Uta Noppeney*, Samuel A. Jones, Tim Rohe and Ambra Ferrari

See what you hear – How the brain forms representations across the senses

https://doi.org/10.1515/nf-2017-A066

Abstarct: Our senses are constantly bombarded with a myriad of signals. To make sense of this cacophony, the brain needs to integrate signals emanating from a common source, but segregate signals originating from the different sources. Thus, multisensory perception relies critically on inferring the world's causal structure (i. e. one common vs. multiple independent sources). Behavioural research has shown that the brain arbitrates between sensory integration and segregation consistent with the principles of Bayesian Causal Inference. At the neural level, recent functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) studies have shown that the brain accomplishes Bayesian Causal Inference by dynamically encoding multiple perceptual estimates across the sensory processing hierarchies. Only at the top of the hierarchy in anterior parietal cortices did the brain form perceptual estimates that take into account the observer's uncertainty about the world's causal structure consistent with Bayesian Causal Inference.

Zusammenfassung: Unsere Sinne werden fortwährend mit den unterschiedlichsten Signalen bombardiert. Um dieses Sinneschaos zu verstehen, muss das Gehirn Sinnesreize integrieren, wenn sie von einer Quelle kommen, aber separate verarbeiten, wenn sie von unterschiedlichen Quellen kommen. Somit beruht multisensorische Wahrnehmung entscheidend auf dem Erfassen des kausalen Struktur, die die Sinnesreize erzeugt hat. Verhaltensstudien legen nahe, dass das Gehirn zwischen Integration and Segregation wie von normativen Modellen der Bayesianischen kausalen Inferenz vorhergesagt abwägt. Neueste funktionelle Magnetresonanztomographie (fMRI) und Elektroenzephaligraphie (EEG) Studien zeigten, dass das Gehirn Bayesianische kausal Inferenz durchführt, indem es mehrere Wahrnehmungsschätzwerte dynamisch auf verschiedenen Ebenen der corticalen Hierarchie der Sinnesverabeitung enkodiert. Erst an der Spitze der Hierarchie in anterioren parietalen Arealen formt das Gehirn Wahrnehmungsschätzwerte, die die Ungewissheit des Beobachters über die kausal Struktur der Umgebung berücksichtigt, wie von Modellen der Bayesianischen kausalen Inferenz vorhergesagt.

Keywords: Perception, multisensory perception, multisensory integration, audiovisual, Bayesian, computational model, perceptual inference, binding problem, causal inference

Schlüsselwörter: Wahrnehmung, multisensorische Wahrnehmung, multisensorische Integration, audiovisuell, Bayesianisch, Computermodell, Modellierung, perzeptuelle Inferenz, Bindungsproblem, Kausale Inferenz

***Corresponding authors Uta Noppeney,** Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, UK, E-mail: U.Noppeney@bham.ac.uk
Samuel A. Jones, Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, UK, E-mail: saj409@student.bham.ac.uk
Tim Rohe, Department of Psychiatry and Psychotherapy, University of Tübingen, Germany, E-mail: Tim.Rohe@med.uni-tuebingen.d
Ambra Ferrari, Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, UK, E-mail: axf548@student.bham.ac.uk

Computational challenges in multisensory perception

In everyday life our senses are constantly bombarded with many different signals: the motor noise of the trucks, a sparkling motor-bike passing by at high speed, the smell of smoke and fumes and the sight of other pedestrians. How does the human brain transform this sensory cacophony into a veridical percept of the world? To misperceive the looming truck as talking and shiny, and your companion as roaring and smelly could be disastrous! This illustrates that multisensory integration and segregation is critical for our daily interactions. Information integration increases the salience of sensory signals thereby allowing us to detect and respond faster and more accurately to important events, such as an approaching truck (Diederich & Colonius, 2004; Frassinetti et al., 2002; Gillmeister & Eimer, 2007; Noesselt et al., 2008). Further, combining complementary (for example object shape by viewing it from the front and touching it from the back) or redundant (for instance object location by vision and audition)

information across the senses enables a more robust and reliable percept (Ernst & Bülthoff, 2004).

The Bayesian framework in neuroscience posits that the brain forms a probabilistic generative model of the sensory inputs that is inverted during perceptual inference (Kersten et al., 2004; Kersten & Yuille, 2003; Knill & Pouget, 2004). Bayesian probability theory offers a precise formulation of how observers should combine uncertain information to form a representation of the world. Critically, multisensory perception relies on solving two fundamental computational challenges. First, the brain needs solve the so-called 'causal inference problem' and infer whether or not signals come from a common source and should be integrated (Shams & Beierholm, 2010). Second, if signals come from a common source, the brain should integrate them into the most reliable or precise (i. e. least variable or noisy) percept of the environment by weighting them according to their relative reliabilities (Alais & Burr, 2004; Ernst & Banks, 2002). Bayesian Causal Inference models account for these two challenges by explicitly modelling the causal structure of the world (Körding et al., 2007; see also Deroy et al., 2016; Rohe & Noppeney, 2015a, 2015b; Shams & Beierholm, 2010; Wozny et al., 2010).

Let us focus on one simple example: Imagine you are an enthusiastic ornithologist prowling through the forest at dawn in order to gather the best photos and sound recordings of birds. Suddenly, you spot a little robin sitting on a branch and you hear a little robin singing in the bush. How should you direct your camera and your microphone? Should you integrate information from vision and audition in order to obtain a more reliable estimate of the bird's location? Or should you use the information only from vision for directing your camera and only from audition when directing your microphone? The answer to this question depends on the hidden causal structure of the world. There are two hypotheses or models that the brain should entertain.

First, there may be one bird sitting on the branch that is the same bird that you hear singing in the bush. In this 'common source' case, you should indeed integrate signals from vision and audition weighted by their sensory reliabilities. This is the classical 'forced or mandatory fusion' model that has dominated the field of multisensory integration and cue combination over the past two decades (Alais & Burr, 2004; Ernst & Banks, 2002; Hillis et al., 2004). As described by maximum likelihood estimation (MLE) an observer obtains the most precise estimate in this common source case if s/he integrates signals weighted by their relative reliabilities, which is the inverse of variance or noise in the signal. For instance, you would assign a smaller weight to a weak unreliable visual signal at dawn than to a strong and clear visual signal during daylight. Critically, multisensory integration according to MLE principles should lead to a variance reduction of the multisensory relative to the least variable unisensory percept, which is greatest (i. e. by a factor of 2) when the variances of the two unisensory signals are equal. Indeed, several psychophysics studies have shown that human observers integrate signals that are likely to come from a common source near-optimally, close to the predictions of maximum likelihood estimation (Alais & Burr, 2004; Bresciani et al., 2006; Ernst & Banks, 2002; Hillis et al., 2004; Jacobs, 1999; Knill & Saunders, 2003). Yet, evidence is not unequivocal. Accumulating research has also highlighted situations where human observers overweight the sensory modality (Battaglia et al., 2003; Burr et al., 2009; Butler et al., 2010; Rosas et al., 2005) that is usually more reliable in everyday life for a particular task and property (Battaglia et al., 2003) or show a smaller multisensory variance reduction than predicted by MLE (Battaglia et al., 2011; Bentvelzen et al., 2009).

Yet there is a second hypothesis about the signal's causal structure: there may be two birds, one that you can see sitting on the branch and one that you can hear singing in the bush. In this 'independent source' or 'full segregation' case, information integration would be detrimental. Instead, you should use only the auditory information for directing your microphone and the visual information for directing your camera.

Critically, the individual sensory signals do not directly inform the brain whether they arise from common or independent events. Instead, we must actively infer the 'hidden' causal structure from a range of multisensory correspondences such as signals happening at the same time ('temporal coincidence or correlations': Lee & Noppeney, 2011a; Lewis & Noppeney, 2010; Magnotti et al., 2013; Maier et al., 2011; Munhall et al., 1996; Noesselt et al., 2007; Parise & Ernst, 2016; Parise et al., 2012; van Wassenhove et al., 2007), same space ('spatial colocation': Lewald & Guski, 2003; Slutsky & Recanzone, 2001; Spence, 2013), semantic (Adam & Noppeney, 2010; Bishop & Miller, 2011; Kanaya & Yokosawa, 2011; Lee & Noppeney, 2011b; Noppeney et al., 2010), metaphoric (Sadaghiani et al., 2009; Parise & Spence, 2009) and other higher-order statistical or learnt congruency cues. Yet some uncertainty about the world's causal structure will remain. To account for this causal uncertainty, the brain computes a final spatial estimate by combining the estimates from the two causal structures using one of various decision functions (for details see: Wozny et al., 2010). For instance, using the computational strategy called model averaging, it should estimate the location for directing the microphone by com-

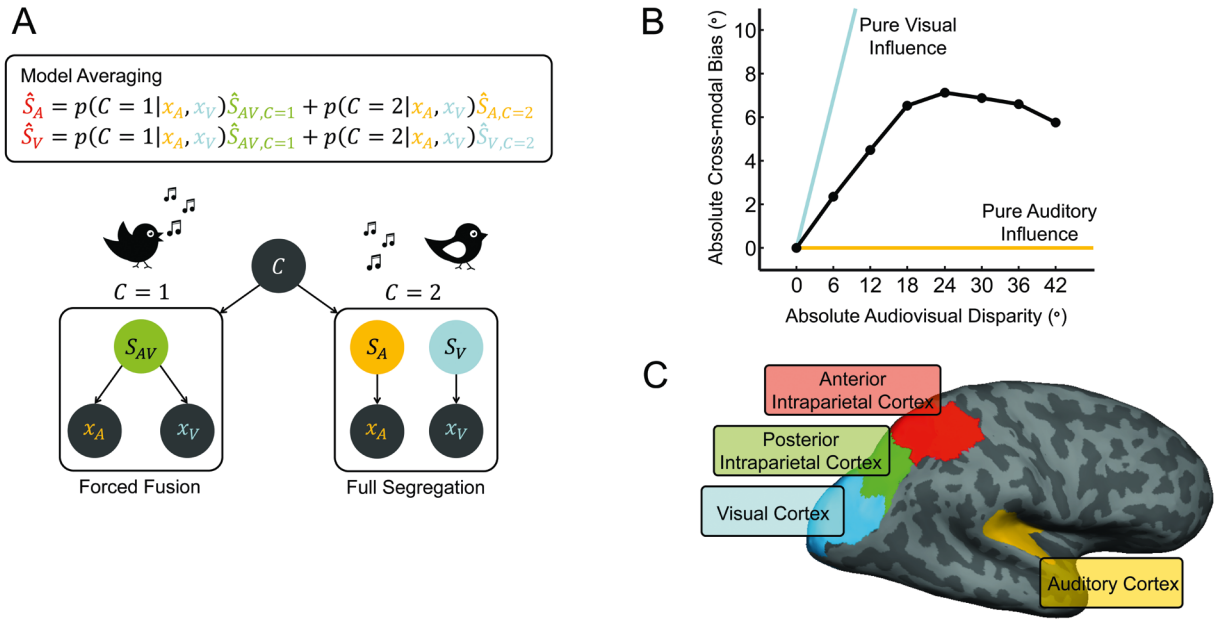

Figure 1: A. Bayesian Causal Inference model: The generative model of Bayesian Causal Inference for spatial localization determines whether the 'sight of the bird' and the 'singing' are generated by common (C=1) or independent (C=2) sources (Körding et al., 2007). For common source, the 'true' audio-visual location (SAV) is drawn from one prior spatial distribution. For independent sources, the 'true' auditory (SA) and 'true' visual (SV) locations are drawn independently from this prior spatial distribution. We then introduce independent sensory noise to generate auditory (XA) and visual (XV) inputs.
B. Visual bias on perceived sound location as a function of audio-visual spatial disparity. As predicted by Bayesian Causal Inference the audio-visual spatial bias depends non-linearly on spatial location. For small spatial disparities, the observer integrates auditory and visual spatial estimates weighted approximately in proportion to their relative reliabilities. For large spatial disparities audio-visual interactions and biases are reduced (Rohe & Noppeney, 2015b).
C. Bayesian Causal Inference within the cortical hierarchy: Primary sensory areas represent predominantly the location of their preferred sensory signals (for example sound location in auditory regions). Posterior intraparietal cortex integrates sensory signals weighted by their reliabilities approximately according to forced fusion principles. Anterior intraparietal sulcus computes the final Bayesian Causal Inference estimate that takes into account the observer's uncertainty about the causal structures that could have generated the sensory signals (Rohe & Noppeney, 2015a).

bining spatial estimates that are computed by the 'forced fusion' and the 'full segregation' models, weighted by the posterior probabilities that audio-visual signals were more likely caused by one single or two different birds (Körding et al., 2007).

Accumulating evidence suggests that human observers arbitrate between sensory integration and segregation qualitatively in line with the principles of Bayesian causal inference (Beierholm et al., 2009; Bertelson & Radeau, 1981; Landy et al., 1995; Roach et al., 2006; Shams & Beierholm, 2010; Wallace et al., 2004). In the laboratory this has been shown in particular for spatial localization (Körding et al., 2007; Rohe & Noppeney, 2015a, 2015b; Wozny et al., 2010) and speech recognition tasks (Magnotti & Beauchamp, 2017; Magnotti et al., 2013). In spatial localization experiments, observers are presented concurrently with auditory signals (for example, bursts of white noise) and visual signals (for instance flashes) at the same or different locations with variable audio-visual spatial disparities. On each trial observers report the location of the flash and/or the location of the noise burst. The results show that an observer's perceived sound location is shifted towards a spatially displaced but synchronous visual flash and vice versa depending on the relative auditory and visual reliabilities. In line with Bayesian Causal Inference these audio-visual spatial biases are attenuated or even abolished for large audio-visual spatial disparities when it is unlikely that auditory and visual signals come from a common source. In other words, audio-visual spatial disparity is a critical cue that observers use to determine whether or not to integrate sensory signals (Körding et al., 2007; Rohe & Noppeney, 2015a, 2015b; Wozny et al., 2010).

The audio-visual spatial bias that emerges for small spatial disparities is in fact the so-called ventriloquist effect (Bertelson & Radeau, 1981; Bonath et al., 2007; Driver, 1996), a perceptual illusion that was used for re-

ligious purposes already in ancient times and later for entertainment at travelling fun fairs (Vox, 1981). To create the ventriloquist illusion the puppeteer speaks without making any articulatory movements. Further, he holds the puppet close to his own face and moves the lips of the puppet in synchrony with his own speech. As a result of the temporal correlations of the auditory (i. e. puppeteer's speech) and visual (i. e. puppet's facial movements) signals the observer infers that auditory and visual signals are generated by a common source and integrates them into a coherent percept weighted by the relative auditory and visual reliabilities. As the spatial reliability of sound perception is usually inferior to the precise visual spatial estimates, the observer most commonly misallocates the speech of the puppeteer to the puppet (Alais & Burr, 2004). In short, the ventriloquist illusion tricks the brain based on the computational principles of Bayesian Causal Inference: it artificially brings auditory and visual signals into spatial conflict while maintaining temporal synchrony to enable integration (see excursion box 1 for perceptual illusions in multisensory perception).

Multisensory interactions are ubiquitous in neocortex

Traditionally, it was thought that multisensory integration is deferred until later processing stages in higher order association areas such as parietal or prefrontal cortices (Avillac et al., 2007; Barraclough et al., 2005; Beauchamp et al., 2004; Calvert et al., 2000; Driver & Noesselt, 2008; Ghazanfar et al., 2008; Macaluso et al., 2003; Miller & D'Esposito, 2005; Sadaghiani et al., 2009; Schroeder & Foxe, 2002; Stevenson & James, 2009). However, over the past two decades neuroimaging in humans (Foxe et al., 2002; Lee & Noppeney, 2011a, 2014; Lehmann et al., 2006; Martuzzi et al., 2007; Molholm et al., 2002; Noesselt et al., 2007; Werner & Noppeney, 2010a), neurophysiology in non-human primates or rodents (Atilgan et al., 2018; Bieler et al., 2017; Bizley et al., 2006; Bizley & King, 2009; Foxe & Schroeder, 2005; Ghazanfar et al., 2005; Ibrahim et al., 2016; Iurilli et al., 2012; Lakatos et al., 2007; Schroeder & Foxe, 2002; Kayser & Logothetis, 2007); and neuroanatomical (Falchier et al., 2002; Rockland & Ojima, 2003; Schroeder et al., 2003) research have accumulated evidence suggesting that multisensory integration emerges already in early sensory and even primary sensory areas and then progressively increases across the cortical hierarchy. Provocatively, it was even proposed that 'the entire neocortex is multisensory' (Ghazanfar & Schroeder, 2006).

In support of low-level integration, numerous fMRI and EEG studies in humans have shown that multisensory interactions can be observed in primary sensory areas and at early processing stages even before 100 ms post-stimulus (Besle et al., 2008; Foxe et al., 2000; Molholm et al., 2002; Molholm et al., 2004). Likewise, neurophysiological recordings in non-human primates (Kayser et al., 2008, 2010; Lakatos et al., 2009; Schroeder & Foxe, 2005) or rodents (Atilgan et al., 2018; Bieler et al., 2017; Bizley & King, 2009; Bizley et al., 2007) revealed that the response to the preferred stimulus in sensory areas, for instance auditory belt and parabelt areas, can be enhanced or suppressed or gain in information content by a concurrent stimulus in a non-preferred sensory modality. While multisensory interactions in low-level sensory areas may be due to top-down influences from superior temporal or parietal cortices (Seltzer & Pandya, 1994), they may also be mediated via thalamo-cortical mechanisms (for example pulvinar) or direct connectivity between sensory areas (Musacchia et al., 2014; Schroeder et al., 2003). Indeed, neuroanatomical tracer studies have shown sparse direct connectivity from early or even primary auditory to visual cortices and vice versa in primates (Falchier et al., 2002; Rockland & Ojima, 2003) and rodents (Bizley et al., 2007; Budinger et al., 2006; Campi et al., 2009; Ibrahim et al., 2016).

This ubiquity of multisensory interplay at all stages of cortical processing challenges traditional hierarchical models of late integration. It suggests that multisensory interactions emerge at multiple cortical levels and within several circuitries including thalamo-cortical, cortico-cortical and higher order association cortices (Musacchia & Schroeder, 2009; Schroeder et al., 2003). Hence, we need to move beyond identifying multisensory regions towards characterizing their functional properties and behavioural relevance.

In primary and low-level sensory cortices, previous research has described driving and modulatory multisensory influences (Atilgan et al., 2018; Bieler et al., 2017; Bizley & King, 2009; BIzley et al., 2007; Kayser et al., 2008; Lakatos et al., 2009; Meijer et al., 2017; Meredith & Allman, 2015). First, both neuroimaging studies in humans (Leitão et al., 2012; Werner & Noppeney, 2011) and electrophysiology in rodents (Ibrahim et al., 2016; Iurilli et al., 2012) have suggested that unisensory stimuli induce deactivations or synaptic inhibition in non-corresponding sensory cortices. For instance, visual stimuli have been shown to elicit a negative BOLD-response in auditory cortices, while auditory stimuli induce synaptic inhibition and fMRI deactivations in visual cortices (Ibrahim et al., 2016; Iurilli et al., 2012; Leitão et al., 2012). Second, a stimulus of a

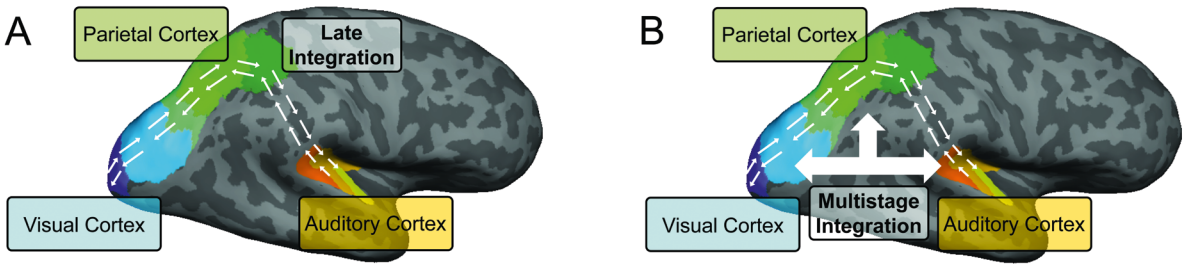

Figure 2: Late and multistage integration models: Traditionally it was thought that multisensory integration emerges at later processing stages in association cortices. We propose that different types of multisensory interactions occur at multiple stage of the cortical processing hierarchy.

non-preferred sensory modality may not necessarily elicit a reliable response in itself, but instead modulate the response to a stimulus of the preferred sensory modality. For instance, auditory core and belt areas are predominantly responsive to auditory rather than visual signals, yet their auditory response and information can be modulated by a concurrent visual input (Kayser et al., 2010). Lakatos and colleagues suggested that these modulatory interactions may rely on mechanisms of phase resetting of theta oscillations (Lakatos et al., 2009; see also Sieben et al., 2012 for related research in rodents). Because in our natural environment the visual signal often precedes the auditory signal (for instance facial articulatory movements often precede speech output), it can modulate the sound-induced activity by resetting the phase of ongoing oscillations (Lakatos et al., 2009; Schroeder et al., 2008). This may be an important mechanism whereby multisensory integration can increase the salience of multisensory events and facilitate their detection. In support of a temporally-sensitive mechanism, recent neurophysiological (Kayser et al., 2010) and fMRI studies (Lewis & Noppeney, 2010; Werner & Noppeney, 2011) have also shown that audio-visual interactions in Heschl's gyrus and planum temporale were sensitive to temporal coincidence or correlations over time. Critically, multisensory response enhancements at the primary cortical level were then gated into higher order association cortices (for instance ventral object vs. dorsal motion recognition system) depending on task context, suggesting that low-level integration effects can propagate to influence higher-order processing to guide behavioural responses (Lewis & Noppeney, 2010).

Multisensory interactions in higher-order association areas such as superior temporal or parietal sulcus are usually less sensitive to the exact timing of the sensory inputs and are characterized by larger temporal binding windows (Werner & Noppeney, 2011). Rather than salience detection, these areas may thus be involved in integrating signals into task-relevant representations (for example spatial, object, speech etc.). In support of this conjecture, the profile of multisensory interactions in STS and IPS directly predicted whether human observers benefitted from multisensory integration: the greater their multisensory enhancement in superior temporal and parietal cortices, the greater was observers' audio-visual benefit for object categorization (Werner & Noppeney, 2010a, 2010b).

Finally, even if sensory signals cannot be integrated into a unified percept because they are incongruent, they can still interact at a decisional level and influence response selection. Using selective intersensory attention tasks a myriad of studies have demonstrated that a task-irrelevant yet incongruent visual stimulus can interfere with observers' decisions on a task-relevant auditory stimulus (Noppeney et al., 2008) and vice versa (Krugliak & Noppeney, 2015; Marks, 1987). Combining a Compatibility Bias model and fMRI we have previously suggested that the prefrontal cortex accumulates sensory evidence from multiple senses until a decisional threshold is reached and a response elicited (Noppeney et al., 2010). Further, in situations of congruent sensory signals the prefrontal cortex shows suppressed responses to audio-visual relative to unisensory signals in line with response facilitation at the decisional level (Sugihara et al., 2006; Werner & Noppeney, 2010a). Interestingly, in line with research in rodents showing multisensory interactions predominantly in transition zones between sensory cortices (Wallace et al., 2004), the suppressive interactions were predominantly found in border zone between auditory- and visual dominant regions (Werner & Noppeney, 2010a).

In summary, accumulating evidence suggests that multisensory integration is a multifaceted process emerg-

ing at multiple stages across the cortical hierarchy. While some sensory interactions take place early and even in primary sensory areas, it is likely that other information may also propagate to higher cortical levels prior to being integrated. Potentially, multisensory interactions in low-level sensory areas serve to amplify the signal strength and salience of multisensory events, which in turn impacts representational integration processes and decision-making in higher levels of the cortical hierarchy (Werner & Noppeney, 2010a).

How the brain performs causal inference and reliability weighted integration

At the beginning we discussed that the brain faces two critical challenges in a multisensory world. First, it needs to identify and bind signals that come from a common source based on a range of correspondence cues such as temporal coincidence or spatial colocation. Second, if signals come from a common source they should be integrated, weighted in proportion to their relative reliabilities. While section 1 summarized the behavioural evidence, in the following we will review neurophysiology and neuroimaging research that provides insight into the underlying neural basis.

Since the seminal work by Stein and colleagues on multisensory integration in the superior colliculus (Meredith & Stein, 1983; Wallace et al., 1996; Stein & Meredith, 1993) a vast number of neurophysiological and neuroimaging studies have shown that the multisensory interactions depend on spatial colocation, temporal synchrony and correlations, as expected for causal inference (Stein & Stanford, 2008). More specifically, Stein and others demonstrated that audio-visual interactions were superadditive (i. e. neural response for the audio-visual stimulus was greater than the sum of the unisensory responses) for spatially collocated audio-visual signals, but turned additive, subadditive or even suppressive when auditory and visual signals were presented at different locations and one signal fell outside the receptive field for the other signal (Stanford, 2005; Stanford & Stein, 2007; Wallace et al., 1996). Thus, organization into receptive fields may enable causal inference based on spatial correspondence cues.

Similar to the role of spatial concordance, audio-visual interactions of transient signals in superior colliculus were limited to a temporal window of approximately 500 ms. Recent modelling approaches suggest that temporal binding windows for naturalistic continuous signals such as speech may rely on detecting multisensory correlations based on the Hassenstein-Reichardt-detector as a relatively simple, yet physiologically plausible model component (Parise & Ernst, 2016). One interesting question that has recently been asked is whether temporal binding may be related to the brain's internal rhythms, i. e. neural oscillations. First, cross-modal phase resetting was put forward as a temporal mechanism that would allow a signal from one sensory modality to modulate the processing of another sensory signal as a function of oscillatory cycle length and audio-visual asynchrony (Lakatos et al., 2009). Second, more recent studies even suggested that the time-varying cycle length of the alpha oscillations in individual observers may determine their audio-visual binding window. Human observers with faster oscillations were associated with a smaller temporal binding window (Cecere et al., 2015; Samaha & Postle, 2015). While the idea that oscillation cycles may serve a similar function for temporal binding windows as receptive fields do for spatial binding windows is intriguing, future studies and more detailed and specific theoretical models are needed.

Recent neurophysiological studies in non-human primates focused on how single neurons and neuronal populations integrate signals weighted by their reliabilities. In a visuo-vestibular heading discrimination task Fetsch and colleagues demonstrated that macaques showed similar near-optimal performance to human observers (Fetsch et al., 2012). Concurrent recording from neurons in dorsal motion area MSTd showed that congruent neurons combined the visual and vestibular inputs subadditively (Gu et al., 2008) and weighted by their relative reliabilities (Fetsch et al., 2012), giving a higher weight to the more reliable sensory signal (Fetsch et al., 2012) on a trial-by-trial basis. As predicted by maximum likelihood estimation under forced fusion assumptions, neurons were more sensitive to heading direction under visuo-vestibular than unisensory stimulation (Gu et al., 2008) in line with behavioural performance (see also Nikbakht et al., 2018 for a related study in rodents). Further, neural population decoding revealed neural sensory weights that corresponded closely to the sensory weights computed from the monkeys' behavioural performance (Fetsch et al., 2012). Additional electrical microstimulation and chemical inactivation of MSTd provided a causal link between the neural computations in MSTd and behavioural performance in a heading discrimination task (Gu et al., 2012). Collectively, this elegant and extensive body of work suggests that neuronal populations and single neurons in MSTd integrate visual and vestibular signals weighted by their relative

reliabilities in representations of heading direction that guide behavioural decisions. While these computations are in line with maximum likelihood estimation, they can be obtained through mechanisms of divisive normalization (Ohshiro et al., 2011, 2017), a canonical neural computation that has previously been proposed for visual processing and attentional modulation (Carandini & Heeger, 2012). Moreover, divisive normalization can also explain a response enhancement that is maximal when the strength of individual signals is weak – a principle referred to as inverse effectiveness since the seminal studies by Stein and colleagues (Stein & Meredith, 1993). Consistently across species and methodologies research has shown that the operational (i. e. super- vs. subadditive) modes of multisensory integration depend on the signal strength as well as a neuron's or voxel's response to unisensory stimuli (Kayser et al., 2008; Siemann et al., 2015; Stanford et al., 2005; Stein & Meredith, 1993; Stein et al., 2014; Werner & Noppeney, 2010b).

At the neural systems level functional imaging studies in humans have shown that higher order association cortices such as parietal or superior temporal sulci integrate sensory signals weighted by their reliabilities in speech recognition (Nath & Beauchamp, 2011), spatial localization (Rohe & Noppeney, 2018) and shape discrimination tasks (Beauchamp et al., 2010; Helbig et al., 2012). Two recent studies moved beyond reliability-weighted integration and investigated how the human brain performs Bayesian Causal Inference in a spatial ventriloquist paradigm at the neural systems level using functional imaging (Rohe & Noppeney, 2015a, 2016). Inside the scanner, observers were presented with audio-visual signals that varied in the spatial disparity and visual reliability. On each trial, they either located the sound or the visual stimulus. Combining psychophysics, fMRI, Bayesian modelling and multivariate decoding the study showed that the brain accomplishes Bayesian Causal Inference by encoding multiple spatial estimates across the cortical hierarchy. At the bottom of the hierarchy, auditory areas encoded predominantly the location of the sound and visual areas the location of the visual stimulus (= segregation). In posterior intra-parietal sulcus, location is estimated under the assumption that the two signals are from a common source (= forced fusion). Only at the top of the hierarchy, in anterior intraparietal sulcus, is the uncertainty about the world's causal structure taken into account. As predicted by Bayesian Causal Inference, location is estimated by combining the segregation and the forced fusion estimates weighted by the posterior probabilities of common and independent sources. Thus, anterior IPS forms a spatial estimate that gracefully transitions from integration to segregation as a function of spatial disparity (Rohe & Noppeney, 2015a, 2016).

Conclusions

In our natural environment our senses are constantly bombarded with many different signals. Ideally, the brain should integrate signals weighted by their reliabilities when they come from a common source, but process them independently when they come from different sources. Human observers have been shown to arbitrate between integration and segregation in line with Bayesian Causal Inference. At the neural level neurophysiological studies in non-human primates and other species have unravelled how the brain integrates signals from common events weighted by their relative reliabilities into a unified representation. Initial neuroimaging studies in human observers suggest that the brain integrates sensory signals in line with Bayesian Causal Inference by encoding multiple perceptual estimates along the cortical hierarchy. Future research combining psychophysics, computational modelling, neurophysiology and neuroimaging across different species is needed to bridge the gap between neural mechanisms, computational operations and behaviour and explore the functional consequences of multisensory integration.

Excursion: Multisensory binding as a mechanism for perceptual illusions

The computations of our perceptual system are optimised for effective interactions with our natural environment. In the laboratory, we can play tricks on observers' perception by placing them in situations that violate the natural statistics for which their perceptual system has been optimized. Particularly, in multisensory integration we can bring sensory signals artificially into conflict along one particular dimension (for example space, time, number, phoneme), while providing sufficient multisensory correspondence cues along another dimension. Thereby, we can persuade the brain to integrate conflicting signals into one unified illusory percept. Multisensory integration has been used to create a myriad of perceptual illusions. In the following we will highlight the most prominent examples.

In the double-flash illusion (Shams et al., 2000) observers are presented with a single flash of light temporally

sandwiched between two beeps. In such cases observers will usually report seeing two flashes, indicating that their final percept placed more weight on the temporally-precise sound signal than on the relatively unreliable flash. Thus, while the ventriloquist illusion exploits the spatial uncertainty of hearing (we can in most circumstances locate an object better by vision than audition), the double-flash illusion exploits the temporal uncertainty of vision (the ears are more reliable than the eyes at determining when something happened). In that respect the double-flash illusion may be considered a temporal equivalent of the ventriloquist effect. But, of course, temporal and spatial dimensions are not quite comparable. While spatial ventriloquism illustrates how the brain estimates the spatial location of an event (i. e. estimation task), the double-flash illusion reveals how it determines the number of events (i. e. detection task).

Multisensory speech signals can also be manipulated to produce illusory percepts. In the so-called McGurk-McDonald illusion (Mcgurk & Macdonald, 1976) the observer is presented in synchrony with a video clip of a speaker articulating /ga/ and a sound recording of the phoneme /ba/. Because of the synchrony cues the observer integrates these conflicting audiovisual phonemes into an illusory /da/ percept. This illusory percept can again be explained by reliability-weighted integration. Using a speech synthesizer, one can generate an artificial 'phoneme' dimension and progressively morph from a /ba/ to a /ga/ phoneme. The perception of 'ba' – 'da' – 'ga' phoneme categories emerges as a result of human categorical perception (Liberman et al., 1957).

The rubber-hand illusion (Botvinick & Cohen, 1998) is an example of our own bodily perception being tricked. In order to track our body's position in space, humans rely on proprioception – the sense that allows us to, for instance, clap with our eyes closed. The rubber-hand illusion overrides this sense by utilising visual and proprioceptive cues that conflict with our understanding of our body's position. The participant is seated with one hand placed on a table. This hand is then concealed from them using a divider, and a replacement rubber hand placed in clear view. The person performing the illusion then proceeds to simultaneously stroke both real and rubber hands with paint brushes, taking care to match the strokes as closely as possible. As the participant continues to see the strokes on the rubber hand but feel them on their real hand, the brain is presented with increasing evidence that these signals are perfectly temporally matched and should be integrated, despite conflicting information from the proprioceptive system that their hand is actually far to the right. The result, in the majority of participants, is a growing belief that the rubber hand has replaced their real one. Such demonstrations often conclude with the experimenter unexpectedly hitting the rubber hand with a hammer.

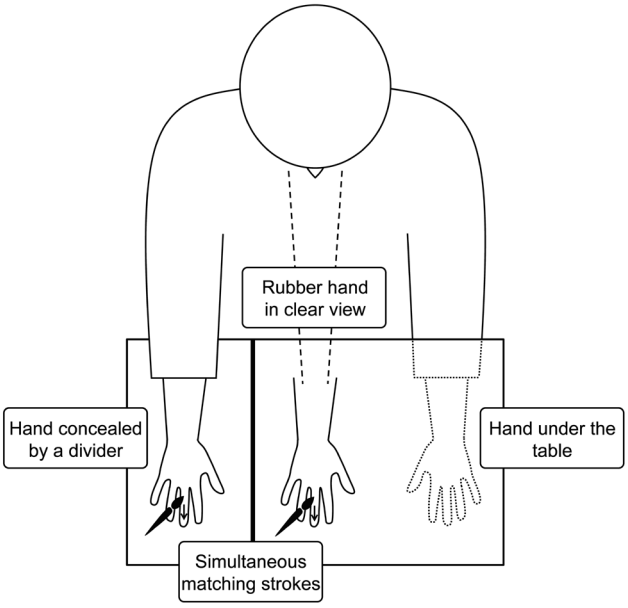

Figure 3: The rubber hand illusion. The participant is seated at a table, with their right hand obscured by a divider, and looks at a rubber hand. The experimenter strokes the participant's hand and the rubber hand simultaneously with paintbrushes, using varied but matching strokes that suggest these haptic and visual signals have the same source and should be integrated. The resulting illusory percept usually manifests as a sense that the rubber hand is a part of one's own body.

Finally, even our sense of taste is not exempt from multisensory illusions. Professor Charles Spence specialises in the sensory perception of food, and his lab has demonstrated a variety of ways in which other senses can influence what we taste. The weight and material of cutlery (Harrar & Spence, 2013), the colour and shape of the plate (Piqueras-Fiszman et al., 2012), and the texture of packaging (Piqueras-Fiszman et al., 2012) have all been shown to influence our experience of food. In 2008, Professor Spence (alongside his colleague Massimiliano Zampini) was presented with the Ig Nobel prize for the novel demonstration that digital sound manipulations can make potato crisps seem crunchier (Zampini & Spence, 2004).

Acknowledgment: This research was funded by ERC-2012-StG_20111109 multsens.

References

Adam, R., & Noppeney, U. (2010). Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. Neuroimage, *52* (4), 1592–1602.

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. Curr. Biol. *14* (3), 257–262.

Atilgan, H., Town, S., Wood, K., Jones, G., Maddox, R., Lee, A., & Bizley, J. K. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. Neuron. *97* (3), 640–655.

Avillac, M., Ben Hamed, S., & Duhamel, J.-R. (2007). Multisensory Integration in the Ventral Intraparietal Area of the Macaque Monkey. J. Neurosci. *27* (8), 1922–1932.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of Visual and Auditory Information by Superior Temporal Sulcus Neurons Responsive to the Sight of Actions. J. Cogn. Neurosci. *17* (3), 377–391.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. J. Opt. Soc. Am. A. *20* (7), 1391–1397.

Battaglia, P. W., Kersten, D., & Schrater, P. R. (2011). How haptic size sensations improve distance perception. PLoS Comput. Biol. *7* (6).

Beauchamp, M. S., Lee, K., Argall, B., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. *41*, 809–823.

Beauchamp, M. S., Pasalar, S., & Ro, T. (2010). Neural substrates of reliability-weighted visual-tactile multisensory integration. Front. Syst. Neurosci. *4*, 1–11.

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. J. Vis. *9* (5), 23.

Bentvelzen, A., Leung, J., & Alais, D. (2009). Discriminating audiovisual speed: Optimal integration of speed defaults to probability summation when component reliabilities diverge. Perception. *38* (7), 966–987.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Percept. Psychophys. *29* (6), 578–584.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.-H. (2008). Visual Activation and Audiovisual Interactions in the Auditory Cortex during Speech Perception: Intracranial Recordings in Humans. J. Neurosci. *28* (52), 14301–14310.

Bieler, M., Sieben, K., Cichon, N., Schildt, S., Röder, B., & Hanganu-Opatz, I. L. (2017). Rate and Temporal Coding Convey Multisensory Information in Primary Sensory Cortices. eNeuro. *4* (2), ENEURO-0037.

Bishop, C. W., & Miller, L. M. (2011). Speech cues contribute to audiovisual spatial integration. PLOS ONE. *6* (8).

Bizley, J. K., & King, A. J. (2009). Visual influences on ferret auditory cortex. Hear. Res. *258* (1–2), 55–63.

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., & King, A. J. (2006). Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb. Cortex. *17* (9), 2172–2189.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. a. (2007). Neural Basis of the Ventriloquist Illusion. Curr. Biol. *17* (19), 1697–1703.

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. Nature. *391* (6669), 756.

Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. J. Vis. *6* (5), 2.

Budinger, E., Heil, P., Hess, A., & Scheich, H. (2006). Multisensory processing via early cortical stages: connections of the primary auditory cortical field with other sensory systems. Neuroscience. *143* (4), 1065–1083.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. Exp. Brain Res. *198* (1), 49–57.

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. J. Vis. *10* (11), 23.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. *10* (11), 649–657.

Campi, K. L., Bales, K. L., Grunewald, R., & Krubitzer, L. (2009). Connections of auditory and visual cortex in the prairie vole (Microtus ochrogaster): evidence for multisensory processing in primary sensory areas. Cereb. Cortex. *20* (1), 89–108.

Carandini, M., & Heeger, D. (2012). Normalization as a caonical neural computation. Nat. Rev. Neurosci. *13* (1), 51–62.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. Curr. Biol. *25* (2), 231–235.

Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in Multisensory Perception. Trends Cogn. Sci. *20* (10), 736–747.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. Percept. Psychophys. *66* (8), 1388–1404.

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

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

Driver, J., & Noesselt, T. (2008). Multisensory Interplay Reveals Crossmodal Influences on "Sensory-Specific" Brain Regions, Neural Responses, and Judgments. Neuron. *57* (1), 11–23.

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

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. Trends Cogn. Sci. *8* (4), 162–169.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. J. Neurosci. *22* (13), 5749–5759.

Fetsch, C. R., Pouget, A., Deangelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. Nat. Neurosci. *15* (1), 146–154.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somato-sensory interactions in early cortical processing revealed by high-density electrical mapping. Cogn. Brain Res. *10* (1), 77–83.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. NeuroReport. *16* (5), 419–423.

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., Ritter & W., Murray, M. M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J Neurophysiol. *88* (1), 540–543.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp. Brain Res. *147* (3), 332–343.

Ghazanfar, A. A., Chandrasekaran, C., & Logothetis, N. K. (2008). Interactions between the Superior Temporal Sulcus and Auditory Cortex Mediate Dynamic Face/Voice Integration in Rhesus Monkeys. J. Neurosci. *28* (17), 4457–4469.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. J. Neurosci. *25* (20), 5004–5012.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? Trends Cogn. Sci. *10* (6), 278–285.

Giard, M. H., & Peronnet, F. (1999). Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. J. Cogn. Neurosci. *11* (5), 473–490.

Gillmeister, H., & Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. Brain Res. *1160* (1), 58–68.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. Nat. Neurosci. *11* (10), 1201–1210.

Gu, Y., DeAngelis, G. C., & Angelaki, D. E. (2012). Causal Links between Dorsal Medial Superior Temporal Area Neurons and Multisensory Heading Perception. J. Neurosci. *32* (7), 2299–2313.

Harrar, V., & Spence, C. (2013). The taste of cutlery: how the taste of food is affected by the weight, size, shape, and colour of the cutlery used to eat it. Flavour. *2* (1), 21.

Helbig, H. B., Ernst, M. O., Ricciardi, E., Pietrini, P., Thielscher, A., Mayer, K. M., … Noppeney, U. (2012). The neural mechanisms of reliability weighted integration of shape information from vision and touch. NeuroImage. *60* (2), 1063–1072.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. J. Vis. *4* (12), 1.

Hollensteiner, K. J., Pieper, F., Engler, G., König, P., & Engel, A. K. (2015). Crossmodal integration improves sensory detection thresholds in the ferret. PloS One, *10* (5), e0124952.

Ibrahim, L. A., Mesik, L., Ji, X. Y., Fang, Q., Li, H. F., Li, Y. T., Zingg, B., Zhang, L. I., & Tao, H. W. (2016). Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. Neuron, *89* (5), 1031–1045.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Bonfenati, F., & Medini, P. (2012). Sound-driven synaptic inhibition in primary visual cortex. Neuron, *73* (4), 814–828.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. Vis. Res. *39* (21), 3621–3629.

Kanaya, S., & Yokosawa, K. (2011). Perceptual congruency of audio-visual speech affects ventriloquism with bilateral visual stimuli. Psychon. Bull. Rev. *18* (1), 123–128.

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? Brain Struct. Funct. *212* (2), 121–132.

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Visual Enhancement of the Information Representation in Auditory Cortex. Curr. Biol. *20* (1), 19–24.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. Cereb. Cortex. *18* (7), 1560–1574.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. Annu. Rev. Psychol. *55*, 271–304.

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. Curr. Opin. Neurobiol. *13* (2), 150–158.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. Trends Neurosci. *27* (12), 712–719.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? Vis. Res. *43* (24), 2539–2558.

Körding, K. P., Beierholm, U. R., Ma, W. J., Quartz, S. R., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. PloS One. *2* (9), e943.

Krugliak, A., & Noppeney, U. (2015). Synesthetic interactions across vision and audition. Neuropsychologia. *88*, 65–73.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. Neuron. *53* (2), 279–292.

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The Leading Sense: Supramodal Control of Neurophysiological Context by Attention. Neuron. *64* (3), 419–430.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. Vis. Res. *35* (3), 389–412.

Lee, H., & Noppeney, U. (2011a). Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc. Nat. Acad. Sci. U.S.A. *108* (51), E1441–50.

Lee, H., & Noppeney, U. (2011b). Physical and Perceptual Factors Shape the Neural Mechanisms That Integrate Audiovisual Signals in Speech Comprehension. J. Neurosci. *31* (31), 11338–11350.

Lee, H., & Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices. Curr. Biol. *24* (8), R309–R310.

Lehmann, C., Herdener, M., Esposito, F., Hubl, D., di Salle, F., Scheffler, K., … Seifritz, E. (2006). Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. NeuroImage. *31* (1), 294–300.

Leitão, J., Thielscher, A., Werner, S., Pohmann, R., & Noppeney, U. (2012). Effects of parietal TMS on visual and auditory processing at the primary cortical level–a concurrent TMS-fMRI study. Cereb. Cortex. *23* (4), 873–884.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. Cogn. Brain Res. *16* (3), 468–478.

Lewis, R., & Noppeney, U. (2010). Audiovisual Synchrony Improves Motion Discrimination via Enhanced Connectivity between Early Visual and Auditory Areas. J. Neurosci. *30* (37), 12329–12339.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. *54* (5), 358–368.

Macaluso, E., Driver, J., & Frith, C. D. (2003). Multimodal Spatial Representations Engaged in Human Parietal Cortex during

Both Saccadic and Manual Spatial Orienting. Curr. Biol. *13* (12), 990–999.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. Science. *289* (5482), 1206–1208.

Magnotti, J. F., & Beauchamp, M. S. (2017). A Causal Inference Model Explains Perception of the McGurk Effect and Other Incongruent Audiovisual Speech. PLoS Comput. Biol. *13* (2), e1005229.

Magnotti, J. F., Ma, W. J., & Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. Front. Psychol. *4*, 1–10.

Maier, J. X., Di Luca, M., & Noppeney, U. (2011). Audiovisual Asynchrony Detection in Human Speech. J. Exp. Psychol. Hum. Percept. Perform. *37* (1), 245–256.

Marks, L. E. (1987). On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination. J. Exp. Psychol. Hum. Percept. Perform. *13* (3), 384–394.

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. Cereb. Cortex. *17* (7), 1672–1679.

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

Meijer, G. T., Montijn, J. S., Pennartz, C. M., & Lansink, C. S. (2017). Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. J. Neurosci. *37* (36), 8783–8796.

Meredith, M. A., & Allman, B. L. (2015). Single-unit analysis of somatosensory processing in the core auditory cortex of hearing ferrets. Eur. J. Neurosci. *41* (5), 686–698.

Meredith, M., & Stein, B. (1983). Interactions among converging sensory inputs in the superior colliculus. Science. *221* (4608), 389–391.

Miller, L. M., & D'Esposito, M. (2005). Perceptual Fusion and Stimulus Coincidence in the Cross-Modal Integration of Speech. J. Neurosci. *25* (25), 5884–5893.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual-Auditory Object Recognition in Humans: A High-density Electrical Mapping Study. Cereb. Cortex. *14* (4), 452–465.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn. Brain Res. *14* (1), 115–128.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. Percept. Psychophys. *58* (3), 351–362.

Musacchia, G., Large, E. W., & Schroeder, C. E. (2014). Thalamocortical mechanisms for integrating musical tone and rhythm. Hear. Res. *308*, 50–59.

Musacchia, G., & Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. Hear. Res. *258* (1–2), 72–79.

Nath, A. R., & Beauchamp, M. S. (2011). Dynamic Changes in Superior Temporal Sulcus Connectivity during Perception of Noisy Audiovisual Speech. J. Neurosci. *31* (5), 1704–1714.

Noesselt, T., Bergmann, D., Hake, M., Heinze, H. J., & Fendrich, R. (2008). Sound increases the saliency of visual events. Brain Res. *1220*, 157–163.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., & Driver, J. (2007). Audiovisual Temporal Correspondence Modulates Human Multisensory Superior Temporal Sulcus Plus Primary Sensory Cortices. J. Neurosci. *27* (42), 11431–11441.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. Cereb. Cortex. *18* (3), 598–609.

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual Decisions Formed by Accumulation of Audiovisual Evidence in Prefrontal Cortex. J. Neurosci. *30* (21), 7434–7446.

Nikbakht, N., Tafreshiha, A., Zoccolan, D., & Diamond, M. E. (2018). Supralinear and supramodal integration of visual and tactile signals in rats: psychophysics and neuronal mechanisms. Neuron. *97* (3), 626–639.e8.

Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2011). A normalization model of multisensory integration. Nat. Neurosci. *14* (6), 775–782.

Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2017). A Neural Signature of Divisive Normalization at the Level of Multisensory Integration in Primate Cortex. Neuron. *95* (2), 399–411.

Parise, C. V., & Spence, C. (2009). "When birds of a feather flock together": Synesthetic correspondences modulate audiovisual integration in non-synesthetes. PloS One. *4* (5), e5664.

Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. Nature Commun. *7* (12), 11543.

Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. Curr. Biol. *22* (1), 46–49.

Piqueras-Fiszman, B., Alcaide, J., Roura, E., & Spence, C. (2012). Is it the plate or is it the food? Assessing the influence of the color (black or white) and shape of the plate on the perception of the food placed on it. Food Qual. Pref. *24* (1), 205–208.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. Proc. Biol. Sci. *273* (1598), 2159–68.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. Int. J. Psychophysiol. *50* (1–2), 19–26.

Rohe, T., & Noppeney, U. (2015a). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. PLOS Biol. *13* (2), e1002073.

Rohe, T., & Noppeney, U. (2015b). Sensory reliability shapes Bayesian Causal Inference in perception via two mechanisms. J. Vis. *15*, 1–38.

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. Curr. Biol. *1* (4), 509–514.

Rohe, T., & Noppeney, U. (2018). Reliability-Weighted integration of audiovisual signals can be modulated by top-down control. eNeuro. *5* (1), ENEURO-0315.

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. J. Opt. Soc. Am. A. *22* (5), 801–809.

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, Metaphoric, and Linguistic Auditory Direction Signals Have

Distinct Influences on Visual Motion Processing. J. Neurosci. *29* (20), 6490–6499.

Samaha, J., & Postle, B. R. (2015). The Speed of Alpha-Band Oscillations Predicts the Temporal Resolution of Visual Perception. Curr. Biol. *25* (22), 2985–2990.

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn. Brain Res. *14* (1), 187–198.

Schroeder, C. E., & Foxe, J. J. (2005). Multisensory contributions to low-level, "unisensory" processing. Curr. Opin. Neurobiol. *15* (4), 454–458.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. *12* (3), 106–113.

Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., & Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. Int. J. Psychophysiol. *50* (1–2), 5–17.

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipita projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. J. Comp. Neurol. *343* (3), 445–463.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. Trends Cogn. Sci. *14* (9), 425–432.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. Nature. *408*, 2000.

Sieben, K., Röder, B., & Hanganu-Opatz, I. L. (2013). Oscillatory entrainment of primary somatosensory cortex encodes visual control of tactile processing. J. Neurosci, *33* (13), 5736–5749.

Siemann, J. K., Muller, C. L., Bamberger, G., Allison, J. D., Veenstra-VanderWeele, J., & Wallace, M. T. (2015). A novel behavioral paradigm to assess multisensory processing in mice. Front. Behav. Neurosci. *8*, 456.

Slutsky, D. a, & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. NeuroReport. *12* (1), 7–10.

Soto-Faraco, S., Kingstone, A., & Spence, C. (2006). Integrating motion information across sensory modalities: The role of top-down factors. Progress Brain Res. *155*, 273–286.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. Ann. NY Acad. Sci. *1296* (1), 31–49.

Stanford, T. R. (2005). Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. J. Neurosci. *25* (28), 6499–6508.

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: Putting the computation in context. NeuroReport. *18* (8), 787–792.

Stein, B. E., & Meredith, M. A. (1993). The merging of the senses (The MIT Press).

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. Nat. Rev. Neurosci. *9* (4), 255–266.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. Nat. Rev. Neurosci. *15* (8), 520.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. NeuroImage. *44* (3), 1210–1223.

Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M. (2006). Integration of Auditory and Visual Communication Information in the Primate Ventrolateral Prefrontal Cortex. J. Neurosci. *26* (43), 11138–11147.

Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. Neuron. *43* (2), 271–282.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. Neuropsychologia. *45* (3), 598–607.

Vox, V. (1981). I can see your lips moving: The history and art of ventriloquism (Kaye & Ward).

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. a. (2004). Unifying multisensory signals across time and space. Exp. Brain Res. *158* (2), 252–258.

Wallace, M. T., Ramachandran, R., & Stein, B. E. (2004). A revised view of sensory cortical parcellation. Proc. Nat. Acad. Sci., *101* (7), 2167–2172.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. J. Neurophysiol. *76* (2), 1246–1266.

Werner, S., & Noppeney, U. (2010a). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. *30* (7), 2662–2675.

Werner, S., & Noppeney, U. (2010b). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb. Cortex. *20* (8), 1829–1842.

Werner, S., & Noppeney, U. (2011). The contributions of transient and sustained response codes to audiovisual integration. Cereb. Cortex. *21* (4), 920–931.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability Matching as a Computational Strategy Used in Perception. PLoS Comput. Biol. *6* (8), e1000871.

Zampini, M., & Spence, C. (2004). The role of auditory cues in modulating the perceived crispness and staleness of potato chips. J. Sens. Stud. *19* (5), 347–363.

Article note: German version available at https://doi.org/10.1515/nf-2017-0066

Bionotes

Uta Noppeney
u.noppeney@bham.ac.uk
+44 121 414 7211
Computational Neuroscience and Cognitive Robotics Centre
School of Psychology
University of Birmingham
Edgbaston
Birmingham
B15 2TT
UK
E-Mail: u.noppeney@bham.ac.uk

Uta Noppeney is Professor of Computational Neuroscience and director of the Computational Neuroscience and Cognitive Robotics Centre at the University of Birmingham, UK. She received a degree in medicine (1997, Freiburg University), a doctorate in medicine (1998, Freiburg University) and a PhD in neuroscience (2004, University College London, UK). After training in neurology at the University Hospital in Aachen, she conducted neuroscience research at Magdeburg University and subsequently at the Wellcome Trust Centre for Neuroimaging, London. In 2005, she became research group leader at the Max Planck Institute for Biological Cybernetics in Tübingen. She combines psychophysics, functional imaging (M/EEG, fMRI, TMS) and computational modelling to investigate how the human brain integrates information across the senses into a coherent percept of the environment.

Tim Rohe is a post-doctoral researcher at the clinic for psychiatry and psychotherapy of the University Hospital Tübingen. He obtained his PhD between 2010 and 2014 at the Max Planck Institute for biological cybernetics in Tübingen. Before this he studied psychology at the University of Freiburg (2004-2010).

Samuel A. Jones received a BSc in Psychology and an MSc in Psychological Research from Bangor University, Wales. He is now in the final year of a PhD at the University of Birmingham. He uses fMRI, psychophysics and computational modelling to explore the impact of brain ageing on multisensory integration.

Ambra Ferrari studied Cognitive Neuroscience and Neuropsychology at the University of Trento, Italy. She is currently undertaking a PhD in Neuroscience at the University of Birmingham. She combines psychophysics, computational modelling and MRI techniques to study the relationship between multisensory integration and high-level cognitive processes such as attention and reward learning.