Review

Cell
PRESS

The multifaceted interplay between attention and multisensory integration

Durk Talsma[1,2], Daniel Senkowski[3], Salvador Soto-Faraco[4] and Marty G. Woldorff[5]

[1] Department of Cognitive Psychology and Ergonomics, University of Twente, P.O. Box 215, 7500 AE, Enschede, The Netherlands
[2] Department of Experimental Psychology, Ghent University, Faculty of Psychology and Educational Sciences, Department of Experimental Psychology, Henri Dunantlaan 2, B-9000 Gent, Belgium
[3] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany
[4] ICREA & Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain
[5] Center for Cognitive Neuroscience, Duke University, Box 90999, Durham, NC 27708-099, USA

Multisensory integration has often been characterized as an automatic process. Recent findings indicate that multisensory integration can occur across various stages of stimulus processing that are linked to, and can be modulated by, attention. Stimulus-driven, bottom-up mechanisms induced by crossmodal interactions can automatically capture attention towards multisensory events, particularly when competition to focus elsewhere is relatively low. Conversely, top-down attention can facilitate the integration of multisensory inputs and lead to a spread of attention across sensory modalities. These findings point to a more intimate and multifaceted interplay between attention and multisensory integration than was previously thought. We review developments in the current understanding of the interactions between attention and multisensory processing, and propose a framework that unifies previous, apparently discordant, findings.

Bidirectional influences between multisensory integration and attention

Our brains are continuously inundated with stimulation arriving through our various sensory pathways. The processes involved in synthesizing and organizing this multisensory deluge of inputs are fundamental to effective perception and cognitive functioning. Although the combination of information across the senses has been investigated since psychology became an experimental discipline [1], the past decade has seen a sharp increase of interest in this question. Within this context, the issue of how attentional shifts in one modality can affect orienting in other modalities has inspired a great deal of this recent research [2], but, perhaps surprisingly, the role that attention plays during multisensory integration itself has been largely overlooked.

Multisensory integration generally refers to the set of processes by which information arriving from the individual sensory modalities (e.g. vision, audition, touch) interacts and influences processing in other sensory modalities, including how these sensory inputs are combined together to yield a unified perceptual experience of multisensory events (Box 1). Seminal studies investigating the response

patterns of single neurons [3–5] in anesthetized animals identified several key stimulus-driven factors, most notably the temporal and spatial concordance of cross-sensory inputs, as major determinants for multisensory integration [3,6] (see [3,7,8] for a detailed discussion). This work has inspired theoretical and empirical research on multisensory integration across a variety of species (including humans), brain structures (including cortical and subcortical networks) [9–14] and methodological perspectives [2,15–19].

A key question within the scope of this review is how these stimulus-driven multisensory interactions interface with attentional mechanisms during the processes that lead to perception. Attention is a relatively broad cognitive concept that includes a set of mechanisms that determine how particular sensory input, perceptual objects, trains of thought, or courses of action are selected for further processing from an array of concurrent possible stimuli, objects, thoughts and actions (Box 2) [20]. Selection can occur in a *top-down* fashion, based on an item's relevance to the goals and intentions of the observer, or in a *bottom-up* fashion, whereby particularly salient stimuli can drive shifts of attention without voluntary control [21]. Interestingly, temporally and spatially aligned sensory inputs in different modalities have a higher likelihood to be favored for further processing, and thus to capture an individual's attention, than do stimuli that are not aligned [22–25]. This indicates that attention tends to orient more easily towards sensory input that possesses multisensory properties. Indeed, recent studies have shown that multisensory bottom-up processes can lead to a capture of attention [24]. From these findings, one might infer that the output of multisensory integration precedes attentional selection, and thus that it operates in a preattentional and largely automatic fashion.

Complementing these findings, however, are behavioral, electrophysiological [8,26,27] and neuroimaging studies [28] in humans that have identified several higher level factors, such as voluntarily oriented spatial attention and semantic congruency, that can robustly influence how integration of information across the senses occurs (Box 1) [29–31]. Some of these factors were originally hinted at by the *modality appropriateness hypothesis* [29]. This hypothesis posits that in the case of bimodal stimulation, the

E-mail addresses: d.talsma@utwente.nl, Durk.Talsma@ugent.be

Glossary

Attentional orienting: Attention involves mechanisms whereby processing resources are preferentially allocated toward particular locations, features or objects. Attentional orienting refers to the process responsible for moving the focus of attention from one location, feature or object, to another. Orienting can occur covertly, that is, in the absence of movements of the eyes or other sensory receptor surfaces (e.g. ears), as well as overtly, where the shift is accompanied by a reorienting of the sensory receptors (e.g. by a head turn) to the newly attended location or object.

Bottom-up: A form of stimulus-driven (see ''Stimulus driven'') selection that is mainly determined by the ability of sensory events in the environment to summon processing resources. This type of selection is invoked relatively independently of voluntary control; rather, stimulus salience (see ''Salience'') is the driving factor. Particularly salient stimuli (i.e. sudden motion in an otherwise still visual scene or loud sounds in an otherwise quiet room), or other stimuli for which an individual has a low detection threshold (e.g. one's own name), attract attention in a bottom-up fashion.

McGurk effect: Audiovisual illusion in which an auditory phoneme (e.g. /b/) dubbed onto incongruent visual lip movements (e.g. /g/) tends to lead to illusory auditory percepts that are typically intermediate between the actual visual and auditory inputs (i.e., /d/), are completely dominated by the visual input (i.e., /g/), or are a combination of the two (i.e., /bg/). The McGurk effect occurs in the context of isolated syllables, words or even whole sentences.

Modality appropriateness: This proposed principle about multisensory integration is based on the fact that some stimulus characteristics are processed more accurately in one sensory modality than in another. For instance, vision in general has a higher spatial resolution than audition, whereas audition has a higher temporal resolution than vision. According to this framework, information from visual stimuli tends to dominate the perceptual outcome of the spatial characteristics of audiovisual events (sometimes causing a shift of the apparent location of an auditory stimulus toward the location of the visual event). Conversely, the perceived temporal characteristics of an audiovisual event tend to be dominated by those of the auditory component.

Pop out: A perceptual phenomenon whereby a stimulus with a particularly distinctive feature relative to its surrounding background triggers quick attentional orienting and leads to rapid detection. It is often used to describe the fact that finding such particularly distinctive objects within a visual display is highly efficient and not affected by the amount of distractor elements in the scene. High stimulus salience (see ''Salience'') leads to pop out.

Resolving: The process of extracting relevant information from an attended stimulus.

Retinotopic: Spatial organization of a group of neurons based on a topographical arrangement whereby their responses map stimulus locations in the retina in a more or less orderly fashion across a brain area. In a retinotopically organized brain area, neurons involved in processing adjacent parts of the visual field are also located adjacently. This organization is most clearly seen in early (i.e. lower-level) areas of the visual pathway, but many higher-order cortical areas involved in processing visual information also show rough retinotopic organization.

Salience: Refers to a characteristic of an object or event that makes it stand out from its context. Visual objects are said to be highly salient when they have a particularly distinctive feature with respect to the neighboring items and the background, or if they occur suddenly. A bright light spot within an otherwise empty, dark context has a high saliency. Salience is often associated with being more likely to capture attention (see 'bottom-up' and ''pop out'').

Sound-induced double-flash illusion: An audiovisual illusion in which a single flash of light, presented concurrently with a train of various (two or three) short tone pips, is perceived as two (or more) flashes. This phenomenon is an example of the tendency of auditory stimuli to dominate in the perception of the temporal characteristics of an audiovisual event.

Stimulus congruence: The match of one or more features across two stimuli, stimulus components or stimulus modalities. Congruence can be defined in terms of temporal characteristics, spatial characteristics or higher-level informational content (such as semantics). In audiovisual speech perception, congruency typically refers to the matching or mismatching of a sequence of auditory speech sounds with respect to lip movements being concurrently presented. Incongruence is at the base of some multisensory phenomena, such as the McGurk illusion and the ventriloquist effects (Box 1).

Stimulus-driven: A process is stimulus driven if it is triggered or dominated by current sensory input; stimulus-driven mechanisms are a defining feature of bottom-up processing (see 'bottom-up').

Top-down: A mode of attentional orienting whereby processing resources are allocated according to internal goals or states of the observer. It is often used to refer to selective processing and attentional orienting directed in a voluntary fashion.

Ventriloquist effect: An audiovisual illusion in which an auditory stimulus is perceived as occurring at or towards the location of a spatially disparate visual stimulus that occurs at the same time.

---

sensory system with the higher acuity with respect to the most crucial aspects of a task plays a predominant role in how multisensory inputs are integrated [32–36]. For example, visual stimuli are dominant in the processing of spatial characteristics, and auditory of temporal ones (Box 1). Additional factors are related to the *unity assumption* [30,31], which corresponds to the degree to which observers infer (consciously or not) that two sensory inputs originate from a single common cause [29,31] (Box 1). These and other findings that show robust influences of

---

Box 1. Multisensory integration

In non-human single-cell recording studies, multisensory integration has been commonly characterized in terms of the firing pattern of neurons that are responsive to concurrent input from more than one sensory modality [23,80]. Multisensory responses in these particular neurons (i.e. when input in both modalities is present) are typically strongest when the crossmodal inputs are spatially aligned, presented in approximate temporal synchrony, and evoke relatively weak responses to unisensory presentations. Interestingly, response depression has also been described in this context, following spatially misaligned inputs. As such, multisensory integration reflects relative modulations in efficacy with which sensory inputs are processed and can compete for processing resources [23]. Neurons exhibiting these properties have been most extensively studied in the superior colliculus, a midbrain structure that contains multisensory convergence zones [4,5] and is involved in the reflexive orienting of attention, both covert and overt, towards salient stimuli.

In keeping with animal physiology, human behavioral, electrophysiological and neuroimaging studies have shown that spatiotemporally concordant inputs to multiple modalities often result in nonlinear (super- or sub-additive) brain potentials and faster response latencies [8,26]. In addition, several other aspects of multisensory integration have been identified in human research. Some of these aspects were originally hinted at by the *modality appropriateness hypothesis* [29] and the *unity assumption* [30,31]. The modality appropriateness hypothesis posits that, in the case of bimodal stimulation, the sensory system with the higher acuity with respect to the relevant task plays a

dominant role in the outcome of multisensory integration [32–36]. For example, the visual system usually dominates audiovisual *spatial* processes because it has a higher spatial acuity than the auditory system, whereas the auditory tends to impart more crossmodal influence in terms of *temporal* analysis. Recent approaches based on Bayesian integration have allowed the formalisation of this idea under a computational framework [15,16]. The unity assumption, on the other hand, relates to the degree to which an observer infers (not necessarily consciously) that two sensory inputs refer to a single unitary distal object or event [29,31]. Both noncognitive factors (temporal and spatial coincidence) and cognitive factors (prior knowledge and expectations) are thought to contribute to this process [30,31,66,81].

Other human behavioral studies have focused on multisensory illusions that result from incongruency across the sensory modalities (i.e. crossmodal conflict; [82]). The ventriloquist phenomenon refers to the well-known illusory displacement of a sound toward the position of a spatially disparate visual stimulus [33,83,84]; a dramatic effect that has been applied in a wide range of settings, within and beyond a purely scientific enquiry [85]. Another example is the McGurk effect [53], an illusion that occurs when speech sounds do not match the sight of simultaneously seen lip movements of the speaker, leading to a perception of a phoneme that is different from both the auditory and visual inputs. These illusions underscore the strong tendency to bind auditory and visual information that under normal (congruent) circumstances helps reduce stimulus ambiguity.

Box 2. Attention

Attention is an essential cognitive function that allows humans and other animals to continuously and dynamically select particularly relevant stimuli from all the available information present in the external or internal environment, so that greater neural resources can be devoted to their processing. One relatively straightforward framework encapsulating many aspects of attention is that of *biased competition* [86]. According to this framework, attention is considered to be the process whereby the competing neural representations of stimulus inputs are arbitrated, either because of the greater intrinsic salience of certain stimuli or because they match better to the internal goals of the individual. Thus, attention can be oriented in a top-down fashion [87,88] in which a selective biasing of the processing of events and objects that match the observer's goals is induced. By contrast, bottom-up, or stimulus-driven, attentional control refers to a mostly automatic mechanism in which salient events in the environment tend to summon processing resources, relatively independent of higher-level goals or expectations [21,89].

A fronto-parietal network of brain areas has been shown to be involved in allocating and controlling the direction of top-down attention by sending control signals that modulate the sensitivity of neurons in sensory brain regions [90–94]. Stimulus-driven attention employs various parts of the same network, seeming to operate in concert with subcortical networks that include the superior colliculus [92,95] and regions in the right temporal-parietal junction [92].

It is still not well understood how top-down attention is controlled or allocated across the sensory modalities. Considerable overlap can be found in brain areas responsible for the top-down orienting of visual [91] and auditory [96] attention. In addition, attending to a specific location in one modality also affects the processing of stimuli presented at that location in another sensory modality, indicating that spatial attention might tend to be directed in a supramodal, or at least modality-coordinated fashion [56,59,62,97,98]. These multisensory links indicate that there is flexible attentional deployment across sensory modalities [99], making it unlikely that there are completely independent control mechanisms for each modality. It has been proposed that top-down spatial selective attention operates by increasing the sensitivity of neurons responsive to the attended stimulus feature [100,101]. This mechanism might explain why attentional orienting can be operated in parallel across modalities, whereas attentional resolving – that is, the processing of relevant information within each modality – can be carried out more or less independently within each modality [102–104].

---

cognitive factors on multisensory integration indicate a more flexible process than previously thought, whereby attention can affect the effectiveness of multisensory integration in a top-down fashion [2,37–45].

How can these apparently opposing ideas regarding the interplay between attention and multisensory integration (i.e. bottom-up, highly automatic, multisensory integration processes that can perhaps capture attention vs. top-down attentional mechanisms that can modulate multisensory processing) have evolved in the literature? One possible reason is that studies investigating the role of top-down influences of attention on multisensory integration, and studies investigating the bottom-up influence of multisensory integration on the orienting of attention have been conducted more or less independently of each other. This disconnection could thus have led to opposing, seemingly mutually exclusive, conclusions. Considering recent findings from the fields of cognitive psychology and cognitive neuroscience, we propose a framework aimed at unifying these seemingly discordant findings (Figure 1).

A framework for interactions between attention and multisensory integration

As noted above, and as reviewed in the sections to follow, the evidence in favor of the existence of bidirectional influences between attention and multisensory integration is considerable. Under certain circumstances the co-occurrence of stimuli in different modalities can lead preattentively and automatically to multisensory integration in a bottom-up fashion, which then makes it more probable that the resulting event will capture attention and thus capitalize available processing resources [24]. In other situations, however, top-down directed attention can influence integration processes for selected stimulus combinations in the environment [40,46]. Based on the pattern of findings in the literature, we propose a framework in which a key factor that determines the nature and directionality of these interactions is the stimulus complexity of the environment, and particularly the ongoing competition between the stimulus components within it.

More specifically, we propose (Figure 1a) that multisensory integration can, and will, tend to occur more or less preattentively in a scene when the amount of competition between stimuli is low. So, for example, under circumstances in which the stimulation in a task-irrelevant sensory modality is sparse (Figure 1b, case 1), the stimuli in that modality will tend to be intrinsically more salient solely because they occur infrequently. Because of their increased salience, these stimuli will tend to enhance the perceptual processing of corresponding (i.e. spatially and/or temporally aligned) sensory events in a concurrently active task-relevant modality. This stimulus-driven (i.e. bottom-up) crossmodal effect will tend to capture attention and processing resources, thereby facilitating the ability for attentional selection and/or attentional 'resolving' (i.e. discrimination; Box 2) of stimuli in the task-relevant modality. This multisensory-based enhancement can occur even in a context where attentional selection in the task-relevant sensory modality is challenging, such as in a difficult visual search task [24]. By contrast, when multiple stimuli within each modality are competing for processing resources, and thus the saliency for individual stimuli within the potentially facilitating modality is low, top-down selective attention is likely to be necessary in order for multisensory integration processes between the appropriately associated stimulus events to take place [41,42,46] (Figure 1b, case 2).

In the following sections we focus on three main forms under which these interactions between multisensory integration and attention can manifest themselves: (i) the stimulus-driven influences of multisensory integration on attention, (ii) the influence of top-down directed attention on multisensory integration and (iii) the spreading of attention across modalities, a process that contains aspects of both top-down and bottom-up mechanisms.

Stimulus-driven influences of multisensory integration on attention

A clear demonstration of the possible involvement of stimulus-driven multisensory integration on attentional
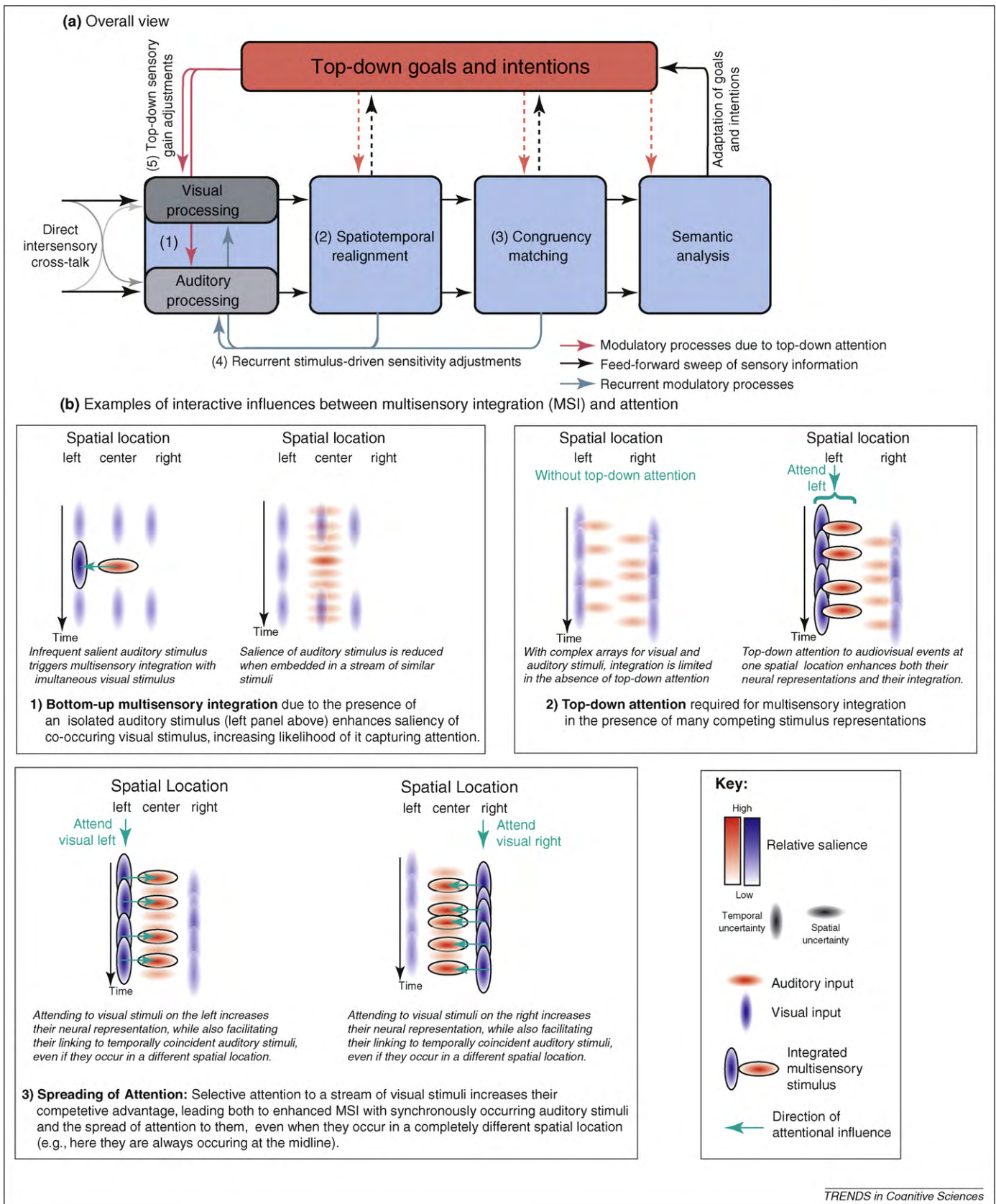
Figure 1. **A framework for the interactions between multisensory integration and attention. (a)** Sequence of Processing Steps. Inputs from the sense organs are thought to interact crossmodally at multiple phases of the processing pathways, including at very early stages [8]. Stimuli can be integrated automatically if several conditions are satisfied: (1) If *initial saliency* of one of the stimuli is at or above a crucial threshold, then preprocessing stages will attempt to spatiotemporally realign this stimulus with one of lesser salience. (2) *Spatiotemporal realignment*. The stimulus stream will then be monitored for congruence in stimulus patterns of the matched streams. (3) *Congruency detection*. If the realignment and/or congruency matching processes succeed, then the neural responsiveness of brain areas in charge of processing the input streams will be increased. (4) *Recurrent stimulus-driven sensitivity adjustments* to sustain the integration process. If stimuli cannot be realigned or when incongruency is detected, the sensory gain would tend to be decreased. Note that we consider these potential gain adjustments to be mainly stimulus driven, and therefore a reflection of
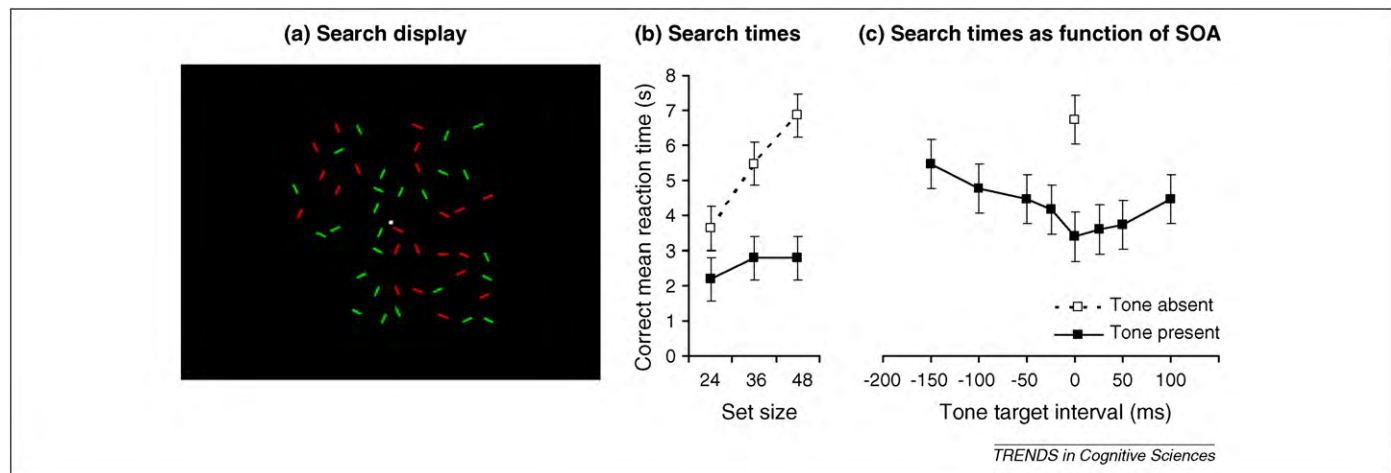
Figure 2. **Multisensory integration mechanisms affecting visual attention in a bottom-up, driven fashion. (a)** Cluttered displays containing a variable number of short line segments were presented. Display elements continuously changed color from green to red (or vice versa) at random moments and locations. A short tone pip could be presented simultaneously with the color change of the target element. Participants were required to detect and report the orientation of the target element, consisting of a horizontal or vertical line among $\pm45°$ tilted line distractors. **(b)** In the absence of a sound, search times increased linearly with the number of distractor items in the display (white squares). By contrast, when the sound was present, search times became much shorter and independent of set size, indicating that the target stimulus popped out of the background (black squares). **(c)** Search times as a function of the relative stimulus onset asynchrony (SOA) between color change of the target and onset of the sound. Negative SOAs indicate that the tone preceded the visual target event, and positive SOAs indicate that the target event preceded the tone. (Data from condition with set size fixed at 48 elements). Maximum search efficiency at positive SOAs indicates multisensory integration. Adapted, with permission, from Ref. [24].

selection was recently provided using a difficult visual search task [24]. In this study visual targets were presented among an array of similar distractor stimuli (Figure 2). Search times for the visual target increased with an increasing number of display items (distractors), as has typically been found in demanding unisensory visual search tasks [47,48]. In these displays, targets and distractors changed color at random moments and irrelevantly with respect to the task. However, when a brief, spatially uninformative sound was presented concurrently with a target color change, search times became much faster and relatively independent of the number of distractor items, providing evidence that the auditory co-occurrence made the visual target pop out from the background elements.

At first glance these findings might seem to contradict those of another visual-search study in which auditory stimuli failed to facilitate the detection of a uniquely corresponding (temporally matched) visual event among visual distractors [49]. Here, however, not only the visual stimuli but also the accessory acoustic events changed at relatively high frequencies, and thus were all consequently of low saliency. These two findings indicate that a co-occurring sound renders it probable that a visual target pops out only when the sound is relatively rare and therefore a salient event by itself [50]. This conclusion is also consistent with data from another study that showed that a sound can affect the processing of a synchronized visual input when it is a single salient event, but not when it is part of an ongoing stream [51].

The framework we are proposing here accounts for such bottom-up multisensory integration mechanisms because

a relatively salient stimulus presented in one modality can elicit a neural response that is strong enough to be automatically linked to a weaker neural response to a stimulus in another modality (Figure 1b, case 1). We hypothesize that multisensory processes triggered by temporally aligned auditory input can affect the strength of the representation of the co-occurring visual stimulus, making it pop out of a background of competing stimuli. By contrast, when multiple stimuli are competing for processing recources, their neural representations may require top-down attention for the relevant crossmodally corresponding stimuli to be readily integrated and for their processing to be effectively facilitated.

Influence of top-down directed attention on multisensory processing

As mentioned earlier, recent studies have provided evidence for the influence of top-down attention on multi-sensory integration processes. In a recent EEG study, for example, it was shown that spatial attention can strongly influence multiple stages of multisensory processing, beginning as early as 80 ms post-stimulus [40]. Complementarily, a recent fMRI study [39] showed increased activity in multiple brain areas, including the superior temporal sulcus, striate visual cortex, extrastriate visual cortex and the superior colliculus, when attention was directed to visual lip movements that matched an auditory spoken sentence, in contrast to when attention was directed to simultaneously presented but non-matching lip movements (Figure 3). Thus, these results again provide evidence for the ability of top-down attention to influence multisensory integration processes. They are

the bottom-up driven shift part of the interaction between multisensory integration and attention. If none of the stimuli is of sufficient saliency, top-down attention might be necessary to set up an initial selection of to-be-integrated-stimuli. (5) *Top-down sensory gain adjustments*. The resulting boost in sensory sensitivity due to a top-down gain manipulation might then be sufficient to initiate the processes of spatial temporal alignment and congruency matching that would otherwise not have occurred. In addition, top-down attention can modulate processing at essentially all of these stages of multisensory processing and integration. **(b)** Three examples of interactive influences between multisensory integration and attention: (1) bottom-up multisensory integration that can then drive a shift of attention, (2) the need for top-down attention for multisensory integration in the presence of many competing stimulus representations and (3) the spreading of attention across space and modality (visual to auditory).
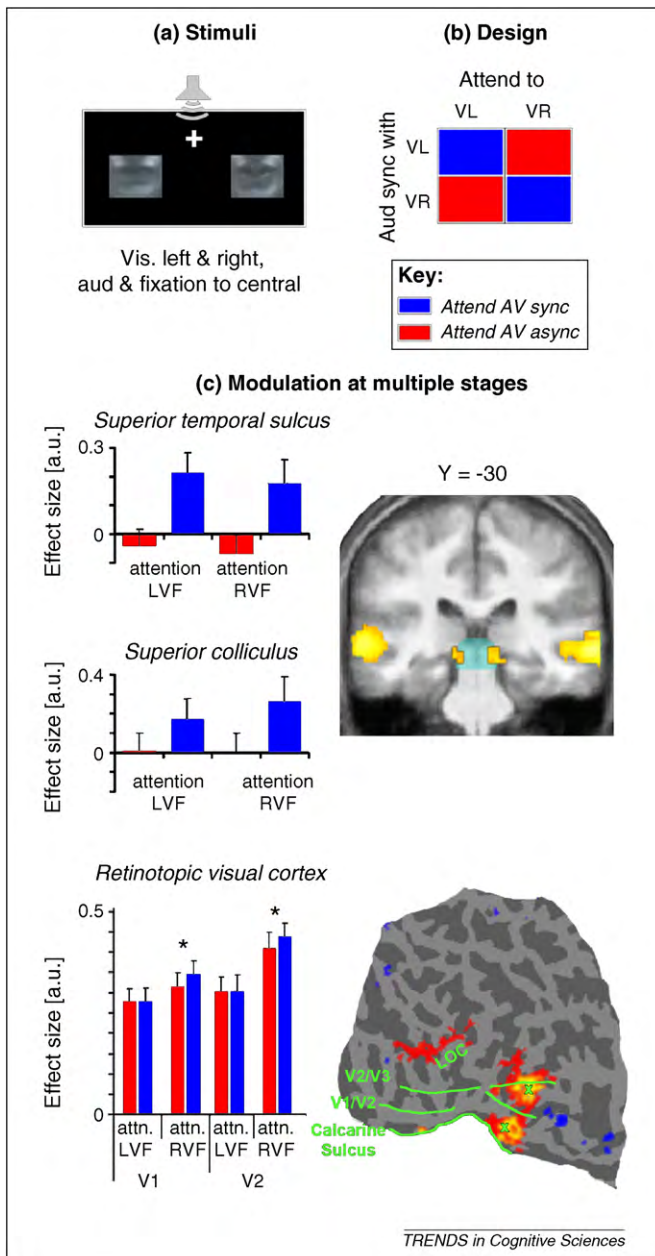
Figure 3. **Effects of top-down spatial attention on audiovisual speech perception.**
(a) Layout of the experiment. Sounds were presented from one central location, while two visually and laterally presented streams of lip-moments were played. One of these streams was congruent with the auditory speech signals while the other stream was incongruent. **(b)** Attention was selectively oriented to either the left or right visual streams, either of which could, in turn, be congruent or incongruent with the auditory speech stimuli. **(c)** Attending to the congruent stimuli resulted in increases in activation in several brain areas that are typically associated with multisensory integration. These areas included the superior temporal sulcus, and superior colliculus (central panel), as well as large parts of the retinotopically organized visual areas V1 and V2. a.u. = arbitrary units. Adapted, with permission, from Ref. [39].

also consistent with EEG data from another study showing that allocating visual attention towards irrelevant lip movements in the visual scene interferes with the recognition of audiovisual speech signals from an attended speaker [52]. Together, these results demonstrate that top-down attention can modulate multisensory processing leading to either facilitation (in the case of rhymically congruent inputs) or interference (in the case of rhythmically incongruent inputs).

Further evidence for the impact of top-down attention on multisensory processing derives from studies using audiovisual speech illusion paradigms [42–44]. For instance, diverting attention to a secondary task reduces susceptibility to the McGurk effect [53], whereby an auditory phoneme dubbed onto incongruent visual lip movements leads to an illusory auditory percept (Box 1). Interestingly, this reduced susceptibility to the McGurk effect due to the allocation of processing resources to another task was observed regardless of the sensory modality (visual, auditory or tactile) in which the secondary task was performed [42,43]. This result suggests that an attentional manipulation, which diverts limited processing resources away from one sensory modality's input for a multisensory event, can also result in reduced integration of that modality's input with the input from other sensory modalities [54]. Thus, when competition between different stimulus sequences is relatively low, we propose that audiovisual speech stimuli may integrate relatively automatically. However, when a secondary task is introduced that involves a diversion of attention from the event, competition ensues between the neural resources devoted to the secondary task and those resources necessary for processing the audiovisual speech stimuli. This, in turn, may reduce the automaticity of the congruency detection and consequently of the multisensory integration process.

It is important to distinguish these described top-down attentional effects from some earlier behavioral [55–59] and electrophysiological [60–62] work on supramodal attention effects. In particular, earlier studies showed that top-down spatial attention in one modality not only induces enhanced processing of stimuli in that modality at the attended location, but also of stimuli in other (task-irrelevant) modalities at that location – for example, attending visually to a location in space results in the processing of auditory stimuli at that location also being enhanced. Other studies have shown that salient stimuli in one modality in a particular location can capture spatial attention and affect the processing of stimuli in another modality that occur shortly afterward at that same location [56,59]. These latter findings suggest that spatial attention tends to be coordinated across modalities, both when voluntarily directed and when induced in a bottom-up cueing fashion. In both cases, however, the stimuli in the different modalities do not need to be presented concurrently to be enhanced, rather, they only need to occur in the spatially attended location. Thus, although such effects reflect fundamental mechanisms by which top-down factors modulate processing across modalities, they contrast with the studies mentioned above in which top-down attention modulates the actual integration of multisensory inputs that occur aligned in time [63].

Crossmodal spreading of attention
Another way that attention can interactively influence the processing of stimulus inputs in different sensory modalities that *do* occur at the same time has been termed crossmodal spreading of attention, which has been found to occur even when the visual and auditory stimuli arise from different spatial locations [37]. More specifically, it has been shown that a task-irrelevant, centrally presented,

auditory stimulus can elicit a different brain response when paired in time with an attended versus an unattended visual stimulus presented laterally (Figure 4) [37]. This effect appeared as a prolonged, negative-polarity, fronto-central ERP deflection, beginning at ~200 ms post-stimulus and lasting for hundreds of milliseconds, as well as a corresponding enhancement of fMRI activity in auditory cortex. The ERP effect resembled an activation known as 'late processing negativity,' a hallmark of the enhanced processing of attended *unisensory* auditory

inputs [64]. This auditory enhanced processing effect was particularly striking because it was induced solely by whether a simultaneously occurring visual stimulus in a different spatial location was attended versus unattended. Such a result is thus consistent with a spread of attention across modalities and space to the temporally coincident auditory stimulus, occurring via a systems-level cascade of top-down and bottom-up influences. This cascade begins with top-down visuospatial attention that determines which stimulus in the visual modality is to be selectively
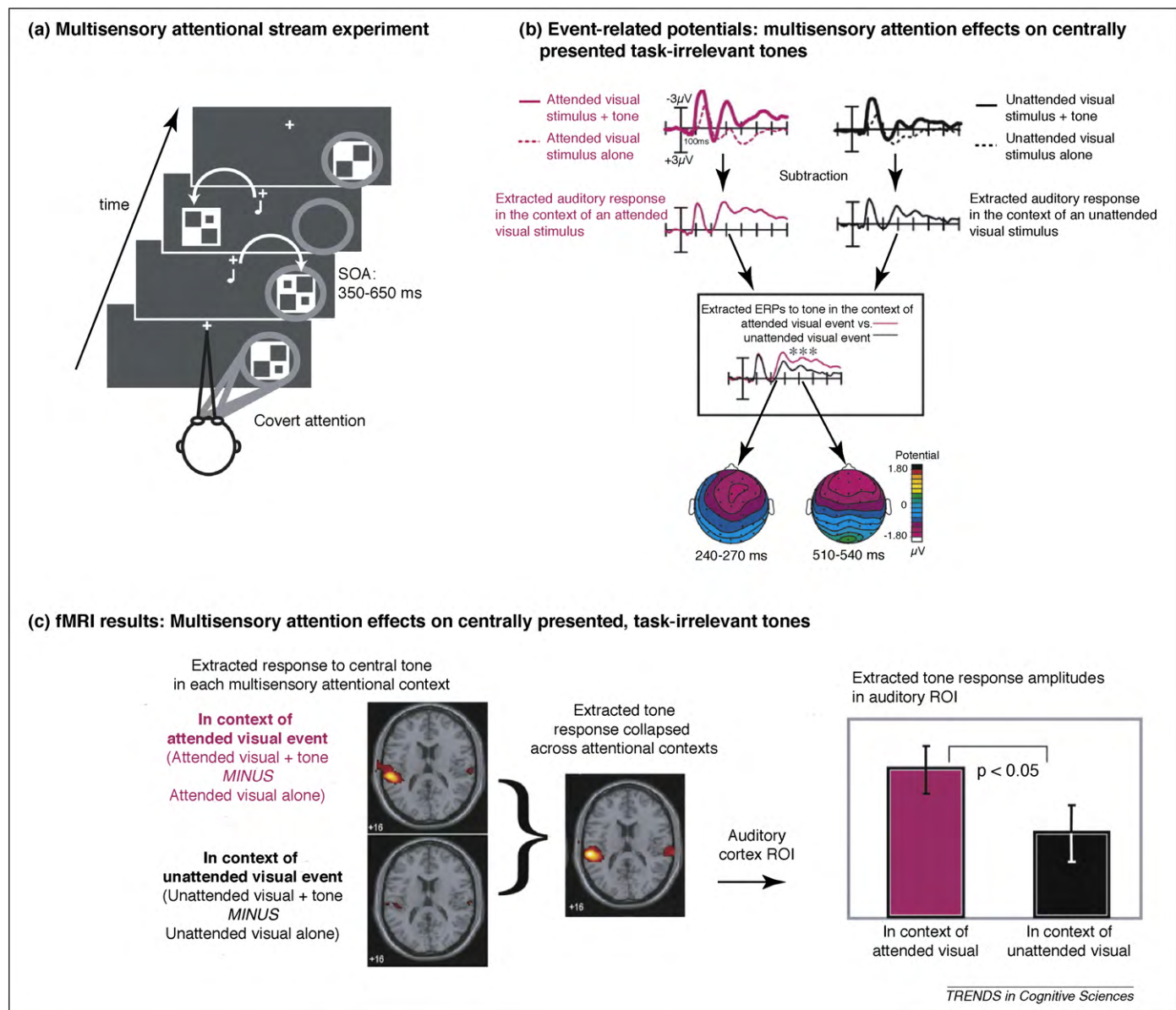

Figure 4. **Spreading of attention across a multisensory object. (a)** *Experimental Design.* Visual stimuli were flashed successively, and in random order, in the left and right visual fields. On half of the trials, a task-irrelevant central tone was presented synchronously with the lateral visual stimulus. Participants were instructed to visually attend selectively to only one of the two locations. **(b)** *ERP subtraction procedure and key results.* ERPs elicited by attended visual stimuli that occurred alone were subtracted from ERPs elicited by attended visual stimuli that were accompanied by a tone, yielding the extracted ERP response to the central tones when they occurred in the context of an attended (lateral) visual stimulus. A similar subtraction procedure was applied to unattended-visual trials to yield the extracted ERP response to the central tones when they occurred in the context of an unattended (lateral) visual stimulus. An overlay of these two extracted ERPs showed that tones presented in the context of an attended visual stimulus elicited a prolonged enhanced negative wave over fronto-central scalp areas beginning at around 200 ms poststimulus. This effect resembles the late auditory processing negativity that is a hallmark neural effect elicited during intramodal auditory attention. **(c)** *fMRI results from the same paradigm.* These results, extracted using an analogous contrast logic, showed that auditory stimuli presented in the context of an attended auditory stimulus yielded an increase in brain activity in the auditory cortex, compared with the activation elicited by the same tone when it was presented in the context of an unattended visual stimulus. The enhanced processing of task-irrelevant auditory stimuli that occur simultaneously with an attended visual stimulus, even one occurring in a different location, indicates that the visual attention has spread across the components of the multisensory object to encompass the auditory part. Adapted, with permission, from Ref. [37]. Copyright © 2005, The National Academy of Sciences.

processed. Then, presumably by means of an automatic (bottom-up) linking mechanism derived from the temporal coincidence of the multisensory components, attention spreads across modalities to encompass the auditory component, despite it being completely task-irrelevant and not even in the same spatial location.

Variants of this effect have recently been observed in other studies [38,41,65]. Moreover, a distinction has been made between the object-based multisensory spread of attention that occurs when stimuli are associated simply due to their co-occurrence in time, and a representation-based spread of attention for naturally associated stimulus pairs [38]. On the other hand, in some circumstances a task-irrelevant auditory stimulus that is semantically or representationally in conflict with a task-relevant visual stimulus can serve as an attention capturing distractor and increase attentional spread [66,67]. Regardless, the currently available data suggest that the spatial-temporal linking of visual and auditory stimulation is initially organized at an early processing stage [37,41,68], but that higher-level cognitive representations can also modulate the crossmodal spreading of attention [38,66,67,69].

In some ways, the spread of attention [37] can be considered to be a reversed version of the sound-induced (i.e. bottom-up) visual pop-out effect described earlier [24]. In the spread-of-attention case [37], it is visuospatial attention that determines what becomes more salient in the auditory world (Figure 1b, case 3). In the sound-induced visual pop-out case [24], an auditory stimulus determines what becomes more salient in the visual world (Figure 1b, case 1). Regardless, stimuli in one modality, or attention to stimuli in one modality, selectively influence the processing of co-occurring stimuli in another modality. Thus, these studies underscore the multifaceted and bidirectional ways by which attention and multisensory processing can interact [70].

Concluding remarks and future directions

Although it is generally acknowledged that multisensory integration processes can influence the bottom-up orienting of attention to salient stimuli [23], the role of top-down attention and the interplay between bottom-up and top-down mechanisms during multisensory integration processes remains a matter of ongoing debate. On the one hand, there is relatively little effect of top-down attention on multisensory integration under conditions of low competition between stimuli [33]. On the other hand, robust top-down effects on multisensory integration processes have been observed under higher degrees of competition between successive or concurrent inputs to different modalities [40,46]. Accordingly, we suggest that the degree of competition between neural representations of stimulus inputs is a major determinant for the necessity of top-down attention, thereby helping to explain the contradictory findings in the literature.

As such, we identify several important directions and challenges for future research (Box 3). One promising direction will be to develop multisensory attention paradigms that systematically modulate the degree of competition between stimuli within an experiment, for instance by manipulating perceptual load [71] or stimulus delivery

---

Box 3. Questions for future research

- What is the effect of stimulus competition on the need for top-down attention in multisensory integration?
- Is there a differential influence of spatial and nonspatial attention on multisensory integration, and if so, which stimulus properties are affected?
- What are the neural mechanisms by which competition influences multisensory integration and how can attention influence these mechanisms?
- Does the complexity of a perceptual task influence the requirement of attention in multisensory integration, or is the attentional requirement only a function of the complexity of the stimulus input array regardless of the task?
- What are the neural mechanisms that control whether the processing of a task-relevant stimulus in one modality is enhanced versus diminished by a task-irrelevant stimulus in a second modality?
- What is the relationship between the spread of attention and sound-induced visual pop-out effects?
- What role does attention play in crossmodal processing deficits in populations with neuropsychological or psychiatric disorders, such as autism and schizophrenia?
- How does the relationship between attention and multisensory integration change across the lifespan?

---

rates, while simultaneously manipulating top-down attentional factors. Such an approach would probably be helpful for determining more precisely which aspects of multisensory integration can be affected by the ongoing interplay between sensory input competition and top-down attentional influence, and the mechanisms by which these interactions occur.

Another challenge lies in more fully and precisely delineating the circumstances in which attention affects multisensory integration, and the mechanisms by which it does so. For instance, top-down visuo-spatial attention can have a strong impact on the processing of concurrent auditory stimuli, as shown by the visual-to-auditory spreading-of-attention effect. However, it seems that spatial attention in the auditory modality has a much less pronounced effect on the integration of visual stimuli that are not presented at the same location. Auditory inputs, and attention to them, can nevertheless affect the processing of visual stimuli, but particularly in terms of their temporal characteristics. This tendency for temporal dominance of the auditory modality, presumed due to the importance of temporal processing in auditory processing, is reflected by the sound-induced double-flash illusion [44] (Box 1), as well as other data indicating that auditory attention tends to temporally align the processing of concurrent visual stimuli [72].

We note that the vast majority of studies on multisensory processing and attention reviewed here have used audio-visual stimulus material. Although vision and audition have by far been the most frequently investigated senses in multisensory research, it will be important in the future to delineate which aspects of the findings and framework presented here can be generalized to other combinations of sensory modalities (e.g. touch, smell, taste or proprioception).

Finally, research on the relationship between multisensory processing and attention could also have crucial implications for our understanding of multisensory

integration deficits in neurological and psychiatric populations, such as in those with hemispatial neglect, schizophrenia [73], and autism [74,75], as well as multisensory integration decrements in the aging population [76,77]. For example, many clinical symptoms in schizophrenia seem to be the result of a pathological connectivity between cortical networks, including between sensory processing networks, during cognitive tasks [78]. It is therefore possible, perhaps even probable, that an altered interplay between attention and multisensory processes will be found in this group of patients [79]. Thus, the study of these interactions in clinical populations is important for the understanding of the cognitive deficits in these groups, as well as for advancing our understanding of the underlying neural mechanisms.

We live in a multisensory world in which we are continuously deluged with stimulus input through multiple sensory pathways. For effective cognitive functioning, we must continually select and appropriately integrate together those inputs that are the most relevant to our behavioral goals from moment to moment. Thus, the dynamic and bidirectional interplay between attentional selection and multisensory processing is fundamental to successful behavior. This review has aimed at organizing the present state of the research in this important area of cognitive science, and at developing a conceptual framework that we hope will be helpful for future research.

References

1 Urbantschisch, V. (1880) Über den Einfluss einer Sinneserrugun auf die übrigen Sinnesempfindungen. *Arch. Gesch. Psych.* 42, 155–175
2 Spence, C. and Driver, J. (2004) *Cross-Modal Space and Cross-Modal attention*, Oxford University Press
3 Stein, B. and Meredith, M.A. (1993) *The merging of the senses*, MIT Press
4 Wallace, M.T. *et al.* (1998) Multisensory integration in the superior colliculus of the alert cat. *J. Neurophys.* 80, 1006–1010
5 Meredith, M.A. (2002) On the neuronal basis for multisensory convergence: A brief overview. *Cogn. Brain Res.* 14, 31–40
6 Welch, R.B. and Warren, D.H. (1986) Intersensory Interactions. In *Handbook of Perception and Human Performance Volume 1: Sensory Processes and Perception* (Kauffman, K.R. and Thomas, J.P., eds), pp. 1–36, Wiley
7 Holmes, N.P. (2009) The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain Topogr.* 21, 168–176
8 Giard, M.H. and Peronnet, F. (1999) Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490
9 Foxe, J.J. and Schroeder, C.E. (2005) The case for feedforward multisensory convergence during early cortical processing. *NeuroReport* 16, 419–423
10 Schroeder, C.E. and Foxe, J. (2005) Multisensory contributions to low-level, 'unisensory' processing. *Curr. Opin. Neurobiol.* 15, 454–458
11 Lakatos, P. *et al.* (2007) Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron.* 53, 279–292
12 Angelaki, D.E. *et al.* (2009) Multisensory integration: psychophysics, neurophysiology, and computation. *Curr. Opin. Neurobiol.* 19, 452–458
13 Driver, J. and Noesselt, T. (2008) Multisensory Interplay Reveals Crossmodal Influences on 'Sensory-Specific' Brain Regions. Neural Responses, and Judgments. *Neuron.* 57, 11–23
14 Kayser, C. and Logothetis, N.K. (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* 212, 121–132
15 Ernst, M.O. and Bülthoff, H.H. (2004) Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169
16 Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433
17 Calvert, G.A. (2001) Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123
18 Beauchamp, M.S. (2005) Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3, 93–113
19 Colonius, H. and Diederich, A. (2004) Multisensory interaction in saccadic reaction time: A time-window-of- integration model. *J. Cogn. Neurosci.* 16, 1000–1009
20 Pashler, H. (1999) *The Psychology of Attention*, Cambridge University Press
21 Theeuwes, J. (1991) Exogenous and Endogenous Control of Attention: the Effect of Visual Onsets and Offsets. *Percep. Psychophys.* 49, 83–90
22 Driver, J. (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381, 66–68
23 Stein, B.E. *et al.* (2005) Multisensory Integration in Single Neurons of the Midbrain. In *The Handbook of Multisensory Processes* (Calvert, G. *et al.*, eds), pp. 243–264, MIT press
24 Van der Burg, E. *et al.* (2008) Pip and Pop: Nonspatial Auditory Signals Improve Spatial Visual Search. *J. Exp. Psychol. Hum. Percep. Perform.* 34, 1053–1065
25 Van der Burg, E. *et al.* (2009) Poke and pop: Tactile-visual synchrony increases visual saliency. *Neurosci. Lett.* 450, 60–64
26 Molholm, S. *et al.* (2002) Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128
27 Fort, A. *et al.* (2002) Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb. Cortex* 12, 1031–1039
28 Calvert, G.A. *et al.* (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657
29 Welch, R.B. and Warren, D.H. (1980) Immediate perceptual response to intersensory discrepancy. *Psych. Bull.* 88, 638–667
30 Vatakis, A. and Spence, C. (2007) Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Percept. Psychophys.* 69, 744–756
31 Welch, R.B. (1999) Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In *Cognitive Contributions to the Perception of Spatial and Temporal Events* (Aschersleben, G. *et al.*, eds), pp. 371–387, Elsevier
32 Navarra, J. *et al.* (2010) Assessing the role of attention in the audiovisual integration of speech. *Information Fusion* 11, 4–11
33 Bertelson, P. *et al.* (2000) The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* 62, 321–332
34 Vroomen, J. *et al.* (2001) The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept. Psychophys.* 63, 651–659
35 Stekelenburg, J.J. *et al.* (2004) Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neurosci. Lett.* 357, 163–166
36 Bonath, B. *et al.* (2007) Neural Basis of the Ventriloquist Illusion. *Curr. Biol.* 17, 1697–1703
37 Busse, L. *et al.* (2005) The spread of attention across modalities and space in a multisensory object. *Proc. Nat. Acad. Sci. U. S. A.* 102, 18751–18756
38 Fiebelkorn, I.C. *et al.* (2010) Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cereb. Cortex* 20, 109–120

39 Fairhall, S.L. and Macaluso, E. (2009) Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur. J. Neurosci.* 29, 1247–1257

40 Talsma, D. and Woldorff, M.G. (2005) Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.* 17, 1098–1114

41 Talsma, D. *et al.* (2007) Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17, 679–690

42 Alsius, A. *et al.* (2005) Audiovisual integration of speech falters under high attention demands. *Curr. Biol.* 15, 839–843

43 Alsius, A. *et al.* (2007) Attention to touch weakens audiovisual speech integration. *Exp. Brain Res.* 183, 399–404

44 Mishra, J. *et al.* (2010) Effect of Attention on Early Cortical Processes Associated with the Sound-Induced Extra Flash Illusion. *J. Cogn. Neurosci.* 22, 1714–1729

45 Soto-Faraco, S. and Alsius, A. (2007) Conscious access to the unisensory components of a cross-modal illusion. *NeuroReport* 18, 347–350

46 Van Ee, R. *et al.* (2009) Multisensory congruency as a mechanism for attentional control over perceptual selection. *J. Neurosci.* 29, 11641–11649

47 Treisman, A.M. and Gelade, G. (1980) A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136

48 Wolfe, J.M. (2003) Moving towards solutions to some enduring controversies in visual search. *Trends Cogn. Sci.* 7, 70–76

49 Fujisaki, W. *et al.* (2006) Visual search for a target changing in synchrony with an auditory signal. *Proc. R. Soc. B Biol. Sci.* 273, 865–874

50 Van der Burg, E. *et al.* (2010) Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE* 5, E10664

51 Vroomen, J. and De Gelder, B. (2000) Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1583–1590

52 Senkowski, D. *et al.* (2008) Look who's talking: The deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *NeuroImage* 43, 379–387

53 McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264, 746–748

54 Hugenschmidt, C.E. *et al.* (2009) Suppression of multisensory integration by modality-specific attention in aging. *NeuroReport* 20, 349–353

55 Santangelo, V. *et al.* (2009) Interactions between voluntary and stimulus-driven spatial attention mechanisms across sensory modalities. *J. Cogn. Neurosci.* 21, 2384–2397

56 Spence, C. and Driver, J. (1996) Audiovisual Links in Endogenous Covert Spatial Attention. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1005–1030

57 Spence, C. and Driver, J. (1997) Audiovisual links in exogenous covert spatial orienting. *Percept. Psychophys.* 59, 1–22

58 Spence, C. *et al.* (2000) Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention. *Percept. Psychophys.* 62, 410–424

59 Koelewijn, T. *et al.* (2009) Auditory and Visual Capture During Focused Visual Attention. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1303–1315

60 Teder-Sälejärvi, W.A. *et al.* (1999) Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. *Brain Res. Cogn. Brain Res.* 8, 327–343

61 Eimer, M. and Schöger, E. (1998) Effects of intermodal attention and cross-modal attention in spatial attention. *Psychophysiology* 35, 313–327

62 Hillyard, S.A. *et al.* (1984) Event-related brain potentials and selective attention to different modalities. In *Cortical Integration* (Reinoso-Suarez, F. and Aimone-Marsan, F., eds), pp. 395–413, Raven Press

63 McDonald, J.J. *et al.* (2001) Multisensory integration and crossmodal attention effects in the human brain. *Science* 292, 1791

64 Näätänen, R. (1982) Processing negativity: An evoked-potential reflection. *Psychol. Bull.* 92, 605–640

65 Degerman, A. *et al.* (2007) Human brain activity associated with audiovisual perception and attention. *NeuroImage* 34, 1683–1691

66 Zimmer, U. *et al.* (2010) The electrophysiological time course and interaction of stimulus conflict and the multisensory spread of attention. *Eur. J. Neurosci.* 31, 1744–1754

67 Zimmer, U. *et al.* (2010) Multisensory conflict modulates the spread of visual attention across a multisensory object. *NeuroImage* DOI: 10.1016/j.neuroimage.2010.04.245

68 Molholm, S. *et al.* (2007) Object-based attention is multisensory: Co-activation of an object's representation in ignored sensory modalities. *Eur. J.Neurosci.* 26, 499–509

69 Iordanescu, L. *et al.* (2008) Characteristic sounds facilitate visual search. *Psychon. Bull. Rev.* 15, 548–554

70 Soto-Faraco, S. and Alsius, A. (2009) Deconstructing the McGurk-MacDonald Illusion. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 580–587

71 Lavie, N. (1995) Perceptual Load as a Necessary Condition for Selective Attention. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 451–468

72 Talsma, D. *et al.* (2009) Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. *Exp. Brain Res.* 198, 313–328

73 Ross, L.A. *et al.* (2007) Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr. Res.* 97, 173–183

74 Magnée, M.J.C.M. *et al.* (2008) Atypical processing of fearful face-voice pairs in Pervasive Developmental Disorder: An ERP study. *Clin. Neurophysiol.* 119, 2004–2010

75 Magnée, M.J.C.M. *et al.* (2008) Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *J. Child Psychol. Psychiatry Allied Disci.* 49, 995–1000

76 Hugenschmidt, C.E. *et al.* (2009) Preservation of crossmodal selective attention in healthy aging. *Exp. Brain Res.* 198, 273–285

77 Peiffer, A.M. *et al.* (2009) Aging and the interaction of sensory cortical function and structure. *Hum. Brain Mapp.* 30, 228–240

78 Jardri, R. *et al.* (2009) Neural functional organization of hallucinations in schizophrenia: Multisensory dissolution of pathological emergence in consciousness. *Conscious. Cogn.* 18, 449–457

79 Foxe, J. and Molholm, S. (2009) Ten Years at the Multisensory Forum: Musings on the Evolution of a Field. *Brain Topogr.* 21, 149–154

80 Meredith, M.A. and Stein, B.E. (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221, 389–391

81 Laurienti, P.J. *et al.* (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp. Brain Res.* 166, 289–297

82 Bertelson, P. (1998) Starting from the ventriloquist: The perception of multimodal events. In *Advances in Psychological Science. Vol. 2: Biological and Cognitive Aspects* (Sabourin, M. and Fergus, C., eds), pp. 419–439, Psychology Press

83 Pick, H.L. *et al.* (1969) Sensory Conflict in Judgements of Spatial Direction. *Percept. Psychophysiol.* 6, 203–205

84 Platt, B.B. and Warren, D.H. (1972) Auditory Localization: The importance of Eye-Movements and a textured visual environment. *Percept. Psychophysiol.* 12, 245–248

85 Connor, S. (2000) *Dumbstruck: A cultural history of Ventriloquism*, Oxford University Press

86 Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222

87 Serences, J.T. and Boynton, G.M. (2007) Feature-Based Attentional Modulations in the Absence of Direct Visual Stimulation. *Neuron* 55, 301–312

88 Theeuwes, J. *et al.* (2000) On the Time Course of Top-Down and Bottom-Up Control of Visual Attention. In *Attention and Performance* (18) Monsel, S. and Driver, J.,eds In pp. 105–124, MIT Press

89 Yantis, S. and Jonides, J. (1984) Abrupt Visual Onsets and Selective Attention - Evidence from Visual-Search. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 601–621

90 Yantis, S. and Serences, J.T. (2003) Cortical mechanisms of space-based and object-based attentional control. *Curr. Opin. in Neurobiol.* 13, 187–193

91 Woldorff, M.G. *et al.* (2004) Functional Parcellation of Attentional Control Regions of the Brain. *J. Cogn. Neurosci.* 16, 149–165

92 Corbetta, M. and Shulman, G.L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215

93 Moore, T. *et al.* (2003) Visuomotor origins of covert spatial attention. *Neuron* 40, 671–683

94 Grent-'t-Jong, T. and Woldorff, M.G. (2007) Timing and sequence of brain activity in top-down control of visual-spatial attention. *PLoS Biol.* 5, 0114–0126

95 Kim, Y.H. *et al.* (1999) The large-scale neural network for spatial attention displays multifunctional overlap but differential asymmetry. *NeuroImage* 9, 269–277

96 Wu, C.T. *et al.* (2007) The neural circuitry underlying the executive control of auditory spatial attention. *Brain Res.* 1134, 187–198

97 Eimer, M. and Driver, J. (2001) Crossmodal links in endogenous and exogenous spatial attention: Evidence from event-related brain potential studies. *Neurosci. Biobehav. Rev.* 25, 497–511

98 McDonald, J.J. *et al.* (2003) Neural substrates of perceptual enhancement by cross-modal spatial attention. *J. Cogn. Neurosci.* 15, 10–19

99 Eimer, M. *et al.* (2001) Cross-modal links in endogenous spatial attention are mediated by common external locations: Evidence from event-related brain potentials. *Exp. Brain Res.* 139, 398–411

100 Hillyard, S.A. *et al.* (1998) Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1257–1270

101 Khayat, P.S. *et al.* (2004) Visual information transfer across eye movements in the monkey. *Vision Res.* 44, 2901–2917

102 Talsma, D. *et al.* (2006) Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology* 43, 541–549

103 Klemen, J. *et al.* (2009) Perceptual load interacts with stimulus processing across sensory modalities. *Eur. J. Neurosci.* 29, 2426–2434

104 de Jong, R. *et al.* (2010) Dynamic crossmodal links revealed by steady-state responses in auditory-visual divided attention. *Int. J. Psychophysiol.* 75, 3–15