BRILL *Multisensory Research (2016)* DOI:10.1163/22134808-00002528 brill.com/msr

The Curious Incident of Attention in Multisensory Integration: Bottom-up *vs.* Top-down

Emiliano Macaluso [1,*], **Uta Noppeney** [2,*], **Durk Talsma** [3,*], **Tiziana Vercillo** [4,*], **Jess Hartcher-O'Brien** [5,**,***] **and Ruth Adam** [6,**,***]

[1] Neuroimaging Laboratory, Fondazione Santa Lucia, Rome, Italy
[2] Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, UK
[3] Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium
[4] Department of Psychology, University of Nevada, Reno, NV, USA
[5] Sorbonne Universites, UPMC Univ Paris, 06, UMR 7222, ISIR, F-75005, Paris, France
[6] Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität LMU, Munich, Germany

Received 23 September 2015; accepted 24 January 2016

Abstract

The role attention plays in our experience of a coherent, multisensory world is still controversial. On the one hand, a subset of inputs may be selected for detailed processing and multisensory integration in a top-down manner, i.e., guidance of multisensory integration by attention. On the other hand, stimuli may be integrated in a bottom-up fashion according to low-level properties such as spatial coincidence, thereby capturing attention. Moreover, attention itself is multifaceted and can be described *via* both top-down and bottom-up mechanisms. Thus, the interaction between attention and multisensory integration is complex and situation-dependent. The authors of this opinion paper are researchers who have contributed to this discussion from behavioural, computational and neurophysiological perspectives. We posed a series of questions, the goal of which was to illustrate the interplay between bottom-up and top-down processes in various multisensory scenarios in order to clarify the standpoint taken by each author and with the hope of reaching a consensus. Although divergence of viewpoint emerges in the current responses, there is also considerable overlap: In general, it can be concluded that the amount of influence that attention exerts on MSI depends on the current task as

[*] Equal contribution (ordered alphabetically).
[**] Equal contribution.
[***] To whom correspondence should be addressed.
E-mail: hartcher@isir.upmc.fr; Ruth.Adam@med.uni-muenchen.de

DOI:10.1163/22134808-00002528

well as prior knowledge and expectations of the observer. Moreover stimulus properties such as the reliability and salience also determine how open the processing is to influences of attention.

Keywords
Bayesian causal inference, endogenous, predictive coding, salience, stimulus-driven

1. Introduction

The interplay between attention and multisensory integration (MSI) is a complex and controversial topic. This may be due, in part, to the fact that attention and MSI interact at multiple levels. Moreover, both attention and MSI are complex, multifaceted processes that contribute to the control of sensory processing and, ultimately, to behaviour. In the current context MSI is defined as the merging of information across two or more sensory modalities in order to obtain a coherent, robust percept. MSI describes the interaction between sensory signals: first, when sensory signals are redundant and second when there is sensory combination with non-redundant cues. Redundant sensory signals arise from within the same coordinate system (e.g., both visual and auditory information can be transformed into craniotopic coordinates) and pertain to the same environmental property (e.g., Ernst and Bülthoff, 2004), whereas sensory combination refers to multisensory interactions for sensory signals that are not redundant, may be coded in different coordinate systems and have potentially different units (e.g., Hecht and Reiner, 2008). Both processes are referred to in the current discussion under the umbrella of MSI. Attention is primarily defined as a guiding process in which relevant inputs are being selected for detailed processing and perceptual awareness out of the inflow of all incoming information (Adam *et al.*, 2014; Marois and Ivanoff, 2005; Talsma *et al.*, 2010). Top-down, endogenous attention can be voluntarily allocated toward a stimulus, a sensory modality or a specific region of space in order to achieve task goals (Li *et al.*, 2004; Spence and Driver, 2004; Wolfe *et al.*, 2003). Attention can also be involuntarily captured 'bottom-up' by external events, even though the attention capturing signals are unrelated to the current goal-directed activity (Öhman *et al.*, 2001; Wolfe *et al.*, 2003; Zhang *et al.*, 2012).

The neural mechanisms that underlie endogenous and stimulus-driven processes have been studied extensively in the visual modality. In the field of visuo-spatial attention control, a relatively straightforward view concerns the distinction between endogenous (internal) control in the dorsal fronto-parietal regions and stimulus-driven (external) control in the right ventral fronto-parietal network (Corbetta and Shulman, 2002). These two attentional control systems are thought to work together influencing the 'responsiveness' of the occipital visual cortex (sensory modulation), e.g., by boosting the processing of visual stimuli at the attended location, and controlling the orienting of attention towards relevant and/or unexpected visual stimuli (Corbetta *et al.*, 2008). Several imaging studies indicated that these two control systems also operate in situations involving non-visual stimuli. For example the dorsal fronto-parietal network has been found to be activated when subjects focused endogenous attention to discriminate either auditory or tactile targets (Hill and Miller, 2010; Krumbholz *et al.*, 2009; Macaluso *et al.*, 2003; Yantis *et al.*, 2002); while the ventral network was found to be activated when participants re-oriented attention to discriminate these targets presented at an unattended location (Downar *et al.*, 2000; Macaluso *et al.*, 2002a). The findings of modality independent responses in the fronto-parietal attention networks is consistent with supramodal mechanisms of attentional control (Farah *et al.*, 1989; Macaluso and Driver, 2005), which provides us with a first link between attention and the processing of multisensory stimuli.

The interaction between MSI and attention has previously been explained both in terms of bottom-up and top-down mechanisms. According to the account of pre-attentive automatic integration, stimuli are integrated spontaneously at the early stage of processing and this integration itself may capture attention. The audio-visual ventriloquist illusion, in which a spatially discrepant sound is perceived to arise from the vicinity of a synchronous visual stimulus, exemplifies integration that is independent of both endogenous (Bertelson *et al.*, 2000) as well as exogenous unisensory attention (Vroomen *et al.*, 2001a). This illusion further enhances spatial attention to speech sounds (Driver, 1996; though see Jack *et al.*, 2013), again suggesting that multisensory binding has occurred automatically and before auditory attentive selection. Similarly, Van der Burg and colleagues (2008) have demonstrated that a sound decreases search times for a synchronized visual object and that detection accuracy is related to an early ERP component (Van der Burg *et al.*, 2011), supporting the idea that the automatic integration of multisensory stimuli can recruit attention. Furthermore, sounds can capture visual attention in cases of limited resources as demonstrated with the attentional blink paradigm (Olivers and Van der Burg, 2008).

Alternatively, attention can limit or boost MSI, even at relatively early processing stages (Karns and Knight, 2009; Senkowski *et al.*, 2005). Attending to an object feature in one modality can direct attention to another modality (Busse *et al.*, 2005; Molholm *et al.*, 2007) and the attentional focus of subjects affects the unisensory weights and extent of integration with, e.g., multimodal attention, as opposed to attending to a single modality, facilitating integration (Oruc *et al.*, 2008; Vercillo and Gori, 2015; although see Bertelson *et al.*, 2000). Also, high level processes such as task goals (Donohue *et al.*, 2015) or prior knowledge (Adam and Noppeney, 2014) can enhance integration. On the other hand, the McGurk effect, an illusory auditory perception generated by in-congruent audio-visual speech stimuli, is considerably reduced by a secondary task suggesting that the high attentional load reduces multisensory processing (Alsius *et al.*, 2005, 2007).

At first sight, these current findings are not consistent and even appear contradictory. We have asked four researchers, Emiliano Macaluso, Uta Noppeney, Durk Talsma and Tiziana Vercillo, who have contributed to research in this field and participated in the IMRF 2015 symposium 'The Curious Incident of Attention in Multisensory Integration: Bottom-up and Top-down' to provide their opinions on this issue. The question and answer format of the current paper was designed to allow different perspectives on attention's role in MSI to be brought together. Specifically, we have restricted the discussion to the role of attention on MSI and attention's modulatory elements in the non-chemical senses.

2. The Role of Attention on MSI

Question 1. What kind of role does attention play in MSI and how much of this role can be accounted for by low-level perceptual processes and how much by top-down influences?

TV: The relation between attention and MSI is complex and results from the interaction between top-down attentional modulations of multisensory processing and bottom-up attentional capture from automatically integrated multisensory inputs. Indeed, on the one hand concurrent sensory stimuli tend to be automatically integrated and processed to form a single coherent percept (Bertelson *et al.*, 2000; Vroomen *et al.*, 2001a) highlighting the importance of bottom-up processes for multisensory integration. On the other hand, several studies have reported top-down effects of attention on multisensory perception, for example factors such as the specific task goal (Donohue *et al.*, 2015) can enhance integration. It is fairly clear that attention and MSI affect each other, both at the level of behavioural outcome and neural processing as both MSI and attention are characterized by multiple mechanisms that occur at different stages of processing.

Talsma *et al.* (2010) proposed that the stimulus complexity of the environment, and particularly the ongoing competition between the stimulus components within it, determines the nature and directionality of these interactions. For instance, these authors have suggested that MSI tends to occur automatically and pre-attentively. However, the modulatory effect of top-down attention seems to be required when multiple stimuli with low saliency within each modality are competing for processing resources. Another possibility is that attentional resources are required to integrate near threshold stimuli while the integration of supra-threshold stimuli may occur automatically and pre-attentively.

EM: The lack of a detailed understanding of the many sub-processes involved in attention control and in MSI is a major obstacle for the understanding of the interactions between these two processes. Nonetheless, here I will attempt to provide a conceptual framework of where to place attention and MSI, and my answers will be within this framework. The basic notion is simple and well acknowledged in the attention literature: the external world stimulates the brain with a vast amount of sensory input and some mechanism(s) must decide to what extent each signal will be processed and, eventually, contribute to determining behaviour. Moreover, any such 'decision' must take into account not only the external input, but also signals that are generated internally and that reflect information stored within the brain. In this framework we can draw a distinction between bottom-up stimulus-driven processes (i.e., related to the external input) and endogenous effects (i.e., related to information stored in the brain).

In this framework one may consider any multisensory input as a source of stimulus-driven attentional control. That is, the presentation of any sensory input will always generate some stimulus-driven attentional signal, which can be linked to the activation of the sensory cortices responding to the stimulation. As noted in the introduction, endogenous attention also contributes to the activity of these regions (see also Kastner *et al.*, 1999) and, therefore, the level of activity in these sensory areas can be interpreted as the outcome of the combined effects of endogenous and stimulus-driven signalling. Concerning MSI, we may ask whether multisensory stimuli can influence activity within these early stages of processing or, rather, MSI requires that multisensory signals propagate extensively in the brain with interactions taking place in higher-level associative regions.

We can think about two extreme examples. One case would involve the sensory input activating only relatively low-level sensory areas, without any effect on behaviour and no interactions between the two modalities in any area of the brain. I would argue that even in this condition the multisensory input is generating some stimulus-driven attentional control signal; that is, the signal represented within the sensory areas. However these signals do not travel much in the brain, for example because the subject is focusing voluntary attention to some other stimulus. The opposite case would involve situations when the interaction between the multisensory stimuli is so relevant that it ends up determining behaviour. This may involve conditions with near-threshold stimuli that can be detected only when attention is fully focused on the stimuli and where specific aspects of the multisensory stimuli (e.g., spatial and/or temporal alignment) determine what the subject perceives. In this case the sensory signals travel extensively in the brain, interacting with each other in many different brain regions, including low-level sensory areas and high-level associative areas involved in attentional selection, decision making and possibly motor control (e.g., Fairhall and Macaluso, 2009; Noppeney *et al.*, 2010). I assume that in most experiments, as well as many everyday life situations, the degree of interaction between multisensory signals will be somewhere between these two extreme cases. This may involve MSI contributing to the allocation of processing resources, without fully governing subjective perception and behaviour. This provides us with a vast range of possibilities to investigate the interplay between attention and MSI.

But what would be the specific contribution of endogenous *versus* stimulus-driven signalling in any given multisensory situation? I believe that a major issue here is that, while it is relatively easy to experimentally manipulate any stimulus-driven effect (e.g., by changing the physical characteristics of the stimuli delivered to the subject), it can be difficult to firmly establish what endogenous signals may be associated with any specific experimental setup. One way to experimentally manipulate endogenous attention involves using dual task procedures, where one can change the load/difficulty of a primary task (e.g., a central discrimination task), while — at the same time — asking the participants to process some multisensory stimulation (Alsius *et al.*, 2005; Santangelo *et al.*, 2009; see also Zimmer and Macaluso, 2007). Studies using this approach have shown a variety of results, ranging from the modulation of behavioural effects under high-load conditions (Alsius *et al.*, 2005) to no effect of load on multisensory processing (Zimmer and Macaluso, 2007). Many different factors may have contributed to these differences (see also Spence, 2010), but I believe that one point to notice is that it is very difficult to know what strategy the participants use in these dual task conditions. Participants may systematically prioritize one or another task, switch between tasks or attempt to perform the two tasks in parallel (see Fischer and Plessow, 2015 for a recent review on cognitive control during dual-task performance). Another situation that is likely to involve endogenous control to an extent that is difficult to monitor relates to the use of stimulus material tapping into pre-existing associations. Examples of this would include studies using audio-visual speech, or real objects and their associated sounds. The role of such ‘content/semantics-related’ associations has been the target of many studies and it is well acknowledged in the MSI literature (Doehrmann and Naumer, 2008). Nonetheless, unlike other low-level stimulus characteristics (e.g., timing, position, etc.) these effects rely on pre-existing ‘internal knowledge’ and it is difficult to exactly know whether/how the participants make use of this knowledge to strategically address and solve any specific task.

UN: The extent to which attention influences MSI remains controversial. According to the account of pre-attentive automatic integration MSI increases the bottom-up stimulus saliency. This account is supported by a vast body of neurophysiological or imaging research demonstrating MSI in anaesthetized non-human primates (e.g., superior colliculus, primary sensory areas — e.g., Kayser *et al.*, 2005; Stanford *et al.*, 2005; Stein and Meredith, 1993). It is also supported by psychophysics studies in humans suggesting that the ventriloquist illusion is immune to endogenous and exogenous spatial attention (Bertelson *et al.*, 2000; Vroomen *et al.*, 2001a) and induces a ‘pre-attentive’ mismatch negativity response (Stekelenburg *et al.*, 2004) comparable to a truly displaced sound. Yet, despite this extensive evidence for automaticity of MSI, more recent studies have also revealed profound influences of attention on MSI. Thus, modality-specific attention was shown to alter the sensory weights in audio-visual integration (Vercillo and Gori, 2015; but see Helbig and Ernst, 2008). Moreover, the McGurk illusion falters when attention is diverted to a secondary task (Alsius *et al.*, 2005; Munhall *et al.*, 2009). Particularly, fMRI and EEG research in humans have highlighted attentional influences on MSI (Talsma *et al.*, 2007, 2010). Attention modulated the amplification of the blood oxygenation level dependent (BOLD) response for congruent audio-visual (AV) speech signals in superior colliculi, primary sensory and association cortices (Fairhall and Macaluso, 2009). Similarly, in EEG AV interactions (e.g., saliency effects) were influenced by modality-specific (or spatial) attention already at ⩽100 ms poststimulus (Talsma and Woldorff, 2005; Talsma *et al.*, 2007). A concurrent visual signal presented in one hemifield induced a lateralization of auditory event-related potentials (ERPs) as a function of spatial attention (Donohue *et al.*, 2011).

Collectively, these seemingly contradicting results suggest a complex relationship between attention and MSI. For instance, MSI in primary sensory areas and superior colliculus may automatically integrate signals to increase their bottom-up salience, which is critical for detection tasks. By contrast, attention is likely to play a critical role in higher order association areas where signals are predominantly integrated into multisensory representations (e.g., spatial or semantic representation that provide estimates to the where and what questions), which are important for estimation, discrimination or categorization tasks (Macaluso and Driver, 2005; Werner and Noppeney, 2010).

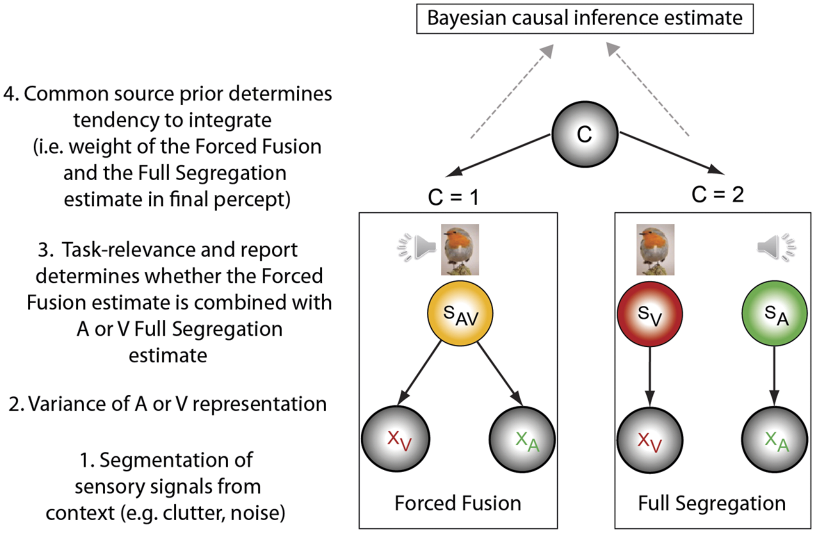
In the following, I will explain how attention may influence this latter representational integration from the perspective of Bayesian causal inference (Körding *et al.*, 2007; Rohe and Noppeney, 2015a, b; Shams and Beierholm, 2010). Bayesian causal inferences has recently been proposed as a normative model that describes how the brain should integrate and segregate sensory signals in the face of uncertainty about the causal structure of the world. Basically, the brain needs to solve two computational challenges: first, it needs to determine which signals emanate from a common source and should be integrated based on them co-occurring in time (e.g., temporal synchrony) and space (e.g., spatial coincidence). Second, it needs to integrate signals from a common source into a statistically optimal percept by weighting them in proportion to their reliabilities. Bayesian causal inference proposes that an ideal observer solves this problem by computing several perceptual estimates. More specifically, it computes perceptual estimates under the forced fusion (i.e., signals being caused by a common event) and full segregation assumptions (i.e., signals being caused by independent events) and finally combines these two estimates into a final so-called Bayesian causal inference estimate.

Attention and task relevance may influence this process *via* multiple mechanisms:

First, attention may facilitate the segmentation of sensory inputs from background clutter and the abstraction of representations (e.g., spatial, phonological, semantic) from the unisensory inputs, which may be a critical prerequisite for representational integration across the senses (Olivers and Van der Burg, 2008). For instance, in the McGurk illusion attention may enable the brain to extract phonological information from the visual facial movements (i.e., visemes), which can then in turn influence auditory speech recognition (Alsius *et al.*, 2005).

Second, modality-specific or spatial attention may increase the reliability of the attended auditory or visual inputs by reducing sensory noise *via* gain modulatory mechanisms. This will in turn change the relative weights of the sensory inputs when being integrated into a coherent percept.

Third, the task-relevance of a sensory modality determines whether the forced fusion estimate is combined with the auditory or the visual full segregation estimate into the final Bayesian causal inference estimate. For instance, when sound location is attended and reported, the observer will combine the forced fusion audio-visual estimate with the auditory estimate under the assumption of full segregation. By contrast, when visual location is attended and reported, the observer will combine the forced fusion audio-visual estimate with the pure visual estimate under the assumption of full segregation. Collectively, this will lead to a different influence of the auditory or visual signal of the perceived stimulus location. In other words, because the brain combines the forced fusion estimate with the auditory estimate for auditory attention and with the visual estimate for visual attention, auditory and visual attention conditions will be associated with different estimates (see Fig. 1 for a graphical explanation).


Bayesian causal inference estimate
4. Common source prior determines tendency to integrate (i.e. weight of the Forced Fusion and the Full Segregation estimate in final percept)
3. Task-relevance and report determines whether the Forced Fusion estimate is combined with A or V Full Segregation estimate
2. Variance of A or V representation
1. Segmentation of sensory signals from context (e.g. clutter, noise)
C
C = 1
C = 2
S_AV
S_V
S_A
X_V
X_A
X_V
X_A
Forced Fusion
Full Segregation

Adapted from Rohe and Noppeney (2015a)

Figure 1. The role of attention in multisensory integration and segregation from the perspective of Bayesian causal inference. Given the uncertainty about the causal structure of the world the observer may compute a full segregation estimate under the assumption of independent sources and a forced fusion estimate under the assumption of one common source. The final Bayesian causal inference estimate takes into account the uncertainty about the causal structure of the world by averaging the task-relevant unisensory auditory (S_{A}) and visual (S_{V}) estimates under full segregation ($C = 2$) with the forced-fusion estimate (S_{AV}) under full integration ($C = 1$), weighted by the posterior probability of each causal structure (for a common source: $p(C = 1|x_{\mathrm{A}}, x_{\mathrm{V}})$; and for independent sources: $1 - p(C = 1|x_{\mathrm{A}}, x_{\mathrm{V}})$. This figure is published in colour in the online version.

Fourth, modality-specific attention (e.g., report the location of the sound) may generally reduce participants' prior assumptions that two signals are caused by a common source. As a consequence, the influence of the forced fusion estimate in their final Bayesian causal inference estimate is reduced. In other words, modality-specific attention will attenuate MSI and make participants rely more on the unisensory estimates (modality-specific attention may modify participants' common source prior or expectations, which in turn influences their integration tendency, see Fig. 1).

DT: There are probably at least two aspects relevant to addressing this question. The first concerns the question of how attention can influence sensory processing. The second concerns the question to what degree selective attention and MSI fulfil similar roles. Several lines of research that can be dated back to the 1970s have indicated that selective attention can influence sensory processes. For instance, event-related potential (ERP) studies on both auditory and visual perception have shown that attention modulates the ongoing sensory processes at several processing stages, including relatively early stages of sensory analysis (Hink *et al.*, 1977; Picton and Hillyard, 1974). For the visual modality, spatial selective attention has been found to enhance the relatively early latency P1 and N1 ERP components (Hillyard *et al.*, 1998), suggesting that attention serves as a gain controller of early sensory processes. For the auditory modality, top-down selective attention has been found to affect either the early latency N1 component itself (Picton and Hillyard, 1974) or an early latency endogenous component that partially overlaps the N1 component, which was labelled 'processing negativity' (Näätänen, 1982). Though the exact functional role of these ERP modulations has subsequently been debated (Luck *et al.*, 1994), this work does indicate that attention can modulate early sensory processing, and thus contribute to enhancing perceptual clarity and reducing stimulus ambiguity.

The question of whether and how attention relates to MSI dates back a number of decades. For instance, Bertelson and colleagues (Bertelson *et al.*, 2000; Vroomen *et al.*, 2001a, b) have argued that selective attention does not influence the ventriloquist effect. Likewise, Driver (1996) argued that the enhancement of speech perception through lip reading is essentially a pre-attentive process. This notion has subsequently been challenged by several findings. For instance Alsius and colleagues (2005) showed that the McGurk effect is sensitive to task-demands (i.e., top-down attention). Concurrently, we (Talsma and Woldorff, 2005) showed that top-down selective attention affected several event-related brain potentials related to multisensory processing. Although the former two studies show that attention can affect MSI, it should also be noted that (pre-attentive) multisensory interactions can also influence attentional processes. For instance, Van der Burg *et al.* (2008) demonstrated that spatially uninformative sounds could increase the saliency of visual stimuli in a visual search task. Taken together, these findings highlight that multisensory processing and attention interact in a complex multifaceted manner (Talsma *et al.*, 2010).

Although much research is still needed to elucidate the finer details, the interaction between attention and multisensory processing can possibly be explained by adopting the predictive coding framework (see Talsma, 2015 for a recent review). According to this framework stochastic models of the environment exist somewhere in the brain, which are updated on the basis of processed sensory information (see Klemen and Chambers, 2012 for a review). These stochastic models thus provide the brain areas lower in the sensory processing hierarchy with predictions that can adjust the processing of ongoing sensory input. A strong mismatch between the prediction and the actual sensory input will then result in a major update of the internal model. Viewed within this context, the internal representation of our external environment is constantly updated on the basis of sensory input (i.e., through feedforward connections) and sensory processing is updated on the basis of predictions provided by these active representations (i.e., through feedback connections). We can therefore argue that feedback connections from the higher-order to the lower-order brain areas embody the causal structure of the external world while anatomical feedforward connections provide feedback regarding prediction errors to higher areas. Anatomical forward connections are thus functional feedback connections, and *vice versa* (Friston, 2005). Prediction errors will then result in strong adjustments in the internal representation and thus in strong top-down functional feedforward (or anatomical feedback) signals. Seen this way, attention could be considered a form of predictive coding; a process that establishes an expectation of the moments in time when the relevant, to be integrated stimulus inputs are to arrive (Klemen and Chambers, 2012).

Summary: We all agree that the interplay between attention and MSI is manifold and that there is large body of evidence for both bottom-up and top-down influences. The conclusion to this question is that the role attention plays in MSI is situation-dependent and definition-dependent. First, before addressing the nature of the interactions between attention and MSI, a deeper understanding of each of these processes in isolation is required. Second, the observed interactions and contradictory findings can be explained by considering the exact stimuli (e.g., integration of near-threshold stimuli require attention), the task goals (e.g., semantic task will involve top-down attentional influences) as well as the exact brain region which is being investigated (e.g., brain regions which are lower at the cortical hierarchy will exhibit more bottom-up like processes) (see Fig. 2 for an illustration of how the factors raised in the current discussion modulate the effect of attention on MSI). Bayesian causal inference as well as predictive coding provide the computational framework to address the interplay between those two complex processes and additional experiments, especially those which will manipulate top-down attention, are still required.

3. Modulatory Factors

Question 2. Does the role of attention in MSI change according to the encountered modalities, for example, that one modality will capture attention more than the others?

TV: Yes, the role of attention in modulating MSI can be affected by the physical attributes of the sensory events.

Van der Burg and colleagues (2010) reported that auditory stimuli can improve visual target identification when presented synchronous with the visual stimulus, supporting the idea that the auditory modality can attract spatial attention in a bottom-up process. In the orthogonal cuing paradigm (Spence

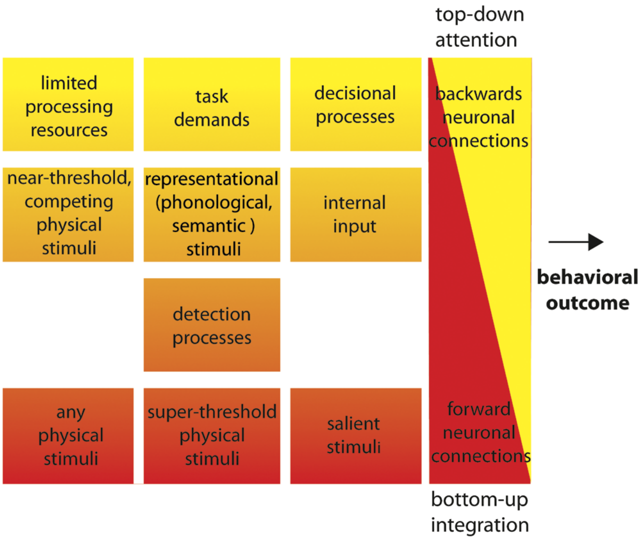

top-down
attention
limited processing resources
task demands
decisional processes
backwards neuronal connections
near-threshold, competing physical stimuli
representational (phonological, semantic) stimuli
internal input
detection processes
any physical stimuli
super-threshold physical stimuli
salient stimuli
forward neuronal connections
bottom-up
integration
behavioral outcome

Figure 2. The various factors which influence the interplay between multisensory integration and attention. The bidirectional influence of multisensory integration and attention is determined by the input, task at hand and cognitive state. This figure is published in colour in the online version.

and Driver, 1994), where uninformative lateralized auditory or visual cues are followed by an unpredictable target, Mazza *et al.* (2007) found faster discriminations for visual targets following ipsilateral auditory, suggesting that peripheral auditory stimuli vigorously capture attention. However, further investigation revealed that the cuing effect, supposed to be an automatic and pre-attentive process, is apparently reduced when spatial attention is focused elsewhere (Santangelo *et al.*, 2007; Spence, 2010; Van der Lubbe and Postma, 2005) suggesting that exogenous orienting of spatial attention might act as a truly automatic mechanism only under certain specific circumstances (Santangelo and Spence, 2008). The top-down attentional modulation, as in the case of high perceptual load, can reduce the effect of multisensory stimuli on exogenous spatial attention.

DT: Several influential biologically plausible theories, such as the framework developed by Corbetta and Shulman (2002), propose that attention operates by modulating the sensitivity of neurons in perceptual areas. More direct evidence for such modulations has been provided using single cell recordings in nonhuman primates (Motter, 1993). MSI is — at least in part — believed to operate on the basis of nonlinear responses of specific neurons in the superior colliculus (Stanford *et al.*, 2005; but see Holmes, 2009). Interestingly, temporally and spatially aligned sensory inputs in different modalities have a higher likelihood to be favoured for further processing, and thus to capture an individual's attention, than do stimuli that are not aligned (Driver, 1996; Stein *et al.*, 2004; Van der Burg *et al.*, 2008). This indicates that attention tends to orient more easily towards sensory input that possesses multisensory properties. These results suggest that MSI and attention operate on the basis of similar modulatory principles that appear to regulate the firing rate of perceptual neurons. Although it is currently unknown how these modulatory processes generalize across modalities, results from cross-modal attention studies suggest that spatial attention tends to be directed in a modality coordinated fashion (Eimer and Driver, 2001; Koelewijn *et al.*, 2009; McDonald *et al.*, 2003; Spence and Driver, 1996, 2004 for a review).

EM: Following my earlier proposal that MSI is part of the mechanism that determines the representation and the propagation of sensory signals in the brain, one way to address the issue of 'modality specificity' is to consider the characteristics of the neural representations associated with the different modalities. These sensory representations will not only determine 'what' about a given sensory input is registered in the brain, but will also constrain what type of multisensory interactions can take place. Each sensory modality is tuned to specific 'features' and the corresponding sensory brain areas are organized specifically to process these features. Examples of this concern the existence of specialized areas to process colour in the visual cortex (Lueck *et al.*, 1989) or the tonotopic organization of the auditory cortex (Langers *et al.*, 2007). Such neural machinery specialized to process modality-specific features implies that stimulus-driven and endogenous signalling about a given feature can exist only in a subset of sensory modalities, which in turn will constrain possible crossmodal and attentional effects based on that feature. For example, Matusz and colleagues (2015) showed that presenting an irrelevant red-shape together with the spoken word 'red' can influence search performance, indicating that the semantic level audio-visual crossmodal correspondence can affect the deployment of spatial attention. While these effects will entail a strong contribution of endogenous signalling related to semantic knowledge, such effects must also rely on the visual brain processing colour information, so that this can interact with any information arising from the auditory modality.

More subtle specificities may be expected when the crossmodal interactions involve spatial or temporal information. The precision to encode the location of the stimuli is higher for vision that for audition, while the opposite applies to the processing of temporal information (Recanzone, 2009). These differences can have an impact on how the two modalities interact with each other, generating various types of asymmetries in spatial/temporal attention and spatial/temporal judgment tasks. While, early behavioural studies suggested that visual or auditory non-predictive spatial cues can speed up responses to lateralized visual targets but not to auditory targets (Buchtel and Butter, 1988), further investigation of these asymmetries demonstrated that specific aspects of the task — such as the relevance of spatial information — can crucially determine whether specific crossmodal cueing effects do or do not occur (e.g., McDonald and Ward, 1999). The latter highlights the role of secondary task-features and endogenous strategic factors in these seemly 'pure' stimulus-driven paradigms (cf. interplay between endogenous and stimulus-driven signalling, above; see also Ward *et al.*, 2000). These differences in 'modality specificity' have been formally accounted for by Bayesian models that weight the contribution of each modality to multisensory processing according to the reliability of the unisensory input (Alais and Burr, 2003; Ernst and Banks, 2002). Nonetheless these models are primarily based on stimulus-driven characteristics of the sensory input, while as noted in the sections above, in most real-life situations knowledge and expectations are likely to play a relevant role. In the case of multisensory processing, expectations may refer to whether two unisensory input are caused by a single (multisensory) source or two separate unisensory sources. These aspects can be formally accounted for by models of 'causal inference' (Shams and Beierholm, 2010) that have been recently applied both to behavioural and to neuroimaging data of audio-visual processing (Rohe and Noppeney, 2015a, b).

In sum, the characteristics of the sensory representations of the different modalities in the brain are likely to play a major role in how multisensory signals interact with each other, as well as constraining any effect of these on mechanisms of attentional control. In a stimulus-driven perspective, the representations of specific sensory features (e.g., colour for vision) and the accuracy/precision of the sensory representations (e.g., spatial resolution) will determine what kind of multisensory interactions can take place and how the signals in the different modalities will be weighted upon multisensory stimulation (e.g., dominance of one or another modality). I believe that the characteristics of the sensory representations also contribute to shaping the connectivity between lower-level sensory areas and higher-level associative regions (e.g., the fronto-parietal attention systems), thus determining how signals propagate in the brain. In general, I expect a strong relationship between 'how' a given signal is represented in the brain and the type of attentional effects that we may observe when that signal is presented as a component of a multisensory stimulus.

UN: The role of attention will depend interactively on the sensory modality and the particular representation (e.g., spatial *vs.* phonological) that is being integrated. For instance, speech recognition relies predominantly on the auditory sense in everyday life, while lip reading plays only a facilitatory role in challenging situations such as a noisy pub. Hence, we would expect that the brain can fairly automatically extract phonological information from sound, but requires attentional resources to extract phonological information from lip movements (e.g., visemes). As a consequence attention to vision should be required for the McGurk illusion to occur. Indeed, Alsius *et al.* (2005) have demonstrated that the McGurk illusion falters when attention is diverted to a secondary task. By contrast, in everyday life spatial localization is usually more reliably performed by the visual than the auditory sense. We would therefore expect that the ventriloquist illusion (i.e., the bias of a spatially disparate visual signal on the perceived sound location) will be increased when participants are less attentive and would therefore automatically rely more on the visual senses. Indeed, unpublished anecdotal evidence from our lab seems to suggest that the ventriloquist illusion is enhanced when participants are less attentive. In conclusion, limited attentional resources during demanding tasks impact observers' auditory percept differently in the McGurk and the ventriloquist illusion. The ventriloquist illusion is increased, because participants will 'automatically' rely more on their visual sense for spatial localization. By contrast, the McGurk illusion is decreased, because participants naturally rely more on their auditory sense for speech recognition. In other words, it is not the modality in itself that determines the role of attention in MSI. Instead, the role of attention depends jointly on the sensory modality and the particular representation that needs to be integrated.

Summary: Yes, attentional capture depends on the encountered modalities. In general, stimuli capture attention (bottom-up processes) based on their physical properties such that temporal and/or spatial coincidences of inputs recruit attention. The extent to which attention will be captured by each of the experienced modalities depends on the current task; each modality is tuned to and is more reliable for coding certain properties of the environment. For example, spatial tasks will boost the importance of visual inputs whereas temporal tasks will increase the weight of auditory stimuli. These sensory predispositions and coding differences will affect the way the stimuli are being integrated and can be accounted for by Bayesian models. In cases of task-based top-down attentional selection, or when observers have pre-existing expectations regarding the cause of the multisensory stimuli, the effect of the physical properties on bottom-up attentional capture will be reduced. In addition, the neural representation of the properties of the environment/stimuli can shape the multisensory interactions for attentional control.

Question 3. How much does the effect of attention depend on the stimulus properties (from perceptual inputs such as motion-direction to complex semantic or linguistic information)?

TV: The characteristics of the environment apparently regulate the effect of attention on MSI with temporal and spatial coincidence facilitating integration. In general, sensory events presented close together, e.g., in time, are more likely to be bound together automatically and pre-attentively (Aller *et al.*, 2015). On the other hand, the importance of such coincidence seems to be task dependent. Particularly, spatial proximity seems to be relevant in tasks involving spatial attention and requiring orienting responses (see Spence, 2013 for a review).

However, the salience of the stimuli also plays an important role in MSI, where salient stimuli are usually linked together while competitive stimuli require an attentional modulation (Talsma *et al.*, 2010).

The attentional modulation of MSI is not only affected by the complexity of the environment, but also by the complexity of the stimulus. For instance, complex stimuli such as linguistic information seem to be much more sensitive to the top-down process of attentional modulation. Alsius *et al.* (2005) described that the McGurk illusion is considerably attenuated when participants have to perform a dual task paradigm. In support of the idea that attention has a strong effect on MSI of high level stimuli, Fairhall and Macaluso (2009) have found that spatial attention enhances BOLD response in different brain areas, such as the superior temporal sulcus, the visual cortex and the superior colliculus for audio-visual speech stimuli. Senkowski and colleagues (2008) have found that in a multisensory speech recognition task, where subjects are presented with competing audiovisual stimuli, the shift of visual spatial attention toward distractor stimuli interferes with speech recognition performance. These results support the hypothesis that attention modulates the processing of multisensory speech stimuli. However, it is unclear whether is the complexity of the task or of the stimulus itself to determine the attentional regulation of MSI.

DT: Traditionally, the literature has been divided between studies using relatively simple stimuli, such as beeps and flashes on the one hand, and more complex, meaningful stimuli on the other. Studies using simple, abstract stimuli have predominantly focused on determining stimulus-driven effects, such as their relative timing or location (Noesselt *et al.*, 2010; Stein and Stanford, 2008; Van Wassenhove *et al.*, 2007) or their relative intensity (Holmes, 2009; Rach *et al.*, 2010) on multisensory processing. Furthermore the intrinsic processing capacities of each individual sensory system (Vatakis and Spence, 2007; Welch and Warren, 1980) have been identified to contribute to multisensory processing. The simultaneous and congruent stimulation of two or more senses has been shown to result in increases in brain activity (Calvert *et* *al.*, 2000; Fairhall and Macaluso, 2009), increased physiological signals from brain areas responding to these stimuli (Molholm *et al.*, 2002) or greater selectivity of relevant stimulus material (Staufenbiel *et al.*, 2011; Van der Burg *et al.*, 2008, 2011). Behavioural and event-related potential (ERP) studies have shown that an object that is simultaneously detected by several sensory systems has a greater potential for capturing one's attention (Spence, 2010; Van der Burg *et al.*, 2008, 2011). This further suggests that when a sensory modality is processing a stimulus simultaneously with one presented to another modality, these concurrently presented stimuli have a natural tendency to be processed in greater depth than stimuli that are either non-concurrent in time. Taken together, these results suggest that stimulus driven (or bottom-up) processes have a major influence on multisensory processing.

It should be emphasised, however, that the results discussed above only reflect a small subset of multisensory processing results, namely the ones that have been obtained under conditions where there is relatively little competition for processing resources. Studies using more naturalistic, meaningful stimuli, such as speech fragments and movie clips have indicated that semantic congruence (Cappe *et al.*, 2012; McGurk and Macdonald, 1976; Tuomainen *et al.*, 2005) between visual and auditory stimuli also strongly influences multisensory processing. On the basis of the latter studies, it has been argued that audio-visual speech perception is a special form of multisensory processing (Tuomainen *et al.*, 2005; Vatakis *et al.*, 2008). Given the wide range of discrepancies between these different approaches in multisensory speech perception, however, it remains to be seen whether that is really the case. Regardless, however, the vastly different sets of results that have been obtained using these simple *vs.* complex stimuli indicate that the type of stimulus involved in multisensory processing does affect MSI.

EM: Above, I suggested that the stimulus properties, or better 'how' the stimulus properties are represented in the brain, ought to be a major determinant of crossmodal interactions in attention control. However, one should also consider other situations, where attention may work in a supramodal manner, irrespective of the specific sensory characteristics of the input. One example of this entails the integration of spatial representations across modalities. In a set of imaging experiments (Macaluso *et al.*, 2002b, 2003), we asked participants to direct voluntary attention to one side of space and to discriminate either visual or tactile targets on the attended side. When vision was the relevant modality we found the expected 'within-modality' effect of spatial attention, with activation of the regions of the visual occipital cortex that represent the attended visual hemifield. However, activity in the same regions also increased when the subjects attended to touch on the same side, indicating that the task-related endogenous signal modulating the response in these visual regions conveyed information about the relevant side/location irrespective of the specific modality to be judged (see Eimer and Driver, 2001 for a review of related effects in ERP). These effects suggest that the interplay between attention and multisensory processing enables integrating spatial information across anatomically separated representations of external space (see Macaluso and Driver, 2005). Thus, crossmodal integration may not operate only to 'bind' redundant sensory signals, but supramodal mechanisms of attentional selection can also integrate how 'abstract' spatial information is represented in the brain (see also Macaluso *et al.*, 2003 for relevant crossmodal effects in preparatory attention, i.e., in the absence of any sensory input).

UN: The role of attention will depend on the complexity of the sensory signals, the context in which they are presented and the representation to be integrated. For instance, when auditory signals are presented in complex multi-tone masks during informational masking, attention will then play a critical role to segregate the auditory signal from the complex scene, which is a necessary precondition for it to be integrated with signals other sensory modalities. In these cases, attention is critical even for low-level integration processes that amplify stimulus salience and facilitate detection (Giani *et al.*, 2015; Olivers and Van der Burg, 2008). For stimuli that are more easily segmented from sensory noise, low-level MSI processes based on temporal coincidence may be more automatic. For instance, we observed activation increases in primary sensory areas for synchronous relative to asynchronous stimuli irrespective of task-context. These low-level synchrony effects propagated then into higher order motion and shape areas depending on the attentional context, i.e., whether participants focused on the motion or shape properties of the stimuli. These results suggest that low-level temporal properties of the stimuli may determine MSI in a more automatic fashion, while higher order representational integration (e.g., motion, shape) may be more sensitive to top-down effects (Lee and Noppeney, 2014; Lewis and Noppeney, 2010).

Summary: Stimulus properties do affect the observed interplay between MSI and attention. Temporally and/or spatially coincident *simple* stimuli will induce stimulus-driven, bottom-up influences on attention while *complex* linguistic and semantic inputs will affect MSI *via* top-down attentional control mechanisms. However, it is important to note that stimulus complexity cannot be separated from the complexity of the task and the environment. An environment in which the stimuli are easily discriminated from noise will induce bottom-up effects. A task that directs attention to multiple stimuli can induce top-down effects irrespective of the presented stimuli.

4. Conclusion

In the current discussion we have attempted to characterize the role of attention on MSI to indicate how much of sensory integration can be accounted for by bottom-up stimulus driven factors and how much by top-down processes such as semantic, contextual, and dual task components (Fig. 2). The amount of influence that attention exerts on MSI depends on the task and the goal of the organism, and the predictions and expectations about the encountered stimuli. Moreover stimulus factors such as the reliability (i.e., the inverse of the variability in response to the stimulus) of a stimulus and its salience also determine how open the processing is to influences of attention. Computational models are useful for explaining such intertwined interactions, e.g., Bayesian causal inference. A further important consideration would be to observe multisensory-attention interactions in both well-controlled experiments and more naturalistic settings. We all agree that the interaction between MSI and attention remains a complex issue that requires further investigation. In the following, the final statements of each author as outlined in the discussion will be summarised.

TV: Both bottom-up and top-down processes drive integration depending primarily on the structure of the stimuli, i.e., complex or simple; salience or near-threshold. Top-down attention seems to facilitate integration when multiple stimuli with low saliency within each modality are competing for processing resources, or in case of near-threshold stimuli. Moreover, spatial attention reduces pre-attentive MSI effects. The integration of supra-threshold stimuli may, however, occur automatically and pre-attentively.

UN: Attention affects MSI at multiple processing stages and cortical hierarchical levels. First, it enables different signals coming from a common source to be segmented from clutter and background noise in order to be integrated. Second, attention may increase the bottom-up salience and sensory reliability. Third, from the perspective of Bayesian causal inference, the task-relevance of a sensory modality influences whether the forced fusion estimate is combined with the full segregation of auditory or visual estimates. Finally, multisensory attention may influence participants' tendency to integrate or segregate sensory signals by modulating their prior assumptions of two signals coming from a common source. Conversely, MSI automatically enhances the bottom-up salience thereby enabling sensory signals to grasp participants' attention.

DT: Both MSI and attention modulate the firing rate of perceptual neurons. The predictive coding framework can be used to explain the interaction between MSI and attentional control: attention helps us to shape our expectations regarding the environment and modulates integration accordingly.

EM: I believe that the complexity of the mechanisms controlling the allocation of processing resources makes it difficult to answer the question about the role of attention in MSI. My personal perspective is that the two processes should not be seen as separate entities, but rather they should be considered within a single framework: that of the combination of stimulus-driven and endogenous signalling for the selection of relevant information and control of overt behaviour. This puts the emphasis on understanding the neural mechanisms associated with the processing of multisensory stimuli and how multisensory signals propagate in the brain. The latter will be determined by the type of sensory feature representation, as well as by prior knowledge and goals. The development of new theoretical and mathematical approaches (e.g., Bayesian causal inference, Rohe and Noppeney, 2015a) will help us with the interpretation of these neuronal effects. In addition, I believe that together with sophisticated and well-controlled experiments, it is important to look also into more naturalistic and life-like multisensory conditions. These could reveal aspects of multisensory processing and attentional control that may be concealed in standard experimental paradigms. In particular, I am referring to the role of endogenous signals associated with prior knowledge and expectations. These are likely to play a major role in everyday life situations and may differ from any task-related, strategic signals that characterize standard experiments in the laboratory.

5. Summary

The current aim was to provide insights into whether and how bottom-up factors or top-down modulation characterise the interaction between attention and MSI. While the interaction between MSI and attention has previously been explained in terms of both bottom-up and top-down mechanisms, two primary components emerged from the current discussion as characterizing the outcome of attentional influences on MSI: context (including observer goal and task) and priors, i.e., the knowledge and expectations that the observer has built over development about the current stimuli and their causes. Bottom-up factors featured less strongly in the current discussion and included the stimulus reliability (inverse noise) as well as spatial and temporal co-location of multisensory inputs that modulate the involvement of attention. But how do these factors change our understanding of the role of attention in MSI? We would suggest that the relative sensitivity of MSI to attentional control depends upon the robustness of the sensory features to noise and perturbations in neural processing. When we consider context related factors we directly refer to the way in which patterns and features of the environment (stimulus properties) are coded by the brain as a function of past behavioural success (goal of the observer given the inputs), which in turn build the observer's priors. For example, in the case of the audio-visual ventriloquist illusion, the audiovisual stimulus combination experienced by the observer is integrated because audiovisual signals are often emitted by a single event (prevalent pattern inherent in our environment) in a manner robust to attentional manipulations. Moreover, the magnitude of the observed 'mislocalization of the sound source' depends on how noisy each sensory input is. However, the overall probability of integration may also be affected by the observer's priors, i.e., the previous behavioural success and meaning associated with the current context (see Purves *et al.*, 2011). This can be seen in the assumption of a common cause for spatially co-located stimuli that predicts the success of the ventriloquist illusion. From this example it is easy to see that a number of factors determine the amount of influence attention has on MSI. As the context and goal/reward of the observer change, so does the role of attention in MSI.

Future Directions

1. Does the influence of attention on sensory processing differ for multisensory *vs.* multiple unisensory inputs?
2. What neuronal networks promote the interplay between attention and MSI?
3. Which computation models can best explain the interaction between attention and multisensory integration?
4. How much learning is involved in shaping the role of attention in MSI?
5. Can we generalize from the known role of attention in MSI to other cognitive phenomenon such as emotion and awareness/consciousness?

Acknowledgements

EM was supported by the Italian Ministry of Health (Grant no. GR-2010-2311171). UN was funded by the European Research Council (ERC-multsens).

References

Adam, R. and Noppeney, U. (2014). A phonologically congruent sound boosts a visual target into perceptual awareness, *Front. Integr. Neurosci.* **8**, 70. DOI:10.3389/fnint.2014.00070

Adam, R., Schönfelder, S., Forneck, J. and Wessa, M. (2014). Regulating the blink: cognitive reappraisal modulates attention, *Front. Psychol.* **5**, 143. DOI:10.3389/fpsyg.2014.00143

Alais, D. and Burr, D. (2003). The 'flash-lag' effect occurs in audition and cross-modally, *Curr. Biol.* **13**, 59–63.

Aller, M., Giani, A., Conrad, V., Watanabe, M. and Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness, *Front. Integr. Neurosci.* **9**, 16. DOI:10.3389/fnint.2015.00016

Alsius, A., Navarra, J., Campbell, R. and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands, *Curr. Biol.* **15**, 839–843.

Alsius, A., Navarra, J. and Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration, *Exp. Brain Res.* **183**, 399–404.

Bertelson, P., Vroomen, J., de Gelder, B. D. and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention, *Percept. Psychophys.* **62**, 321–332.

Buchtel, H. A. and Butter, C. M. (1988). Spatial attentional shifts: implications for the role of polysensory mechanisms, *Neuropsychologia* **26**, 499–509.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H. and Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object, *Proc. Natl Acad. Sci. USA* **102**, 18751–18756.

Calvert, G. A., Campbell, R. and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex, *Curr. Biol.* **10**, 649–657.

Cappe, C., Thelen, A., Romei, V., Thut, G. and Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration, *J. Neurosci.* **32**, 1171–1182.

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

Corbetta, M., Patel, G. and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind, *Neuron* **58**, 306–324.

Doehrmann, O. and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration, *Brain Res.* **1242**, 136–150.

Donohue, S. E., Roberts, K. C., Grent-'t-Jong, T. and Woldorff, M. G. (2011). The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events, *J. Neurosci.* **31**, 7982–7990.

Donohue, S. E., Green, J. J. and Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli, *Front. Integr. Neurosci.* **9**, 32. DOI:10.3389/fnint.2015.00032

Downar, J., Crawley, A. P., Mikulis, D. J. and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment, *Nat. Neurosci.* **3**, 277–283.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading, *Nature* **381**(6577), 66–68.

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. DOI:10.1016/S0149-7634(01)00029-X

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature* **415**(6870), 429–433.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept, *Trends Cogn. Sci.* **8**, 162–169.

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.

Farah, M. J., Wong, A. B., Monheit, M. A. and Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: modality-specific or supramodal? *Neuropsychologia* **27**, 461–470.

Fischer, R. and Plessow, F. (2015). Efficient multitasking: parallel *versus* serial processing of multiple tasks, *Front. Psychol.* **6**, 1366. DOI:10.3389/fpsyg.2015.01366

Friston, K. (2005). A theory of cortical responses, *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 815–836.

Giani, A. S., Belardinelli, P., Ortiz, E., Kleiner, M. and Noppeney, U. (2015). Detecting tones in complex auditory scenes, *NeuroImage* **122**, 203–213.

Hecht, D. and Reiner, M. (2008). Sensory dominance in combinations of audio, visual and haptic stimuli, *Exp. Brain Res.* **193**, 307–314.

Helbig, H. B. and Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention, *J. Vis.* **8**, 21. DOI:10.1167/8.1.21

Hill, K. T. and Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening, *Cereb. Cortex* **20**, 583–590.

Hillyard, S. A., Vogel, E. K. and Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence, *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **353**, 1257–1270.

Hink, R. F., Van Voorhis, S. T., Hillyard, S. A. and Smith, T. S. (1977). The division of attention and the human auditory evoked potential, *Neuropsychologia* **15**, 597–605.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations, *Brain Topogr.* **21**, 168–176.

Jack, B. N., O'Shea, R. P., Cottrell, D. and Ritter, W. (2013). Does the ventriloquist illusion assist selective listening? *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1496–1502.

Karns, C. M. and Knight, R. T. (2009). Intermodal auditory, visual, and tactile attention modulates early stages of neural processing, *J. Cogn. Neurosci.* **21**, 669–683.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R. and Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation, *Neuron* **22**, 751–761.

Kayser, C., Petkov, C. I., Augath, M. and Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex, *Neuron* **48**, 373–384.

Klemen, J. and Chambers, C. D. (2012). Current perspectives and methods in studying neural mechanisms of multisensory interactions, *Neurosci. Biobehav. Rev.* **36**, 111–133.

Koelewijn, T., Bronkhorst, A. and Theeuwes, J. (2009). Auditory and visual capture during focused visual attention, *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 1303–1315.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. and Shams, L. (2007). Causal inference in multisensory perception, *PLoS ONE* **2**, e943. DOI:10.1371/journal.pone.0000943

Krumbholz, K., Nobis, E. A., Weatheritt, R. J. and Fink, G. R. (2009). Executive control of spatial attention shifts in the auditory compared to the visual modality, *Hum. Brain Mapp.* **30**, 1457–1469.

Langers, D. R. M., Backes, W. H. and Van Dijk, P. (2007). Representation of lateralization and tonotopy in primary *versus* secondary human auditory cortex, *NeuroImage* **34**, 264–273.

Lee, H. and Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices, *Curr. Biol.* **24**, R309–R310.

Lewis, R. and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination *via* enhanced connectivity between early visual and auditory areas, *J. Neurosci.* **30**, 12329–12339.

Li, W., Piëch, V. and Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex, *Nat. Neurosci.* **7**, 651–657.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P. and Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection, *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 887–904.

Lueck, C. J., Zeki, S., Friston, K. J., Deiber, M.-P., Cope, P., Cunningham, V. J., Lammertsma, A. A., Kennard, C. and Frackowiak, R. S. J. (1989). The colour centre in the cerebral cortex of man, *Nature* **340**(6232), 386–389.

Macaluso, E. and Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain, *Trends Neurosci.* **28**, 264–271.

Macaluso, E., Frith, C. D. and Driver, J. (2002a). Supramodal effects of covert spatial orienting triggered by visual or tactile events, *J. Cogn. Neurosci.* **14**, 389–401.

Macaluso, E., Frith, C. D. and Driver, J. (2002b). Directing attention to locations and to sensory modalities: multiple levels of selective processing revealed with PET, *Cereb. Cortex* **12**, 357–368.

Macaluso, E., Eimer, M., Frith, C. D. and Driver, J. (2003). Preparatory states in crossmodal spatial attention: spatial specificity and possible control mechanisms, *Exp. Brain Res.* **149**, 62–74.

Marois, R. and Ivanoff, J. (2005). Capacity limits of information processing in the brain, *Trends Cogn. Sci.* **9**, 296–305.

Matusz, P. J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R. and Scerif, G. (2015). Multimodal distraction: insights from children's limited attention, *Cognition* **136**, 156–165.

Mazza, V., Turatto, M., Rossi, M. and Umiltà, C. (2007). How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia* **45**, 514–522.

McDonald, J. J. and Ward, L. M. (1999). Spatial relevance determines facilitatory and inhibitory effects of auditory covert spatial orienting, *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 1234–1252.

McDonald, J. J., Teder-Sälejärvi, W. A., Russo, F. D. and Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention, *J. Cogn. Neurosci.* **15**, 10–19.

McGurk, H. and Macdonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**(5588), 746–748.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E. and Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study, *Cogn. Brain Res.* **14**, 115–128.

Molholm, S., Martinez, A., Shpaner, M. and Foxe, J. J. (2007). Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities, *Eur. J. Neurosci.* **26**, 499–509.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli, *J. Neurophysiol.* **70**, 909–919.

Munhall, K. G., MacDonald, E. N., Byrne, S. K. and Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate, *J. Acoust. Soc. Am.* **125**, 384–390.

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

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J. and Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity, *J. Neurosci.* **30**, 13609–13623.

Noppeney, U., Ostwald, D. and Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex, *J. Neurosci.* **30**, 7434–7446.

Öhman, A., Flykt, A. and Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass, *J. Exp. Psychol. Gen.* **130**, 466–478.

Olivers, C. N. L. and Van der Burg, E. (2008). Bleeping you out of the blink: sound saves vision from oblivion, *Brain Res.* **1242**, 191–199.

Oruc, I., Sinnett, S., Bischof, W. F., Soto-Faraco, S., Lock, K. and Kingstone, A. (2008). The effect of attention on the illusory capture of motion in bimodal stimuli, *Brain Res.* **1242**, 200–208.

Picton, T. W. and Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of attention, *Electroencephalogr. Clin. Neurophysiol.* **36**, 191–200.

Purves, D., Wojtach, W. T. and Lotto, R. B. (2011). Understanding vision in wholly empirical terms, *Proc. Natl Acad. Sci. USA* **108**(Suppl. 3), 15588–15595.

Rach, S., Diederich, A. and Colonius, H. (2010). On quantifying multisensory interaction effects in reaction time and detection rate, *Psychol. Res.* **75**, 77–94.

Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time, *Hearing Res.* **258**, 89–99.

Rohe, T. and Noppeney, U. (2015a). Cortical hierarchies perform Bayesian causal inference in multisensory perception, *PLoS Biol.* **13**, e1002073. DOI:10.1371/journal.pbio.1002073

Rohe, T. and Noppeney, U. (2015b). Sensory reliability shapes perceptual inference *via* two mechanisms, *J. Vis.* **15**, 22. DOI:10.1167/15.5.22

Santangelo, V. and Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies, *Consc. Cogn.* **17**, 989–1015.

Santangelo, V., Olivetti Belardinelli, M. and Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged, *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 137–148.

Santangelo, V., Belardinelli, M. O., Spence, C. and Macaluso, E. (2009). Interactions between voluntary and stimulus-driven spatial attention mechanisms across sensory modalities, *J. Cogn. Neurosci.* **21**, 2384–2397.

Senkowski, D., Talsma, D., Herrmann, C. S. and Woldorff, M. G. (2005). Multisensory processing and oscillatory gamma responses: effects of spatial selective attention, *Exp. Brain Res.* **166**, 411–426.

Senkowski, D., Saint-Amour, D., Gruber, T. and Foxe, J. J. (2008). Look who's talking: the deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions, *NeuroImage* **43**, 379–387.

Shams, L. and Beierholm, U. R. (2010). Causal inference in perception, *Trends Cogn. Sci.* **14**, 425–432.

Spence, C. (2010). Crossmodal spatial attention, *Ann. N. Y. Acad. Sci.* **1191**, 182–200.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule, *Ann. N. Y. Acad. Sci.* **1296**, 31–49.

Spence, C. J. and Driver, J. (1994). Covert spatial orienting in audition: exogenous and endogenous mechanisms, *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 555–574.

Spence, C. and Driver, J. (1996). Audiovisual links in endogenous covert spatial attention, *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 1005–1030.

Spence, C. and Driver, J. (2004). *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxford, UK.

Stanford, T. R., Quessy, S. and Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus, *J. Neurosci.* **25**, 6499–6508.

Staufenbiel, S. M., van der Lubbe, R. H. J. and Talsma, D. (2011). Spatially uninformative sounds increase sensitivity for visual motion change, *Exp. Brain Res.* **213**, 457–464.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press, Cambridge, MA, USA.

Stein, B. E. and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron, *Nat. Rev. Neurosci.* **9**, 255–266.

Stein, B. E., Jiang, W. and Stanford, T. R. (2004). Multisensory integration in single neurons of the midbrain, in: *The Handbook of Multisensory Processes*, Vol. 15, G. A. Calvert, C. Spence and B. E. Stein (Eds), pp. 243–264. MIT Press, Cambridge, MA, USA.

Stekelenburg, J. J., Vroomen, J. and de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity, *Neurosci. Lett.* **357**, 163–166.

Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind, *Front. Integr. Neurosci.* **9**, 19. DOI:10.3389/fnint.2015.00019

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

Talsma, D., Doty, T. J. and Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* **17**, 679–690.

Talsma, D., Senkowski, D., Soto-Faraco, S. and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration, *Trends Cogn. Sci.* **14**, 400–410. DOI:10.1016/j.tics.2010.06.008

Tuomainen, J., Andersen, T. S., Tiippana, K. and Sams, M. (2005). Audio-visual speech perception is special, *Cognition* **96**, B13–B22.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W. and Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search, *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1053–1065.

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J. and Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events, *PLoS ONE* **5**, e10664. DOI:10.1371/journal.pone.0010664

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C. and Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects, *NeuroImage* **55**, 1208–1218.

Van der Lubbe, R. H. J. and Postma, A. (2005). Interruption from irrelevant auditory and visual onsets even when attention is in a focused state, *Exp. Brain Res.* **164**, 464–471.

Van Wassenhove, V., Grant, K. W. and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception, *Neuropsychologia* **45**, 598–607.

Vatakis, A. and Spence, C. (2007). Crossmodal binding: evaluating the 'unity assumption' using audiovisual speech stimuli, *Percept. Psychophys.* **69**, 744–756.

Vatakis, A., Ghazanfar, A. A. and Spence, C. (2008). Facilitation of multisensory integration by the 'unity effect' reveals that speech is special, *J. Vis.* **8**, 14. DOI:10.1167/8.9.14

Vercillo, T. and Gori, M. (2015). Attention to sound improves auditory reliability in audio-tactile spatial optimal integration, *Front. Integr. Neurosci.* **9**, 34. DOI:10.3389/fnint.2015.00034

Vroomen, J., Bertelson, P. and de Gelder, B. D. (2001a). The ventriloquist effect does not depend on the direction of automatic visual attention, *Percept. Psychophys.* **63**, 651–659.

Vroomen, J., Bertelson, P. and de Gelder, B. (2001b). Directing spatial attention towards the illusory location of a ventriloquized sound, *Acta Psychol. (Amst.)* **108**, 21–33.

Ward, L. M., McDonald, J. J. and Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting, *Percept. Psychophys.* **62**, 1258–1564.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy, *Psychol. Bull.* **88**, 638–667.

Werner, S. and Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization, *J. Neurosci.* **30**, 2662–2675.

Wolfe, J. M., Butcher, S. J., Lee, C. and Hyle, M. (2003). Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons, *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 483–502.

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J. and Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts, *Nat. Neurosci.* **5**, 995–1002.

Zhang, X., Zhaoping, L., Zhou, T. and Fang, F. (2012). Neural activities in V1 create a bottom-up saliency map, *Neuron* **73**, 183–192.

Zimmer, U. and Macaluso, E. (2007). Processing of multisensory spatial congruency can be dissociated from working memory and visuo-spatial attention, *Eur. J. Neurosci.* **26**, 1681–1691.