Bayesian causal inference: A unifying neuroscience theory

Ladan Shams [a],[*], Ulrik Beierholm [b]

[a] Departments of Psychology, BioEngineering, and Neuroscience Interdepartmental Program, University of California, Los Angeles, USA
[b] Department of Psychology, University of Durham, UK

ABSTRACT

Understanding of the brain and the principles governing neural processing requires theories that are parsimonious, can account for a diverse set of phenomena, and can make testable predictions. Here, we review the theory of Bayesian causal inference, which has been tested, refined, and extended in a variety of tasks in humans and other primates by several research groups. Bayesian causal inference is normative and has explained human behavior in a vast number of tasks including unisensory and multisensory perceptual tasks, sensorimotor, and motor tasks, and has accounted for counter-intuitive findings. The theory has made novel predictions that have been tested and confirmed empirically, and recent studies have started to map its algorithms and neural implementation in the human brain. The parsimony, the diversity of the phenomena that the theory has explained, and its illuminating brain function at all three of Marr's levels of analysis make Bayesian causal inference a strong neuroscience theory. This also highlights the importance of collaborative and multi-disciplinary research for the development of new theories in neuroscience.

1. Introduction

Neuroscience has been one of the fastest growing areas of science in the last two decades. While the proliferation of empirical findings in this area of research has been dizzying, parsimonious and unifying theories that probe the principles of cognitive function have been far and few in between. Here we propose Bayesian causal inference as a parsimonious and unifying theory in cognitive neuroscience and examine its evolution, successes, and limitations in that context.

In mid 2000s, in an attempt to account for auditory-visual perceptual phenomena ranging from integration to segregation a Bayesian model, known as Bayesian Causal Inference (Bayesian CI) model (Körding et al., 2007), was proposed that involved a competition between two hypotheses, a common cause and independent causes (see Box 1). In the last 15 years, the model has been extended, refined, and adapted to account for a large number of perceptual and sensorimotor phenomena, and vast amounts of behavioral data. More recently the neural underpinnings of Bayesian CI have been the subject of extensive research. The model is mathematically similar to a few other models that had been proposed in other domains, and the core computation involved in Bayesian CI appears to be at work in a number of diverse perceptual and sensorimotor domains in humans and other species (Shams and Beierholm, 2010). Bayesian causal inference is a computation that appears to

be frequently employed in a variety of cognitive tasks and domains, and appears to have a long-standing evolutionary root (Shams and Beierholm, 2010). We therefore, henceforth refer to the core computation, which involves competitive priors, as Bayesian Causal Inference theory. We will refer to the Bayesian causal inference model of multisensory perception as a model within that theory which has been most extensively studied, provides an apt example of Bayesian Causal Inference theory, and will be the main focus of this review (Fig. 1).

Here we will describe Bayesian CI using two examples in the perceptual domain, however, the same core computations apply to other tasks and domains of processing as described later.

When faced with sensory stimuli, the nervous system has to estimate the events/sources of sensory inputs and has to overcome two challenges: (a) determining the causal structure of the signals, and (b) the identity of the source(s) that gave rise to the stimuli. However, these two challenges are intertwined, and cannot be addressed separately and independently.

For example, if we see a talking face and hear speech, the speech perception system has to determine whether the two signals originated from the same source (the auditory and visual signals were produced by the same person), or whether they came from different sources (e.g., the muted video from TV, and the sound of someone speaking in the room). The degree of similarity/discrepancy (in time, space, content) between

Box 1

Hierarchical Bayesian Causal Inference.

The Bayesian Causal Inference model of multisensory perception is a statistical model that essentially infers the more likely of two causal structures, given sensory inputs and assumptions about the structures. As an example, here we will use the bisensory perception of Fig. 2a which is simple and has been studied most extensively. Assume that a set of sensory inputs X_a and X_v can have been generated by either of two causal structures, here denoted $C=1$ or $C=2$. The inference about which structure generated the sensory data can be an be statistically described (using Bayes rule (Bayes, 1763)) through the posterior probability

$$P(C = 1|X_a, X_v) = P(X_a, X_v|C = 1)P(C = 1)/P(X_a, X_v)$$

where the prior $P(C=1)$ specifies the expectation of a common cause by the nervous system a priori, before receiving the sensory input, while the likelihood $P(X_a, X_v|C=1)$ encodes how likely it is that the sensory inputs have been generated from a common cause structure.

The estimation of the latent variables depends on the probabilities of the causal structures, and again can be computed using Bayes Rule. Assuming a mean-squared error cost function (i.e. that the subjects want on average to avoid large errors), the optimal estimate of the source S_a (e.g., the location of a sound) will be:

$$\widehat{S_a} = P(C = 1|X_a, X_v)\widehat{S}_{a,C=1} + P(C = 2|X_a, X_v)\widehat{S}_{a,C=2}$$

The estimate is a mixture of the estimate from each causal structure, weighted by how probable the structure is. As the posterior probability of the causal structure is a non-linear function of the sensory inputs, X_a and X_v, the optimal estimate is also non-linear function of the sensory cues. This is in contrast to the linear combination of sensory cues in the reliability-weighted (or maximum likelihood) models of cue combination (Alais and Burr, 2004; Clark and Yuille, 1990; Ernst and Banks, 2002). While the mean-squared error of latent variable S_a results in a model averaging strategy as the optimal estimate of S_a, other cost functions can lead to other estimates arising from, for example selecting the estimate arising from the more likely causal structure (model selection) or stochastically sampling from the distribution (probability matching, (Wozny et al., 2010)).

The competing causal structures can be thought of as priors, because they are models that have been learned by experience or over the course of evolution. For example, by experience individuals can learn that both circles and ovals can produce retinal projections that are oval shaped, giving rise to the two competing causal structures involved in slant perception (see Fig. 2c). Similarly, experience can teach the perceptual system that a given voice and given face can have either the same cause (produced by the same person) or two different causes (a speaker who is not visible and a visible person who is not speaking). Because these competing causal structures are models of the world and exist prior to any sensory experience, they are "priors" and hence the term "competitive priors" framework.
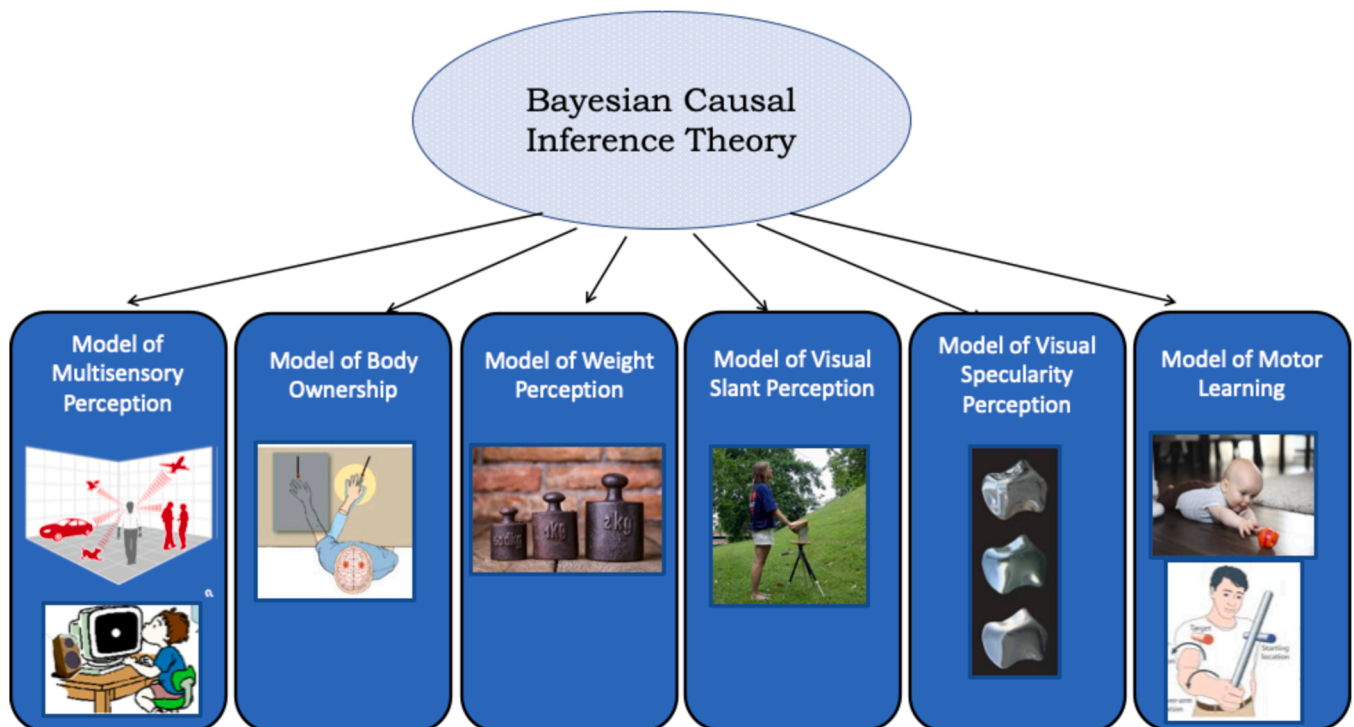
Fig. 1. Bayesian Causal Inference theory. Various models employing Bayesian inference and competitive priors on the causal structure have been proposed to account for a variety of tasks. The models are essentially computationally equivalent and fall under Bayesian Causal Inference Theory.
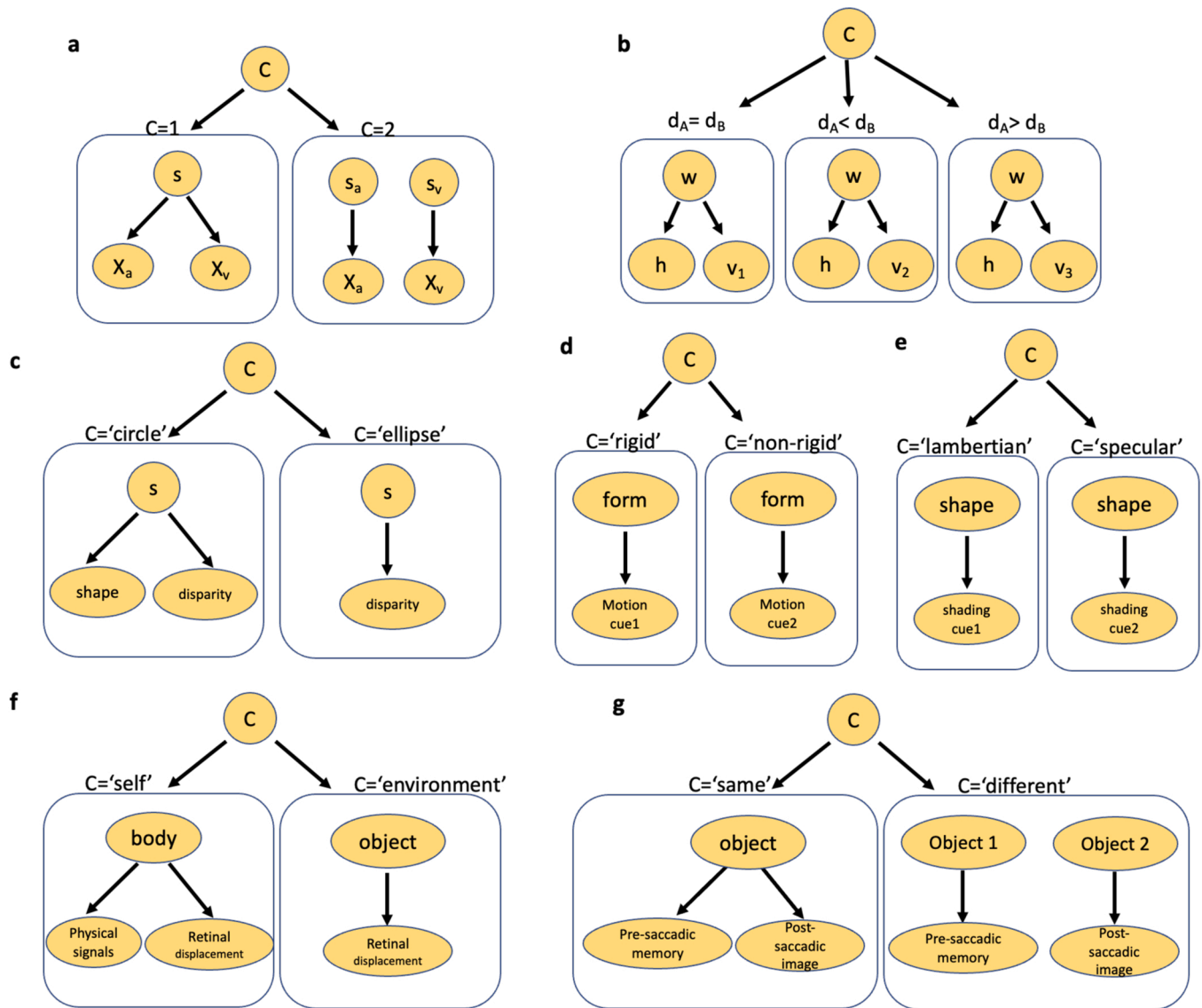
Fig. 2. Generative model of seven different Bayesian CI models. a) The generative model of Bayesian CI model of bisensory perception (Körding et al., 2007). b) The generative model of weight perception in size-weight illusion paradigm (Peters et al., 2016). c) The generative model of visual slant perception (Knill, 2007). d) The generative model of object form perception in the structure-from-motion paradigm (Yuille and Clark, 1994). e) The generative model of shape perception in the shape-from-shading paradigm (Yuille and Bulthoff, 1996). f) The generative model of postural control (self vs. environment motion) (Dokka et al., 2010). g) The generative model of visual stability in face of saccades (Atsma et al., 2016).

the auditory and visual inputs as well as our prior expectations about the structure of the world affect the interpretation of the causal structure (whether they originate from the same source or not), but likewise the causal structure is important as to whether the sensory inputs should be integrated or not (sensations with different causes should not be integrated), and thus, to the estimation of perceptual variables (the content of speech).

In this simple example, there are only two causal structures (or hypotheses) that need to be examined, the common cause and independent causes (see Fig. 2a). These two hypotheses can be thought of as two priors (or models of the world) that compete to explain the sensory data. For this reason, Bayesian CI model of multisensory perception is a form of competitive priors model (Colas et al., 2010; Yuille and Bulthoff, 1996; Yuille and Clark, 1994), and mathematically almost equivalent to competitive priors models that had previously been explored in vision science (Knill, 2007; Stocker and Simoncelli, 2006; Yuille and Bulthoff, 1996; Yuille and Clark, 1994).

As another example, consider the task of acting on objects in the environment which sometimes requires determining the slant of objects

and surfaces. Determining the slant of an oval shape in the visual field requires the visual system to first infer the shape of the object, whether it was an oval object or a circular object that has given rise to the oval projection (see Fig. 2c). The degree of similarity/discrepancy between the binocular disparity cue and shape cue, as well as our prior expectations about the world (how common oval vs. circular objects are) affect the interpretation of the causal structure (oval vs. circle), but likewise the causal structure is important as to whether the binocular disparity and shape signals should be integrated or not (aspect ratio of the oval shape is only informative about slant if the source is a circle). Therefore, here again causal structure (object form) priors/hypotheses compete to explain the sensory data, and inference based on sensory stimuli leads to the estimation of causal structure, and conversely, the inferred causal structure influences the estimation of perceptual variables (slant).

Bayesian CI is a normative framework that addresses both the problem of causal inference (which causal structure/hypothesis generated the stimuli) and the problem of integration/estimation of hidden variables, in a unified and coherent fashion using Bayesian inference.

The probability of each causal structure (e.g., *C=1*, see Fig. 2a or *C="circle"*, see Fig. 2c) is computed based on the prior probability of that causal structure (e.g., the expectation of a common cause, or the expectation of a circle) and the similarity/congruency between the sensory inputs, according to Bayes Rule. The estimate of each sensory source (e.g., what was said in the video/audio, or the slant of the object) is computed based on the probability of the causal structures (which determine the extent to which the sensations should be integrated) and the reliability of the sensory measurements (the more precise/reliable cue would play a more important role), again according to Bayes rule (see Box 1 for a detailed description).

Therefore, the core computations that constitute the Bayesian Causal Inference theory can be summarized as follows: (a) competition between priors (causal structures) to account for sensory data, (b) estimation of hidden perceptual/cognitive variables based on the inferred causal structure, (c) computation of both (a) and (b) using Bayes rule.

In the following sections, we will review studies of Bayesian CI within the framework suggested by David Marr (Marr, 1982), assessing the theory at a computational level, followed by algorithmic/representational level, followed by implementational level. While the initial studies and the vast majority of studies to date have focused on the computational level of analysis, in the last few years there has been an increasing number of studies that have probed the representations and brain mechanisms of Bayesian CI.

2. Computational level of analysis

A natural starting point for understanding human perception is to hypothesize that the human nervous system has evolved a strategy similar to Bayesian CI in solving perceptual/cognitive/sensorimotor problems. This hypothesis can be tested by comparing the behavior of Bayesian CI with that of human observers in specific tasks (Fig. 2, Fig. 3).

Below, we review the studies that have done exactly that, quantitatively or qualitatively comparing human observer data with predictions of Bayesian CI in a variety of tasks, and in a variety of sensory/sensorimotor conditions in each task. We classify the studies by the nature of the task they have tackled, and briefly summarize the findings in each section.

2.1. Spatial perception

Spatial localization of objects, as well as auditory-visual interactions in this process as exemplified by the ventriloquist illusion (Thurlow and Jack, 1973; Warren et al., 1981), have been studied extensively (e.g., Alais and Burr, 2004; Bertelson et al., 2000; Choe et al., 1975; Jack and Thurlow, 1973; Recanzone, 2003; Slutsky and Recanzone, 2001;

Wallace et al., 2004). Early computational models of multisensory localization had focused on the limited range of cases in which the senses deviated little from each other, and therefore perceptual cues were always fused into a single unified percept. However, when the auditory and visual stimuli differ in location substantially there is an absence of integration and human data exhibits a spectrum of phenomena ranging from segregation to integration. The Bayesian CI model explains this by the integration of perceptual cues when they are inferred to be causally linked (e.g. when proximal), and segregation of the cues when likely to originate from different sources (e.g. when far apart). Multiple studies of auditory-visual spatial localization have now been performed in which data from observers was compared with Bayesian CI (Beierholm et al., 2009; Körding et al., 2007; Odegaard et al., 2016, 2015; Odegaard and Shams, 2016; Wozny et al., 2010), and a Bayesian CI model with only 4 free parameters could account for the observers' data remarkably well (e.g., accounting for 97% of variance in 250 data points) (Körding et al., 2007).

In addition to accounting well for the observers' perceived location, Bayesian CI also makes predictions about the judgment of unity (common cause). This allows Bayesian CI to also account (Körding et al., 2007) for a phenomenon that had been previously considered counter-intuitive and puzzling, namely, "negative bias" (Wallace et al., 2004), by showing how splitting data based on reports of unity can lead to bias in location responses. Since trials with perceived independent causes tend to be trials in which a larger disparity exists between the encoding of the auditory and visual inputs, limiting the analysis to these trials would lead to an apparent negative bias.

In a more recent study (Rohe and Noppeney, 2015a), observers were asked to report not only the auditory location, but also their judgment of common/independent cause. In accordance with Bayesian CI, the more reliable visual stimuli sharpened the window of common cause perception, and increased the visual bias of auditory localization when the discrepancy between the two was not large. Furthermore, in a later study it was found that the reliability of the visual stimuli was itself estimated over time, as modeled through the Bayesian CI model (Beierholm et al., 2020).

While these studies investigated spatial processing along the azimuth, a recent study examined the perception of verticality which relies on visual and vestibular information. The study performed model comparison and reported that perception of verticality by human observers was best accounted for by the Bayesian Causal Inference model (de Winkel et al., 2018). Furthermore, Bayesian CI has been used to explain the combination of visual and tactile stimuli, where stimuli were presented spatially along the arm of the participants (Verhaar et al., 2021).

Slightly different variants of Bayesian CI used by different research
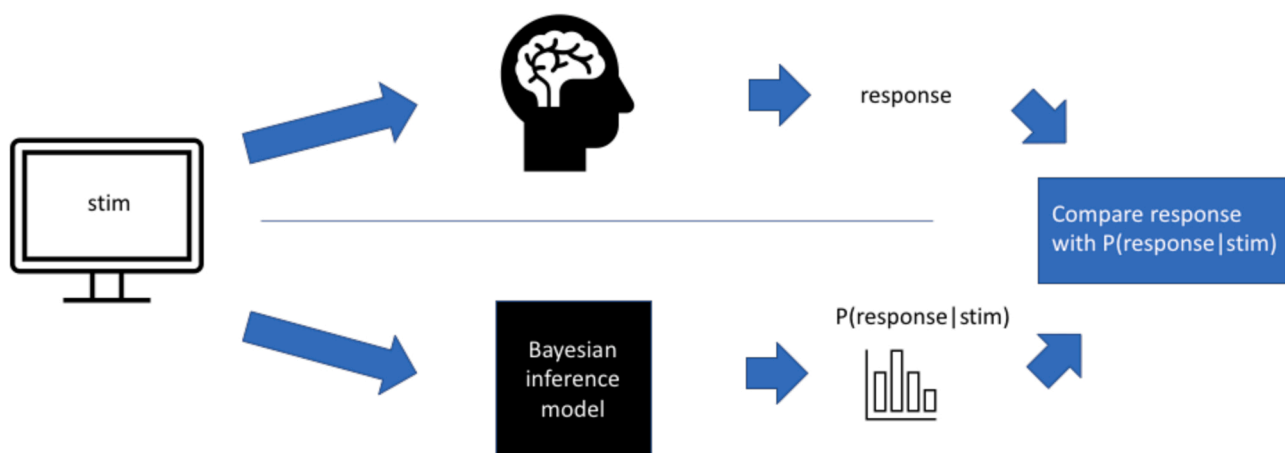
Fig. 3. The general approach used to gain insight into the computations used by the nervous system to solve a problem (to perform a given task). This approach has been successful in shedding light on the computations and goals of different brain systems.

groups or in perceptual tasks that differed from the original localization task used by (Beierholm et al., 2009; Körding et al., 2007; Odegaard et al., 2015; Wozny et al., 2010) showed that Bayesian CI could account for observer data in spatial tasks remarkably well (Hospedales and Vijayakumar, 2008; Mohl et al., 2020; Rohe and Noppeney, 2015a; Sato et al., 2007), for both humans and monkeys.

Subsequently, several studies have used Bayesian CI as a quantitative tool to explore the influence of selective attention (Odegaard et al., 2016), frame of reference (Odegaard et al., 2019), aging (Jones et al., 2019), adaptation (Wozny and Shams, 2011), and stability and generalization of auditory-visual binding (Odegaard and Shams, 2016) in human spatial processing, again confirming that Bayesian CI captures human spatial processing remarkably well.

2.2. Temporal perception

Bayesian CI applies to temporal as well as spatial processing. A simple non-hierarchical (see Box 2) variant of Bayesian CI was originally developed to account for behavioral data in a temporal numerosity judgment task (Shams et al., 2005). This is the task of reporting the number of flashes and beeps presented simultaneously to observers, in which the sound-induced flash illusion often occurs when the number of flashes and beeps are not the same (Shams et al., 2002, 2000), for example, when a single flash is accompanied with two brief beeps, leading to the percept of two flashes.

The hierarchical model (Körding et al., 2007) (Fig. 2a) was later tested also on the temporal numerosity task and shown to account for the data very well in bisensory conditions (Beierholm, 2007; Odegaard et al., 2016; Odegaard and Shams, 2016).

A related task to temporal numerosity judgment is temporal rate discrimination or categorization. A non-hierarchical variant of Bayesian CI was shown to account for observers' auditory-visual rate discrimination (Roach et al., 2006; Shams and Beierholm, 2010). In a recent study (Cao et al., 2019) using auditory-visual temporal rate categorization, the participants' responses were compared with predictions of three models, including the traditional forced fusion model, and the Bayesian CI. Bayesian CI could qualitatively and quantitatively account for observers' data and outperformed the other models.

Although not yet quantitatively studied, given that the sound-induced flash illusion has been reported to occur in other species such as rodents (Ito et al., 2019), it appears that Bayesian CI also operates in the sensory processing systems of lower mammals. Future research needs to examine this quantitatively.

Auditory-visual temporal processing has also been studied using speech stimuli and temporal asynchrony detection task, showing that in addition to the actual physical time discrepancy between the two signals, the perceptual experience of synchrony depends on the temporal acuity of the observers and the prior expectation of a common cause, and can be accounted for by a Bayesian CI model (Magnotti et al., 2013).

Temporal processing in a unisensory setting also requires an inference about the grouping of the stimuli. For example, the perception of the time intervals in an auditory sequence would depend on the inferred causal structure (whether three brief sounds belong to the same event or not). Sawai et al. (Sawai et al., 2012) demonstrated that a Bayesian CI model can account for observers' perceptions of sound sequences.

2.3. Spatio-temporal perception

The need for causal inference is not limited to the situations where information about a given sensory variable is available from multiple cues. There are numerous situations both in unisensory and multisensory settings wherein the nervous system has to consider multiple causal structures (or competitive priors) for determining and estimating the attributes of objects (including one's own body) and events.

For example, interpretation of retinal displacement requires model inference. Yuille, Bulthoff and colleagues were the first to note that the perceptual system has to solve a model inference problem (Clark and Yuille, 1990; Yuille and Bulthoff, 1996). They proposed a competitive priors model (Yuille and Clark, 1994) that accounted for the different interpretation of visual motion cue for structure depending on whether the object is rigid or non-rigid (see Fig. 2d).

More recently, studies of motion perception have shown that human perception can be accounted for by Bayesian models incorporating heavy-tailed priors (Lu et al., 2010; Stocker and Simoncelli, 2006), which have been shown to be computationally very similar to Bayesian CI (Shams and Beierholm, 2010). The perceptual system has to determine whether two local patches of motion were caused by the same object or different objects in order to determine whether or not to integrate the information across space and apply the same constraints to both patches.

2.4. Weight perception

Bayesian CI has been successful in explaining weight perception. A classic example is the size-weight illusion in which when lifting two objects that are identical in shape, mass, and apparent material, but different in size, the smaller object is perceived to be heavier than the larger object (Charpentier, 1891; Koseleff, 1957). This illusion had evaded theoretical explanation for decades and had been considered to be an "anti-Bayesian" illusion (Brayanov and Smith, 2010; Ernst, 2009). Indeed a simple non-competitive prior Bayesian model would predict the opposite of this phenomenon: the smaller object should be perceived as lighter.

Peters et al. (Peters et al., 2016) showed that a competitive prior model, equivalent to Bayesian CI, in which priors on density relationship between the two objects compete to explain the sensory data, can qualitatively and quantitatively account for the illusion. Furthermore, the model accounts for findings of other studies such as the effect of training with small heavy objects on the illusion. Importantly, the model made a novel prediction that the prior expectations of the density relationship across individuals should correlate with the degree of illusion experienced, subsequently confirmed in experiments (Peters et al., 2016).

Similarly in the material-weight illusion, when lifting two objects that have the same size, shape, and mass, the object that appears to be made of low-density material (such as styrofoam) is perceived to be heavier (Harshfield and DeHardt, 1970). Peters et al. (Peters et al., 2018) showed that the same Bayesian model with competing priors on density relationship can also explain this illusion.

2.5. Body ownership perception

Research in the field of body perception over the last two decades has revealed that even the perception of body ownership and attributes are remarkably malleable and involves continuous processing of multisensory information such as visual, proprioceptive, tactile, and vestibular inputs (Blanke, 2012; Blanke et al., 2002; Botvinick and Cohen, 1998; Ehrsson, 2007; Ehrsson et al., 2004; Hoort et al., 2011; Lenggenhager et al., 2007; Petkova and Ehrsson, 2008; Tsakiris and Haggard, 2005).

In the Rubber-hand illusion (Botvinick and Cohen, 1998), the observer experiences ownership over a fake rubber hand positioned where one's own hand typically would be when the real hand is out of the view and stroked simultaneously (tactile input) with the visible stroking of the rubber hand. This illusion also produces a recalibration of the proprioceptive perception of the real hand.

Samad et al. (Samad et al., 2015) offered the first computational account of the rubber-hand illusion through a Bayesian CI model with two competing causal structures, one in which all the sensory signals are caused by the same object, the observer's hand, and one in which the proprioceptive and tactile signals are caused by the real hand, and the visual signal is caused by the rubber hand. The outcome of causal inference computation depends on the spatial and temporal discrepancy

among the sensory measurements and the prior probability of a common cause for the sensory signals. The model provided a parsimonious explanation for the known phenomena, but also predicted that the illusion could also be experienced in the absence of tactile input (stroking) confirmed through both subjective reports and skin conductance data. The model also predicted that the Rubber-hand illusion and accompanying proprioceptive drift would not occur if the rubber hand is placed at 30 cm or more from the real hand, consistent with behavioral data (Lloyd, 2007). An analogous body ownership phenomenon, the Rubber-foot illusion, can similarly be accounted for by Bayesian CI (Schürmann et al., 2019). A conceptualization based on Bayesian CI has been proposed to account for several experimental findings regarding body ownership illusions (Kilteni et al., 2015).

Another phenomenon that is related to the perception of self is the sense of agency, in which one perceives one's own actions to have caused an observed event. A Bayesian CI model was shown to qualitatively account for empirical findings related to the sense of agency (Legaspi and Toyoizumi, 2019).

The illusions of body ownership have also been observed in monkeys, and Bayesian CI has been shown to account for behavioral responses from monkeys as well as those of human participants (Fang et al., 2019a). These findings again suggest that the Bayesian causal inference is not unique to the human brain and has a longer evolutionary history. The study of the monkeys obviously could not probe subjective experience of body ownership and had to rely on measures of proprioceptive drift. Some studies have reported imperfect correlation between this measure and subjective reports of ownership (Rohde et al., 2011) Therefore, additional research is needed to examine body ownership further with more explicit measures of ownership, as well as manipulations of sensory uncertainty.

2.6. Sensori-motor processing

At any given moment, the visual system has to perform causal inference to determine how many causes exist for any changes in the retinal image, and accordingly, estimate the movement of objects in the scene and one's own eye/body movements (e.g., so that we can navigate accurately and not bump into things). Human postural behavior has been accounted for well by a causal inference model with competing priors of self vs. environment as causes of retinal displacement (Dokka et al., 2010).

Despite the continual changes in the retinal image caused by frequent saccades, we perceive the world to be stable and generally not moving. Atsma et al. (Atsma et al., 2016) showed that as quantitatively predicted by Bayesian CI, spatial constancy depends on the degree of consistency between pre-saccadic object location memory and post-saccadic visual input, and the two sensory inputs are integrated and/or segregated based on the posterior probability that they refer to the same position in the world (common cause).

When the eyes move, the motor system sends a copy of the command sent to the muscles to the sensory system, a signal known as the "efference copy." (Jeannerod, 2003) When we are passively moved–for example, riding in a moving car–the retinal image changes and there is no efference copy of motor commands available for spatial updating. In a study of passive self motion, Perdreau et al. (Perdreau et al., 2018) found that as predicted by Bayesian CI, the perceptual system weighs the integration of the internally updated target position and of the visual feedback by the posterior probability that they correspond to a common position in the world. The data were accounted for by Bayesian CI and the account was superior to those of alternative models.

Another important perceptual task for mobile organisms is to determine the direction of heading during movement. Because visual and vestibular information may not always be originating from the same source (body movement), similarly to examples above, the nervous system has to first determine if the two sensory inputs have a common source, and if so, to integrate them optimally to best estimate the heading direction and velocity. It has recently been shown that Bayesian CI can account for human heading perception (de Winkel et al., 2017), as well as heading perception in monkeys (Acerbi et al., 2018; Dokka et al., 2019).

Recent studies also support the idea that the process of determining the source of an error in any action is governed by Bayesian causal inference. The motor system typically receives feedback from the visual system. When faced with an error (e.g. the deviation from the target in reaching), the motor system has to determine whether the error is due to the motor system or due to other sources (e.g., change in the environment/target, etc.). If the error is due to the motor system, it needs to be corrected, but not otherwise. A study by Wei and Körding (Wei and Körding, 2009) in which the visual feedback of the observed error was manipulated reported a pattern of motor learning/correction consistent with that predicted by a model similar to the Bayesian CI. Participants showed no correction if the error was large (and thus the error was not attributed to the motor system; but instead attributed to the experimenter, for example). The largest correction occurred for the largest size error that could still be attributed to the motor system (not too large relative to the variability of the motor system). Motor adaptation has also been examined in a setting that allowed change in both the environment and the motor system (e.g., due to fatigue, etc.) and shown to follow Bayesian CI (Berniker and Kording, 2008). Participants' behavior in several studies in which participants' movements were perturbed experimentally have been accounted for by Bayesian CI as well (see (Wei and Kording, 2012)).

2.7. Other perceptual and sensorimotor tasks

Visual slant perception has been accounted for by Bayesian models utilizing mixed priors (Knill, 2007, 2003) or heavy-tailed likelihoods (Girshick and Banks, 2009) that have been shown to be equivalent or very similar to Bayesian CI (Shams and Beierholm, 2010) (see Box 2).

The oddity detection task is a task that can involve stimuli in one sensory modality or in multiple sensory modalities. Hospedales et al. (Hospedales and Vijayakumar, 2009) showed that a Bayesian CI-style model can nicely account for both unisensory and multisensory oddity detection findings that had been previously posed as a challenge to forced fusion maximum likelihood models of sensory integration. The human observer data showed that the detection of oddity depended on the degree of discrepancy between the components in each stimulus, and thus whether or not they are bound together.

Bayesian CI has also been employed to account for behavioral data in tasks related to speech perception, such as judgment of asynchrony between auditory and visual speech tokens (Magnotti et al., 2013), the McGurk effect (Magnotti et al., 2020, 2018; Magnotti and Beauchamp, 2017), and the identification of phonetic categories across speakers (Kleinschmidt and Jaeger, 2015).

Research on animal navigation has indicated that many animals, including rats, hamsters, honeybees, and spiders, can exploit multiple sources of information by utilizing path integration (the ability to keep track of distance and direction of path traversed) and landmarks. The pattern of processing of the two cues qualitatively follows Bayesian causal inference, in that when the discrepancy between the two cues is small, a bigger weight is given to the landmark cue than path integration, and when the conflict between the two cues is large, the landmark cue seems to be ignored (Etienne et al., 1990; Shettleworth and Sutton, 2005). This is consistent with a process of causal inference, deciding whether the landmark cue corresponds to the target or to another location (Shams and Beierholm, 2010).

2.8. Learning, adaptation, recalibration, and attention

While psychologists and neuroscientists have studied adaptation, recalibration, perceptual learning, and selective attention for decades, an understanding of how computation and neural processing is affected

Box 2

Non-hierarchical Bayesian Causal Inference and alternative models.

Hierarchical models are not the only way to explain causal inference. Even within the Bayesian CI, if the causal structure is not explicitly of interest (e.g. if the latent variable S_a is the only variable to report) then it is possible to marginalize over the structure (integrate out C). Non-hierarchical versions do not involve a hierarchy of inference: inference about the structure followed by inference about the sensory variables. Instead of explicit representation and inference of causal structure, they assume a prior on the sensory variables that would capture the two causal scenarios by the mixture of the two priors on the sensory variables:

In the simplest case (Fig. 2a) this marginalization can be shown to be equivalent to using a prior that is a mixture of components from the two structures, which when using Normal (Gaussian) distributions reduces to

$$P(S_a, S_v) = P(C = 1)P(S = S_a = S_v) + P(C = 2)P(S_a)P(S_v) = w*N(S_a - S_v, \sigma^2) + (1 - w)*N(S_a, \sigma_a^2)N(S_v, \sigma_v^2)$$

where σ^2, σ_a^2, σ_v^2 are respectively the variances of the joint, auditory and visual priors.

Framing the prior in this way makes it easier to compare to other models, such as the similar competitive 2D priors model (Roach et al., 2006), or the heuristic coupling prior (Ernst, 2007). See Shams and Beierholm (Shams and Beierholm, 2010) for a graphical comparison.

When computing the full hierarchical Bayesian CI becomes computationally intractable (see Box 4) and even a non-hierarchical Bayesian CI model has too many parameters, using a heuristic model (e.g. using a two dimensional Gaussian prior) can be a reasonable alternative. A non-hierarchical heuristic approximation to Bayesian CI model was shown to account for trisensory conditions (Wozny et al., 2008) capturing interactions among all combinations of flashes, beeps, and taps on the finger. In a study of visual-tactile interactions in temporal numerosity judgment where the deviation between number of flashes and taps was limited to one a non-hierarchical heuristic approximation to Bayesian CI was shown to account well for the observers responses (Bresciani et al., 2006). However when it is possible to compute the hierarchical Bayesian CI, it tends to perform better than the heuristic approximations (Körding et al., 2007).

by these processes has remained elusive. Because Bayesian models, such as Bayesian CI, allow quantitative and rigorous characterization of the various components of perceptual and sensorimotor processing in each individual subject, the change in each of these components can be examined quantitatively in various settings and scenarios.

Odegaard et al. (Odegaard et al., 2016) investigated the influence of selective attention to visual or auditory stimuli in a spatial localization task and a temporal numerosity judgment task. Surprisingly, they found that selective attention only benefits the sensory modality that is already "good" (or reliable) at the task, and not the sensory modality that is weaker (or less reliable). Specifically, using Bayesian CI model fitting they found that in the spatial task, visual precision improves and in the temporal task auditory precision improves as a result of selective attention. The tendency to bind the stimuli (or perceive a common cause) did not seem to be affected by selective attention to a specific sensory modality. A more recent study (Badde et al., 2020) also used Bayesian CI in an attempt to characterize the effect of selective attention to visual and tactile modalities on processing components.

The spatial recalibration of auditory map by vision is known as the ventriloquist aftereffect, and happens after repeated exposure to simple visual and auditory stimuli that are presented to the observer at a fixed spatial discrepancy. The same outcome can be due to either a shift in auditory representations of space (likelihood functions), or the prior distribution of the stimuli (priors), or a combination of the two. Wozny et al. (Wozny and Shams, 2011) investigated ventriloquist aftereffect in human observers using Bayesian CI and showed that the observer responses were most consistent with a shift in the likelihood functions, therefore, supporting a very low-level neural representation phenomenon.

Using the same experimental paradigm, however, manipulating both the spatial and temporal discrepancy between the visual and auditory stimuli during the adaptation phase, Odegaard et al. (Odegaard et al., 2017) investigated if there are sensory exposures that can modify the priors rather than likelihoods in the spatial localization task. Surprisingly, they discovered that repeated exposure to large auditory-visual spatial discrepancy resulted in an *increase*–instead of a decrease–in the prior expectation of a common cause. Using the Bayesian CI framework, they explain why such an adaptation exposure can enhance the tendency to bind, an effect that could potentially be exploited in clinical and

educational applications. Similarly, Tong and colleagues (Tong et al., 2020) used different levels of congruency between visual and auditory stimuli to manipulate the prior expectation of a common cause.

The recalibration paradigm described above involves passive exposure to auditory-visual stimuli, no task, and no feedback during the exposure phase. In contrast, in perceptual learning paradigms, the observer actively performs a task during a training phase, and usually receives feedback on the accuracy of their responses. In a study using perceptual learning paradigm, McGovern et al. (McGovern et al., 2016) trained observers in an audio-visual simultaneity task, and examined their AV integration in a spatial localization task before and after the training session. They found two effects of training: the window of integration narrowed and there was an overall reduction in integration across the whole range of spatial discrepancies. They showed that a Bayesian CI model that included both spatial and temporal variables (an extension of Kording et al. (Körding et al., 2007) model) could quantitatively and qualitatively account for the findings well, by indicating an increase in temporal precision and a reduction in the tendency to bind subsequent to training.

Different sensory modalities encode space in different frames of reference (e.g. vision in eye-centered, audition in head-centered) and yet our perception of space is unified and coherent and independent of the modality of origin. However, several studies of spatial perception including those discussed above have suggested that the human perceptual system utilizes priors in the estimation of auditory-visual location. Therefore, this begs the question of which frame of reference the prior expectation of space is encoded in. Odegaard et al. (Odegaard et al., 2019) recently investigated this question using Bayesian CI and manipulating the direction of gaze of human observers in an auditory-visual localization task. The results of their quantitative Bayesian CI modeling suggested that the frame of reference is a combination of eye-centered and head-centered frames.

In a recent study Rohlf et al. (Rohlf et al., 2020) showed that children as young as 5 years old were able to perform multisensory integration in a spatial localization task, and their performance was best explained by the Bayesian CI. However, the children were not able to recalibrate when visual and auditory stimuli were presented with a consistent discrepancy, implying that the updating of the parameters of the Bayesian CI occurs at a different time scale from the development of the

CI itself. Older adults, who frequently exhibit lower sensory reliabilities, also appear to integrate sensory information consistent with Bayesian CI (Jones et al., 2019).

In summary, Bayesian CI has not only accounted for a vast number of perceptual and sensorimotor phenomena, but has also been successfully used as a framework to quantitatively characterize and elucidate effects of modulations due to processes such as attention, adaptation, aging, and learning. Still, important questions regarding the nature of the priors involved in Bayesian CI models, and the computational complexity of Bayesian CI in natural environments and the approximations, heuristics or other mechanisms that may be required to make this computation feasible in the human nervous system remain unanswered (see Box D).

3. Algorithmic level of analysis

The studies discussed so far have collectively provided compelling evidence that Bayesian CI accounts for the computations carried out by the perceptual and sensorimotor systems in the human brain. These studies have shown that human (and other primates) behavior in a vast range of settings and tasks is consistent with the behaviors exhibited by Bayesian causal inference, both qualitatively and quantitatively. Importantly, Bayesian CI has been shown to be not overly flexible or powerful and has passed the various tests of parsimony, specificity, model comparison, and model prediction testing (see Box 3). Therefore, all-in-all, the evidence for the Bayesian causal inference as a governing computation carried out by the human nervous system in the wide range of tasks and domains discussed above has proved to be compelling.

The approach used in these studies is to compare the human observers' behavior in a given task and for a given set of stimuli with that of a Bayesian CI model, and examine whether the responses qualitatively and/or quantitatively are consistent with each other. Therefore, this approach can be characterized as comparing two black "boxes" with each other (see Fig. 3). When providing the same input to the boxes, they produce the same output, suggesting that the two boxes carry out the same computation. While this approach is informative about the computation involved, it does not shed light on how the computation is carried out. For example, if the brain took the average of the visual and auditory location as the best guess of the location, there would nevertheless be many ways that this computation could be performed, e.g. by combining the auditory and visual estimates together, and then dividing it by two ($(Xa+Xv)/2$) or by adding half the difference between the visual and auditory signals to the auditory signal ($Xa+(Xv-Xa)/2$), etc. In other words, the same computation can be carried out in a lot of different ways, both at an algorithmic level, and at an implementation level (For example, one box may use a calculator, whereas the other box may use an abacus).

In order to shed light on the representations or algorithm used by the human nervous system in computing Bayesian CI, Beierholm et al. (Beierholm et al., 2009) asked whether priors and likelihoods are represented independently in human auditory-visual spatial processing. To approach this question, the visual stimulus contrast was manipulated to lead to two different levels of reliability/precision, and hence two different likelihood functions. The study examined whether the estimate of the priors changed as a result of change in the stimulus precision (and in turn, change in likelihoods). The results suggested that priors remained the same despite a substantial change in likelihoods. Therefore, this study provided evidence against representations and algorithms that would rely on lookup tables of posterior estimates (or subject's responses), and provided support for the idea that the nervous system achieves the optimal Bayesian causal inference by encoding likelihoods and prior distributions independently of each other and then combining them according to Bayes Rule (Fig. 4a). These findings are supported by a recent study (Tong et al., 2020) that used a converse manipulation. In this study, the authors aimed to manipulate the prior expectation of a common cause by extended exposure to either congruent or incongruent audio-visual stimuli, and reported no change in auditory and visual precisions (suggesting unchanged likelihoods),


Box 3
How meaningful is a good fit to the data?.

Some have criticized Bayesian models of cognition by questioning their flexibility due to the potential use of ad hoc priors or a large number of free parameters. It is indeed possible for a model to be excessively flexible/powerful and account for any set of data and thus, not shed light on the true underlying computations at work. This warrants examination of the nature, structure, and flexibility of the model.

1. **Bayesian CI is not an ad hoc theory that was concocted to fit the data.** Bayesian CI models are based on generative models that reflect the causal structure of events in the environment (or the body) leading to the stimuli that are observed by the nervous system. Bayesian CI is normative and represents the optimal way of solving the problems of causal inference and source estimation, problems that the nervous system has to solve in a large number of processing domains from early sensory stages all the way to cognitive and motor stages.
2. **Bayesian CI accounts for data with few free parameters.** Bayesian CI models typically have a small number of free parameters (usually 3–4) accounting for large sets of data (usually more than 500 data points). Moreover, fitting the parameters using a subset of the data (not used in the subsequent parameter-free model prediction testing) (Shams et al., 2005; Wozny et al., 2008) or showing that the model fails to account for scrambled data (Shams et al., 2005) provide additional evidence that the model is not overly flexible, and its account is selective to the observer data.
3. **Bayesian CI models can account for patterns of behavior even with no free parameters.** Bayesian CI makes qualitative predictions about the pattern of behavior in a range of stimulus conditions that have been repeatedly confirmed by empirical data in a variety of tasks, populations, and processing domains. This includes complete fusion and a large bias by the more reliable/precise cue when the disparity is small, complete segregation of the cues when the discrepancy is large, or 'partial integration' (assuming model averaging) when the discrepancy is moderate. No other model had been able to account for this pattern of behavior.
4. **Bayesian CI models outperform alternative models.** In some studies (e.g., (de Winkel et al., 2018; Girshick and Banks, 2009; Körding et al., 2007)) Bayesian CI was compared with other models of comparable complexity (same or larger number of free parameters) and was shown to provide a better account for the data.
5. **Bayesian CI's predictions–including some counter-intuitive predictions– have been empirically confirmed** (Peters et al., 2018; Samad et al., 2015). This is generally considered the ultimate test of a theory, and so far Bayesian CI has successfully passed this test.

These facts and findings collectively provide compelling support for Bayesian CI as computation governing the perceptual and sensorimotor processing as discussed in Section 2.
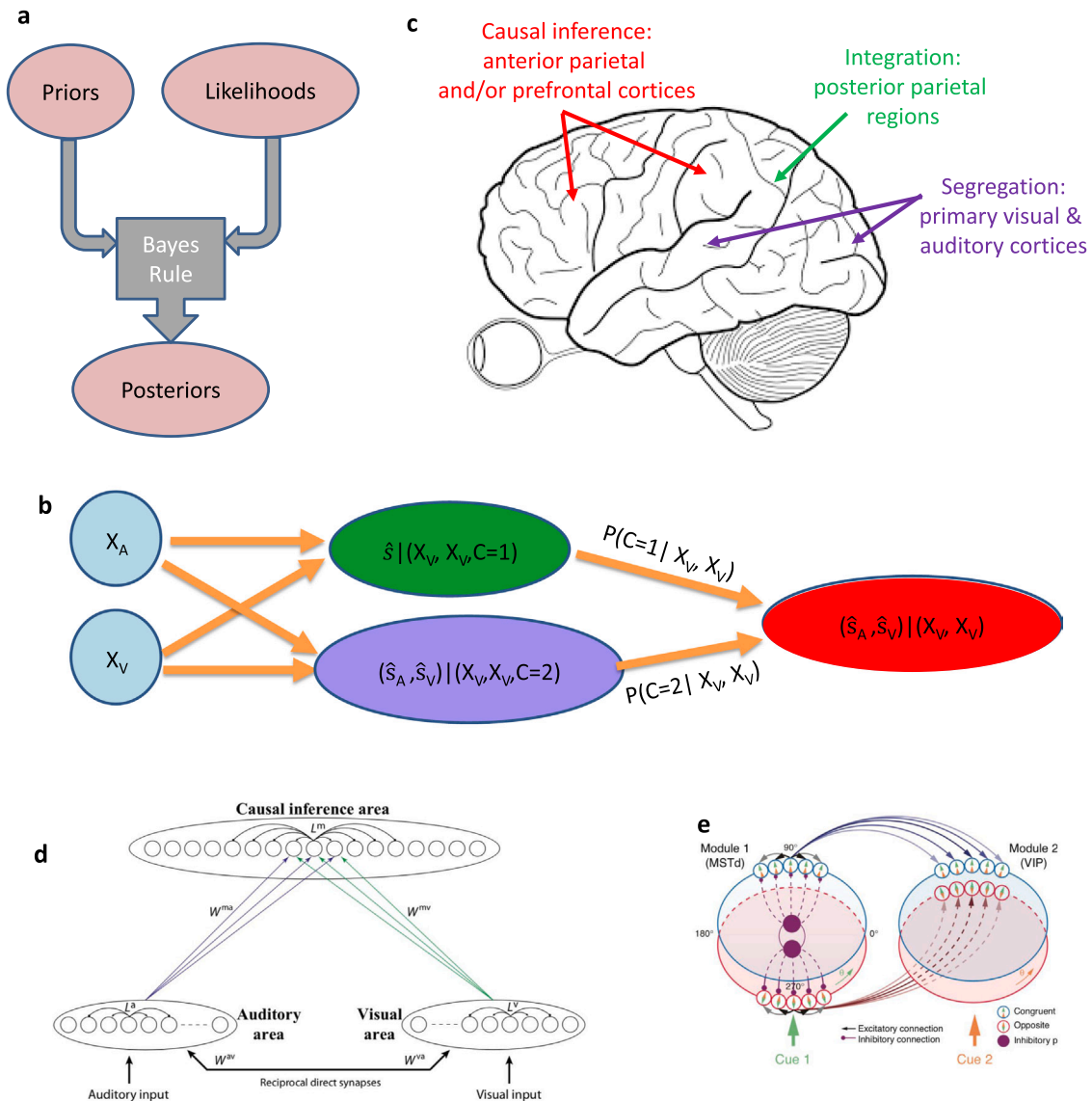
Fig. 4. a) One algorithm for carrying out Bayesian inference in the brain. b) An algorithm for carrying out Bayesian CI computations. c) Possible neural architecture of Bayesian CI in sensory tasks in human brain (Cao et al., 2019; Rohe et al., 2019; Rohe and Noppeney, 2015b). d) A proposed neural circuitry for Bayesian CI in auditory-visual spatial tasks, reproduced from (Cuppini et al., 2017). e) A proposed neural circuitry for Bayesian CI in a vestibular visual heading task, reproduced from (Zhang et al., 2019a).

providing additional support for the independence of priors and likelihoods (Fig. 4a).

More recently, multiple studies using a variety of neuroimaging techniques have probed the representations and processing of Bayesian CI in human observers. Despite the significant variety in tasks used, and the methods of study, all of these studies have suggested that the human brain carries out Bayesian CI in a sequential fashion (Fig. 4b), by first representing the unisensory estimates (e.g., X_A and X_V), followed by computing and representing the reliability-weighted fusion of the unisensory estimates, followed by estimates of causal structure, followed by the combination of the fusion and segregation estimates according to the probability of their respective causal structure to produce the Bayesian CI estimates of the variable of interest (time, space, etc.) (Aller and Noppeney, 2019; Cao et al., 2019; Rohe et al., 2019; Rohe and Noppeney, 2015b). Therefore, these studies confirmed what Beierholm et al.'s (Beierholm et al., 2009, p. 200) study had suggested, which is Bayesian CI is not only a good computational model of multisensory perception, but also a good process model in that it captures the representations and operations used by the nervous system to achieve the final outcome

(Maloney and Mamassian, 2009).

While recent studies have shed light on the representations and algorithms involved in Bayesian CI, additional research is needed to shed light on questions about the nature of the information that is retained and manipulated by the nervous system in the process of inference (e.g., distributions vs. point estimates) as discussed in more detail in Box 5.

4. Implementational level of analysis

The studies discussed above shed light on the representations and processes involved in Bayesian CI. The next important question is how this algorithm is carried out by the machinery of the brain, networks of neurons.

Some of the studies discussed above, in addition to shedding light on the representations and algorithm used by the human nervous system, have also examined the specific brain areas involved in each of these processes. These studies have used physiological measurements of brain activity together with the predictions of Bayesian CI to gain insight into brain areas involved in encoding the different components of Bayesian

CI computation.

For example, using fMRI and a spatial localization task, Rohe & Noppeney (Rohe and Noppeney, 2015b) investigated the role of different brain areas along the human visual pathway in representing various relevant variables and computations of Bayesian causal inference. They found a hierarchical and sequential evolution of computation (as described above in Section 3) with representation of unisensory auditory and visual estimates of location in primary visual and auditory cortical areas, reliability weighted integration of the two signals in the anterior parietal region, and the final combination of location estimate from the two causal scenarios in the posterior parietal region.

A similar approach (Rohe et al., 2019), however using the temporal numerosity judgment task (counting number of flashes and beeps) and EEG measurements and analyzing the temporal dynamics, showed the same style of processing. Sensory cortical areas appeared to produce unisensory estimates of the number of flashes and beeps, followed by higher-up representation of multisensory estimates of numerosity under the common cause vs. independent causes assumption, followed by the final Bayesian CI estimate. Another study using EEG to investigate temporal dynamics of Bayesian CI, but utilizing a spatial localization task, reported the same hierarchical processing of information as found in the temporal numerosity task (Aller and Noppeney, 2019).

Furthermore, a study probing the neural mechanisms of Bayesian CI by investigating the temporal dynamics using MEG employed a somewhat different temporal task (temporal rate categorization) and yet reported the same pattern of hierarchical evolution of the computation from sensory cortical areas advancing to higher areas of processing including parietal regions and prefrontal regions of the brain (Cao et al., 2019). While this study also showed parietal regions to be involved in processing multisensory estimates, the arbitration between the common cause vs. independent cause hypotheses, and the combination of the two estimates appeared to take place in the prefrontal cortex.

A recent study sought to illuminate the implementation of Bayesian CI at the level of individual neurons and populations of neurons within a brain area in the context of body ownership (Fang et al., 2019a). The activity of neurons in monkey premotor cortex was recorded in unisensory and proprioceptive-visual conditions in a behavioral paradigm analogous to the Rubber-hand illusion. The study found neuronal activities (both at individual neuron level and population level) correlated with various components of Bayesian CI, namely segregation, integration of the two modalities, and posterior probability of a common cause.

These studies provide compelling evidence that the human and primate nervous systems implement Bayesian CI for a variety of tasks across a variety of sensori-motor pathways. However, how exactly these brain areas and neurons accomplish these computations remains unclear.

Previous work had shown how theoretically neurons could perform cue integration through Probabilistic Population Coding (Beck et al., 2008; Ma et al., 2006) (PPC), while segregation relies only on potential parameter transformation. It has also been proposed that the dynamic properties of single neurons, single synapses, and sensory receptive fields in effect carry out the non-linear computations involved in causal inference (Lochmann and Deneve, 2011). The challenging question that had not been probed until recently was how the posterior over the common cause would be computed by networks of neurons, and how to use this to non-linearly combine the integrated and segregated estimates. In the last several years, suggestions have been made to address these questions.

Ma & Rahmati (Ma and Rahmati, 2013) used the vocabulary provided by PPC to build a 'neural circuit' that would be able to perform the exact computation required for estimating the posterior probability of the common cause. While they were able to devise such a circuit, they acknowledged that the biological plausibility of this method was questionable due to the complexity of the circuit.

Yamashita and colleagues (Yamashita et al., 2013) developed a network, consisting of a single layer with lateral connections, that

performed an implicit calculation of the probability of a single cause and was able to replicate the observed behavior. However, as a consequence of just using a single layer, access to unisensory information would be lost.

In contrast, Cuppini, Ursino and collaborators created a neural network based on two parallel layers representing two unisensory modalities as well as a crossmodal layer receiving inputs from the unisensory layers (Cuppini et al., 2017; Shi and Griffiths, 2009; Yu et al., 2016). After unsupervised learning of the stimulus statistics presented to the network, the connectivity within, and across layers, produced behavior very similar to the non-linear cue combination of causal inference, with cues being pulled together when close together, but unaffected when further apart. A three layer network designed to perform predictive coding proposed by Spratling (Spratling, 2016) also produces the same behavioral phenomena, while Tong et al. (Tong et al., 2018) expanded this approach to also allow recalibration across trials.

Neurophysiological studies of heading direction in monkeys have reported neurons that appear to encode full integration of the sensations and neurons that appear to be tuned to opposite directions (known as "opposite neurons"). Inspired by these findings Zhang et al. (Zhang et al., 2019b, 2019a) proposed a model consisting of integration neurons and opposite neurons that can estimate the probability of a common cause, and account for behavioral data on heading direction. A limitation to this model is however that it only applies to circular variables (e.g. direction) as it relies on the properties of neurons with opposite direction tuning.

Using an alternative approach Yu and colleagues (Yu et al., 2016) used importance sampling as a proposed method of neural computations, an approximation that has previously been shown to be biologically feasible (Shi and Griffiths, 2009). By combining importance sampling with principles from PPC they were able to recreate the posterior probability of common cause in the multisensory Bayesian CI model. An intriguing possibility explored in the paper is to what degree this method can be used to generalize to multiple stimuli (>2), a computationally difficult problem (see Box 1). In a later paper (Fang et al., 2019b) these ideas were extended to allow the same neural circuit to perform both the causal inference estimation, as well as the integration of cues.

These models have used simple biological mechanisms and connectivity patterns (e.g., long lateral inhibition) that are known to exist in sensory cortices of primate brains, to either explicitly or implicitly encode the probabilities required for the algorithms in Bayesian CI. Overall, while there are still uncertainties about how the brain implements a causal inference strategy, there are several proposals of ways it could happen. Future experimental work, in collaboration with more modeling, is needed to narrow down these hypotheses.

In summary, it appears that Bayesian CI is implemented in a distributed fashion across processing domains. At least in some domains that have been examined empirically, the implementation appears to be hierarchical and spanning brain areas across cortical lobes (e.g., occipital, parietal, frontal). Future research can shed light on what kind of circuitry is involved in implementation of Bayesian CI, and whether the same type of circuitry (e.g., network connectivity) and mechanisms (e.g., long-range inhibition, short-range excitation, or opposite neurons, etc.) are involved in the implementation of Bayesian CI across tasks and processing domains. The advent of new experimental techniques, such as cellular imaging and optogenetic manipulations, also raise enticing possibilities for answering these types of questions in rodents. These methods may allow probing the neural implementation of Bayesian CI at the level of single neurons and neuronal circuits, and the causal relationships between brain activity and computational mechanisms as opposed to a correlational relationship between the two.

Future research can also investigate whether these mechanisms are hardwired or learned by experience.

5. Discussion

5.1. The emergence of Bayesian CI as a neuroscience theory

Up until early 2000s, the predominant model of multisensory processing and cue combination was maximum likelihood estimation that implicitly assumed the sensory signals stem from the same source, and therefore, should completely get fused in the estimation of the source. Research on the Bayesian CI model started by making the observation that sensory signals don't always get fused (Shams et al., 2002; Wallace et al., 2004); sometimes they get fused (full integration leading to illusions if there is a discrepancy), sometimes they are segregated, and sometimes they interact partially (Shams et al., 2005). It was shown that a normative Bayesian model that allowed both integration and segregation could account for these observations (Körding et al., 2007; Shams et al., 2005) and also explain counter-intuitive phenomena (Körding et al., 2007) such as negative bias (Wallace et al., 2004). Later, the same theory was shown to account for other multisensory phenomena, while making novel and unexpected predictions in different tasks which in turn were also empirically tested and confirmed (Peters et al., 2018; Samad et al., 2015).

The Bayesian CI theory involves competition amongst hypotheses, aka, priors. For this reason, this theory can also be referred to as a "competitive priors" theory. The Bayesian CI model that was proposed in the context of multisensory perception is mathematically equivalent to the competitive prior model that had been proposed earlier in the context of visual perceptual tasks (Knill, 2007; Yuille and Clark, 1994). As discussed in Section 2, Bayesian CI models have successfully accounted for human (and other mammalians) behavior in a number of different tasks ranging in processing domain (spatial, temporal, spatio-temporal, speech, etc.), ranging in combination of sensory modalities (unisensory visual, unisensory auditory, multisensory with various combinations of sensory modalities), perception of the world as well as perception of self (body perception and ownership), sensori-motor, and motor processing. It appears that evolution discovered this powerful computational mechanism and employed it in a variety of processes in the nervous system.

### 5.2. The validity of Bayesian CI as a neuroscience theory

A strong neuroscience theory should be able to account for existing data, and make predictions that can be tested and verified. It should be simple, yet able to explain a diversity of phenomena (Gorini, 2003). And lastly, it should be verifiable at all three levels of analysis as proposed by Marr[3].

Bayesian CI is simple and normative (see Box 3) and has accounted for a wide variety of phenomena (see Section 2). It is a parsimonious, unifying theory of perceptual and sensorimotor processing, governing the realms of multisensory perception, unisensory perception, body ownership perception, as well as sensorimotor processing (Fig. 5).

The research summarized in Section 2 is the computational level of analysis that was carried out in a distributed fashion by several research groups around the globe (but see Box 4). Once the multisensory Bayesian CI model was established as a successful computational model, research groups started exploring the algorithmic/representational and implementation aspects of it by investigating neuronal architectures and circuitries that can carry out the computation in the nervous system (Gorini, 2003) (see Box 5).

As summarized in Sections 3 and 4, a variety of recent studies employing sophisticated neuroscientific methods, for example, EEG, fMRI, MEG, and single-electrode recordings, and using machine learning methods of analysis, have reported a hierarchical architecture in the human (and primate) cortex that appears to implement Bayesian CI in multiple processing domains. Moreover, several studies by different research groups have proposed specific and biologically plausible neural circuits that could carry out Bayesian CI in the context of spatial or temporal processing tasks. Thus, Bayesian CI has been tested on all three of Marr's levels, emerging as a strong neuroscience theory.

We argue that this hierarchical and systematic approach to understanding brain function that is based on a sequence of observation of behavior, developing theoretical ideas, testing predictions, exploring generality, and investigating algorithms and potential neural implementation provides a template for a successful approach to understanding the brain function. It also demonstrates the effectiveness of the scientific method (Gorini, 2003; Toomer, 1964) and scientific
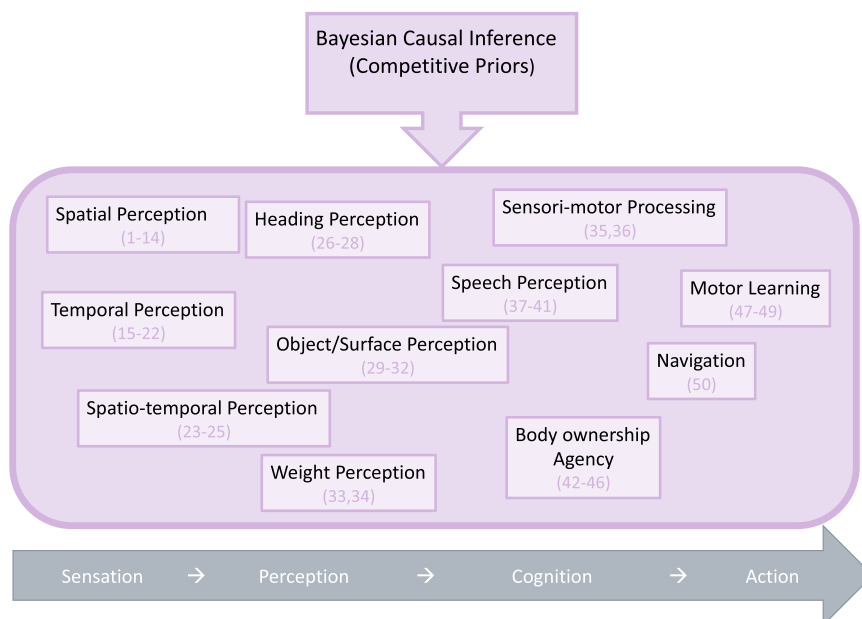
Fig. 5. Bayesian Causal Inference is a unifying theory. Each box represents a domain of brain processing and the numbers in each box are references to some of the studies that have shown Bayesian CI accounts for a task in that domain as listed below. Bayesian CI accounts for empirical findings in a large number of studies probing diverse tasks ranging from sensation to action, and a vast number of phenomena that appear completely unrelated, or appear counter-intuitive. References: 1) Körding et al., 2007. 2) Beierholm et al., 2009. 3) Wozny et al., 2010. 4) Mohl et al., 2020. 5) Odegaard et al., 2019. 6) Odegaard & Shams, 2016. 7) Odegaard et al., 2016. 8) Odegaard et al., 2015. 9) Rohe & Noppeney, 2015a. 10) Beierholm et al. 2020. 11) de Winkel et al., 2018. 12) Sato et al., 2007. 13) Hospedales & Vijayakumar, 2008. 14) Jones et al., 2019. 15) Odegaard & Shams, 2016. 16) Odegaard et al., 2016. 17) Shams et al., 2005. 18) Beierholm 2007. 19) Roach et al., 2006. 20) Cao et al., 2019. 21) Magnotti et al., 2013. 22) Sawai et al., 2012. 23) Yuille & Clark, 1994. 24) Stocker & Simoncelli, 2006. 25) Lu et al., 2010. 26) de Winkel et al., 2017. 27) Acerbi et al., 2018. 28) Dokka et al., 2010. 29) Knill, 2007. 30) Knill, 2003. 31) Girshick & Banks, 2009. 32) Hospedales & Vijayakumar, 2009. 33) Peters et al., 2016. 34) Peters et al., 2018. 35) Atsma et al. 2016. 36) Perdreau et al., 2018. 37) Magnotti & Beauchamp, 2017. 38) Magnotti et al. 2020. 39) Magnotti et al., 2013. 40) Magnotti et al., 2018. 41) Kleinschmidt & Jaeger, 2015. 42) Samad et al., 2015. 43) Schumann et al., 2019. 44) Kilteni et al. 2015. 45) Legaspi & Toyoizumi, 2019. 46) Fang et al., 2019. 47) Wei & Körding, 2012. 48) Wei & Körding, 2009. 49) Berniker & Körding, 2008. 50) Shams & Beierholm, 2010.

Box 4

Limitations and outstanding questions: *Computational level of analysis*.

1. **How are Bayesian CI computations carried out in natural environments?** The nervous system typically is faced with a large number of sensory (or sensorimotor) measurements at any given moment. These situations would require the nervous system to choose from a large number of causal structures (Gershman and Niv, 2010), and hence the problem of combinatorial explosion. As the number of parameters increases faster than exponential, approximations are needed when multiple cues are present (Wozny et al., 2008) or more clever generative models are also possible (Yates et al., 2017) that allow for the number of parameters to grow with the stimulus size using non-parametric Bayesian inference. It is also possible that instead of computing exact probabilities, the nervous system can overcome the computational intractability by instead computing approximations of the probabilities (Sanborn and Chater, 2016). Are there strategies, constraints, heuristics, or approximations utilized by the nervous system to make the probabilistic inference involved in Bayesian CI more computationally efficient and tractable?

2. **Is there a difference in the way high-level priors vs. low-level priors influence the computation?** Our current knowledge of the priors involved in Bayesian CI is limited. It is clear that biases can influence the inference in the various tasks explored at multiple levels of processing. For example, in a spatial localization task, the prior expectation that the visual and auditory signals have a common cause plays an important role in the perception of location, however, this prior bias can be induced by high-level cognitive factors such as experimenter's instructions, as well as low-level biases that are due to the statistics of the environment and likely encoded at early sensory stages of processing.

3. **What is the nature of the priors?** While some studies have started investigating the properties of Bayesian priors (Beierholm et al., 2009; Odegaard et al., 2019; Odegaard and Shams, 2016), there are still some basic questions about the nature, plasticity, and other characteristics of priors that are not explored and understood. For example, in a multisensory setting, are there different priors for unisensory vs. multisensory conditions? In a similar vein, in situations where sensory information is available from multiple senses, does Bayesian CI in each sensory modality (e.g., the various visual cues for depth, or object identification) take place prior to inference across the senses (e.g., auditory and visual depth cues, or cues to object identity), or do they occur in parallel?

4. **What determines the loss function?** Normative models of the brain function assume that evolution or experience has resulted in brain mechanisms that minimize a certain type of loss or cost (e.g., a certain kind of error). This is known as the loss function. There is some evidence that at least in some basic spatial and temporal tasks the loss function is not uniform across individuals, and may not be uniform even within an individual across time (Odegaard and Shams, 2016; Wozny et al., 2010). What are the factors that determine the loss function for any given task, individual, situation?

Box 5

Limitations and outstanding questions: Representational and implementation levels of analysis.

1. **To what extent is the inferred causal structure accessible to consciousness?** While the probability of each causal structure is computed in Bayesian CI and contributes to the estimation of the variables of interest (e.g., spatial location, time, speech, etc.), it is not clear whether this inference is accessible to consciousness, and whether the nervous system commits to a given causal structure at any stage of processing. If we do not explicitly ask participants to report their perceived causal structure (e.g., common cause or independent causes), would the nervous system bother with an explicit/conscious encoding of these probabilities?

2. **Does the nervous system encode probability distributions or some summary statistics of the distributions?** For example, it is possible that the system only encodes the mean (or max) and/or variance of the probability distributions. For a complicated probability distribution it becomes impossible for the brain to explicitly encode it. In such cases does the brain rely on other approximations, such as sampling (Sanborn and Chater, 2016)?

3. **Is Bayesian CI hardwired or learned by the human brain?** It is possible that it is hardwired for some basic tasks and learned for others. If it is learned, what are the mechanisms for learning, at what age do they emerge, and how fast are they learned?

4. **What are the neural mechanisms of Bayesian CI in the human brain?** Very little is known about the implementation of Bayesian inference in the brain in general. It is not clear whether and how probability distributions are represented in the nervous system, how prior knowledge is combined with likelihood functions (or sensory input). The studies using fMRI, EEG, MEG and single electrode recordings discussed in Sections 3 and 4 have shed light on the architecture and brain areas where some of the Bayesian CI computations may take place, however how the computations are carried out at the level of circuitry and network of neurons is still unclear. The neural network models discussed in Section 4 provide a proof of concept that such implementation is possible, however which mechanism is employed by the biological neural networks remains a topic of future research.

Interestingly, gaining insight into the neural mechanisms of Bayesian CI may shed light on the computational strategies of Bayesian CI for realistic complex situations with several sensory inputs, and the problem of combinatorial explosion discussed above. It is likely that evolution has found a trick to make this computationally intractable task feasible, and investigating the neural implementations of Bayesian CI may reveal simple approximations or heuristics that make this problem computationally tractable.

collaboration, and the power of multi-disciplinary approaches to brain research.

Acknowledgements

Declarations of Interest

None.

References

Acerbi, L., Dokka, K., Angelaki, D.E., Ma, W.J., 2018. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. PLoS Comput. Biol. 14, e1006110 https://doi.org/10.1371/journal.pcbi.1006110.

Alais, D., Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. Curr. Biol. 14, 257–262. https://doi.org/10.1016/j.cub.2004.01.029.

Aller, M., Noppeney, U., 2019. To integrate or not to integrate: temporal dynamics of hierarchical Bayesian causal inference. PLoS Biol. 17, e3000210 https://doi.org/10.1371/journal.pbio.3000210.

Atsma, J., Maij, F., Koppen, M., Irwin, D.E., Medendorp, W.P., 2016. Causal inference for spatial constancy across saccades. PLoS Comput. Biol. 12, e1004766 https://doi.org/10.1371/journal.pcbi.1004766.

Badde, S., Navarro, K.T., Landy, M.S., 2020. Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. Cognition 197, 104170. https://doi.org/10.1016/j.cognition.2019.104170.

Bayes, T., 1763. Essay towards solving a problem in the doctrine of chances. Philos. Trans. R. Soc. Lond. 53 https://doi.org/10.1098/rstl.1763.0053.

Beck, J.M., Ma, W.J., Kiani, R., Hanks, T., Churchland, A.K., Roitman, J., Shadlen, M.N., Latham, P.E., Pouget, A., 2008. Probabilistic population codes for Bayesian decision making. Neuron 60, 1142–1152. https://doi.org/10.1016/j.neuron.2008.09.021.

Beierholm, U., Quartz, S., Shams, L., 2009. Bayesian priors are encoded independently from likelihoods in human multisensory perception. J. Vis. 9, 1–9. https://doi.org/10.1167/9.5.23.

Beierholm, U., Rohe, T., Ferrari, A., Stegle, O., Noppeney, U., 2020. Using the past to estimate sensory uncertainty. eLife 9, e54172. https://doi.org/10.7554/eLife.54172.

Beierholm, U.R., 2007. Bayesian modeling of sensory cue combinations (phd). California Institute of Technology. https://doi.org/10.7907/K89B-XW75.

Berniker, M., Kording, K., 2008. Estimating the sources of motor errors for adaptation and generalization. Nat. Neurosci. 11, 1454–1461. https://doi.org/10.1038/nn.2229.

Bertelson, P., Vroomen, J., De Gelder, B., Driver, J., 2000. The ventriloquist effect does not depend on the direction of deliberate visual attention. Percept. Psychophys. 62, 321–332. https://doi.org/10.3758/BF03205552.

Blanke, O., 2012. Multisensory brain mechanisms of bodily self-consciousness. Nat. Rev. Neurosci. 13, 556–571. https://doi.org/10.1038/nrn3292.

Blanke, O., Ortigue, S., Landis, T., Seeck, M., 2002. Stimulating illusory own-body perceptions. Nature 419, 269–270. https://doi.org/10.1038/419269a.

Botvinick, M., Cohen, J., 1998. Rubber hands 'feel' touch that eyes see, 756–756 Nature 391. https://doi.org/10.1038/35784.

Brayanov, J.B., Smith, M.A., 2010. Bayesian and "Anti-Bayesian" biases in sensory integration for action and perception in the size–weight illusion. J. Neurophysiol. 103, 1518–1531. https://doi.org/10.1152/jn.00814.2009.

Bresciani, J.-P.P., Dammeier, F., Ernst, M.O., 2006. Vision and touch are automatically integrated for the perception of sequences of events. J. Vis. 6, 554–564. https://doi.org/10.1167/6.5.2.

Cao, Y., Summerfield, C., Park, H., Giordano, B.L., Kayser, C., 2019. Causal inference in the multisensory brain. Neuron 102, 1076–1087.e8. https://doi.org/10.1016/j.neuron.2019.03.043.

Charpentier, A., 1891. Analyse experimentale: De quelques elements de la sensation de poids [Experimental study of some aspects of weight perception]. Arch. Physiol. Norm. Pathol. 3, 122–135.

Choe, C.S., Welch, R.B., Gilford, R.M., Juola, J.F., 1975. The "ventriloquist effect": visual dominance or response bias. Percept. Psychophys. 18, 55–60.

Clark, J.J., Yuille, A.L., 1990. Data Fusion for Sensory Information Processing Systems. Kluwer Academic Pulishers, Boston.

Colas, F., Diard, J., Bessière, P., 2010. Common Bayesian models for common cognitive issues. Acta Biotheor. 58, 191–216. https://doi.org/10.1007/s10441-010-9101-1.

Cuppini, C., Shams, L., Magosso, E., Ursino, M., 2017. A biologically inspired neurocomputational model for audiovisual integration and causal inference. Eur. J. Neurosci. 46, 2481–2498. https://doi.org/10.1111/ejn.13725.

Dokka, K., Kenyon, R.V., Keshner, E.A., Kording, K.P., 2010. Self versus environment motion in postural control. PLoS Comput. Biol. 6, e1000680 https://doi.org/10.1371/journal.pcbi.1000680.

Dokka, K., Park, H., Jansen, M., DeAngelis, G.C., Angelaki, D.E., 2019. Causal inference accounts for heading perception in the presence of object motion. Proc. Natl. Acad. Sci. USA 116, 9060–9065. https://doi.org/10.1073/pnas.1820373116.

Ehrsson, H.H., 2007. The experimental induction of out-of-body experiences. Science 317, 1048. https://doi.org/10.1126/science.1142175.

Ehrsson, H.H., Spence, C., Passingham, R.E., 2004. That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. Science 305, 875–877. https://doi.org/10.1126/science.1097011.

Ernst, M.O., 2007. Learning to integrate arbitrary signals from vision and touch. J. Vis. 7, 1–14. https://doi.org/10.1167/7.5.7.Introduction.

Ernst, M.O., 2009. Perceptual learning: inverting the size–weight illusion. Curr. Biol. 19, R23–R25. https://doi.org/10.1016/j.cub.2008.10.039.

Ernst, M.O., Banks, M.S., 2002. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415, 429–433. https://doi.org/10.1038/415429a.

Etienne, A.S., Teroni, E., Hurni, C., Portenier, V., 1990. The effect of a single light cue on homing behaviour of the golden hamster. Anim. Behav. 39, 17–41. https://doi.org/10.1016/S0003-3472(05)80723-7.

Fang, W., Li, J., Qi, G., Li, S., Sigman, M., Wang, L., 2019. Statistical inference of body representation in the macaque brain. Proc. Natl. Acad. Sci. USA 116, 20151–20157. https://doi.org/10.1073/pnas.1902334116.

Fang, Y., Yu, Z., Liu, J.K., Chen, F., 2019. A unified neural circuit of causal inference and multisensory integration. Neurocomputing 358, 355–368. https://doi.org/10.1016/j.neucom.2019.05.067.

Gershman, S.J., Niv, Y., 2010. Learning latent structure: carving nature at its joints. Curr. Opin. Neurobiol. 20, 251–256. https://doi.org/10.1016/j.conb.2010.02.008.

Girshick, A.R., Banks, M.S., 2009. Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts, 8–8 J. Vis. 9. https://doi.org/10.1167/9.9.8.

Gorini, R., 2003. Al-Haytham the man of experience: first steps in the science of vision. J. Int. Soc. Hist. Islam. Med. 2, 53–55.

Harshfield, S.P., DeHardt, D.C., 1970. Weight judgment as a function of apparent density of objects. Psychon. Sci. 20, 365–366. https://doi.org/10.3758/BF03335692.

Hoort, B., van der, Guterstam, A., Ehrsson, H.H., 2011. Being Barbie: the size of one's own body determines the perceived size of the world. PLoS One 6, e20195. https://doi.org/10.1371/journal.pone.0020195.

Hospedales, T., Vijayakumar, S., 2009. Multisensory oddity detection as Bayesian inference. PLoS One 4, e4205. https://doi.org/10.1371/journal.pone.0004205.

Hospedales, T.M., Vijayakumar, S., 2008. Structure inference for Bayesian multisensory scene understanding. IEEE Trans. Pattern Anal. Mach. Intell. 30, 2140–2157. https://doi.org/10.1109/TPAMI.2008.25.

Ito, Y., Sato, R., Tamai, Y., Hiryu, S., Uekita, T., Kobayasi, K.I., 2019. Auditory-induced visual illusions in rodents measured by spontaneous behavioural response. Sci. Rep. 9, 19211. https://doi.org/10.1038/s41598-019-55664-z.

Jack, C.E., Thurlow, W.R., 1973. Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. Percept. Mot. Skill 37, 967–979.

Jeannerod, M., 2003. Action monitoring and forward control of movements. In: The Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge, MA, pp. 83–85.

Jones, S.A., Beierholm, U., Meijer, D., Noppeney, U., 2019. Older adults sacrifice response speed to preserve multisensory integration performance. Neurobiol. Aging 84, 148–157. https://doi.org/10.1016/j.neurobiolaging.2019.08.017.

Kilteni, K., Maselli, A., Kording, K.P., Slater, M., 2015. Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. Front. Hum. Neurosci. 9. https://doi.org/10.3389/fnhum.2015.00141.

Kleinschmidt, D.F., Jaeger, T.F., 2015. Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. Psychol. Rev. 122, 148–203. https://doi.org/10.1037/a0038695.

Knill, D.C., 2007. Robust cue integration: a Bayesian model and evidence from cue-conict studies with stereoscopic and gure cues to slant. J. Vis. 7, 1–24. https://doi.org/10.1167/7.7.5.Introduction.

Knill, D.C.C., 2003. Mixture models and the probabilistic structure of depth cues. Vis. Res. 43, 831–854. https://doi.org/10.1016/S0042-6989(03)00003-8.

Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B., Shams, L., 2007. Causal inference in multisensory perception. PLoS One 2, e943. https://doi.org/10.1371/journal.pone.0000943.

Koseleff, P., 1957. Studies in the perception of heaviness. I.1.2: some relevant facts concerning the size-weight-effect (SWE). Acta Psychol. 13, 242–252. https://doi.org/10.1016/0001-6918(57)90023-9.

Legaspi, R., Toyoizumi, T., 2019. A Bayesian psychophysics model of sense of agency. Nat. Commun. 10, 4250. https://doi.org/10.1038/s41467-019-12170-0.

Lenggenhager, B., Tadi, T., Metzinger, T., Blanke, O., 2007. Video ergo sum: manipulating bodily self-consciousness. Science 317, 1096–1099. https://doi.org/10.1126/science.1143439.

Lloyd, D.M., 2007. Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand. Brain Cogn. 64, 104–109. https://doi.org/10.1016/j.bandc.2006.09.013.

Lochmann, T., Deneve, S., 2011. Neural processing as causal inference. Curr. Opin. Neurobiol. 21, 774–781. https://doi.org/10.1016/j.conb.2011.05.018.

Lu, H., Lin, T., Lee, A., Vese, L., Yuille, A., 2010. Recovering the functional form of the slow-and-smooth prior in global motion perception, 819–819 J. Vis. 10. https://doi.org/10.1167/10.7.819.

Ma, W.J., Rahmati, M., 2013. Towards a neural implementation of causal inference in cue combination. Multisens. Res. 26, 159–176. https://doi.org/10.1163/22134808-00002407.

Ma, W.J., Beck, J.M., Latham, P.E., Pouget, A., 2006. Bayesian inference with probabilistic population codes. Nat. Neurosci. 9, 1432–1438. https://doi.org/10.1038/nn1790.

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, e1005229 https://doi.org/10.1371/journal.pcbi.1005229.

Magnotti, J.F., Ma, W.J., Beauchamp, M.S., 2013. Causal inference of asynchronous audiovisual speech. Front. Psychol. 4 https://doi.org/10.3389/fpsyg.2013.00798.

Magnotti, J.F., Smith, K.B., Salinas, M., Mays, J., Zhu, L.L., Beauchamp, M.S., 2018. A causal inference explanation for enhancement of multisensory integration by co-articulation. Sci. Rep. 8, 18032. https://doi.org/10.1038/s41598-018-36772-8.

Magnotti, J.F., Dzeda, K.B., Wegner-Clemens, K., Rennig, J., Beauchamp, M.S., 2020. Weak observer–level correlation and strong stimulus-level correlation between the McGurk effect and audiovisual speech-in-noise: a causal inference explanation. Cortex 133, 371–383. https://doi.org/10.1016/j.cortex.2020.10.002.

Maloney, L.T., Mamassian, P., 2009. Bayesian decision theory as a model of human visual perception: testing Bayesian transfer. Vis. Neurosci. 26, 147–155. https://doi.org/10.1017/S0952523808080905.

Marr, D., 1982. Vision. W. H. Freeman, San Francisco, CA.

McGovern, D.P., Roudaia, E., Newell, F.N., Roach, N.W., 2016. Perceptual learning shapes multisensory causal inference via two distinct mechanisms. Sci. Rep. 6, 24673. https://doi.org/10.1038/srep24673.

Mohl, J.T., Pearson, J.M., Groh, J.M., 2020. Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli. J. Neurophysiol. 124, 715–727. https://doi.org/10.1152/jn.00046.2020.

Odegaard, B., Shams, L., 2016. The brain's tendency to bind audiovisual signals is stable but not general. Psychol. Sci. 27, 583–591. https://doi.org/10.1177/0956797616628860.

Odegaard, B., Wozny, D.R., Shams, L., 2015. Biases in visual, auditory, and audiovisual perception of space. PLoS Comput. Biol. 11, e1004649 https://doi.org/10.1371/journal.pcbi.1004649.

Odegaard, B., Wozny, D.R., Shams, L., 2016. The effects of selective and divided attention on sensory precision and integration. Neurosci. Lett. 614, 24–28. https://doi.org/10.1016/j.neulet.2015.12.039.

Odegaard, B., Wozny, D.R., Shams, L., 2017. A simple and efficient method to enhance audiovisual binding tendencies. PeerJ 5, e3143. https://doi.org/10.7717/peerj.3143.

Odegaard, B., Beierholm, U.R., Carpenter, J., Shams, L., 2019. Prior expectation of objects in space is dependent on the direction of gaze. Cognition 182. https://doi.org/10.1016/j.cognition.2018.10.011.

Perdreau, F., Cooke, J.R.H., Koppen, M., Medendorp, W.P., 2018. Causal inference for spatial constancy across whole body motion. J. Neurophysiol. 121, 269–284. https://doi.org/10.1152/jn.00473.2018.

Peters, M.A.K., Ma, W.J., Shams, L., 2016. The size-weight illusion is not anti-Bayesian after all: a unifying Bayesian account. PeerJ 4, e2124. https://doi.org/10.7717/peerj.2124.

Peters, M.A.K., Zhang, L.-Q., Shams, L., 2018. The material-weight illusion is a Bayes-optimal percept under competing density priors. PeerJ 6, e5760. https://doi.org/10.7717/peerj.5760.

Petkova, V.I., Ehrsson, H.H., 2008. If i were you: perceptual illusion of body swapping. PLoS One 3, e3832. https://doi.org/10.1371/journal.pone.0003832.

Recanzone, G.H., 2003. Auditory influences on visual temporal rate perception. J. Neurophysiol. 89, 1078–1093.

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. R. Soc. B Biol. Sci. 273, 2159–2168. https://doi.org/10.1098/rspb.2006.3578.

Rohde, M., Luca, M.D., Ernst, M.O., 2011. The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. PLoS One 6, e21659. https://doi.org/10.1371/journal.pone.0021659.

Rohe, T., Noppeney, U., 2015a. Sensory reliability shapes perceptual inference via two mechanisms. J. Vis. 15, 22. https://doi.org/10.1167/15.5.22.

Rohe, T., Noppeney, U., 2015b. Cortical hierarchies perform Bayesian causal inference in multisensory perception. PLoS Biol. 13, e1002073 https://doi.org/10.1371/journal.pbio.1002073.

Rohe, T., Ehlis, A.-C., Noppeney, U., 2019. The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. Nat. Commun. 10, 1907. https://doi.org/10.1038/s41467-019-09664-2.

Rohlf, S., Li, L., Bruns, P., Röder, B., 2020. Multisensory integration develops prior to crossmodal recalibration. Curr. Biol. 30, 1726–1732.e7. https://doi.org/10.1016/j.cub.2020.02.048.

Samad, M., Chung, A.J., Shams, L., 2015. Perception of body ownership is driven by Bayesian sensory inference. PLoS One 10, e0117178. https://doi.org/10.1371/journal.pone.0117178.

Sanborn, A.N., Chater, N., 2016. Bayesian brains without probabilities. Trends Cogn. Sci. 20, 883–893. https://doi.org/10.1016/j.tics.2016.10.003.

Sato, Y., Toyoizumi, T., Aihara, K., 2007. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. Neural Comput. 19, 3335–3355. https://doi.org/10.1162/neco.2007.19.12.3335.

Sawai, K., Sato, Y., Aihara, K., 2012. Auditory time-interval perception as causal inference on sound sources. Front. Psychol. 3 https://doi.org/10.3389/fpsyg.2012.00524.

Schürmann, T., Vogt, J., Christ, O., Beckerle, P., 2019. The Bayesian causal inference model benefits from an informed prior to predict proprioceptive drift in the rubber foot illusion. Cogn. Process. 20, 447–457. https://doi.org/10.1007/s10339-019-00928-9.

Shams, L., Beierholm, U.R., 2010. Causal inference in perception. Trends Cogn. Sci. 14, 1–8. https://doi.org/10.1016/j.tics.2010.07.001.

Shams, L., Kamitani, Y., Shimojo, S., 2000. What you see is what you hear, 788–788 Nature 408. https://doi.org/10.1038/35048669.

Shams, L., Kamitani, Y., Shimojo, S., 2002. Visual illusion induced by sound. Cogn. Brain Res. 14, 147–152. https://doi.org/10.1016/S0926-6410(02)00069-1.

Shams, L., Ma, W.J.W.J., Beierholm, U., 2005. Sound-induced flash illusion as an optimal percept. Neuroreport 16, 1923–1927.

Shettleworth, S.J., Sutton, J.E., 2005. Multiple systems for spatial learning: dead reckoning and beacon homing in rats. J. Exp. Psychol. Anim. Behav. Process. 31, 125–141. https://doi.org/10.1037/0097-7403.31.2.125.

Shi, L., Griffiths, T., 2009. Neural implementation of hierarchical Bayesian inference by importance sampling. Adv. Neural Inf. Process. Syst. 22, 1669–1677.

Slutsky, D.A., Recanzone, G.H., 2001. Temporal and spatial dependency of the ventriloquism effect. Neuroreport 12, 7–10.

Spratling, M.W., 2016. A neural implementation of Bayesian inference based on predictive coding. Connect. Sci. 28, 346–383. https://doi.org/10.1080/09540091.2016.1243655.

Stocker, A.A., Simoncelli, E.P., 2006. Noise characteristics and prior expectations in human visual speed perception. Nat. Neurosci. 9, 578–585. https://doi.org/10.1038/nn1669.

Thurlow, W.R., Jack, C.E., 1973. Certain determinants of the ventriloquism effect. Percept. Mot. Skill 36, 1171–1184.

Tong, J., Parisi, G.I., Wermter, S., Röder, B., 2018. Closing the loop on multisensory interactions: a neural architecture for multisensory causal inference and recalibration. ArXiv180206591 Cs Q-Bio.

Tong, J., Li, L., Bruns, P., Röder, B., 2020. Crossmodal associations modulate multisensory spatial integration. Atten. Percept. Psychophys. 82, 3490–3506. https://doi.org/10.3758/s13414-020-02083-2.

Toomer, G.J., 1964. Review of Ibn al-Haythams Weg zur Physik. Isis J. Hist. Sci. Soc. 55, 463–465.

Tsakiris, M., Haggard, P., 2005. The rubber hand illusion revisited: visuotactile integration and self-attribution. J. Exp. Psychol. Hum. Percept. Perform. 31, 80–91. https://doi.org/10.1037/0096-1523.31.1.80.

Verhaar, E., Medendorp, W.P., Hunnius, S., Stapel, J.C., 2021. Bayesian causal inference in visuotactile integration in children and adults. Dev. Sci. e13184 https://doi.org/10.1111/desc.13184.

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 https://doi.org/10.1007/s00221-004-1899-9.

Warren, D.H., Welch, R.B., McCarthy, T.J., 1981. The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. Percept. Psychophys. 30, 557–564.

Wei, K., Kording, K., 2012. Causal inference in sensorimotor learning and control. Sens. Cue Integr. https://doi.org/10.1093/acprof:oso/9780195387247.003.0002.

Wei, K., Körding, K., 2009. Relevance of error: what drives motor adaptation? J. Neurophysiol. 101, 655–664. https://doi.org/10.1152/jn.90545.2008.

de Winkel, K.N., Katliar, M., Bülthoff, H.H., 2017. Causal inference in multisensory heading estimation. PLoS One 12, e0169676. https://doi.org/10.1371/journal.pone.0169676.

de Winkel, K.N., Katliar, M., Diers, D., Bülthoff, H.H., 2018. Causal inference in the perception of verticality. Sci. Rep. 8, 5483. https://doi.org/10.1038/s41598-018-23838-w.

Wozny, D.R., Shams, L., 2011. Computational characterization of visually induced auditory spatial adaptation. Front. Integr. Neurosci. 5 https://doi.org/10.3389/fnint.2011.00075.

Wozny, D.R., Beierholm, U.R., Shams, L., 2008. Human trimodal perception follows optimal statistical inference. J. Vis. 8, 24. https://doi.org/10.1167/8.3.24. Introduction.

Wozny, D.R., Beierholm, U.R., Shams, L., 2010. Probability matching as a computational strategy used in perception. PLoS Comput. Biol. 6, e1000871 https://doi.org/10.1371/journal.pcbi.1000871.

Yamashita, I., Katahira, K., Igarashi, Y., Okanoya, K., Okada, M., 2013. Recurrent network for multisensory integration-identification of common sources of audiovisual stimuli. Front. Comput. Neurosci. 7 https://doi.org/10.3389/fncom.2013.00101.

Yates, T., Larigaldie, N., Beierholm, U., 2017. A non-parametric Bayesian prior for causal inference of auditory streaming. In: Proceedings of the 39th Annual Conference of the Cognitive Science Society. Presented at the Annual meeting of the Cognitive Science Society, pp. 1381–1386.

Yu, Z., Chen, F., Dong, J., Dai, Q., 2016. Sampling-based causal inference in cue combination and its neural implementation. Neurocomputing 175, 155–165. https://doi.org/10.1016/j.neucom.2015.10.045.

Yuille, A.L., Bulthoff, H.H., 1996. Bayesian decision theory and psychophysics. In: Knill, D.C., Richards, W. (Eds.), Perception as Bayesian Inference. Cambridge University Press.

Yuille, A.L., Clark, J.J., 1994. Bayesian models, deformable templates and competitive priors. In: Proceedings of the 1991 York Conference on Spatial Vision in Humans and Robots. Cambridge University Press, USA, pp. 333–347.

Zhang, W.-H., Wang, H., Chen, A., Gu, Y., Lee, T.S., Wong, K.M., Wu, S., 2019a. Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation. eLife 8, e43753. https://doi.org/10.7554/eLife.43753.

Zhang, W.-H., Wu, S., Doiron, B., Lee, T.S., 2019b. A normative theory for causal inference and Bayes factor computation in neural circuits. Adv. Neural Inf. Process. Syst. 32, 3804–3813.