping Yu<sup>10</sup> and David J. Lewkowicz<sup>11</sup>

<sup>1</sup>Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Winston-Salem, NC 27157-1010, USA

<sup>2</sup>Department of Psychology, Università degli Studi di Firenze, Florence, Italy

<sup>3</sup>Department of Radiology, Wake Forest University School of Medicine, Winston-Salem, NC, USA

<sup>4</sup>Department of Anatomy and Neurobiology, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

<sup>5</sup>Biological Psychology and Neuropsychology, University of Hamburg, Hamburg, Germany

<sup>6</sup>Departments of Neurology, Rehabilitation Medicine, and Psychology, Emory University, and Rehabilitation R&D Center of Excellence, Atlanta VAMC, Atlanta, GA, USA

<sup>7</sup>Cognitive Neuroscience and Schizophrenia Program, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA

<sup>8</sup>Department of Psychology, University of California, Los Angeles, CA, USA

<sup>9</sup>Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA

<sup>10</sup>Institute of Cognitive Neuroscience, East China Normal University, Shanghai, China

<sup>11</sup>Department of Psychology, Florida Atlantic University, Boca Raton, FL, USA

**Keywords:** amodal, crossmodal, intersensory, multimodal, supramodal

## Abstract

There is now a good deal of data from neurophysiological studies in animals and behavioral studies in human infants regarding the development of multisensory processing capabilities. Although the conclusions drawn from these different datasets sometimes appear to conflict, many of the differences are due to the use of different terms to mean the same thing and, more problematic, the use of similar terms to mean different things. Semantic issues are pervasive in the field and complicate communication among groups using different methods to study similar issues. Achieving clarity of communication among different investigative groups is essential for each to make full use of the findings of others, and an important step in this direction is to identify areas of semantic confusion. In this way investigators can be encouraged to use terms whose meaning and underlying assumptions are unambiguous because they are commonly accepted. Although this issue is of obvious importance to the large and very rapidly growing number of researchers working on multisensory processes, it is perhaps even more important to the non-cognoscenti. Those who wish to benefit from the scholarship in this field but are unfamiliar with the issues identified here are most likely to be confused by semantic inconsistencies. The current discussion attempts to document some of the more problematic of these, begin a discussion about the nature of the confusion and suggest some possible solutions.

## Introduction

Recently, an article in the New York Times (by Natalie Angier, July 20, 2009) identified a significant problem among behavioral biologists: they can't seem to agree on the definition of 'behavior'. The issue seemed to become even more humorous when botanists entered the fray with questions about whether seed dispersal and phototropism qualify as behaviors.

In fact, the apparent humor masks a serious issue that is at the heart of what we do as scientists. Whatever phenomenon we choose to study, first and foremost we must define it. To do so we must deal with the issue of semantics because different definitions lead to different kinds of research questions. These, in turn, determine how we

construct conceptual frameworks that help us understand those phenomena. One celebrated example of the central importance of semantics is the exchange between Daniel Lehrman and Konrad Lorenz in which they vehemently defended their respective interpretations of the emergence of novel behaviors at birth in the context of the nature–nurture debate. This is a good example to keep in mind, because it illustrates the impact of semantic differences and the difficulty in reaching consensus when it comes to terms with which we have become comfortable. Lehrman (1970) argued that the failure to resolve the nature–nurture debate was largely due to semantic issues related to the way in which developmental scientists on the one hand, and ethologists on the other, interpreted the meaning of particular terms that were key to their theories.

Semantic issues often crop up as we gain sophistication about biological issues and must categorize phenomena, presumably to

Correspondence: Dr Barry E. Stein, as above.

E-mail: bestein@wfsbmc.edu

Received 1 October 2009, revised 21 December 2009, accepted 22 December 2009

enhance the clarity of our understanding. Sophistication about multisensory integration is growing more rapidly now than ever before, and we, like those in other fields, must face the issue of how semantic differences are influencing our discussions. Like behavioral scientists, we have to agree even on the meaning of the term that defines our field. For many neurophysiologists (e.g. see Stein *et al.*, 2009a for a recent discussion), the term ‘multisensory integration’ has a very specific meaning. In contrast, for many behavioral scientists this term has a considerably broader meaning, and for those outside these two domains it has an even broader one. As a result, there is the potential for compromising the clarity of our discussions and the design of the experiments that result from them.

### Using the same terms differently

The semantic confusion concerning the use of ‘multisensory integration’ is perhaps most apparent when considering the development of multisensory processes: an issue of great significance to those interested in how the functional properties of sensory systems become instantiated during early life. Generally, most

developmental models and discussions of these processes are centered on responses to the cross-modal combination of vision and hearing, although studies concerned with the effects of other cross-modal combinations (e.g. vision–somatosensation, audition–somatosensation, vision–vestibular processes, cross-talk between the different special senses and among the different chemical senses, and trisensory information processing) have also been considered by various investigators. The semantic difficulties arise when the term ‘multisensory integration’ is used to refer to different neural processes by investigators whose perspectives and methods are quite different because they use different models and have different assumptions. Thus, it is helpful to examine the approaches and assumptions used by neurophysiologists who use animal subjects and cognitive scientists who use human subjects (though these do not really form discrete groups).

### Animal studies

Neurophysiologists examine the development of the brain’s ability to integrate visual and auditory information by investigating neural

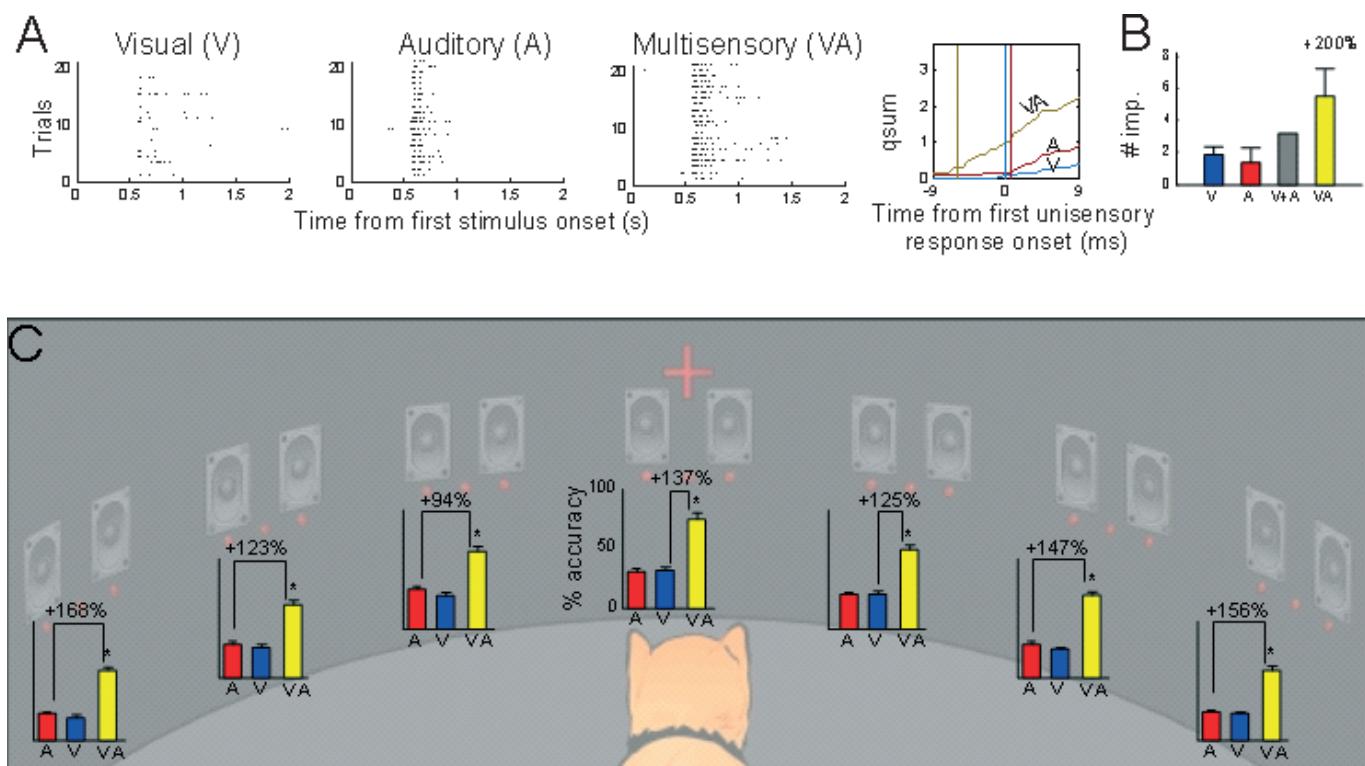


FIG. 1. Multisensory integration at the level of the single superior colliculus (SC) neuron and its manifestation in SC-mediated orientation behavior. (A) This visual-auditory SC neuron was recorded from during the presentation of repeated and interleaved trials consisting of visual (LED), auditory (broadband noise burst) and visual-auditory stimuli. The responses from these trials were then resorted and presented in the three raster displays at the left. Each dot in the raster display represents one impulse, and each row responses to one stimulus presentation. Trials in each display are ordered bottom to top. The middle line graph figure illustrates the mean cumulative impulse count for each response and shows that the cross-modal stimulus also produced a shortening (7 ms) of the response latency, and the computation was clearly superadditive in its initial phase, although this amplification was less obvious later in the response. (B) In this example neuron, overall unisensory response efficacy was sufficiently low to yield a superadditive computation in the response magnitude averaged over the entire response duration. Both figures obtained from data published in Stanford *et al.* (2005). (C) Cats were tested on an orientation/approach task in a 90-cm-diameter perimetry apparatus containing a complex of LEDs and speakers separated by 15°. Each complex consisted of three LEDs and two speakers. Trials consisted of randomly interleaved modality-specific stimuli (a single visual or auditory stimulus) and cross-modal stimuli (a visual-auditory stimulus pair) at each location between ± 45°, as well as ‘catch’ trials in which no stimulus was presented (the animal remained still in response). At every spatial location, multisensory integration produced substantial performance enhancements (94–168%; mean, 137%) that exceeded performance in response to the best modality-specific component stimulus. Errors bars indicate the SEM response accuracy computed over multiple experimental days. Asterisks indicate comparisons that were significantly different ( $\chi^2$  test,  $P < 0.05$ ). In addition, errors (No-Go and Wrong Localization) were significantly decreased as a result of multisensory integration (not shown). Modified from Gingras *et al.* (2009).

activity (i.e. signals generated by individual multisensory neurons or groups of neurons) in the neonates of various animal species.

The results from neurophysiological studies in animals have been used in support of the idea that the development of an ability to engage in multisensory integration requires a prolonged postnatal period because its emergence is dependent, in large part, on sensory experience (Wallace & Stein, 1997, 2001, 2007; Stein, 2005; Yu *et al.* 2010). During early life the brain is exposed to multiple iterations of events that provide cross-modal cues with similar statistics (they generally occur in spatiotemporal concordance). The effects of these sensory experiences appear to be cumulative and to be incorporated gradually into the nervous system. They are then reflected as a capacity to integrate those multiple sensory inputs based on the likelihood that they are derived from the same event or from different events. In the case of the former, the cross-modal stimuli now significantly enhance responses, as shown in Fig. 1A and B. In the case of the latter, there is either no significant change in the response or there is a significantly depressed response. The physiological changes in multisensory responses will then affect behavioral responses to cross-modal events

as shown in Fig. 1C. Results from a developmental study of the emergence of this neurophysiological property are depicted in Fig. 2.

### Human infant studies

Perhaps the best-known method for studying multisensory processing in human infants is the preferential-looking technique, which is designed to assess the perception of equivalencies (or differences) across stimuli from different senses (Spelke, 1976). Typically, infants are presented with side-by-side films of different objects engaged in different actions and a sound that corresponds to one of them is presented at the same time. Usually, infants prefer to look at the object whose actions correspond to the sound, implying that they perceive the audio-visual relationship. The original version of the paired-preference cross-modal matching procedure presents the auditory stimulus concurrently with the visual stimuli and can be seen in Fig. 3. Subsequent studies have used a modified version of this procedure. Here, infants are familiarized with a stimulus in one sensory modality for a short period of time (usually 30–60 s) and then are given

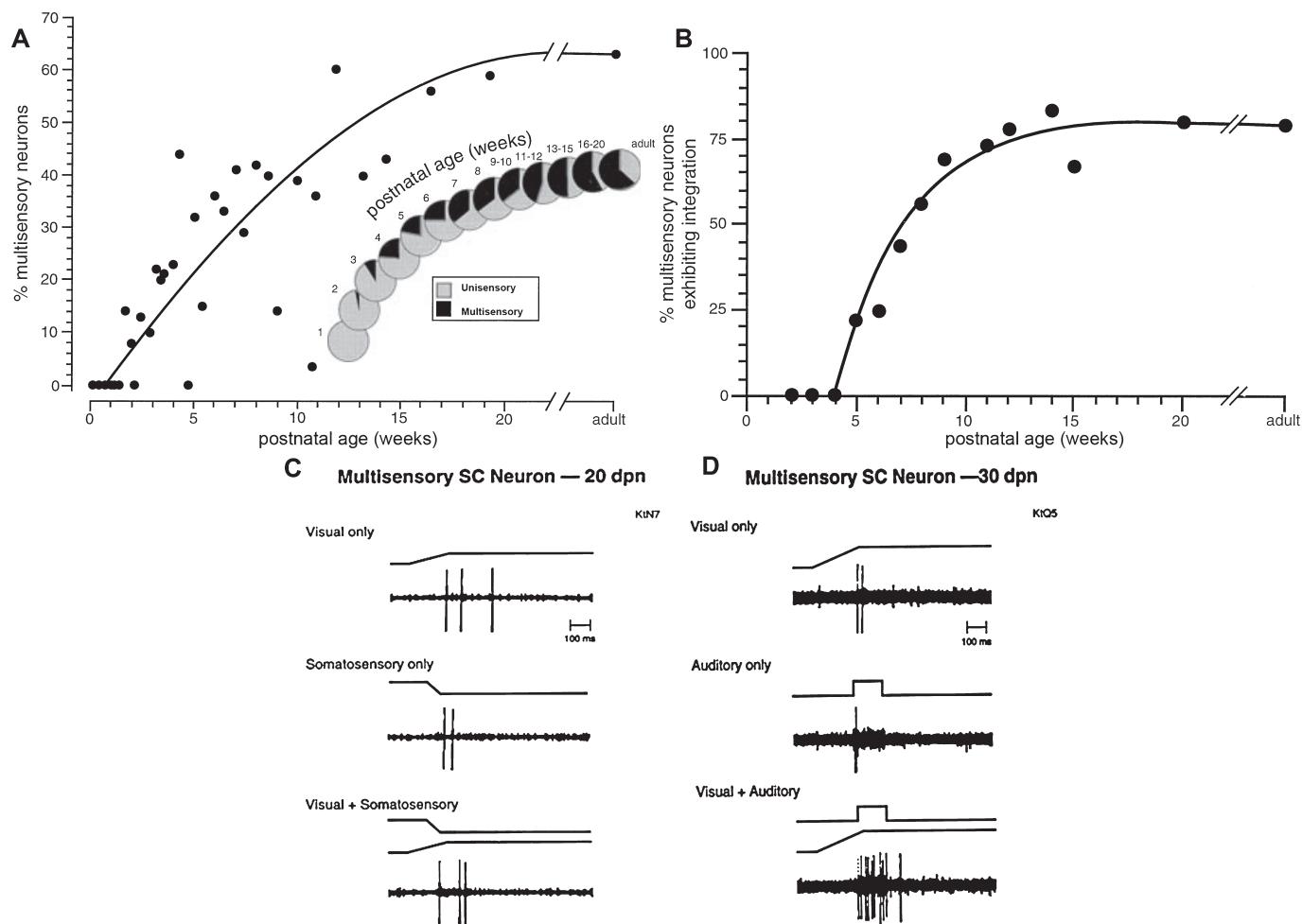


FIG. 2. The development of multisensory neurons and of multisensory neurons capable of multisensory integration takes place over a protracted postnatal period. (A) The increase in cat superior colliculus (SC) multisensory neurons as a function of postnatal age is plotted on the left. The insert shows this as an increasing proportion of the sensory-responsive neurons in the multisensory layers of the structure. (B) The plot on the right shows the late appearance and gradual increase in the proportion of neurons capable of integrating their different sensory inputs. (C and D) Two exemplar multisensory neurons are shown. In all cases the stimuli were presented within a neuron's excitatory receptive fields. The one on the left was recorded in a 20-day-old animal and was incapable of multisensory integration. It typified the immature state in which the response (number of impulses) to the cross-modal stimulus is no greater than the response to the most effective modality-specific component stimulus. In contrast, at 30 days of age some neurons were capable of multisensory integration. The neuron on the right shows that the multisensory response consisted of significantly more impulses than the response to the visual stimulus. Modified from Wallace & Stein (1997).



FIG. 3. The paired-preference cross-modal matching procedure. In the original version of the procedure, infants are seated in front of two side-by-side visual stimuli. These can be faces or objects that can be either static or moving. A sound that corresponds to one of the visual stimuli is presented concurrently. In the picture depicted here, the woman is seen producing two different utterances and the auditory track, which is presented concurrently through centrally placed speakers, corresponds to one of the visible utterances. The infant's looking is monitored via a camera placed in the center. Typically, multiple such trials, each lasting anywhere between 30 s and 1 min, are given and the dependent measure is the amount of time the infant spends looking at each visual stimulus. Matching is inferred when looking at the corresponding visual stimulus is greater than looking at the non-matching one.

paired-preference visual test trials to determine whether they will shift their looking toward the matching visual stimulus (see, e.g. Fig. 4 and Pons *et al.*, 2009). A third technique that has been used to study multisensory processing in infants is the habituation–test procedure (Lewkowicz & Turkewitz, 1980). Here, infants are habituated to a stimulus in one sensory modality and then are presented either with a single stimulus or several different stimuli in another sensory modality. If they perceive the relationship between the habituation stimulus and stimuli in the other modality, they do not exhibit response recovery to those stimuli (because they are equivalent and, thus, familiar), but they do exhibit response recovery to the other stimuli in the other modality (presumably because they are not equivalent and, thus, novel).

Although not all studies of multisensory processing in human infants use preferential looking (for discussion of the habituation–test procedure in multisensory processing, see Lewkowicz, 2002), many do. As noted above, these studies are designed to assess the perception of cross-modal stimulus equivalencies (or differences). In general, published reports have been explicit and careful about the possible inferences that may be drawn given a particular variant of the preferential-looking task. Nonetheless, because the task is inherently multisensory, it is often assumed that cross-modal matching depends on a process of multisensory integration (though a variety of other and equivalent terms have been used as well). Thus, the appearance of cross-modal matching capability early during human infant development

is seemingly inconsistent with the protracted developmental time course observed for the appearance of multisensory integrative capacity in the animal model as assessed at the level of the single neuron. In fact, the conflict is more apparent than real. As discussed below, the relatively common usage of the term ‘multisensory integration’ to refer to all manner of multisensory phenomena tends to obscure the fact that results from human infant studies and animal studies probably reflect the operation of very different underlying processes.

The human infant studies suggest a somewhat more complex developmental picture that incorporates what might be seen as multisensory integration and/or other multisensory processes. On the one hand, like the animal neurophysiological studies, some human infant studies indicate that postnatal experience contributes heavily to the development of multisensory integration (see Neil *et al.*, 2006; Putzari *et al.*, 2007; Bremner *et al.*, 2008a, 2008b; Gori *et al.*, 2008; Lewkowicz & Ghazanfar, 2009). Unfortunately, such data are relatively rare simply because it is difficult to directly manipulate experience in human infants. On the other hand, some human infant studies suggest a developmental pattern wherein some low-level multisensory capabilities appear to be present at birth or emerge shortly thereafter, whereas higher-level multisensory processes emerge much later in infancy. For example, the perception of audio-visual synchrony and cross-modal equivalence of intensity, which does not require the extraction of complex cross-modal relations, has been

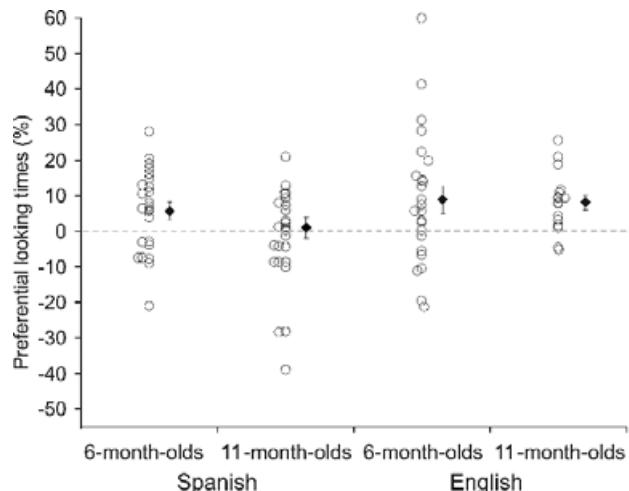


FIG. 4. The familiarization/matching procedure used to study the narrowing of audio-visual speech perception in infancy. Here, during the first two baseline trials, infants saw side-by-side faces of the same person repeatedly uttering a silent/ba/syllable on one side and a silent/va/syllable on the other side for a total of 42 s (with the side of syllables switched after 21 s). During the remaining four test trials, two auditory familiarization trials were interspersed with two test trials to determine whether hearing one of the syllables would shift visual preferences toward the matching visual syllable on the subsequent test trial. Half the infants heard the/ba/syllable and the other half heard the/va/syllable during the two familiarization trials. During each of the two silent test trials that followed each familiarization trial, infants viewed the two visual syllables presented side-by-side (counterbalanced for side across these two test trials) again. Percentage looking times directed to the matching visual syllable were computed separately for the baseline and the test trials, and shown in the figure are the differences between these two scores, with a positive value meaning that infants increased their looking at the matching visual syllable following familiarization to the auditory syllable (open circles represent each infant's difference score and black circles with error bars represent the mean difference score and the SEM for each group). Here, it can be seen that Spanish-learning, monolingual 6-month-old infants exhibited significantly greater looking at the matching visible syllable despite the fact that this phonemic distinction does not exist in Spanish, but that 11-month-old Spanish-learning infants no longer did. This finding indicates that the ability to make cross-modal matches of non-native phonemes is initially present in infancy and that it subsequently declines when experience with this phonetic distinction is absent. In contrast, English-learning infants, who have experience with this phonetic distinction, made cross-modal matches at both ages. From Pons *et al.* (2009).

shown to be present at birth or shortly thereafter (Lewkowicz & Turkewitz, 1980, 1981; Lewkowicz *et al.*, 2010), whereas the perception of higher-level amodal relations, such as affect or gender, emerges many months later (Walker-Andrews, 1986; Patterson & Werker, 2002). The emergence of the latter ability may depend on experience, and may coincide with the appearance of true multisensory integrative processes.

### Defining terms and their underlying computations

As noted above, it is likely that the human infant testing procedures access different underlying processes than those implicated in neurophysiological studies of multisensory integration. From the physiological perspective, integration refers to the creation of a true product (i.e. neural signal) derived from an interaction among two or more different sensory inputs (e.g. see Stein & Meredith, 1993). That integrated neural signal is different (e.g. bigger, smaller, having a different temporal evolution) than each of the presumptive component responses, and cannot readily be deconstructed to yield the unique contributions to its formation. There are many multisensory processes

that do not involve such a product and for which the inferences that are drawn are quite different. For example, the cross-modal matching studies discussed above seek the identification of equivalencies in two objects, scenes or events and, unlike studies of neurophysiological multisensory integration, they require the preservation of the characteristics of the stimulation in each modality. Only then can they be matched.

One way of looking at these differences is to ask about their underlying computations. The key feature in multisensory integration is that the multisensory response is significantly different (e.g. larger, smaller) from the best component response. An integrated multisensory response may be identified by the inequality  $[(C_1C_2) \neq \max(C_1, C_2)]$ , where  $C_1C_2$  refers to the actual multisensory response that is recorded, and  $C_1$  and  $C_2$  refer to the presumptive individual component responses (determined in separate or interleaved tests by presenting the modality-specific stimuli individually). In the case of cross-modal matching, the operation may be described as an evaluation of the equality  $[C_1 = C_2]$ . Both processes are critical features of normal brain function but do not share the same computational platform and could have very different developmental time courses.

Another way of appreciating this is to recognize that cross-modal matching tests whether an association between cross-modal features has been established. This association may occur with any two arbitrary correlated features, or through a common amodal variable. Computationally, if two sensory features are cues to the same amodal variable (e.g. gender, intensity, motion, etc.), these features are statistically dependent and thus associated. While association between any two features in different senses is necessary for multisensory integration, it is not sufficient. For example, the nervous system may quickly learn that when an object moves it makes a sound. However, it may take much longer for the nervous system to learn how to combine the two sources of information in order to obtain a more precise estimate of the speed and direction of an object's motion. While the former type of learning may require a simple associative mechanism, the latter may require the learning of how to encode the uncertainty of each sensory estimate and how to combine the two estimates given their uncertainty. Therefore, it is not unlikely that the two processes would have different developmental onsets and trajectories.

### How to evaluate the comparative development of different multisensory processes

One factor that hinders our understanding of the relative development of each of the distinct multisensory processes under discussion is the diversity of experimental approaches. There is no common experimental model. Although there are some notable exceptions (e.g. neural recordings during human brain surgery), studies of single neurons generally necessitate studies in animals. In these studies, the age of the subject and its experience can be controlled in a way that is impractical (unethical) in studies of human infants. This is an extremely important issue because very short periods of cross-modal experience can have substantial consequences for multisensory processes (Yu *et al.*, 2009, 2010).

This underscores the difficulty in evaluating the impact of prior prenatal and/or postnatal experience on the results of cross-modal matching, or any other multisensory studies in human infants. Nevertheless, this is an exceedingly important issue. One animal model that ostensibly provides a parallel for these studies in human infants, and may be helpful in this context by permitting the investigation of the neurophysiological basis for cross-modal matching, is the monkey. Studies in monkeys have revealed distinct neural

correlates of cross-modal matching that can be assessed using single-neuron neurophysiological techniques.

Monkeys may be trained to remember a visual stimulus (e.g. a letter or oriented grating) and operate a lever when presented with a tactile stimulus that matches it, and *vice versa*. Neurons in the somatosensory cortex (area SII) and visual cortex (area V4) have been shown to discharge more vigorously when a stimulus in one modality appears to match the stimulus in the other modality (Maunsell *et al.*, 1991; Hsiao *et al.*, 1993). In these experiments, the tactile and visual stimuli were presented in sequence, with an intervening delay period. The visual cue did not produce action potentials in SII neurons and the tactile cue did not evoke action potentials in V4 neurons, nor was there any sign of elevated ongoing neural discharges prior to the appearance of the match stimulus. Taken at face value, these findings suggest that the phenomenon was not dependent on the integration of the two signals in the same neuron. Instead, the monkey was able to form an expectation about which stimulus of a different modality would constitute a match. Appearance of the match caused elevated neuronal activity in sensory cortex compared to the identical stimulus presentation when it was not a match. An alternate approach used in monkeys is to train them to use a rule, such as finding the deviant or ‘oddball’ stimulus to identify behavioral targets regardless of modality (Lakatos *et al.*, 2008). In both cases, controlled processing is likely to involve prefrontal cortex. Prefrontal neurons respond selectively to auditory and visual stimuli that have been associated with each other, and remain active during the delay interval that bridges the cue and match presentation, in a sound–color matching task (Fuster *et al.*, 2000). Moreover, prefrontal neurons clearly participate in top-down control of processing, whether it is stimulus- or rule-based (Buschman & Miller, 2009).

As noted above, this technique is an ‘apparent’ parallel to the cross-modal matching technique used with human infants. However, care must be taken here to recognize and control for a potential problem: that an arbitrary association between stimulus 1 and stimulus 2 can also be learned in this circumstance, so that the animal may not be matching what he sees with what he feels as ‘equivalent’ stimulus attributes in different modalities. However, investigators have already begun using the procedures established for examining human infant

cross-modal matching to study cross-modal matching in adult monkeys (Ghazanfar & Logothetis, 2003) and young monkeys (see Zangenehpour *et al.*, 2009; and Fig. 5 for description of this method and results). Adapting the preferential-looking technique to the infant monkey (or any other amenable animal species) would provide a corollary method with which to examine the neural and behavioral manifestations of early cross-modal matching capabilities and make it possible to draw comparisons with methods used to study the development of multisensory integration.

Moreover, the use of cross-modal matching procedures in animals makes it possible to obtain additional information about the development of other multisensory processes that cannot easily be obtained from human infants. Rapid changes in multisensory processes could easily be missed during early development if experience is not tightly controlled. Indeed, the sensitivity of multisensory neurons to short-term experiences is proving to be remarkably acute at all stages of development (see Stein *et al.*, 2009b; Yu *et al.*, 2009, 2010). Approaches to controlling for this possibility in experiments with animals have included delaying the onset of experience with cross-modal cues, or manipulations of the statistics of the sensory environment (see Wallace & Stein, 2007; Yu *et al.*, 2010).

There is, however, a substantial practical problem to behavioral studies in infant monkeys. Rapid fatigue in a particular task makes the accumulation of many trials in a given session quite difficult. This then requires many sessions, between which the animal would have to be deprived of gaining additional experience with cross-modal stimuli. This problem has been solved in cats by maintaining them in a dark room except for the testing sessions. This solution is readily adaptable to studies of monkeys.

### A practical immediate solution

To deal best with our current semantic problem, it might be most advantageous to refer to our field not as one that investigates ‘Multisensory Integration’ but rather one that investigates ‘Multisensory Processing.’ In addition, it might be best to acknowledge that the latter term probably captures the operation of more than a single underlying perceptual or neural process and may take many forms. As such, the term ‘Multisensory Processing’ would be interpreted as a

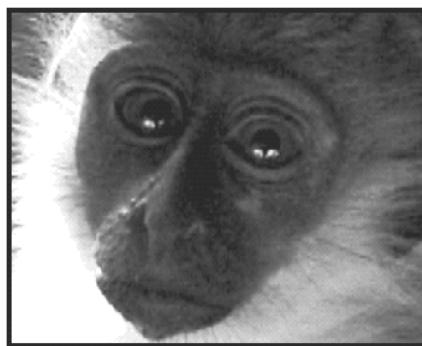
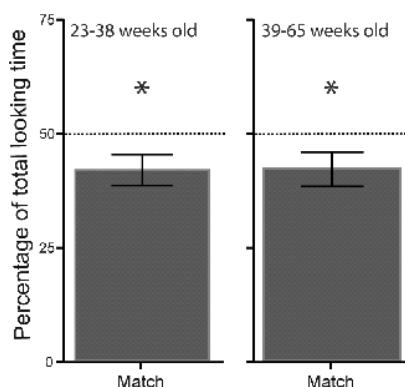


FIG. 5. Looking behavior of infant vervets at dynamic audio-visual presentations of macaque calls. This study assessed the possibility of multisensory perceptual narrowing in a nonhuman infant species. On each trial, the vervets first saw side-by-side videos of the same macaque monkey repeatedly producing a ‘coo’ call on one side of a screen and a ‘grunt’ call on the other side. The duration of their looks to each video was recorded. During the initial part of each trial, the vervets saw the two calls in silence for 4 s and during the subsequent 16 s of the trial they saw them while they also heard one of the two audible calls. The bar graph on left of the figure shows the data for two age groups. It depicts the percentage of time that the vervets looked at the matching call in the presence of the audible calls out of the total amount of time they looked at the matching call in the presence and absence of the audible calls. Here, the vervets looked significantly (\* $P < 0.05$ ) less at the matching call. This preference for the ‘wrong’ call was evidence of cross-modal matching because subsequent experiments suggested that this was due to the fear-inducing nature of the naturalistic macaque calls. The right side of the figure shows a vervet monkey looking at the videos. From Zangenehpour *et al.*, 2009.

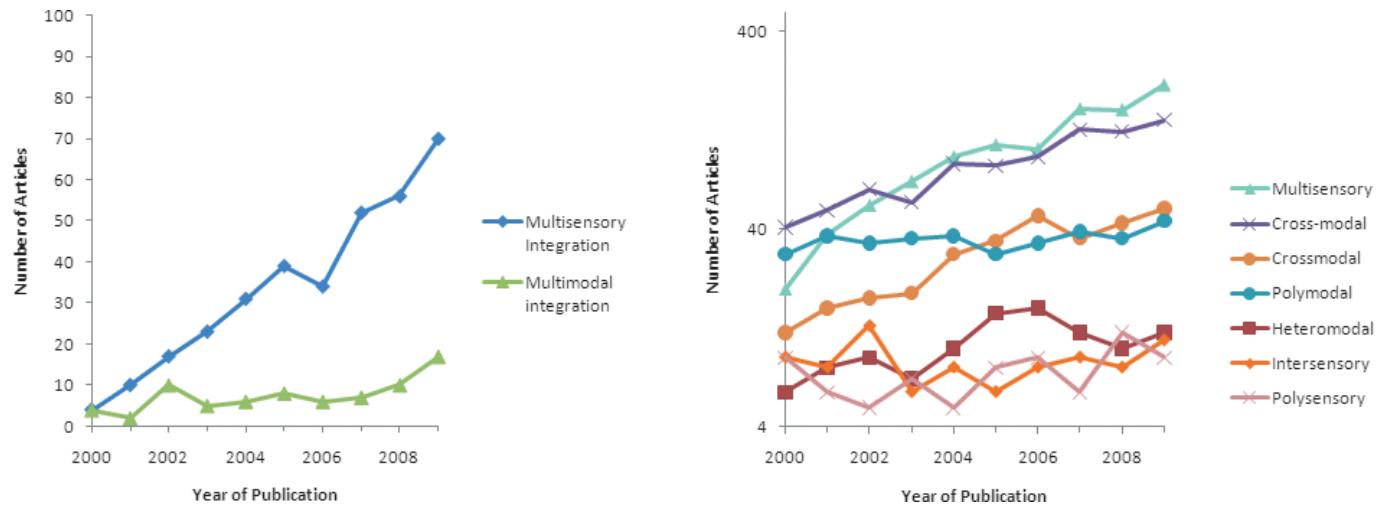


FIG. 6. The use of terms pertaining to multisensory processes over the last decade. The charts depict the number of articles containing each term as an indexed word (appearing in the title, abstract or keywords) in Pubmed journals by year of publication.

generic overarching term describing processing involving more than one sensory modality but not necessarily specifying the exact nature of the interaction between them. This definition is in keeping with this overarching term as used in the recent Handbook summarizing the state of the field in 2004 (see Calvert *et al.*, 2004). However, greater care should then be used in invoking terms that describe a particular subdomain such as multisensory integration or cross-modal matching.

### A related issue

While considering this semantic issue regarding the development of multisensory integration, it may also be appropriate to consider abandoning the multiplicity of terms that are often invoked as synonyms in describing multisensory processes or multisensory brain regions. These include such terms as ‘multimodal’, ‘intersensory’, ‘heteromodal’, ‘polymodal’ and ‘polysensory.’ Often, these terms are combined with the word ‘integration’, and often the same term is used to refer to physical stimuli and neural processes, as in the case of ‘cross-modal stimuli’ and ‘cross-modal integration.’ Although the usages of ‘multisensory’ and ‘cross-modal’ are becoming most common, many other terms are still in use (see Fig. 6).

‘Multimodal’ is perhaps the most problematic among the terms that are still in common use. It has long been used among clinicians, for whom it refers to the multiple treatments for a given disease, and among statisticians, for whom it refers to multiple peaks in a distribution. Additionally, and perhaps most importantly, the term ‘modality’ has a long history of use in referring to components of sensory input within a single sense, for example light tactile activation versus joint proprioception within the somatosensory system. Clearly it is not a unique identifier in our field, and internet searches for ‘multimodal’ turn up much that is irrelevant to our endeavor (the thousands of irrelevant ‘hits’ made it useless in the right-hand portion of Fig. 6). Unfortunately, abandoning terms we have used in publications is often difficult, and, given the absence of a compelling or principled argument for one term over the other (e.g. ‘multisensory’ versus ‘polysensory’), there is a high likelihood that any choice will be looked upon with disfavor by some of us. However, an operational rule that would work well is to separate those terms that refer to stimuli and those that refer to the particular sensory products (perceptual or neural) that they generate.

One possibility for using this approach is to designate ‘cross-modal’ to describe the stimulus complex (e.g. visual and auditory) and ‘modality-specific’ to describe its individual components (visual or auditory). The resultant neural processes, as well as the neurons and the neural circuits involved in these processes, would be referred to as ‘multisensory’ (or ‘unisensory’). The simple division between stimuli

TABLE 1. Glossary: one possible solution to the confusion of terms presently in use, and as an opening for a discussion of possible alternatives

Term*	Definition
Properties of stimuli	
Modality-specific	Describes a stimulus (or stimulus property) confined to a single sensory modality
Cross-modal	Describes a complex of two or more modality-specific stimuli from different sensory modalities
Neural or behavioral properties	
Unisensory	Describes any neural or behavioral process associated with a single sensory modality
Multisensory	Describes any neural or behavioral process associated with multiple sensory modalities
Multisensory integration	The neural process by which unisensory signals are combined to form a new product. It is operationally defined as a multisensory response (neural or behavioral) that is significantly different from the responses evoked by the modality-specific component stimuli
MSI, Multisensory index	The proportionate difference between a multisensory response to a cross-modal stimulus and the unisensory response to the most effective modality-specific component stimulus
Cross-modal matching	A process by which stimuli from different modalities are compared to estimate their equivalence
Multisensory process	A general descriptor of any multisensory phenomenon (e.g. multisensory integration and cross-modal matching)

\*Note that the top two terms describe properties of stimuli and the second two describe neural or behavioral properties (see text for discussion).

and their biological consequences may make these designations easy to remember (see glossary in Table 1). We offer this scheme as one possible solution to the confusion of terms presently in use, and as an opening for a discussion of possible alternatives. This is certainly not an exhaustive treatment of the plethora of terms used in our field. For example, in the same spirit it might pay to consider distinguishing between ‘supramodal’ and ‘amodal.’ Both terms are commonly used to refer to stimulus features that are not specific to a given sensory modality, such as intensity, duration or number. These are not specific to any given sense, but a characteristic of all sensory systems. Once again, the need to make a choice among terms is more compelling than the specificity of that choice.

## Conclusion

The semantic considerations discussed above regarding the processes underlying multisensory integration supersede the particular methods used in its study. Thus, the various neural approaches (e.g. single neuron, local field potentials, event-related potentials, imaging) used to understand its underlying physiological mechanisms, as well as the various approaches to understanding its behavioral, perceptual and emotional consequences, will all benefit from a common lexicon. Progress in the field depends on our ability to utilize the information provided by others using very different techniques to study a common process, and this requires that we share a common nomenclature. By doing so, we can avoid the confusion that arises when different groups of investigators use the same terms to mean different things, or different terms to mean the same thing.

## Acknowledgements

Some of the research described here was funded in part by grants NS036916 and EY016716 (B.E.S.) and HD35849 and BCS-0751888 (D.J.L.).

## References

- Bremner, A.J., Holmes, N.P. & Spence, C. (2008a) Infants lost in (peripersonal) space? *Trends Cogn. Sci.*, **12**, 298–305.
- Bremner, A.J., Mareschal, D., Lloyd-Fox, S. & Spence, C. (2008b) Spatial localization of touch in the first year of life: early influences of a visual spatial code and the development of remapping across changes in limb position. *J. Exp. Psychol. Gen.*, **137**, 149–162.
- Buschman, T.J. & Miller, E.K. (2009) Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron*, **63**, 386–396.
- Calvert, G.A., Spence, C. & Stein, B.E. (2004) *The Handbook of Multisensory Processes*. MIT Press, Cambridge, MA.
- Fuster, J.M., Bodner, M. & Kroger, J.K. (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, **405**, 347–351.
- Ghazanfar, A.A. & Logothetis, N.K. (2003) Facial expressions linked to monkey calls. *Nature*, **423**, 937–938.
- Gingras, G., Rowland, B.A. & Stein, B.E. (2009) The differing impact of multisensory and unisensory integration on behavior. *J. Neurosci.*, **29**, 4897–4902.
- Gori, M., Del Viva, M., Sandini, G. & Burr, D.C. (2008) Young children do not integrate visual and haptic form information. *Curr. Biol.*, **18**, 694–698.
- Hsiao, S.S., O’Shaughnessy, D.M. & Johnson, K.O. (1993) Effects of selective attention on spatial form processing in monkey primary and secondary somatosensory cortex. *J. Neurophysiol.*, **70**, 444–447.
- Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I. & Schroeder, C.E. (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, **320**, 110–113.
- Lehrman, D.S. (1970) Semantic and conceptual issues in the nature-nurture problem. In Aronson, L.R., Lehrman, D.S., Tobach, E. & Rosenblatt, J.S. (eds), *Development and Evolution of Behavior*. W. H. Freeman, San Francisco, pp. 17–52.
- Lewkowicz, D.J. (2002) Heterogeneity and heterochrony in the development of intersensory perception. *Brain Res. Cogn. Brain Res.*, **14**, 41–63.
- Lewkowicz, D.J. & Ghazanfar, A.A. (2009) The emergence of multisensory systems through perceptual narrowing. *Trends Cogn. Sci.*, **13**, 470–478.
- Lewkowicz, D.J. & Turkewitz, G. (1980) Cross-modal equivalence in early infancy: auditory-visual intensity matching. *Dev. Psychol.*, **16**, 597–607.
- Lewkowicz, D.J. & Turkewitz, G. (1981) Intersensory interaction in newborns: modification of visual preferences following exposure to sound. *Child Dev.*, **52**, 827–832.
- Lewkowicz, D.J., Leo, I. & Simion, F. (2010) Intersensory Perception at Birth: newborns Match Non-Human Primate Faces & Voices. *Infancy*, **15**, 46–60.
- Maunsell, J.H., Sclar, G., Nealey, T.A. & DePriest, D.D. (1991) Extraretinal representations in area V4 in the macaque monkey. *Vis. Neurosci.*, **7**, 561–573.
- Neil, P.A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D.J. & Shimojo, S. (2006) Development of multisensory spatial integration and perception in humans. *Dev. Sci.*, **9**, 54–464.
- Patterson, M.L. & Werker, J.F. (2002) Infants’ ability to match dynamic phonetic and gender information in the face and voice. *J. Exp. Child Psychol.*, **81**, 93–115.
- Pons, F., Lewkowicz, D.J., Soto-Faraco, S. & Sebastián-Gallés, N. (2009) Narrowing of intersensory speech perception in infancy. *Proc. Natl. Acad. Sci. USA*, **106**, 10598–10602.
- Putzar, L., Goerendt, I., Lange, K., Rösler, F. & Röder, B. (2007) Early visual deprivation impairs multisensory interactions in humans. *Nat. Rev. Neurosci.*, **10**, 1243–1245.
- Spelke, E.S. (1976) Infants’ intermodal perception of events. *Cogn. Psychol.*, **8**, 553–560.
- Stanford, T.R., Quesy, S. & Stein, B.E. (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J. Neurosci.*, **25**, 6499–6508.
- Stein, B.E. (2005) The development of a dialogue between cortex and midbrain to integrate multisensory information. *Exp. Brain Res.*, **166**, 305–315.
- Stein, B.E. & Meredith, M.A. (1993) *The Merging of the Senses*. MIT Press, Cambridge, MA.
- Stein, B.E., Stanford, T.R., Ramachandran, R., Perrault, T.J. Jr & Rowland, B.A. (2009a) Challenges in quantifying multisensory integration: alternative criteria, models and inverse effectiveness. *Exp. Brain Res.*, **198**, 113–126.
- Stein, B.E., Yu, L. & Rowland, B.A. (2009b) Plasticity of superior colliculus neurons: late acquisition of multisensory integration. *Program No. 847.4. 2009 Neuroscience Meeting Planner*. Chicago, IL, Society for Neuroscience. Online.
- Walker-Andrews, A.S. (1986) Intermodal perception of expressive behaviors: relation of eye and voice? *Dev. Psychol.*, **22**, 373–377.
- Wallace, M.T. & Stein, B.E. (1997) Development of multisensory neurons and multisensory integration in cat superior colliculus. *J. Neurosci.*, **17**, 2429–2444.
- Wallace, M.T. & Stein, B.E. (2001) Sensory and multisensory responses in the newborn monkey’s superior colliculus. *J. Neurosci.*, **21**, 8886–8894.
- Wallace, M.T. & Stein, B.E. (2007) Early experience determines how the senses will interact. *J. Neurophysiol.*, **97**, 921–926.
- Yu, L., Stein, B.E. & Rowland, B.A. (2009) Adult plasticity in multisensory neurons: short-term experience-dependent changes in the superior colliculus. *J. Neurosci.*, **29**, 15910–15922.
- Yu, L., Rowland, B.A. & Stein, B.E. (2010) Initiating the development of multisensory integration by manipulating sensory experience. *J. Neurosci.*, **30**, 4904–4913.
- Zangenehpour, S., Ghazanfar, A.A., Lewkowicz, D.J. & Zatorre, R.J. (2009) Heterochrony and cross-species intersensory matching by infant rhesus monkeys. *PLoS ONE*, **4**, e4302.