CellPress

Review

Metacognition in Multisensory Perception

Ophelia Deroy,[1,*] Charles Spence,[2] and Uta Noppeney[3,*]

Metacognition – the ability to monitor one's own decisions and representations, their accuracy and uncertainty – is considered a hallmark of intelligent behavior. Little is known about metacognition in our natural multisensory environment. To form a coherent percept, the brain should integrate signals from a common cause but segregate those from independent causes. Multisensory perception thus relies on inferring the world's causal structure, raising new challenges for metacognition. We discuss the extent to which observers can monitor their uncertainties not only about their final integrated percept but also about the individual sensory signals and the world's causal structure. The latter causal metacognition highlights fundamental links between perception and other cognitive domains such as social and abstract reasoning.

Trends

To form a coherent percept of our multisensory environment the brain needs to integrate signals caused by a common source (e.g., an event) but segregate those from different sources; natural multisensory perception thus relies inherently on inferring the world's causal structure.

Human observers are known to metacognitively monitor the uncertainty of their perceptual estimates in simple sensory tasks, but it is unclear whether they can monitor their uncertainties about their integrated percept, the individual sensory signals, and the causal structure of complex multisensory environments.

Causal metacognition highlights fundamental links between perception and other cognitive domains such as social and abstract reasoning and may be critical for our understanding of neuropsychiatric diseases such as schizophrenia.

Metacognition: Monitoring One's Own Cognition

'**Metacognition**' (see Glossary) refers to cognitive processes about other cognitive processes, knowing about knowing, or beliefs about one's own beliefs. It describes the formation of second-order representations that allow observers to monitor their first-order representations about objects or events in the real world [1–3]. Metacognitive research investigates the extent to which observers can assess the uncertainty or accuracy of their perceptual representations and judgments. For instance, observers can not only spot a friend in the crowd but also metacognitively evaluate their uncertainty or doubtfulness about their first-order perceptual interpretation (e.g., 'Is this really my friend?'). In a wider sense, however, metacognition characterizes an observer's ability to introspect the perceptual inference processes that led to their first-order world representations [4]. Metacognition can operate in numerous domains including perception [5–7], memory [8,9], collective decision making [10], and social learning [11,12].

Despite a recent surge of interest in metacognition, most perception research to date has focused on simple visual or auditory tasks that were based on a single signal stream [7,13–16]. Yet in our natural environment our senses are constantly bombarded with many different signals. To form a coherent percept of the world, the brain is challenged to integrate signals caused by common events but segregate those caused by independent events. Natural perception thus relies inherently on inferring the world's causal structure. In this review, we focus on the challenges a natural complex environment poses not only for first-order perception but also for second-order metacognition. First, we introduce **Bayesian Causal Inference** as a normative model that describes how an ideal observer should arbitrate between sensory integration and segregation when exposed to multiple sensory signals in our natural environment [17–19]. Next, we discuss whether observers can monitor their uncertainties associated with the different sorts of estimates that Bayesian Causal Inference involves, such as the uncertainties about their final integrated percept, the individual sensory estimates, and the inferred causal structure of the world [2,20,21]. Finally, we ask whether human observers can move beyond the integrated percept and metacognitively introspect these perceptual inference processes. Is multisensory perception encapsulated as an unconscious inference process or is it open to metacognitive

[1] Centre for the Study of the Senses, Institute of Philosophy, University of London, London, UK
[2] Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, Oxford, UK
[3] Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, Birmingham, UK

*Correspondence:
ophelia.deroy@sas.ac.uk (O. Deroy) and
u.noppeney@bham.ac.uk
(U. Noppeney).

http://dx.doi.org/10.1016/j.tics.2016.08.006

Box 1. Monitoring Causal Uncertainty Beyond Perception

Causal inference is critical not only for perception but, more generally, for many other cognitive domains such as inductive, abstract, and social reasoning [82]. If two burglaries occur in the same town on the same day, the police ought to inquire whether they are likely to be performed by the same or different criminal gangs. Likewise, if a patient presents initially with a rash followed by high fever, cough, shortness of breath, and wheezing, the medical doctor needs to infer whether all of these symptoms are caused by measles infection or whether some of them may be caused by a subsequent bacterial (e.g., streptococcal) superinfection that requires antibiotic treatment. These examples highlight that causal inference is pervasive in our everyday lives. Causal metacognition enables observers to monitor their uncertainty about the underlying causal structure and decide whether to seek additional evidence to arbitrate between several potential causal structures. If the medical doctor is in doubt about whether the patient may have incurred an additional streptococcal infection, he/she may order blood tests, chest radiography, etc.

Causal inference is also fundamental for successful communication and interactions across social agents. For instance, if two social agents talk about a person called 'Peter' they usually assume that they refer to the same person as the causal source that generates their thoughts and representations associated with 'Peter'. This shared causal perspective is fundamental for successful collective decision making [10]. Surprises and comic moments may emerge if the agents discover during the course of their conversation that their inference was wrong and they had been referring to two different individuals both called 'Peter'. In other words, they suddenly discovered that their thoughts and representations did not pertain to one common source 'Peter' but to two different individuals.

Causal inference as a process to arbitrate between one or multiple causes for sensory signals, medical symptoms, or mental representations is part of the wider question of how observers can infer hidden structure from statistical correlations in observed data (e.g., correlations between different symptoms). How can they build veridical or at least useful models of the world? As reviewed in more detail in [17,88–90], Bayesian models can be used to accommodate human structure inference across numerous domains including inductive reasoning [82], semantics [91], social cognition [10], and aggregation of information across individuals [92].

introspection? While we focus on multisensory perception and cue combination as prime examples of the integration of information from independent sensory channels [17,22,23], the fundamental challenges and principles apply more generally to situations and tasks that require information integration and segregation in perception and wider cognition (Box 1).

Metacognition enables human and nonhuman observers [24] to act more strategically; for instance, to determine whether to defer a response and acquire more information [20,25]. **Causal metacognition** is particularly critical for situations with information emanating from potentially different sources, not only in perception but also in social and abstract reasoning [17,26].

Metacognition in Perception

In the 19th century, Helmholtz described perception as 'unconscious inference' that maps from noisy sensory inputs to perceptual interpretations and choices under the guidance of prior experience [27]. Likewise, more recent Bayesian statistical models formalize perception as a probabilistic inference process whereby the brain combines prior expectations with uncertain sensory evidence to infer the most likely state of the world [28]. Perception is thus inherently uncertain and error prone. Metacognitive research investigates whether observers can assess their uncertainty about the perceptual representations that are formed on the basis of noisy sensory evidence. Are observers appropriately confident about the accuracy of their perceptual choices and do they eventually use this information to adjust subsequent responses [21,29]? Accumulating evidence based on decisional **confidence rating** [30], **no-loss gambling** [31], and **post-decision wagering** [32,33] demonstrates that human and nonhuman observers can indeed access the uncertainty of their perceptual representations and adjust their decisional confidence accordingly. In some cases, observers even compute their confidence about the correctness of their perceptual judgment (e.g., motion discrimination) in a Bayes-optimal fashion. In other words, their confidence reflects the probability that their perceptual choices are correct given the sensory signals (e.g., motion) [29].

Glossary

Bayesian Causal Inference model: a normative Bayesian model that describes how an observer should arbitrate between information integration and segregation to compute an estimate of an environmental property. Bayesian Causal Inference [17–19,52,66] explicitly models the potential causal structures (i.e., common or independent sources) that could have generated the two signals.

Causal metacognition: monitoring one's own uncertainty about the causal structure underlying certain signals (e.g., sensory signals).

Causal metamers: identical causal structures inferred from signals generated by physically different causal structures.

Confidence rating, post-decision wagering, no-loss gambling [30]: methods to assess an observer's metacognitive insights or awareness. For instance, observers may rate their confidence about the correctness of their decision on a numerical scale. In post-decision wagering, they are asked to bet on the correctness of their reported choices. As a result, observers should place higher wagers when they are more confident about the correctness of their decision to maximize their gains. In no-loss gambling, observers need to choose whether they are given a reward depending on the correctness of their perceptual choice or depending on a lottery with prespecified probabilities. Both post-decision wagering and no-loss gambling provide observers with an incentive to reveal their decisional confidence and subjective probabilities truthfully. However, post-decision wagering may be sensitive to additional biases such as risk aversiveness.

Intersensory correspondences: the observer uses different sorts of correspondences, such as spatial collocation [50–52,58,59], temporal coincidence [56,57,60] and correlations [61,62], or semantic or phonological congruency [63–65], to determine which signals are likely to come from a common source and should be bound during perception.

McGurk illusion: an audiovisual illusion [71,79,81] where observers perceive, for instance, the phoneme [da] when presented with a video of a face articulating <<ga>> and a

Critically, observers' decisional confidence depends on the uncertainty of their first-order perceptual representations (for other influences, see [34]). For instance, when presented with weak motion signals, observers will be close to chance not only when discriminating motion direction but also when judging whether their motion discrimination response was correct. In other words, observers' perceptual sensitivity (e.g., their ability to discriminate left from right motion) constrains their maximally possible metacognitive sensitivity (i.e., their ability to discriminate between their correct and incorrect choices) [14,35]. While d' is used as a signal-theoretic index to quantify observers' perceptual sensitivity, meta-d' has recently been proposed as a signal-theoretic index to quantify observers' metacognitive sensitivity. A large meta-d' indicates that observers can reliably discriminate between their correct and incorrect perceptual judgments. Critically, while meta-d' depends on both the quality of the sensory evidence and its metacognitive assessment, directly comparing the perceptual and the metacognitive d' quantifies observers' metacognitive efficiency [14,35]. It provides insights into observers' ability to evaluate the uncertainty of their perceptual representations and choices. A 'metacognitively ideal observer' (i.e., where meta-d' is equal to d') can access all information that was used for the first-order perceptual judgment for his/her second-order metacognitive evaluation.

Abundant evidence suggests that the brain is able to represent and use estimates of uncertainty for neural computations in perception, learning, and cognition more widely [21–23,36,37]. However, the underlying neural coding principles remain debated. For instance, uncertainty may be represented in probabilistic population codes [38,39] or may rely on sampling-based methods [40]. Likewise, it remains controversial whether metacognitive 'confidence estimates' are directly read out from first-order neural representations [13,20] or formed in distinct 'metacognitive' neural circuitries [7,41,42]. In support of a shared system, or common mechanism, underlying perceptual decisions and confidence, neurophysiological research has demonstrated that the same neurons in a lateral parietal area encode both the perceptual choice of monkeys and its confidence [43,44]. Dissociations between perceptual choice and confidence may emerge when decision confidence is interrogated after the subject committed to a perceptual choice, thereby relying on different sensory evidence [3,13,45]. By contrast, neuropsychological and neuroimaging studies in humans indicate dedicated metacognitive neural circuitries in the prefrontal cortex [7,42,46]. For instance, fMRI revealed that activations in the anterior prefrontal cortex reflect changes in confidence when perceptual performance is held constant [47]. Likewise, patients with anterior prefrontal lesions showed a selective deficit in metacognitive accuracy [42]. Decisional confidence estimates encoded in dedicated circuitries may serve as a common currency and enable direct comparisons across different cognitive tasks [15] or sensory modalities [5].

The Multisensory Challenge: Causal Inference and Reliability-Weighted Integration

Imagine you are packing your shopping items from your trolley into the back of your car, which is parked on a busy street. Suddenly you hear a loud horn. Is this sound coming from a car on the opposite side of the road competing for a parking spot or from a car hidden behind your back indicating that your trolley is blocking the traffic? Or is the sound perhaps coming from one of your shopping items? While the latter suggestion seems rather unlikely, the other two may be valid interpretations of the sensory inputs (Figure 1). This example illustrates the two fundamental computational challenges that the brain faces in our everyday multisensory world. First, it needs to solve the so-called causal inference problem [17–19] and determine whether signals come from common sources and should be integrated. Second, if two signals come from a common source, the brain is challenged to integrate them into the most reliable percept by weighting them optimally in proportion to their **reliabilities** (i.e., the inverse of sensory variance [22,23,48,49]).

voice uttering /ba/. The McGurk illusion is a prime example of a perceptual metamer; that is, the conflicting signals are perceived as identical to a face and voice articulating [da].

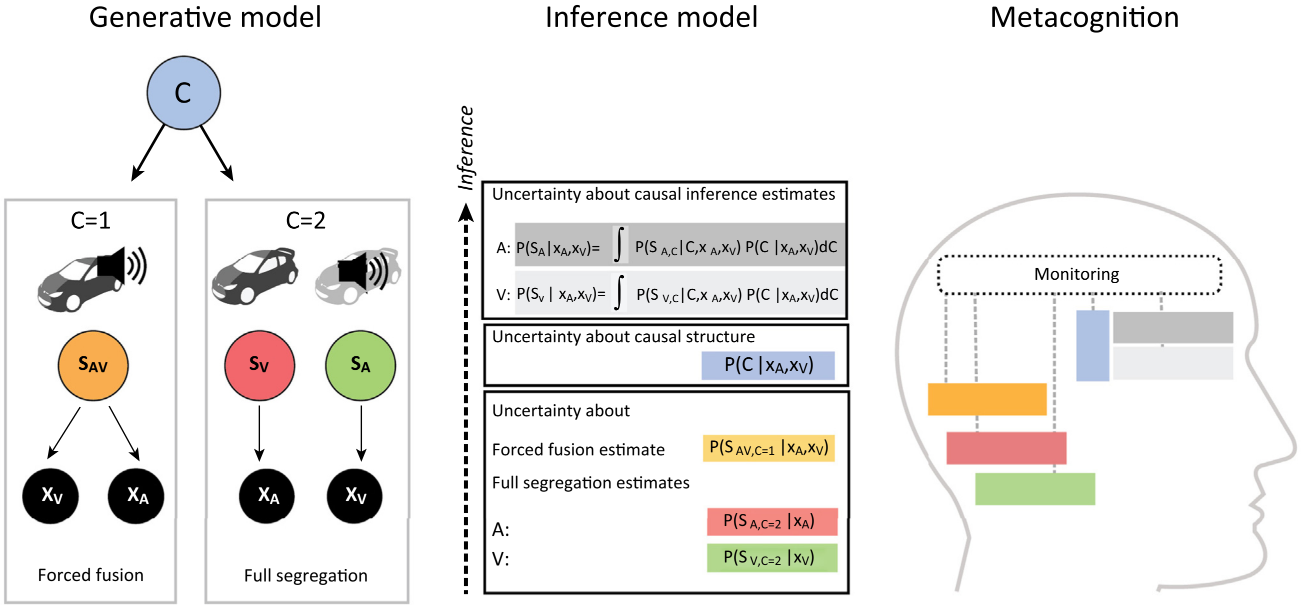
Metacognition: cognitive processes about other cognitive processes or beliefs about one's own beliefs [1–3,24].

Perceptual metamers: identical perceptual (e.g., spatial, phoneme) estimates formed from physically different signals.

Sense of agency: the subjective feeling that one initiates and controls one's own actions [72,73,83].

Sensory reliability: the inverse of sensory variance (or uncertainty). Reliability decreases with the noise of a sensory signal.

Ventriloquist illusion: a multisensory perceptual illusion induced by presenting two signals from different sensory modalities in synchrony but at different spatial locations. In classical audiovisual cases, the perceived location of a sound is shifted towards the actual location of the visual signal and vice versa [18,50–52].



Trends in Cognitive Sciences

Figure 1. Metacognition in Multisensory Perception. Left: Generative model. The generative model of Bayesian Causal Inference for spatial localization determines whether the 'visual car' and the 'sound of the horn' are generated by common (C = 1) or independent (C = 2) sources (for details see [18]). For a common source, the 'true' audiovisual location (S_{AV}) is drawn from one prior spatial distribution. For independent sources, the 'true' auditory (S_A) and 'true' visual (S_V) locations are drawn independently from this prior spatial distribution. We introduce independent sensory noise to generate auditory (x_A) and visual (x_V) inputs [18]. Middle: Bayesian inference model. During perceptual inference the observer is thought to compute three sorts of estimates from the auditory and visual signals for spatial localization: (i) spatial estimates under the assumption of a common source (i.e., forced-fusion estimate: $\widehat{S_{AV,C=1}}$) and independent sources (i.e., full-segregation estimates separately for auditory and visual locations: $\widehat{S_{V,C=2}}$, $\widehat{S_{A,C=2}}$); (ii) estimates of the causal structure; and (iii) the final auditory and visual Bayesian Causal Inference spatial estimates based on model averaging that take into account the observer's causal uncertainty by marginalizing (i.e., integrating) over the different causal structures: $\widehat{S_V}$, $\widehat{S_A}$). Each of those estimates is associated with uncertainties as indicated by the specified probability distributions. Right: Metacognition may be able to access and monitor the three sorts of estimates and their uncertainty: (i) forced-fusion and full-segregation spatial estimates; (ii) the inferred causal structure; and (iii) the final auditory and visual Bayesian Causal Inference spatial estimates. Note that this image is for illustrational purposes only and does not indicate potential locations of neural substrates of metacognition.

In the laboratory, the principles of multisensory integration can be studied by presenting conflicting and non-conflicting signals. For instance, if auditory and visual signals are presented in synchrony but at different spatial locations, the **ventriloquist illusion** emerges. The perceived sound location shifts towards the location of a spatially distant visual signal and vice versa depending on the relative auditory and visual reliabilities. Importantly, spatial biasing is reduced at large spatial disparities when it is unlikely that the two signals come from a common source [50,51]. This attenuation of sensory integration at large spatial disparities is well accommodated by hierarchical Bayesian Causal Inference, which explicitly models the potential causal structures that could have generated the sensory signals; that is, whether auditory and visual signals come from common or independent sources [18,52] (for related models based on heavy-tailed prior distributions, see [17,53,54]). During perceptual inference, the observer is then thought to invert this generative process. Under the assumption of a common signal source, the two unisensory estimates of a physical property are combined and weighted according to their relative reliabilities (i.e., the inverse of variance). For instance, to estimate the location of a singing bird from audition and vision the observer should give a stronger weight to the visual signal during daytime than at night. Under the hypothesis of two different sources, the auditory and visual signals are treated independently. On a particular instance, the brain needs to infer the causal structure of the world (e.g., one or two sources) from the sensory inputs. Multiple sorts of **intersensory correspondences** [55], such as spatiotemporal coincidence (i.e., auditory and visual signals occurring at the same time and location [56–62]), semantic (e.g., the shape and singing of a bird) [63–65] and other higher-order correspondences (e.g., gender: female voice with female face), can inform the brain about whether signals are likely to come from a common source or independent sources. Finally, an estimate of the physical property in question (e.g., auditory location) is obtained by combining the estimates under the two causal structures using different decisional functions [18,52,66]. For instance, using model averaging observers may form a final estimate by averaging the estimates from the two causal structures weighted by their posterior probabilities. Alternatively, they may report the estimate of the most likely causal structure as the final estimate, a decisional strategy referred to as model selection.

Monitoring Uncertainties About the World's Causal Structure and Environmental Properties

The additional complexity of multisensory perception or, more generally, tasks that rely on multiple information channels raises questions and challenges that go beyond metacognition studied with, for example, simple visual discrimination or detections tasks. In particular, it raises the question of whether observers can monitor the different sorts of uncertainties involved in Bayesian Causal Inference.

First, observers may monitor their uncertainty about the causal structure that has generated the sensory signals [18,19,66]. The uncertainty about the causal structure increases with the noise in the sensory channels. For instance, at dawn it is more difficult (i.e., associated with greater uncertainty) to attribute a singing voice to a specific bird in the bush than in bright sunlight. Hence, the uncertainty about the inferred causal structure critically depends on the sensory uncertainty given in all sensory channels [52]. Moreover, causal uncertainty emerges because there is some natural variability in the temporal, spatial, or higher-order (e.g., semantic) relationship of the sensory signals. Even when two signals are generated by a common source, they do not need to be precisely temporally synchronous or spatially collocated. For speech signals, it is well established that visual facial movements often precede the auditory signal to variable degrees at speech onset [67]. Further, differences in the velocities of light and sound induce variability in the arrival times of visual and auditory signals at the receptor level that depends on the distance of the physical source from the observer [68,69]. Likewise, higher-order correspondences such as gender or semantics may relate probabilistically to low-level physical features (e.g., a low-pitched voice is more likely to be associated with a male than a female person). Experimentally, we therefore need to determine whether observers' causal uncertainty reflects the uncertainty determined by the signal-to-noise ratio of the sensory signals and their spatiotemporal and higher-order (e.g., semantic) statistical relationships. Moreover, causal uncertainty may be influenced by participants' prior expectations [70,71] that sensory signals are likely to come from a common external source or be generated by one's own voluntary actions [72,73] (Box 2).

Second, it is well established that observers use the uncertainty associated with the individual cues or sensory signals to assign the appropriate weighting during cue combination or multisensory integration. However, an unresolved question is whether these uncertainty estimates for individual cues are then lost or accessible for metacognition. To approach these questions, future experiments may consider asking observers to explore objects visuohaptically (i.e., via vision and touch) and report both the haptic size they perceived and their uncertainty about their perceptual estimate in the context of the visual information as well as if they had fully ignored the visual information (e.g., they may be asked to imagine that they had closed their eyes and only haptically explored the object). If observers maintain partial access to the unisensory estimates and their associated uncertainties, we would expect that the two reports differ.

Finally, observers may monitor their uncertainty associated with their final perceptual estimate (e.g., the reported location during audiovisual localization tasks). According to Bayesian Causal Inference, these final (e.g., auditory and visual) perceptual estimates are formed by combining

Box 2. Causal Metacognition and Sense of Agency

Causal inference enables the brain to dissociate the sensory effects caused by one's own actions from those caused by other agents or events in the outside world. Previous neuroimaging and neurophysiological studies have suggested that the cerebellum may form a predictive forward model that maps from the action plan to the motor outputs and their sensory consequences. These forward models enable the brain to distinguish between self- and other-generated sensory signals leading to effects such as sensory attenuation (e.g., the predicted outputs of tickling ourselves are not felt as tickling [100]) or intentional binding (e.g., the temporal interval between a voluntary action and its sensory consequences is subjectively compressed [72,73,83]; Figure I). Both effects are considered central to our sense of agency, which is the subjective judgment or feeling that we are causally responsible for changes in the environment. Critically, the temporal-compression effect was increased in patients with schizophrenia, indicating an enhanced sense of agency [85–87]. From the perspective of causal metacognition, we would expect the sense of agency to be related to the degree of confidence about our beliefs that a certain sensory outcome was self- rather than other-generated [84]. Further, manipulating biases in confidence by prior context or instructions may influence sensory attenuation and intentional binding, even when the sensory and motor components are held constant. For instance, if an agent is more confident that he/she has generated certain sensory signals, he/she should experience the same signal as less tickling and the interval between the action and the occurrence of the tickling sensation to be less compressed in time. A critical question for future research is therefore whether the altered sense of agency in patients with schizophrenia [85] may be associated with more general changes in causal metacognition.

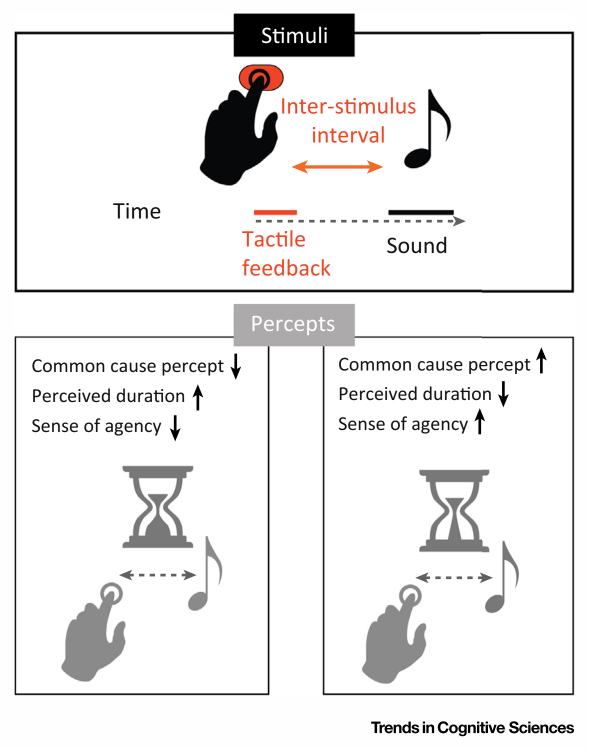


Figure I. Intentional Binding, Sense of Agency, and Causal Metacognition. Observers have been shown to perceive an interval of a certain duration between an action and its sensory consequences (e.g., a 'beep') as temporally compressed when the action was voluntary and associated with a sense of agency – a phenomenon referred to as 'intentional binding' [72]. Causal metacognition may be closely related to the sense of agency by virtue of monitoring uncertainty about the causal relationship between one's own voluntary actions and their sensory consequences.

the estimates under the assumptions of common and independent sources according to various decision functions such as model averaging, probability matching, or model selection [66]. As a result, the uncertainty of these final Bayesian Causal Inference perceptual estimates is dependent on observers' sensory and causal uncertainty. A critical question for future investigation is to determine the extent to which observers' uncertainty about their reported perceptual estimate reflects their perceived causal uncertainty.

A few studies have started to directly tackle the question of metacognitive uncertainty or confidence estimates in multisensory perception, albeit not always with these different sorts of uncertainties in mind. For instance, a recent psychophysical study [74] demonstrated that observers correctly assessed the accuracy of their temporal-order judgments in confidence ratings. These results indicate that the precision of audiovisual temporal relation estimates is accessible to metacognition. Further, a recent study by White and colleagues [75] presented observers with audiovisually non-conflicting (e.g., visual <<ba>> with auditory /ba/) conflicting phonemic cues that could be integrated into a so-called **McGurk percept** (e.g., McGurk: visual <<ga>> with auditory /ba/ resulting in an illusory [da] percept), and conflicting phonemic cues that could not be integrated into one unified percept (i.e., non-McGurk: visual <<pa>> with auditory /ka/). Observers reported their perceived auditory phoneme immediately before providing a second-order confidence rating. The authors demonstrated that observers were less confident about their illusory McGurk percepts than about their auditory percept for conflicting or non-conflicting stimuli. From a Bayesian Causal Inference perspective, observers' lower confidence about their McGurk responses may emerge from an increase in causal uncertainty for McGurk stimuli. While non-conflicting signals are likely to come from a common source and conflicting signals from independent sources, McGurk stimuli introduce an intermediate phonological conflict that introduces uncertainty about the underlying causal structure. This causal uncertainty may indirectly influence and increase observers' uncertainty about their final phoneme percept. However, this is only one of several possible explanations for the observed response profile (see also [76]). It highlights the need for future dual-task paradigms that ask observers concurrently to rate not only their confidence about their phonological percept but also their causal uncertainty about whether sensory signals (e.g., auditory phoneme and facial movements in speech recognition) were generated by a common source.

Perceptual and Causal Metamers

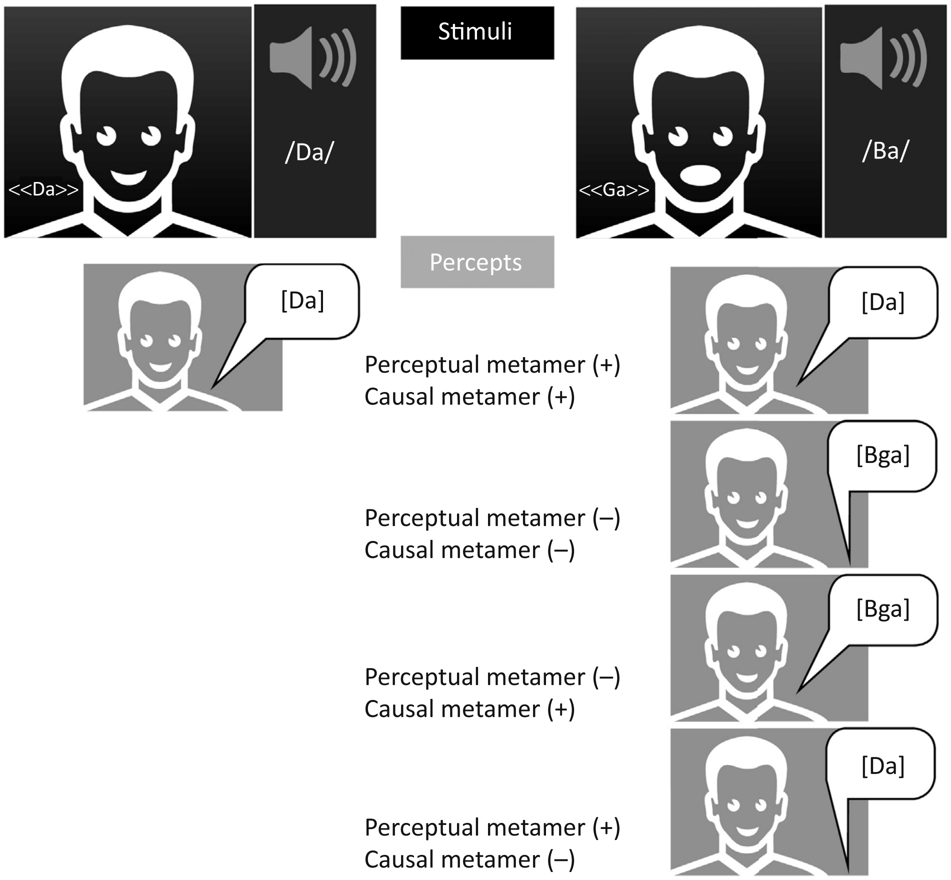
Further insights into whether observers can move beyond the integrated percept and metacognitively monitor the perceptual inference can be obtained from so-called metamers; that is, (near-)identical perceptual interpretations formed from different combinations of sensory signals [77]. Let us assume we present an observer with two signals in synchrony: a brief flash at -2° visual angle (i.e., left) and a spatially equally reliable beep at $+2^\circ$ visual angle (i.e., right). Where will the observer perceive this event? Because of the small audiovisual spatial disparity, the observer may infer that the two signals come from a common source and hence integrate them weighted by their relative reliabilities. As a result he would perceive the audiovisual event at 0° visual angle, where in fact no signal was presented at all. Hence, this conflicting flash–beep event would elicit the same percept as a non-conflicting flash–beep event where both auditory and visual signals are presented at 0° visual angle. In other words, the conflicting and non-conflicting flash–beep events elicit **perceptual metamers**. Moreover, the observer inferred that the auditory and visual signals come from a single event in both situations. Hence, the two cases elicit not only perceptual but also **causal metamers**. The critical question is whether observers may nevertheless be able to discriminate between the conflicting and non-conflicting flash–beep events, indicating that they can metacognitively access additional information about the underlying perceptual inference process.

First, observers would be able to discriminate between the non-conflicting and conflicting signals if they monitor their uncertainty about their perceptual interpretation and causal inference. In the small-conflict case, observers who use Bayesian Causal Inference with model selection may decide that the two signals come from a common source and integrate them weighted by their relative reliabilities. Critically, although they commit to a single event as the more likely causal structure, they should be less certain about their causal inference. In other words, monitoring their causal uncertainty would allow observers to discriminate between conflicting and non-conflicting sensory signals even if they elicit perceptual and causal metamers. Within the framework of Bayesian Causal Inference and depending on decisional functions and biases [66], it is also conceivable that observers may integrate different combinations of auditory and visual signals into the same perceptual (e.g., auditory, visual) estimates and yet report different causal structures. Hence, perceptual metamers may not necessarily imply causal metamers.

Second, observers may be able to go beyond the integrated percept and maintain at least partial access to the individual sensory signals (see discussion above). Again, this partial access would allow them to discriminate between conflicting and non-conflicting flash–beep events. In a wider sense of metacognition, it would demonstrate that multisensory perception is not informationally encapsulated but that observers can introspect and metacognitively monitor the unisensory representations that form the basis for their perceptual inference.

Surprisingly, only a few studies to date have used perceptual metamers as an approach to characterize observers' metacognitive access in cue combination. An intriguing early study by Hillis *et al.* [77] focused on the emergence of perceptual metamers in visual (slant from disparity and texture cues in vision) and visuohaptic (object size from vision and touch; i.e., haptic cues) contexts. In an oddity-judgment task, observers were asked to identify the odd stimulus in a sequence of three stimuli: two identical standard stimuli defined by non-conflicting cues and one odd stimulus defined by conflicting cues that could be fused into a perceptual metamer of the standard stimulus [77,78]. The results revealed that observers lost access to individual cues in the visual but not in the visuohaptic setting: only conflicting visual cues were mandatorily fused into perceptual metamers of the non-conflicting standard stimulus. Yet even in the visual case participants were able to discriminate the conflicting stimulus from the non-conflicting ones for larger conflict sizes, indicating that metamers emerge only for small conflict size. What happened, however, in unisensory cases with larger conflict? As the oddity-judgment task does not explicitly define the dimension according to which participants should compare the stimuli, it remains unclear whether observers identified the conflicting stimulus because they did not integrate the conflicting cues into one unified slant estimate (i.e., into a perceptual metamer of the non-conflicting stimulus) or whether instead they integrated them but were aware that their metameric percepts emerged from different causal structures or at least were associated with different causal uncertainties. Observers may still have fused conflicting signals into approximate perceptual metamers without them being causally metameric to the non-conflicting standard stimulus. In other words, observers may potentially have identified the odd one out because of partial access to the causal structure that has generated the sensory inputs. In line with this conjecture, observers reported a 'weird' percept for larger conflict sizes (M. Ernst, personal communication), indicating that they were aware of the conflict manipulation while still integrating signals into a near-unified percept. This may perhaps be taken as initial evidence that perceptual and causal metamers may be to some extent dissociable. Future studies that explicitly assess the emergence of perceptual and causal metamers are needed to experimentally determine whether participants can form perceptual metamers while recognizing that they are based on different causal structures.

Another approach to dissociate perceptual and causal metamers is to introduce conflicts along multiple dimensions, such as lower temporal and higher-order phonological dimensions. For instance, observers may be presented with conflicting and non-conflicting visual and auditory phonetic cues at multiple audiovisual asynchronies. For small audiovisual asynchronies, conflicting audiovisual signals such as a visual <<ga>> and an auditory /ba/ may be fused into a [da] percept at the phonological level as in the classical McGurk–MacDonald illusion [79] (Figure 2). The critical question is whether the fusion of conflicting audiovisual signals into a [da] percept as a perceptual metamer of a non-conflicting audiovisual [da] emerges in cases where observers inferred that the two signals came from different sources because of their audiovisual asynchrony (i.e., no causal metamer).



Trends in Cognitive Sciences

Figure 2. Perceptual and Causal Metamers in the Audiovisual McGurk Illusion. Left: Observers are presented with non-conflicting audiovisual stimuli; that is, a video of a face articulating <<da>> and a voice uttering /da/. They will perceive the audiovisual signals as coming from one source and integrate them into a [da] percept. Right: Observers are presented with conflicting audiovisual stimuli; that is, a video of a face articulating <<ga>> and a voice uttering /ba/. In the McGurk illusion, they should perceive the audiovisual signals as coming from one source and integrate them into a [da] percept, which would be a causal and perceptual metamer to the estimates formed from the non-conflicting audiovisual signals. However, perceptual and causal inference may result in other outcomes. Observers may potentially perceive a [da] and yet recognize the audiovisual conflict and hence infer that the two signals come from independent sources (i.e., perceptual metamer but no causal metamer).

Research showing that the temporal integration windows that allow the McGurk illusion to emerge mostly correspond to those where observers perceive the audiovisual signals as being synchronous has suggested that the detection of temporal conflicts precludes the emergence of perceptual metamers [80]. However, other evidence suggests that conflicting visual phonetic information influences the perceived auditory phonemes even when observers are able to detect low-level temporal conflicts [81]. In the light of this controversial evidence, future studies are needed to determine whether perceptual metamers at higher representational levels emerge even when lower-level temporal conflicts prevent the emergence of causal metamers.

Concluding Remarks

Accumulating evidence shows that human observers can metacognitively assess the uncertainty of perceptual estimates formed from vision, touch, or audition in unisensory perception. Conversely, research in multisensory perception demonstrates that observers integrate signals from multiple sensory modalities into percepts that take into account the uncertainty about the

Outstanding Questions

To what extent can observers metacognitively monitor the individual signals, the inferred causal structure, and their respective uncertainties in sensory or cue integration? Do their perceptual uncertainties reflect their causal uncertainties and vice versa?

How does causal metacognition in perception relate to metacognition in other cognitive domains such as causal reasoning or social interactions?

What are the benefits of causal metacognition in perception? Do observers adjust their future perceptual interpretations based on their causal metacognitive assessments?

Is the sense of agency grounded in causal metacognition?

Which neural circuitries sustain causal metacognition during perceptual and other cognitive tasks in the human brain?

Is causal metacognition impaired in neuropsychiatric diseases such as schizophrenia?

How does causal metacognition develop during infancy and childhood? Does it emerge later than metacognition about perceptual decisions based on a single information stream?

Nonhuman organisms have been shown to monitor their uncertainties about their perceptual decisions. Can they also monitor their uncertainty about the causal structure of the world?

Box 3. Challenging Causal Naivety Assumptions in Philosophy

The capacity to represent causation is usually granted only on the evidence that explicit causal reasoning and inferences to hidden or distant causes are performed. As Hume's challenge goes, there is a difference between predicting that one event regularly follows another and representing that it was caused by the first event. This view, which started in philosophical discussions [93], is also widespread in psychology [94]. Does causal metacognition challenge this claim, suggesting that we are sensitive to differences between hidden causal structures when we perceive events? How sophisticated do we need to be to monitor the uncertainty of our causal models of the world?

Evidence of causal metacognition in younger children and nonhuman animals should address this question and possibly reveal whether hidden causal structures are accessed and monitored as such even in the absence of more explicit causal reasoning. However, causal metacognition brings a broader challenge to philosophical models of perception. It is widely assumed that we are causally naive when it comes to perceiving the world: perception does not make us aware of objects as causes of our perception [95]. When we perceive a singing bird, we do not see that a physical bird, or light, is causing our perception: we perceive a bird as a mind-independent object not as a likely cause of our percept. The claim that perception rests on a process of causal inference at the subpersonal level [96,97], although widely accepted by cognitive neuroscientists, explains from the outside what the system is set up to do but does not suppose that causes are represented as such, or even less consciously accessed [98,99]. Sensitivity to differences in the causal origin of our integrated percepts offers an intermediate step where the causal character of perception is made manifest.

How this form of causal metacognition fits within causal cognition in general, and whether it is also present in more explicit forms of reasoning, is an open question. While it is common to stress the difference between aggregating information between agents and combining information from different sensory modalities, it might be the case that both are optimal if the uncertainty about the underlying causal model dictating the problem is adequately monitored.

world's causal structure. In this review we have merged these two research fields and discussed the new challenges and questions that metacognition poses for situations where the brain needs to integrate information from multiple channels, such as in multisensory perception and cue combination. Recent developments of hierarchical Bayesian models of multisensory perception raise the possibility that human observers can introspect perceptual inference processes and monitor not only the final integrated percept but also the unisensory estimates and the causal relationship, thereby challenging the long-dominant view in philosophy that observers are causally naive about perceptual inference (Box 3). Future studies in causal metacognition will need to determine the extent to which human observers can accurately assess their uncertainty about the perceptual estimates and the inferred causal structure of the environment. They open new research avenues that link metacognition in perception more tightly with higher-order cognitive capacities such as abstract causal reasoning [82] or the aggregation of information across agents (Box 1) (see Outstanding Questions). Causal metacognition sheds new light on the emergence of the **sense of agency** [83] (Box 2) and will be critical for our understanding of neuropsychiatric diseases such as schizophrenia that affect multisensory binding, causal inference, and metacognitive control [75,84–87].

Acknowledgments

The authors thank Ulrik Beierholm, Peter Dayan, Chris Frith, Nick Shea, David Meijer, Hakwan Lau, Steve Fleming, and Bahador Bahrami for helpful discussions. O.D. and C.S. are funded by AHRC UK (Rethinking the Senses); U.N is funded by the European Research Council (ERC-multsens).

References

1. Flavell, J.H. (1979) Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. *Am. Psychol.* 34, 906–911
2. Fleming, S.M. *et al.* (2012) Metacognition: computation, biology and function. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367, 1280–1286
3. Yeung, N. and Summerfield, C. (2012) Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367, 1310–1321
4. Overgaard, M. and Sandberg, K. (2012) Kinds of access: different methods for report reveal different kinds of metacognitive access. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1287–1296
5. De Gardelle, V. *et al.* (2016) Confidence as a common currency between vision and audition. *PLoS One* 11, e0147901
6. Ais, J. *et al.* (2016) Individual consistency in the accuracy and distribution of confidence judgments. *Cognition* 146, 377–386
7. Fleming, S.M. *et al.* (2012) Prefrontal contributions to metacognition in perceptual decision making. *J. Neurosci.* 32, 6117–6125
8. Ratcliff, R. and Starns, J.J. (2013) Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychol. Rev.* 120, 697–719

9. Rutishauser, U. *et al.* (2015) Representation of retrieval confidence by single neurons in the human medial temporal lobe. *Nat. Neurosci.* 18, 1–12

10. Bahrami, B. *et al.* (2010) Optimally interacting minds. *Science* 329, 1081–1085

11. Goupil, L. *et al.* (2016) Infants ask for help when they know they don't know. *Proc. Natl Acad. Sci. U.S.A.* 113, 3492–3496

12. Heyes, C. (2016) Who knows? Metacognitive social learning strategies. *Trends Cogn. Sci.* 20, 204–213

13. Van Den Berg, R. *et al.* (2016) A common mechanism underlies changes of mind about decisions and confidence. *Elife* 5, e12192

14. Fleming, S.M. and Lau, H.C. (2014) How to measure metacognition. *Front. Hum. Neurosci.* 8, 443

15. de Gardelle, V. and Mamassian, P. (2014) Does confidence use a common currency across two visual tasks? *Psychol. Sci.* 25, 1286–1288

16. Barthelmé, S. and Mamassian, P. (2010) Flexible mechanisms underlie the evaluation of visual confidence. *Proc. Natl Acad. Sci. U.S.A.* 107, 20834–20839

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

18. Kording, K.P. *et al.* (2007) Causal inference in multisensory perception. *PLoS One* 2, e943

19. Rohe, T. and Noppeney, U. (2015) Cortical hierarchies perform Bayesian Causal Inference in multisensory perception. *PLoS Biol.* 13, e1002073

20. Meyniel, F. *et al.* (2015) Confidence as Bayesian probability: from neural origins to behavior. *Neuron* 88, 78–92

21. Pouget, A. *et al.* (2016) Confidence and certainty: distinct probabilistic quantities for different goals. *Nat. Neurosci.* 19, 366–374

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

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

24. Smith, J.D. *et al.* (2012) The highs and lows of theoretical interpretation in animal-metacognition research. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1297–1309

25. Shea, N. *et al.* (2014) Supra-personal cognitive control and metacognition. *Trends Cogn. Sci.* 18, 186–193

26. Tenenbaum, J.B. and Griffiths, T.L. (2003) Theory-based causal inference. *Adv. Neural Inf. Process. Syst.* 15, 35–42

27. von Helmholtz, H. (1867) *Handbuch der Physiologischen Optik,* Leopold Voss (in German)

28. Kersten, D. *et al.* (2004) Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304

29. Aitchison, L. *et al.* (2015) Doubly Bayesian analysis of confidence in perceptual decision-making. *PLoS Comput. Biol.* 11, e1004519

30. Massoni, S. *et al.* (2014) Confidence measurement in the light of signal detection theory. *Front. Psychol.* 5, 1455

31. Dienes, Z. and Seth, A. (2010) Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious. Cogn.* 19, 674–681

32. Persaud, N. *et al.* (2007) Post-decision wagering objectively measures awareness. *Nat. Neurosci.* 10, 257–261

33. Clifford, C.W.G. *et al.* (2008) Getting technical about awareness. *Trends Cogn. Sci.* 12, 54–58

34. Fleming, S.M. *et al.* (2015) Action-specific disruption of perceptual confidence. *Psychol. Sci.* 26, 89–98

35. Maniscalco, B. and Lau, H. (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* 21, 422–430

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

37. Körding, K.P. and Wolpert, D.M. (2004) Bayesian integration in sensorimotor learning. *Nature* 427, 244–247

38. Ma, W.J. *et al.* (2006) Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438

39. Pouget, A. *et al.* (2003) Inference and computation with population codes. *Annu. Rev. Neurosci.* 26, 381–410

40. Fiser, J. *et al.* (2010) Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14, 119–130

41. Middlebrooks, P.G. and Sommer, M.A. (2012) Neuronal correlates of metacognition in primate frontal cortex. *Neuron* 75, 517–530

42. Del Cul, A. *et al.* (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132, 2531–2540

43. Kiani, R. and Shadlen, M.N. (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759–764

44. Fetsch, C.R. *et al.* (2014) Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83, 797–804

45. Murphy, P.R. *et al.* (2015) Neural evidence accumulation persists after choice to inform metacognitive judgments. *Elife* 4, e11946

46. Grimaldi, P. *et al.* (2015) There are things that we know that we know, and there are things that we do not know we do not know: confidence in decision-making. *Neurosci. Biobehav. Rev.* 55, 88–97

47. Lau, H.C. and Passingham, R.E. (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl Acad. Sci. U.S.A.* 103, 18763–18768

48. Helbig, H.B. *et al.* (2012) The neural mechanisms of reliability weighted integration of shape information from vision and touch. *Neuroimage* 60, 1063–1072

49. Fetsch, C.R. *et al.* (2012) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat. Neurosci.* 15, 146–154

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

51. Wallace, M.T. *et al.* (2004) Unifying multisensory signals across time and space. *Exp. Brain Res.* 158, 252–258

52. Rohe, T. and Noppeney, U. (2015) Sensory reliability shapes perceptual inference via two mechanisms. *J. Vis.* 15, 22

53. Roach, N.W. *et al.* (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc. Biol. Sci.* 273, 2159–2168

54. Knill, D.C. (2003) Mixture models and the probabilistic structure of depth cues. *Vision Res.* 43, 831–854

55. Spence, C. and Deroy, O. (2013) How automatic are crossmodal correspondences? *Conscious. Cogn.* 22, 245–260

56. Bonath, B. *et al.* (2014) Audio-visual synchrony modulates the ventriloquist illusion and its neural/spatial representation in the auditory cortex. *Neuroimage* 98, 425–434

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

58. Rohe, T. and Noppeney, U. (2016) Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Curr. Biol.* 26, 509–514

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

60. Vatakis, A. and Spence, C. (2010) Audiovisual temporal integration for complex speech, object–action, animal call, and musical stimuli. In *In Multisensory Object Perception in the Primate Brain*, pp. 95–121, Springer

61. Parise, C.V. *et al.* (2012) When correlation implies causation in multisensory integration. *Curr. Biol.* 22, 46–49

62. Lee, H. and Noppeney, U. (2011) Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl Acad. Sci. U.S.A.* 108, E1441–E1450

63. Hein, G. *et al.* (2007) Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 27, 7881–7887

64. Laurienti, P.J. *et al.* (2004) Semantic congruence is a critical factor in multisensory behavioral performance. *Exp. Brain Res.* 158, 405–414

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

66. Wozny, D.R. *et al.* (2010) Probability matching as a computational strategy used in perception. *PLoS Comput. Biol.* 6, e1000871

67. Chandrasekaran, C. *et al.* (2009) The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5, e1000436

68. Spence, C. and Squire, S. (2013) Multisensory integration: maintaining the perception of synchrony. *Curr. Biol.* 13, R519–R521

69. Fujisaki, W. *et al.* (2004) Recalibration of audiovisual simultaneity. *Nat. Neurosci.* 7, 773–778

70. Gau, R. and Noppeney, U. (2016) How prior expectations shape multisensory perception. *Neuroimage* 124, 876–886

71. Nahorna, O. *et al.* (2012) Binding and unbinding the auditory and visual streams in the McGurk effect. *J. Acoust. Soc. Am.* 132, 1061–1077

72. Haggard, P. (2005) Conscious intention and motor cognition. *Trends Cogn. Sci.* 9, 290–295

73. Wolpe, N. *et al.* (2013) Cue integration and the perception of action in intentional binding. *Exp. Brain Res.* 229, 467–474

74. Keane, B. *et al.* (2015) Metacognition of time perception. *J. Vis.* 15, 814

75. White, T.P. *et al.* (2014) Eluding the illusion? Schizophrenia, dopamine and the McGurk effect. *Front. Hum. Neurosci.* 8, 565

76. Abadi, R.V. and Murphy, J.S. (2014) Phenomenology of the sound-induced flash illusion. *Exp. Brain Res.* 232, 2207–2220

77. Hillis, J.M. *et al.* (2002) Combining sensory information: mandatory fusion within, but not between, senses. *Science* 298, 1627–1630

78. Hospedales, T. and Vijayakumar, S. (2009) Multisensory oddity detection as Bayesian inference. *PLoS One* 4, e4205

79. McGurk, H. and Macdonald, J. (1976) Hearing lips and seeing voices. *Nature* 264, 691–811

80. van Wassenhove, V. *et al.* (2007) Temporal window of integration in auditory–visual speech perception. *Neuropsychologia* 45, 598–607

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

82. Tenenbaum, J.B. *et al.* (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.* 10, 309–318

83. Haggard, P. *et al.* (2002) Voluntary action and conscious awareness. *Nat. Neurosci.* 5, 382–385

84. Adams, R.A. *et al.* (2013) The computational anatomy of psychosis. *Front. Psychiatry* 4, 47

85. Frith, C.D. *et al.* (2000) Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res. Brain Res. Rev.* 31, 357–363

86. Haggard, P. *et al.* (2003) Awareness of action in schizophrenia. *Neuroreport* 14, 1081–1085

87. Voss, M. *et al.* (2010) Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain* 133, 3104–3112

88. Tenenbaum, J.B. *et al.* (2011) How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285

89. Williamson, J. (2005) *Bayesian Nets and Causality: Philosophical and Computational Foundations,* Oxford University Press

90. Spirtes, P. *et al.* (2000) *Causation, Prediction, and Search.* (2nd edn), MIT Press

91. Griffiths, T.L. *et al.* (2007) Topics in semantic representation. *Psychol. Rev.* 114, 211–244

92. Bradley, R. *et al.* (2014) Aggregating causal judgments. *Philos. Sci.* 81, 491–515

93. Hume, D. and Beauchamp, T.L. (1998) *An Enquiry Concerning the Principles of Morals: A Critical Edition,* Oxford University Press

94. Sperber, D. *et al.* (1995) *Causal Cognition: A Multidisciplinary Debate,* Clarendon Press

95. Matthen, M. (ed.) (2015) *Oxford Handbook of Philosophy of Perception*, Oxford University Press

96. Dennett, D.C. (1969) *Content and Consciousness,* Routledge

97. Drayson, Z. (2014) The personal/subpersonal distinction. *Philos. Compass* 9, 338–346

98. Block, N. (2011) Perceptual consciousness overflows cognitive access. *Trends Cogn. Sci.* 15, 567–575

99. Bayne, T. *et al.* (2016) Are there levels of consciousness? *Trends Cogn. Sci.* 20, 405–413

100. Blakemore, S.J. *et al.* (1998) Central cancellation of self-produced tickle sensation. *Nat. Neurosci.* 1, 635–640