Excitation/Inhibition imbalance impairs multisensory causal inference by increasing the proneness to experience the sound-induced flash illusion in the schizophrenia spectrum

Renato Paredes[1,2], Francesca Ferri[3], Vincenzo Romei[4,5] and Peggy Seriès[6,*]

[1]Department of Psychology, Pontifical Catholic University of Peru, Lima, Peru

[2]Instituto de Investigaciones Psicológicas, Facultad de Psicología, Universidad Nacional de Córdoba – CONICET, Córdoba, Argentina

[3]Department of Neuroscience, Imaging and Clinical Sciences, University of Chieti-Pescara, Chieti, Italy

[4]Consciousness Group, Centre for Studies and Research in Cognitive Neuroscience, Department of Psychology "Renzo Canestrari", University of Bologna, Bologna, Italy

[5]Facultad de Lenguas y Educación, Universidad Antonio de Nebrija, Madrid, 28015, Spain

[6]Institute for Adaptive and Neural Computation, University of Edinburgh, Edinburgh, United Kingdom

*Corresponding E-mail: pseries@inf.ed.ac.uk

Abstract

Background and Hypothesis

The spectrum of schizophrenia is characterised by an altered sense of self with known impairments in tactile sensitivity, proprioception, body-self boundaries, and self-recognition. These are thought to be produced by failures in multisensory integration mechanisms, commonly observed as enlarged temporal binding windows during audiovisual illusion tasks. To our knowledge, there is an absence of computational explanations for multisensory integration deficits in patients with schizophrenia and individuals with high schizotypy, particularly at the neurobiological level.

Study Design

We implemented a multisensory causal inference network to reproduce the responses of individuals who scored low in schizotypy in a simulated double flash illusion task. Next, we explored the effects of Excitation/Inhibition imbalance, feedback weights, and synaptic density on the visual illusory responses of the network.

Study Results

Using quantitative fitting to empirical data, we found that an increase in recurrent excitation or cross-modal connectivity in the network enlarges the temporal binding window and increases the overall proneness to experience the illusion, matching the responses of individuals scoring high in schizotypy. Moreover, we found that an increase in the E/I balance by either neural mechanism increases the probability of inferring a common cause from the stimuli.

Conclusions

We propose an E/I imbalance account of reduced temporal discrimination in the SCZ spectrum and discuss possible links with Bayesian theories of schizophrenia. We highlight the importance of adopting a multisensory causal inference perspective to address body-related symptomatology of schizophrenia.

Keywords: Computational Psychiatry, Neural Network, Multisensory Integration, Schizotypy, Double Flash Illusion

1 Introduction

The spectrum of schizophrenia is characterised by a disturbed sense of self. Individuals within this spectrum are known to have impairments in primary somatosensory functions such as tactile sensitivity [1–3] and proprioception [4, 5]. Moreover, there is consistent evidence of higher-order somatosensory impairments observed in body ownership [6–9], bodily-self boundary [10–13] and self-recognition [14, 15] processes within the spectrum.

These impairments are thought to be produced by failures in multisensory integration mechanisms. Multisensory integration is defined as the neural process by which unisensory signals (e.g. visual, tactile, auditory) are combined to form a new feature [16, 17]. Our interaction with the external world is highly dependent on the integration of vestibular, proprioceptive, and interoceptive information of the body with exteroceptive (sensory) information about external stimuli. Hence, multisensory integration is crucial for the origin of a coherent sense of self and consciousness [18, 19].

Impairments in multisensory integration in the schizophrenia spectrum are also observed with non-body-related stimuli, as audiovisual illusion experiments suggest. Patients with schizophrenia are more prone to experience the streaming-bouncing illusion[20], the McGurk effect [21] and the double flash illusion [22]. These impairments have also been observed in high schizotypal individuals for the McGurk [23] and the double flash illusion [24]. Despite the amount of empirical evidence available, there is as yet no theoretical understanding of why multisensory integration is impaired in the schizophrenia spectrum.

These generalised failures in multisensory integration mechanisms could be explained by the enlarged temporal binding window (i.e., the epoch of time in which stimuli from different modalities are more likely to be perceptually bound) (TBW) observed in the schizophrenia spectrum in several paradigms [13, 25–27]. In general, experiments that evaluate the judgment of the relative timing of stimuli with different onset asynchronies reveal that patients with schizophrenia and individuals with high schizotypy show a longer TBW compared to controls [13, 26–28].

For example, in the double flash illusion (DFI) paradigm[1], schizophrenia patients reported the perception of an illusory flash more often than control participants, together with a wider window of temporal integration of auditory and visual stimuli [22]. Similarly, high-schizotypy individuals showed an enlarged temporal window and greater susceptibility to the illusion than those with low schizotypy, with both measures positively correlated [24].

[1]Participants are presented with one visual flash accompanied by two brief auditory stimuli and are asked to report whether they perceived one or two visual flashes.

In general, audiovisual illusions have been consistently accounted for in healthy individuals with causal inference models of multisensory integration [29–33]. These models mainly use the Bayesian inference framework [34], which provides a high-level description (i.e. computational level of analysis according to Marr [35]) of the computations carried out by the brain to integrate unisensory signals. Models of multisensory causal inference using neural networks that provide a more biologically plausible description of this mechanism have only recently been proposed [36–38]. To our knowledge, none of these approaches have been used to explore the neural mechanisms behind the proneness to experience audiovisual illusions in individuals in the schizophrenia spectrum.

In sum, the neural mechanism underlying multisensory integration deficits in the schizophrenia spectrum is unknown. There has not been a mechanistic description of how such an impairment is implemented at the neural level. In particular, a comprehensive explanation of why patients with schizophrenia and individuals with schizotypal traits are more prone to experience audiovisual illusions is lacking. Similarly, there is no clear mechanistic understanding (for a current debate, see [39, 40]) of how an enlarged TBW [13, 25–27] is related to failures in multisensory integration/segregation and causal inference in the schizophrenia spectrum. Moreover, it is unclear how impairments in these neural processes are related to bodily self-aberrations [7, 8, 11, 12, 15] and symptoms of schizophrenia and schizotypal traits.

To this end, we build on recent advances in multisensory causal inference network models in healthy individuals [36] to reproduce the greater proneness to double flash illusion observed in individuals with high schizotypy. Our model is based on the idea that inferring a common cause to audiovisual stimuli increases the occurrence of DFI [41]. In particular, we evaluate the hypothesis that an imbalance of E/I in sensory networks increases temporal binding windows [25], which in turn alters causal inference.

2 Methods

2.1 Experimental task

The computational model simulates the double flash illusion (DFI) experiment carried out in individuals with high schizotypy [24] (see Figure 1 for an illustration). The experiment consisted of presenting one visual flash (lasting for 12 ms) to the participants accompanied by two tones (lasting for 7 ms each). The first tone was always displayed simultaneously with the flash, whereas the second tone could be presented at 15 different stimulus onset asynchronies (SOA) from the first tone (ranging from 36 to 204 ms). The experiment consisted of 150 trials, and the second tone was presented 10 times for each SOA.

In each trial, participants were asked to verbally report whether they perceived one or two visual flashes. They are asked to focus only on visual stimuli and ignore tones (see [24] for an extensive description). The manipulation of the delay of the second auditory stimulus produces the subjective experience that the shorter the delay, the more likely a second illusory visual flash is perceived [42–44]. In the original study, the overall proneness to the illusion (OPI) and the temporal window of illusion (TWI) were compared in two groups separated by their scores on the Schizotypal Personality Questionnaire (SPQ) [24].

2.2 Multisensory Causal Inference network model of DFI

We model the DFI by extending the audiovisual network model proposed by [45]. Following [36], the main modification is the incorporation of a multisensory causal inference layer with feedforward and feedback connectivity to unisensory layers. This multisensory layer computes causal inference: infers whether the stimuli (flashes and beeps) originate from the same source (i.e. they have a common cause) or from different sources (i.e. they have independent causes). Importantly, this structural change allows us to model the DFI as part of a causal inference process described at the neural level.

In short, the model describes two reciprocally connected unisensory areas (auditory and visual) connected with a multisensory area that computes causal inference (see Figure 1 for an illustration). Each of the three areas consists of 30 rate neurons that encode 30°

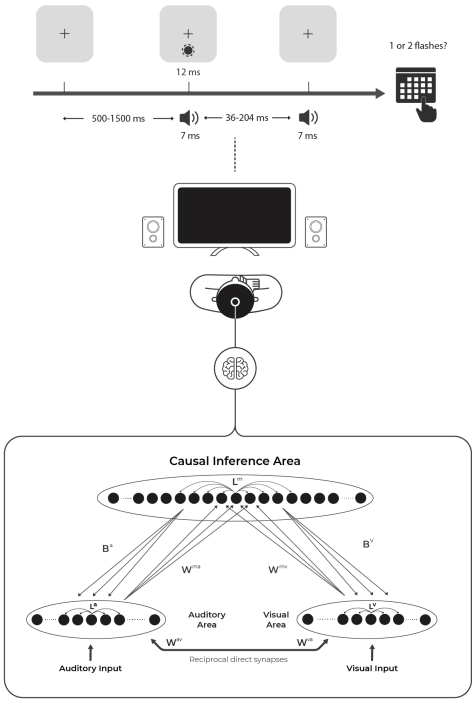

FIGURE 1: Double flash illusion experimental paradigm and Multisensory Causal Inference network model. In this experimental setup, two speakers and a monitor are placed in front of the participant. The black arrow indicates the evolution of the presentation of stimuli in time: one visual flash (lasting 12 ms), two tones (lasting 7 ms each), and an interval (randomly ranging from 500 to 1500 ms). The first tone was always displayed simultaneously with the flash, whereas the second tone could be presented at 15 different stimulus onset asynchronies (SOA) from the first tone (ranging from 36 to 204 ms). In each trial, participants were asked to verbally report whether they perceived one or two visual flashes. The shorter the SOA, the more likely a second illusory visual flash is perceived. The network model is composed of two unisensory areas (auditory and visual) connected with a multisensory causal inference area. These areas are arranged to encode a portion of 30° of the external space.

of external space (see Figure 1 for an illustration).

The neurons within each area are connected through lateral synapses following a Mexican hat pattern of excitation and inhibition (L^a, L^v, L^m). In addition, neurons in the auditory and visual areas are connected to neurons of the other unisensory area through cross-modal synapses W^{av}, W^{va}.

The neurons in the multisensory area are connected to the neurons in the auditory and visual areas that encode the proximal portions of the space, both through feedforward (W^{mv}, W^{ma}) and feedback synapses (B^{vm}, B^{am}). All the inter-areal synapses are symmetrical and followed a Gaussian pattern (see the Supplementary Material for an extensive mathematical description of the model).

Following [46], the network informs two levels of multisensory processing. In the unisensory areas, the model computes the spatial position of the external stimuli (i.e., implicit causal inference). In addition, in the multisensory area the model computes causal inference (i.e. explicit causal inference), mimicking responses of neurons located in parietal-temporal association cortices (e.g., right precuneus and anterior intraperietal sulcus). Furthermore, the network incorporates feedback connectivity that allows iterative computing of causal inference across the network [33].

An extensive description of the model and values of the parameters can be found in the Supplementary Material. The model was implemented in Python using the Scikit-NeuroMSI framework for modelling multisensory integration [47].

2.3 Experiment simulation in the network

To simulate audiovisual stimuli in the DFI, the network receives Gaussian 1-D inputs centred in the middle of the encoded space. Mimicking the experimental paradigm, the inputs last 12 ms in the visual modality and 7 ms in the auditory modality.

During each trial, the network runs for 600 ms. The visual and first auditory stimuli are always presented 16 ms after the onset of the trial. In contrast, the onset of the second auditory stimulus is presented at 15 different delays from the onset of the first stimuli, ranging from 36 to 204 ms.

We model DFI as a multisensory integration process with early and late cross-modal

interactions in distinct cortical hierarchies, as recent reviews suggest [41, 48]. The literature informs that an early interaction arises from cross-modal input between primary sensory cortices observed between 35-90 ms after the onset of the stimuli. These have been observed in Heschl's gyrus and Calcarine cortex [49, 50], and functionally supported by TMS and EEG evidence [51–53]. Next, a delayed interaction arises from the input of feed-forward-feedback connectivity with higher cortical regions observed in the 130-600 ms post-stimulus range [54–56]. These have been observed in multisensory areas such as the angular gyrus and the superior temporal sulcus, where sensory information is integrated with input from the other modality [48]. Critically, brain activity consistent with Bayesian Causal Inference[2] has been observed in anterior parietal areas in the 200-400 ms range [33].

To approximate these temporal dynamics in the network model, we consider a temporal delay of 16 ms for cross-modal interactions (as in [45]) and 95 ms for feedforward and feedback interactions. The justification for establishing the extent of the temporal delays was to model the feedforward and feedback inputs as being at least five times slower than the cross-modal inputs, as indicated by the aforementioned evidence. Moreover, following [45], we included temporal filters for auditory, visual, and multisensory neurons to mimic the temporal evolution of neural input and synaptic dynamics. Such filters consider specific time constants that define the temporal dynamics of each group of neurons (auditory, visual, or multisensory).

The responses of the participants in the DFI experimental paradigm [24] can be simulated by a principled readout of the activity of the network. Visual neurons' peak activity values above 0.15 (where activity is normalised from 0 to 1) were read out as the probability of perceiving a flash. Consistent with the methodology delineated in [36], this threshold was established to indicate that a neuron must exhibit activity up to 15% of its maximal firing rate to achieve a distinguishable perception. The probability of perceiving two flashes within a single trial was defined as the outcome of a combined probability operation: the product of the probability of detecting an initial flash and the probability of

[2]The combination of unisensory and multisensory estimates weighted by the posterior probabilities of common and independent cause models.

detecting a subsequent flash, contingent upon the condition that these probability values exceed the established detection threshold and the valley between them is at least .15 in depth[3]. All noise sources were eliminated for this approach.

To reproduce the sigmoid fit of the average responses of the low schizotypy group [24], we fitted the parameters that define the cross-modal, feedforward and feedback synaptic weights, as well as the auditory, visual, and multisensory time constants (see Supplementary Materials for a full description of the fitting procedure). The purpose of this manipulation was to calibrate the model to reproduce responses that closely resemble those observed in healthy individuals in the empirical study. Hence, in the following this network setup will be named the low schizotypy (L-SPQ) model and will be taken as the baseline for further exploration.

An illustration of the dynamics of the baseline model is shown in Figure 2. For the network readout, we selected the neuron with the maximal response to the stimuli delivered in each layer (winner take all mechanism). We observe that at every SOA, the visual neuron shows two peaks of activity: the first generated by auditory and visual stimuli, and the second as a result of feedback neural input. This feedback input depends on the degree of activation of the multisensory causal inference neuron, which in turn depends on the combined feedforward input from auditory and visual neurons. The dynamics of the model show that the closer the auditory stimuli (lower SOA), the higher the combined feedforward signal, as a consequence of cross-modal neural input.

Given these model dynamics in which the multisensory neuron tends to reach its maximal firing rate at every SOA, the multisensory neurons' activity above .8 was read out as the probability of inferring a cause from unisensory input. The probability of inferring a single cause of the stimuli within a single trial was calculated as the complement of the probability of inferring two different causes (i.e. the product of two peak values of multisensory neurons, contingent upon the condition that these are above the detection threshold and the valley between them is at least .80 in depth).

[3]Considering that participants could only select whether they perceived one or two flashes, when more than two visual peak values above threshold where identified, the average product of all possible combinations of these peak values is computed.

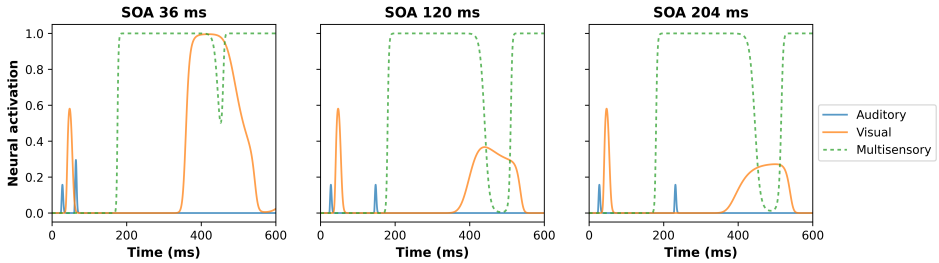

FIGURE 2: Base model (L-SPQ) dynamics. The figure shows the activity of auditory (blue line), visual (orange line) and multisensory (dashed green line) neurons at 15 different stimulus onset asynchronies (SOA). At each SOA, the visual neuron shows two peaks of activity: the first generated by the visual stimulus that is synchronous with the auditory stimulus, and the second as a result of neural feedback. In turn, the strength of such feedback input is modulated by the activation of the multisensory neuron, which depends on the combined feedforward input from auditory and visual neurons. As a consequence of cross-modal neural signals, the closer the auditory stimuli, the higher the combined feedforward signal.

2.4 Modelling the influence of high schizotypy in the network

We evaluated the consequences of three consistently observed impairments in the SCZ spectrum on the sound-induced illusion: excitation/inhibition (E/I) imbalance [57, 58], failures in top-down signalling [59], synaptic density decrease [60].

2.4.1 Modulation of the E/I balance

The modulation of the E/I balance was implemented by increasing or decreasing the strength of the recurrent excitatory connectivity. This was done by modifying the value of the parameter L^c_{0ex} in each area, defined in equation 1:

$$L^c_{jk} = \begin{cases} L^c_{0ex} \cdot exp\left(-\frac{(D_{jk})^2}{2\cdot(\sigma^c_{ex})^2}\right) - L^c_{0in} \cdot exp\left(-\frac{\left(D^c_{jk}\right)^2}{2\cdot\left(\sigma^c_{in}\right)^2}\right), & D_{jk} \neq 0 \\ 0, & D_{jk} = 0 \end{cases} \quad (1)$$

$$c = a, v, m$$

L^c_{jk} denotes the weight of the synapse from the pre-synaptic neuron at position *k* to post-synaptic neuron at position *j* [4]. D_{jk} indicate the distance between the presynaptic neuron and the postsynaptic neurons. The excitatory Gaussian function is defined by the parameters L^c_{0ex} and σ^c_{ex}, whereas the inhibitory function is defined by L^c_{0in} and σ^c_{in}. A null term (i.e. zero) was included in equation 1 to avoid auto-excitation.

Alternatively, we modelled the E/I balance modulation by increasing or decreasing the weights of cross-modal connectivity. This was implemented by manipulating the value of the parameter W^{cd}_0 in both sensory modalities, defined in the equation 2:

$$W^{cd}_{jk} = W^{cd}_0 \cdot exp\left(-\frac{(D_{jk})^2}{2\,(\sigma^{cd})^2}\right), cd = av\ or\ va \tag{2}$$

W^{cd}_0 denotes the highest level of synaptic efficacy and D_{jk} is the distance between the neuron at position j in the postsynaptic unisensory region and the neuron at position k in the presynaptic unisensory region. σ^{cd} defines the width of the cross-modal synapses. These synaptic weights are symmetrically defined ($W^{av}_0 = W^{va}_0$ and $\sigma^{av} = \sigma^{va}$) by the Gaussian function described above.

2.4.2 Failures in top-down signalling

The failures in top-down signalling were implemented by uniformly weakening top-down synaptic weights. This was achieved by changing the value of the parameter B^c_0, defined in equation 3:

$$B^c_{jk} = B^c_0 \cdot exp\left(-\frac{\left(D^s_{jk}\right)^2}{2\cdot(\sigma^c)^2}\right) \tag{3}$$

Here, D^c_{jk} is the distance between the multisensory neuron at position j and the unisensory neuron at position k. B^c_0 denote the value of the feedback when D^c_{jk} is equal to zero, representing the highest synaptic efficacy. σ^c represents the width of the feedback synapses.

[4] Here, *i* and *j* denote the position of the neuron in the area *c*, where *c* can be either *a* (auditory), *v* (visual) or *m* (multisensory). Refer to the Supplementary Material for a detailed description.

2.4.3 Decrease of synaptic density

Following [61–63], the decrease in synaptic density between multisensory and unisensory neurons was implemented by resetting the connection weights of the feedforward synapses that were below a certain threshold $\rho_{W^{mc}}$ to zero. Similarly, the connection weights of the cross-modal synapses with weights below $\rho_{W^{cd}}$ were reset to zero, to decrease the density of the cross-modal synapses between unisensory areas.

3 Results

3.1 Parameter exploration

We systematically varied the parameters governing recurrent excitation (L_{0ex}), the strength of feedback (B_0^c) and cross-modal (W_0^{cd}) weights, and feedforward ($\rho_{W^{mc}}$) and cross-modal ($\rho_{W^{cd}}$) pruning to explore plausible mechanisms behind the changes in the temporal window of illusion (TWI) observed in individuals with H-SPQ (see Figure 3A). Our simulations revealed that either an increase in recurrent excitation or cross-modal weights enlarges the temporal window of illusion. Opposite effects were found for changes in feedback weights and the pruning of feedforward or cross-modal synapses.

Similarly, we found that an increase in lateral excitation, cross-modal or feedback weights leads to an increase in the overall proneness to experience the illusion (OPI) (i.e. the percentage of reports of perceiving two flashes across all SOAs) at the level observed in individuals with H-SPQ (see Figure 3B). The opposite effect was observed when manipulating feedback weights and pruning of feedforward or cross-modal synapses.

Furthermore, we varied these same parameters to explore their impact on the causal inference mechanism of the model. Our simulations revealed that an increase in lateral excitation or cross-modal weights increases the average probability of inferring a single common cause to the stimuli across all SOAs. No clear effect on causal inference was found for changes in feedback weights or pruning.

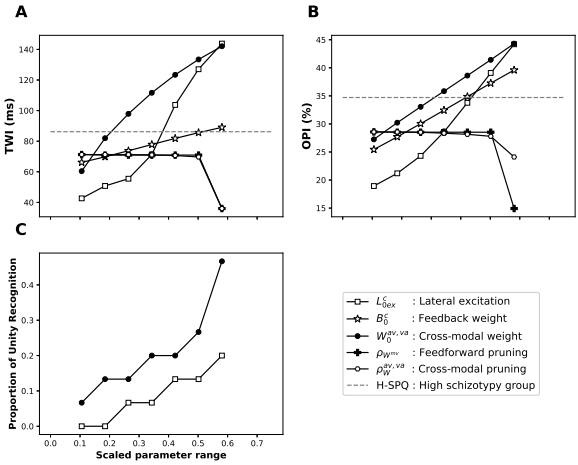

FIGURE 3: Effects of potential neural impairments in the temporal window of illusion, overall proneness to the illusion and causal inference in the network model. **Panels A**, **B** and **C** show the effect of systematic variation of parameters in the range at which they produce sensible responses to stimuli (these ranges were scaled to facilitate visual comparison). The dashed lines represent the average responses of H-SPQ individuals in the experimental study [24]. **Panel A** shows that an increase in the temporal window of illusion observed in H-SPQ could be reproduced by an increase in lateral excitation (L^c_{0ex}) or cross-modal weights (W^{cd}_0). **Panel B** indicates that the overall proneness to experience the double flash illusion observed in H-SPQ could be reproduced by an increase in lateral excitation (L^c_{0ex}), cross-modal (W^{cd}_0) or feedback weights (B^c_0). **Panel C** indicates that both an increase in lateral excitation (L^c_{0ex}) or cross-modal (W^{cd}_0) weights increases the overall probability of inferring a common cause to the presented stimuli.

3.2 Identification of the H-SPQ model

To select the model that provides the best fit to the data of the H-SPQ group [24], we considered the Root Mean Square Error (RMSE) adjusted by the number of free parameters, as in [63].

The models that best fit the data from the H-SPQ group are presented in Figure 4. Our results suggest that the network requires increased lateral excitation in the unisensory and multisensory areas to match the responses of individuals with H-SPQ in the DFI task

(adj RMSE = 32.09). An alternative model based on an increase in cross-modal weights offered an equivalent match to the empirical data (adj RMSE = 32.30) (see Discussion).

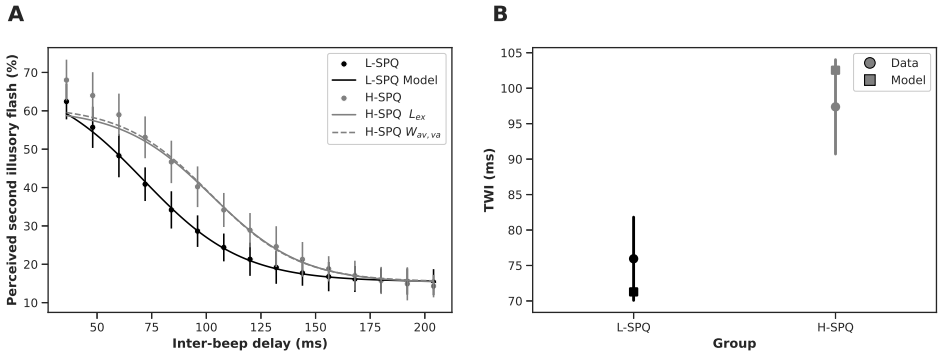

FIGURE 4: L-SPQ and H-SPQ network models identified by the fitting procedure. **Panel A** shows the DFI task responses generated by the L-SPQ and H-SPQ network models. The solid lines depict the sigmoid fit obtained from the data generated by the models. The dots represent the discretisation of the sigmoid fit obtained from the data collected in the experiment [24]. The quantitative fitting of the L-SPQ model (black) using L^c_{0ex} (grey) is sufficient to reproduce the H-SPQ data. Alternatively, a fitting using (W_0^{cd}) (dashed grey) generates a close match to the H-SPQ data. **Panel B** shows the temporal window of illusion (TWI) generated by the L-SPQ and H-SPQ network models. Squares depict the TWI generated by the models, whereas dots and standard error bars depict the TWI observed in the experiment [24].

3.3 Predictions

The responses of the identified models were further explored in a related experimental paradigm. Following [43], we simulated an experiment with four experimental conditions in which a single flash was always presented together with one, two, three or four beeps at a fixed SOA and participants were asked to report the number of perceived flashes. The corresponding average responses of the models in each experimental condition are presented in Figure 5A.

In line with experimental evidence [43], we observed that the baseline network model

(L-SPQ) on average detects a single flash when stimulated with a single beep, while the average number of detected flashes tends to increase as the number of displayed beeps increases, converging at a value close to two. Based on the results of both H-SPQ models, we predict that individuals scoring high on schizotypy would show a slight increase in the average number of perceived flashes under each experimental condition compared to individuals scoring low on schizotypy.

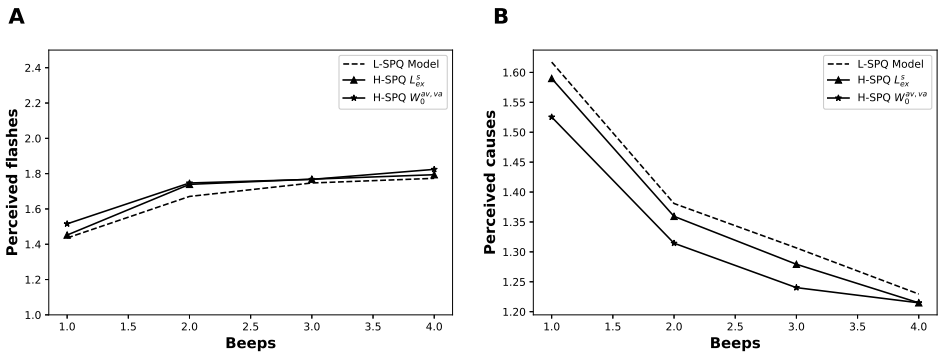

FIGURE 5: Simulation of the average number of perceived flashes and causes in the L-SPQ and H-SPQ network models. **Panels A** and **B** show the average responses of the models in the number of flashes and causes perceived in a simulated experimental task delivering a single flash and up to four beeps. For each experimental condition, the simulation lasts for 750 ms and the SOA among beeps is maintained constant at 48 ms. **Panel A** shows the average perceived flashes for up to four auditory stimuli. Both H-SPQ network models present a slight increase in the average number of perceived flashes under each experimental condition compared to the L-SPQ model. **Panel B** shows the average perceived causes of up to four auditory stimuli. All models show a decrease in the average number of perceived causes as the number of displayed beeps increases. This effect is stronger in the H-SPQ models.

Furthermore, the causal inference responses of the models in each experimental condition are presented in Figure 5B. We found that the models show a decrease in the average number of perceived causes as the number of displayed beeps increases. Based on the causal inference responses of both H-SPQ models, we predict that individuals scoring

high in schizotypy would report a single cause of the stimuli at each experimental condition more frequently than individuals scoring low in schizotypy.

4 Discussion

This study aimed at modelling the neural mechanisms behind the higher proneness to the double flash illusion observed in individuals scoring high in schizotypy [24]. For this purpose, an existing double flash illusion audiovisual network model [45] was adapted to incorporate a multisensory layer with feedforward and feedback connectivity to unisensory layers, allowing iterative causal inference across the network [33, 46]. This adaptation allowed us to quantitatively fit this model to behavioural data for the first time, revealing the crucial role of feedback input to generate sound-induced visual illusions (see Figure 2).

We explored the effects of E/I imbalance, failures in top-down signaling, and synaptic decrease in illusory responses of the model. These neural impairments are compatible with the malfunction of NMDA receptors and GABA neurons together with increased activity of the D_2 receptor, which are key elements of computational accounts of schizophrenia [58, 59, 64]. Fitting to the experimental data (Section 3.2) suggests two neural mechanisms of E/I balance modulation that explain the increased proneness to double flash illusion in individuals with high schizotypy: an increase in the recurrent excitation of unisensory and multisensory neurons or an increase in the strength of cross-modal connectivity at the unisensory level. Our simulations (Section 3.1) reveal that both mechanisms enlarge the temporal window of the illusion (TWI) and increase the overall proneness to experience the illusion (OPI). Moreover, we found that such neural mechanisms increase the overall probability that the network infers a common cause from the stimuli.

Furthermore, we used the H-SPQ network models to predict illusory responses in a related sound-induced flash illusion paradigm [43] (Section 3.3). We predict that high-schizotypy individuals would show a slight increase in the average number of perceived flashes when stimulated with up to 4 beeps and more often attribute them to a single cause

compared to low-schizotypy individuals.

4.1 An E/I imbalance account of reduced temporal discrimination in the SCZ spectrum

Our results are in line with magnetic resonance spectroscopy (MRS) studies that consistently show alterations in glutamatergic and GABAergic neurotransmission in the schizophrenia spectrum [65]. Particularly, a disrupted E/I balance has been associated with a wider temporal window of illusion in individuals with schizotypal traits: concentrations of glutamatergic compounds (Glx) mediated by individual E/I genetic profiles predict multisensory temporal discrimination in the Simultaneity Judgement (SJ) task[5] and cognitive-perceptual scores on the SPQ scale. Higher concentrations of Glx during the SJ task correlate with lower temporal discrimination and higher schizotypy in individuals with a genetic shift toward greater excitation [25].

Specifically, our results expand evidence showing that the concentration of GABA in STS is correlated with the perception of DFI and influences the oscillatory activity of the gamma band [55]. Electrophysiological evidence has consistently shown a reduction in the power of gamma band activity in patients with SCZ [66–69]. In the DFI paradigm, this is observed as a lack of enhancement of beta/gamma band (25-35 Hz) responses in trials where the illusory flash is perceived, suggesting early binding deficits in SCZ [70]. Presumably, the imbalance of E/I could be driving the slower oscillatory patterns observed in key areas of multisensory integration in the SCZ spectrum [28].

Furthermore, our results may be compatible with a recent animal study showing that a reduction in GABAergic inhibition in the audiovisual cortex is sufficient to alter multisensory processing and reduce audiovisual temporal discrimination [71]. Here, wild-type rats received Gabazine (a GABA-A receptor antagonist) in the V2L cortex (an area known to process audiovisual stimuli) and evaluated behaviourally and electrophysiologically during audiovisual stimulation and the Temporal Order Judgement (TOJ) task[6]. The

[5] In the SJ task participants indicate whether auditory and tactile stimuli delivered at varying onset disparities were presented at the same time or not.

[6] In the TOJ task animals were trained to indicate whether auditory or tactile stimuli delivered at different

results indicate that the infusion of gabazine was sufficient to widen the temporal window of audiovisual integration and decrease the perception of time differences between auditory and visual stimuli. Nevertheless, we modelled the perturbation of the E/I balance as an increase in excitatory activity (rather than a diminution of inhibitory mechanisms) between the unisensory regions or within both the unisensory and multisensory areas.

To our knowledge, this is the first study to show computationally at the network level how a biologically plausible deregulation of the E/I balance reduces temporal sensitivity (indexed by the size of the TWI) in the schizophrenia spectrum. Moreover, we show how this impairment causes a higher proneness to experience sound-induced visual illusory phenomena [22, 24]. In line with [63], we suggest that this mechanism causes the decreased sensitivity observed in schizophrenia and high schizotypal individuals in variants of the task (e.g., tactile-induced double flash illusion) [28] and related spatiotemporal discrimination paradigms [13, 26, 27, 72, 73].

4.2 Possible links with Bayesian accounts of SCZ

Our modelling approach could be seen as a way to bridge the neural and Bayesian level of analysis of sound-induced visual illusory phenomena. Our results extend previous network modelling of audiovisual integration showing that cross-modal synapses encode prior probabilities of the co-occurrence of audiovisual stimuli [74, 75]. This aligns with Bayesian Causal Inference modelling of the SJ task, indicating that temporal binding windows are influenced by an a priori bias to bind sensory information [76–78].

Future research could aim to link our results to predictive coding accounts of schizophrenia [57, 59, 79–81]. Our findings on the effects of the E/I imbalance in multisensory integration could be compatible with strong prior accounts of psychotic symptoms [82–84]: The increased proneness to experience sound-induced visual illusory phenomena in the schizophrenia spectrum may be a manifestation of amplified prior sensory expectations for a common cause encoded in cross-modal synapses. However, there are alternative Bayesian accounts of SCZ suggesting weaker or variable priors [85–92]; stronger priors at high levels of the sensory processing hierarchy coupled with weaker

onset asynchronies were presented first

priors at low levels [93, 94]; or circular inference[7] in multisensory integration [58].

We stress the importance of evaluating whether stronger causal priors at early sensory processing stages (indexed by the strength of cross-modal or lateral connectivity in primary sensory cortices) enlarge temporal binding windows and increase the overall proneness to experience sound-induced visual illusory phenomena.

4.3 Multisensory causal inference in bodily self-aberrations

To our knowledge, this is the first study to computationally examine explicit multisensory causal inference in the schizophrenia spectrum [95]. We showed at the network level how deregulation of the E/I balance would increase the overall probability of inferring a common cause from sensory stimuli (see Figures 3C and 5B), resembling the increase in unified evoked responses of the multisensory cortex (V2L) after Gabazine infussion in the aforementioned animal study [71].

Speculatively, our predictions about multisensory causal inference impairments due to E/I imbalance could be related to bodily self-aberrations observed in the SCZ spectrum [4, 13, 15, 96]. Empirical evidence consistently shows a co-occurrence of reduced spatiotemporal discrimination and impaired body ownership indexed by a higher proneness to experience the Rubber Hand Illusion (RHI) in the SCZ spectrum [6–9]. We tentatively suggest that a stronger prior of sensory stimuli (visual and tactile) coming from a common source produced by an imbalance of E/I in multisensory integration networks increases susceptibility to RHI and bodily perceptual symptoms observed in the SCZ spectrum [9, 97, 98].

This view is supported by a recent Bayesian Causal Inference report showing that prior probabilities for a common cause of sensory stimuli in RHI and SJ tasks are correlated [78]. In addition, there is neuroimaging evidence of reduced activity in the intraparietal sulcus (IPS) observed in SCZ [99], a key area of the posterior parietal cortex thought to implement Bayesian causal inference of body ownership [100]. However, direct evidence

[7] Here MSI is understood as a hierarchical process where two parallel hierarchies (auditory and visual) are connected with a common node at the top (representing a common cause). Sensory information climbs the hierarchy moving from unisensory to multisensory areas and can enter the opposite as a prior.

of causal inference for RHI and SJ tasks in SCZ is lacking.

We highlight the relevance of directly examining the role of causal inference in multisensory integration in other multisensory illusion paradigms [4, 14, 15, 96] to clarify our understanding of bodily self-disturbances experienced by SCZ patients.

4.4 Limitations and Future Directions

The network model used in this study is a rough simplification of electrochemical processes and synaptic connectivity of neurons that compute multisensory integration. As a consequence, our network does not account for the dynamics of oscillatory activity observed in illusory phenomena induced by sound or touch [28, 55, 70]. Moreover, our network does not accurately reproduce the responses observed at lower inter-beep delays (see Figure 4A). At best, the network model presented here serves as a tool for theoretical reasoning and should not be taken as equivalent to the real neural mechanism behind audiovisual integration (see [101]). Furthermore, our modelling approach does not consider relevant neural input from frontal areas. This input is crucial considering that cognitive load and attention modulate the proneness to experience the DFI [48] and specific impairments in selective attention and cognitive control have been observed in the schizophrenia spectrum [102, 103]. Furthermore, neuroimaging evidence suggests a critical role for fronto-temporal regions in multisensory integration impairments observed in SCZ [58, 104].

The experimental evidence directly evaluating sound-induced flash illusions in the SCZ spectrum has only begun to accumulate [22, 24, 70, 105]. Our findings are based on the modelling of a single study that employs a specific experimental paradigm with healthy individuals and does not directly evaluate causal inference responses in participants [24]. However, there is a wide array of experimental paradigms measuring sound-induced flash illusions with variations in the stimuli (e.g. number of flashes and beeps) and spatiotemporal manipulations [41, 48]. These available paradigms have not been thoroughly explored in the SCZ spectrum and complicate the interpretation of contradictory findings about the proneness of the illusion in SCZ [105]. Further multisensory illusion studies with a more diverse stimuli setup and explicit measurements of both causal infer-

ence and spatiotemporal sensitivity in the SCZ spectrum are required.

5 Abbreviations

TBW: Temporal Binding Window, TWI: Temporal Window of Illusion, SOA: Stimulus Onset Asynchrony, SCZ: Schizophrenia, L-SPQ: Low schizotypy, H-SPQ: High schizotypy, MSI: Multisensory Integration, DFI: Double Flash Illusion, SJ: Simultaneity Judgment, RHI: Rubber Hand Illusion, TOJ: Temporal Order Judgment.

6 Declarations

6.1 Authors' contribution

RP: Conceptualisation, Methodology, Software, Formal Analysis, Writing - Original Draft. FF: Conceptualisation, Investigation, Writing - Review and Editing. VR: Conceptualisation, Investigation, Writing - Review and Editing. PS: Conceptualisation, Methodology, Supervision, Writing - Reviewing and Editing.

6.2 Competing interests

The authors declare that the research was conducted in the absence of commercial or financial relationships that could be construed as a potential conflict of interest.

6.3 Availability of data and materials

The code used for the implementation of the model can be found at:

https://github.com/renatoparedes/SCZ_DFI_model.

6.4 Consent for publication

Authors have agreed to submit it in its current form for consideration for publication in the Journal.

6.5 Ethics approval and consent to participate

Not applicable. No tests, measurements, or experiments were performed in humans as part of this work.

6.6 Funding

F.F. is supported by the "Departments of Excellence 2023–2027" initiative of the Italian Ministry of Education, University and Research for the Department of Neuroscience, Imaging and Clinical Sciences (DNISC) of the University of Chieti-Pescara, and by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) CUP: D53D23020890001. V.R. is supported by MUR – Ministry of University and Research, Italy (P2022XAKXL and 2022H4ZRSN), Ministerio de Ciencia, Innovación y Universidades, Spain (PID2019111335GA-100) and BIAL Foundation (033/22).

7 Acknowledgements

We acknowledge Denise Robles Soberón (Hey Cometa) for her contribution in the graphics employed in this paper.

References

[1] B. P. Chang and M. F. Lenzenweger. Somatosensory processing in the biological relatives of schizophrenia patients: a signal detection analysis of two-point discrimination. *Journal of Abnormal Psychology*, 110(3):433–442, August 2001.

[2] B. P. Chang and M. F. Lenzenweger. Investigating Graphesthesia Task Performance in the Biological Relatives of Schizophrenia Patients. *Schizophrenia Bulletin*, 30(2):327–334, January 2004.

[3] Bernard P. Chang and Mark F. Lenzenweger. Somatosensory processing and schizophrenia liability: proprioception, exteroceptive sensitivity, and graphesthe-

sia performance in the biological relatives of schizophrenia patients. *Journal of Abnormal Psychology*, 114(1):85–95, February 2005.

[4] Jamie Michael and Sohee Park. Anomalous bodily experiences and perceived social isolation in schizophrenia: An extension of the Social Deafferentation Hypothesis. *Schizophrenia Research*, 176(2-3):392–397, October 2016.

[5] Katharine N. Thakkar, Heathman S. Nichols, Lindsey G. McIntosh, and Sohee Park. Disturbances in Body Ownership in Schizophrenia: Evidence from the Rubber Hand Illusion and Case Study of a Spontaneous Out-of-Body Experience. *PLoS ONE*, 6(10):e27089, October 2011.

[6] Ileana Rossetti, Daniele Romano, Vincenzo Florio, Stefania Doria, Veronica Nisticò, Andreas Conca, Claudio Mencacci, and Angelo Maravita. Defective Embodiment of Alien Hand Uncovers Altered Sensorimotor Integration in Schizophrenia. *Schizophrenia Bulletin*, page sbz050, May 2019.

[7] Marcello Costantini, Anatolia Salone, Giovanni Martinotti, Federica Fiori, Francesca Fotia, Massimo Di Giannantonio, and Francesca Ferri. Body representations and basic symptoms in schizophrenia. *Schizophrenia Research*, 222:267–273, August 2020.

[8] Regine Zopf, Kelsie Boulton, Robyn Langdon, and Anina N. Rich. Perception of visual-tactile asynchrony, bodily perceptual aberrations, and bodily illusions in schizophrenia. *Schizophrenia Research*, 228:534–540, February 2021.

[9] Jingqi He, Honghong Ren, Jinguang Li, Min Dong, Lulin Dai, Zhijun Li, Yating Miao, Yunjin Li, Peixuan Tan, Lin Gu, Xiaogang Chen, and Jinsong Tang. Deficits in Sense of Body Ownership, Sensory Processing, and Temporal Perception in Schizophrenia Patients With/Without Auditory Verbal Hallucinations. *Frontiers in Neuroscience*, 16:831714, April 2022.

[10] Sung-Hyouk Park, Jeonghun Ku, Jae-Jin Kim, Hee Jeong Jang, So Young Kim, Soo Hyun Kim, Chan-Hyung Kim, Hyeongrae Lee, In Young Kim, and Sun I.

Kim. Increased personal space of patients with schizophrenia in a virtual social environment. *Psychiatry Research*, 169(3):197–202, October 2009.

[11] Daphne J. Holt, Emily A. Boeke, Garth Coombs, Stephanie N. DeCross, Brittany S. Cassidy, Steven Stufflebeam, Scott L. Rauch, and Roger B.H. Tootell. Abnormalities in personal space and parietal–frontal function in schizophrenia. *NeuroImage: Clinical*, 9:233–243, 2015.

[12] Giulio Di Cosmo, Marcello Costantini, Anatolia Salone, Giovanni Martinotti, Giuseppe Di Iorio, Massimo Di Giannantonio, and Francesca Ferri. Peripersonal space boundary in schizotypy and schizophrenia. *Schizophrenia Research*, 197:589–590, July 2018.

[13] Giulio Di Cosmo, Marcello Costantini, Ettore Ambrosini, Anatolia Salone, Giovanni Martinotti, Mariangela Corbo, Massimo Di Giannantonio, and Francesca Ferri. Body-environment integration: Temporal processing of tactile and auditory inputs along the schizophrenia continuum. *Journal of Psychiatric Research*, 134:208–214, February 2021.

[14] F. Ferroni, M. Ardizzi, M. Sestito, V. Lucarini, B.D. Daniel, F. Paraboschi, M. Tonna, C. Marchesi, and V. Gallese. Shared multisensory experience affects Others' boundary: The enfacement illusion in schizophrenia. *Schizophrenia Research*, 206:225–235, April 2019.

[15] Karl Erik Sandsten, Julie Nordgaard, Troels Wesenberg Kjaer, Vittorio Gallese, Martina Ardizzi, Francesca Ferroni, Janne Petersen, and Josef Parnas. Altered self-recognition in patients with schizophrenia. *Schizophrenia Research*, 218:116–123, April 2020.

[16] Barry E. Stein and Terrence R. Stanford. Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4):255–266, April 2008.

[17] Barry E. Stein, David Burr, Christos Constantinidis, Paul J. Laurienti, M. Alex

Meredith, Thomas J. Perrault, Ramnarayan Ramachandran, Brigitte Röder, Benjamin A. Rowland, K. Sathian, Charles E. Schroeder, Ladan Shams, Terrence R. Stanford, Mark T. Wallace, Liping Yu, and David J. Lewkowicz. Semantic confusion regarding the development of multisensory integration: a practical solution. *The European journal of neuroscience*, 31(10):1713–1720, May 2010.

[18] Jean-Paul Noel, Olaf Blanke, and Andrea Serino. From multisensory integration in peripersonal space to bodily self-consciousness: from statistical regularities to statistical inference: Multisensory integration and self-consciousness. *Annals of the New York Academy of Sciences*, 1426(1):146–165, August 2018.

[19] Andrea Serino. Peripersonal space (PPS) as a multisensory interface between the individual and the environment, defining the space of the self. *Neuroscience & Biobehavioral Reviews*, 99:138–159, April 2019.

[20] Mikhail Zvyagintsev, Carmen Parisi, and Klaus Mathiak. Temporal processing deficit leads to impaired multisensory binding in schizophrenia. *Cognitive Neuropsychiatry*, 22(5):361–372, September 2017.

[21] Brice Martin, Anne Giersch, Caroline Huron, and Virginie van Wassenhove. Temporal event structure and timing in schizophrenia: Preserved binding in a longer "now". *Neuropsychologia*, 51(2):358–371, January 2013.

[22] Katharina Haß, Christopher Sinke, Tanya Reese, Mandy Roy, Daniel Wiswede, Wolfgang Dillo, Bob Oranje, and Gregor R. Szycik. Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. *Cognitive Neuropsychiatry*, 22(2):145–158, March 2017.

[23] Anne-Marie Muller, Tyler C. Dalal, and Ryan A. Stevenson. Schizotypal personality traits and multisensory integration: An investigation using the McGurk effect. *Acta Psychologica*, 218:103354, July 2021.

[24] Francesca Ferri, Agnese Venskus, Francesca Fotia, Jason Cooke, and Vincenzo Romei. Higher proneness to multisensory illusions is driven by reduced temporal

sensitivity in people with high schizotypal traits. *Consciousness and Cognition*, 65:263–270, October 2018.

[25] Francesca Ferri, Yuliya S Nikolova, Mauro Gianni Perrucci, Marcello Costantini, Antonio Ferretti, Valentina Gatta, Zirui Huang, Richard A E Edden, Qiang Yue, Marco D'Aurora, Etienne Sibille, Liborio Stuppia, Gian Luca Romani, and Georg Northoff. A Neural "Tuning Curve" for Multisensory Experience and Cognitive-Perceptual Schizotypy. *Schizophrenia Bulletin*, 43(4):801–813, July 2017.

[26] Han-yu Zhou, Xin-lu Cai, Michael Weigl, Peter Bang, Eric F.C. Cheung, and Raymond C.K. Chan. Multisensory temporal binding window in autism spectrum disorders and schizophrenia spectrum disorders: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, 86:66–76, March 2018.

[27] Tyler C Dalal, Anne-Marie Muller, and Ryan A Stevenson. The Relationship Between Multisensory Temporal Processing and Schizotypal Traits. *Multisensory Research*, page 19, 2021.

[28] Francesca Fotia, Jason Cooke, Loes Van Dam, Francesca Ferri, and Vincenzo Romei. The temporal sensitivity to the tactile-induced double flash illusion mediates the impact of beta oscillations on schizotypal personality traits. *Consciousness and Cognition*, 91:103121, May 2021.

[29] Yinan Cao, Christopher Summerfield, Hame Park, Bruno Lucio Giordano, and Christoph Kayser. Causal Inference in the Multisensory Brain. *Neuron*, 102(5):1076–1087.e8, June 2019.

[30] Christoph Kayser and Ladan Shams. Multisensory Causal Inference in the Brain. *PLOS Biology*, 13(2):e1002075, February 2015.

[31] John F. Magnotti, Kristen B. Smith, Marcelo Salinas, Jacqunae Mays, Lin L. Zhu, and Michael S. Beauchamp. A causal inference explanation for enhancement of multisensory integration by co-articulation. *Scientific Reports*, 8(1):18032, December 2018.

[32] Jeff T. Mohl, John M. Pearson, and Jennifer M. Groh. Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli. *Journal of Neurophysiology*, 124(3):715–727, September 2020.

[33] Tim Rohe, Ann-Christine Ehlis, and Uta Noppeney. The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications*, 10(1):1907, December 2019.

[34] Konrad P. Körding, Ulrik Beierholm, Wei Ji Ma, Steven Quartz, Joshua B. Tenenbaum, and Ladan Shams. Causal Inference in Multisensory Perception. *PLoS ONE*, 2(9):e943, September 2007.

[35] David Marr. *Vision: a computational investigation into the human representation and processing of visual information*. MIT Press, Cambridge, Mass, 2010. OCLC: ocn472791457.

[36] Cristiano Cuppini, Ladan Shams, Elisa Magosso, and Mauro Ursino. A biologically inspired neurocomputational model for audiovisual integration and causal inference. *European Journal of Neuroscience*, 46(9):2481–2498, 2017.

[37] Ying Fang, Zhaofei Yu, Jian K. Liu, and Feng Chen. A unified neural circuit of causal inference and multisensory integration. *Neurocomputing*, 358:355–368, September 2019.

[38] Reuben Rideaux, Katherine R. Storrs, Guido Maiello, and Andrew E. Welchman. How multisensory neurons solve causal inference. *Proceedings of the National Academy of Sciences*, 118(32):e2106235118, August 2021.

[39] Jason Samaha and Vincenzo Romei. Alpha-band frequency and temporal windows in perception: a review and living meta-analysis of 27 experiments (and counting). *Journal of Cognitive Neuroscience*, pages 1–15, 2024.

[40] Jan-Mathijs Schoffelen, Ugo Giulio Pesci, and Uta Noppeney. Alpha oscillations and temporal binding windows in perception—a critical review and best practice guidelines. *Journal of Cognitive Neuroscience*, pages 1–36, 2024.

[41] Rebecca J. Hirst, David P. McGovern, Annalisa Setti, Ladan Shams, and Fiona N. Newell. What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. *Neuroscience & Biobehavioral Reviews*, 118:759–774, November 2020.

[42] Roberto Cecere, Geraint Rees, and Vincenzo Romei. Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. *Current Biology*, 25(2):231–235, January 2015.

[43] Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Visual illusion induced by sound. *Cognitive Brain Research*, 14(1):147–152, June 2002.

[44] Jason Cooke, Claudia Poch, Helge Gillmeister, Marcello Costantini, and Vincenzo Romei. Oscillatory properties of functional connections between sensory areas mediate cross-modal illusory perception. *Journal of Neuroscience*, 39(29):5711–5718, 2019.

[45] Cristiano Cuppini, Elisa Magosso, Nadia Bolognini, Giuseppe Vallar, and Mauro Ursino. A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage*, 92:248–266, May 2014.

[46] Tim Rohe and Uta Noppeney. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biology*, 13(2):e1002073, February 2015.

[47] Renato Paredes, Peggy Seriès, and Juan Cabral. Scikit-NeuroMSI: a Python framework for multisensory integration modelling. *IX Congreso de Matemática Aplicada, Computacional e Industrial*, 9:545–548, 2023.

[48] Julian Keil. Double Flash Illusions: Current Findings and Future Directions. *Frontiers in Neuroscience*, 14:298, 2020.

[49] Ladan Shams, Sunao Iwaki, Aman Chawla, and Joydeep Bhattacharya. Early modulation of visual cortex by sound: an MEG study. *Neuroscience Letters*, 378(2):76–81, April 2005.

[50] Tommi Raij, Jyrki Ahveninen, Fa-Hsuan Lin, Thomas Witzel, Iiro P. Jääskeläinen, Benjamin Letham, Emily Israeli, Cherif Sahyoun, Christos Vasios, Steven Stufflebeam, Matti Hämäläinen, and John W. Belliveau. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices: Onset timing of audiovisual processing for simple stimuli. *European Journal of Neuroscience*, 31(10):1772–1782, May 2010.

[51] Vincenzo Romei, Micah M Murray, Lotfi B Merabet, and Gregor Thut. Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. *Journal of Neuroscience*, 27(43):11465–11472, 2007.

[52] Céline Cappe, Gregor Thut, Vincenzo Romei, and Micah M Murray. Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. *Journal of Neuroscience*, 30(38):12572–12580, 2010.

[53] Vincenzo Romei, Joachim Gross, and Gregor Thut. Sounds reset rhythms of visual cortex and corresponding human visual perception. *Current biology*, 22(9):807–813, 2012.

[54] J. Mishra, A. Martinez, T. J. Sejnowski, and S. A. Hillyard. Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. *Journal of Neuroscience*, 27(15):4120–4131, April 2007.

[55] Johanna Balz, Julian Keil, Yadira Roa Romero, Ralf Mekle, Florian Schubert, Semiha Aydin, Bernd Ittermann, Jürgen Gallinat, and Daniel Senkowski. GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. *NeuroImage*, 125:724–730, January 2016.

[56] Julian Keil and Daniel Senkowski. Neural Oscillations Orchestrate Multisensory Processing. *The Neuroscientist*, 24(6):609–626, December 2018.

[57] Renaud Jardri, Kenneth Hugdahl, Matthew Hughes, Jérôme Brunelin, Flavie Waters, Ben Alderson-Day, Dave Smailes, Philipp Sterzer, Philip R. Corlett, Pan-

telis Leptourgos, Martin Debbané, Arnaud Cachia, and Sophie Denève. Are Hallucinations Due to an Imbalance Between Excitatory and Inhibitory Influences on the Brain? *Schizophrenia Bulletin*, 42(5):1124–1134, September 2016.

[58] Pantelis Leptourgos, Vincent Bouttier, Sophie Denève, and Renaud Jardri. From hallucinations to synaesthesia: A circular inference account of unimodal and multimodal erroneous percepts in clinical and drug-induced psychosis. *Neuroscience & Biobehavioral Reviews*, 135:104593, April 2022.

[59] Philipp Sterzer, Rick A. Adams, Paul Fletcher, Chris Frith, Stephen M. Lawrie, Lars Muckli, Predrag Petrovic, Peter Uhlhaas, Martin Voss, and Philip R. Corlett. The Predictive Coding Account of Psychosis. *Biological Psychiatry*, 84(9):634–643, November 2018.

[60] Ian Ellison-Wright and Ed Bullmore. Meta-analysis of diffusion tensor imaging studies in schizophrenia. *Schizophrenia Research*, 108(1-3):3–10, March 2009.

[61] R. E. Hoffman and S. K. Dobscha. Cortical Pruning and the Development of Schizophrenia: A Computer Model. *Schizophrenia Bulletin*, 15(3):477–490, January 1989.

[62] R. E. Hoffman and T. H. McGlashan. Using a Speech Perception Neural Network Computer Simulation to Contrast Neuroanatomic versus Neuromodulatory Models of Auditory Hallucinations. *Pharmacopsychiatry*, 39:54–64, February 2006.

[63] Renato Paredes, Francesca Ferri, and Peggy Seriès. Influence of E/I balance and pruning in peri-personal space differences in schizophrenia: A computational approach. *Schizophrenia Research*, 248:368–377, October 2022.

[64] Pablo Lanillos, Daniel Oliva, Anja Philippsen, Yuichi Yamashita, Yukie Nagai, and Gordon Cheng. A review on neural network models of schizophrenia and autism spectrum disorder. *Neural Networks*, 122:338–363, February 2020.

[65] Mounica Reddy-Thootkur, Nina Vanessa Kraguljac, and Adrienne Carol Lahti. The role of glutamate and GABA in cognitive dysfunction in schizophrenia and

mood disorders – A systematic review of magnetic resonance spectroscopy studies. *Schizophrenia Research*, 249:74–84, November 2022.

[66] Gregor Leicht, Valerie Kirsch, Ina Giegling, Susanne Karch, Irmgard Hantschk, Hans-Jürgen Möller, Oliver Pogarell, Ulrich Hegerl, Dan Rujescu, and Christoph Mulert. Reduced Early Auditory Evoked Gamma-Band Response in Patients with Schizophrenia. *Biological Psychiatry*, 67(3):224–231, February 2010.

[67] Daniel Lenz, Susanne Fischer, Jeanette Schadow, Bernhard Bogerts, and Christoph S. Herrmann. Altered evoked gamma-band responses as a neurophysiological marker of schizophrenia? *International Journal of Psychophysiology*, 79(1):25–31, January 2011.

[68] Gregor Leicht, Christina Andreou, Nenad Polomac, Clarissa Lanig, Daniel Schöttle, Martin Lambert, and Christoph Mulert. Reduced auditory evoked gamma band response and cognitive processing deficits in first episode schizophrenia. *The World Journal of Biological Psychiatry*, 16(6):387–397, August 2015.

[69] Julian Keil, Yadira Roa Romero, Johanna Balz, Melissa Henjes, and Daniel Senkowski. Positive and Negative Symptoms in Schizophrenia Relate to Distinct Oscillatory Signatures of Sensory Gating. *Frontiers in Human Neuroscience*, 10, March 2016.

[70] Johanna Balz, Yadira Roa Romero, Julian Keil, Martin Krebber, Michael Niedeggen, Jürgen Gallinat, and Daniel Senkowski. Beta/Gamma Oscillations and Event-Related Potentials Indicate Aberrant Multisensory Processing in Schizophrenia. *Frontiers in Psychology*, 7, December 2016.

[71] Ashley L Schormans and Brian L Allman. An imbalance of excitation and inhibition in the multisensory cortex impairs the temporal acuity of audiovisual processing and perception. *Cerebral Cortex*, 33(18):9937–9953, September 2023.

[72] Huai-Hsuan Tseng, Matthijs G. Bossong, Gemma Modinos, Kuan-Ming Chen, Philip McGuire, and Paul Allen. A systematic review of multisensory cog-

nitive–affective integration in schizophrenia. *Neuroscience & Biobehavioral Reviews*, 55:444–452, August 2015.

[73] Francesca Ferri, Ettore Ambrosini, and Marcello Costantini. Spatiotemporal processing of somatosensory stimuli in schizotypy. *Scientific Reports*, 6:38735, December 2016.

[74] Mauro Ursino, Andrea Crisafulli, Giuseppe di Pellegrino, Elisa Magosso, and Cristiano Cuppini. Development of a Bayesian Estimator for Audio-Visual Integration: A Neurocomputational Study. *Frontiers in Computational Neuroscience*, 11:89, October 2017.

[75] Mauro Ursino, Cristiano Cuppini, Elisa Magosso, Ulrik Beierholm, and Ladan Shams. Explaining the Effect of Likelihood Manipulation and Prior Through a Neural Network of the Audiovisual Perception of Space. *Multisensory Research*, 32(2):111–144, 2019.

[76] John F. Magnotti, Wei Ji Ma, and Michael S. Beauchamp. Causal inference of asynchronous audiovisual speech. *Frontiers in Psychology*, 4, 2013.

[77] Jean-Paul Noel, Ryan A. Stevenson, and Mark T. Wallace. Atypical audiovisual temporal function in autism and schizophrenia: similar phenotype, different cause. *European Journal of Neuroscience*, 47(10):1230–1241, May 2018.

[78] Marie Chancel, H Henrik Ehrsson, and Wei Ji Ma. Uncertainty-based inference of a common cause for body ownership. *eLife*, 11:e77221, September 2022.

[79] Luca Tarasi, Giuseppe di Pellegrino, and Vincenzo Romei. Are you an empiricist or a believer? neural signatures of predictive strategies in humans. *Progress in Neurobiology*, 219:102367, 2022.

[80] Luca Tarasi, Jelena Trajkovic, Stefano Diciotti, Giuseppe di Pellegrino, Francesca Ferri, Mauro Ursino, and Vincenzo Romei. Predictive waves in the autism-schizophrenia continuum: a novel biobehavioral model. *Neuroscience & Biobehavioral Reviews*, 132:1–22, 2022.

[81] Luca Tarasi, Maria Eugenia Martelli, Marta Bortoletto, Giuseppe di Pellegrino, and Vincenzo Romei. Neural signatures of predictive strategies track individuals along the autism-schizophrenia continuum. *Schizophrenia Bulletin*, 49(5):1294–1304, 2023.

[82] Christoph Teufel, Naresh Subramaniam, Veronika Dobler, Jesus Perez, Johanna Finnemann, Puja R. Mehta, Ian M. Goodyer, and Paul C. Fletcher. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences*, 112(43):13401–13406, October 2015.

[83] A. R. Powers, C. Mathys, and P. R. Corlett. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, 357(6351):596–600, August 2017.

[84] Clifford M. Cassidy, Peter D. Balsam, Jodi J. Weinstein, Rachel J. Rosengard, Mark Slifstein, Nathaniel D. Daw, Anissa Abi-Dargham, and Guillermo Horga. A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. *Current Biology*, 28(4):503–514.e4, February 2018.

[85] J.M. Nazimek, M.D. Hunter, and P.W.R. Woodruff. Auditory hallucinations: Expectation–perception model. *Medical Hypotheses*, 78(6):802–810, June 2012.

[86] Katharina Schmack, Ana Gòmez-Carrillo De Castro, Marcus Rothkirch, Maria Sekutowicz, Hannes Rössler, John-Dylan Haynes, Andreas Heinz, Predrag Petrovic, and Philipp Sterzer. Delusions and the Role of Beliefs in Perceptual Inference. *The Journal of Neuroscience*, 33(34):13701–13712, August 2013.

[87] Katharina Schmack, Alexandra Schnack, Josef Priller, and Philipp Sterzer. Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition*, 2(2):72–77, June 2015.

[88] Katharina Schmack, Marcus Rothkirch, Josef Priller, and Philipp Sterzer.

Enhanced predictive signalling in schizophrenia. *Human Brain Mapping*, 38(4):1767–1779, April 2017.

[89] Renaud Jardri, Sandrine Duverne, Alexandra S Litvinova, and Sophie Denève. Experimental evidence for circular inference in schizophrenia. *Nature Communications*, 8(1):14218, January 2017.

[90] Vincent Valton, Povilas Karvelis, Katie L Richards, Aaron R Seitz, Stephen M Lawrie, and Peggy Seriès. Acquisition of visual priors and induced hallucinations in chronic schizophrenia. *Brain*, 142(8):2523–2537, August 2019.

[91] Paul Charles Fletcher and Christoph R. Teufel. The Changing Weight of Expectation: How Shifting Priors Underpin Variability in Hallucination Frequency. *Biological Psychiatry*, 92(10):752–753, November 2022.

[92] Isabella Goodwin, Joshua Kugel, Robert Hester, and Marta I. Garrido. Bayesian accounts of perceptual decisions in the nonclinical continuum of psychosis: Greater imprecision in both top-down and bottom-up processes. *PLOS Computational Biology*, 19(11):e1011670, November 2023.

[93] Genevieve J. Yang, John D. Murray, Xiao-Jing Wang, David C. Glahn, Godfrey D. Pearlson, Grega Repovs, John H. Krystal, and Alan Anticevic. Functional hierarchy underlies preferential connectivity disturbances in schizophrenia. *Proceedings of the National Academy of Sciences*, 113(2), January 2016.

[94] Predrag Petrovic and Philipp Sterzer. Resolving the Delusion Paradox. *Schizophrenia Bulletin*, 49(6):1425–1436, November 2023.

[95] Ranran L French and Gregory C DeAngelis. Multisensory neural processing: from cue integration to causal inference. *Current Opinion in Physiology*, 16:8–13, August 2020.

[96] Maayke Klaver and H. Chris Dijkerman. Bodily Experience in Schizophrenia: Factors Underlying a Disturbed Sense of Body Ownership. *Frontiers in Human Neuroscience*, 10, June 2016.

[97] Andrew Laurin, Nicolas Ramoz, Aurély Ameller, Antoine Dereux, Julie Zajac, Maxime Bonjour, Sarah Tebeka, Yann Le Strat, and Caroline Dubertret. Self-consciousness impairments in schizophrenia with and without first rank symptoms using the moving rubber hand illusion. *Consciousness and Cognition*, 93:103154, August 2021.

[98] Lénie J. Torregrossa and Sohee Park. Body ownership across schizotypy dimensions: A rubber hand illusion experiment. *Psychiatry Research Communications*, 2(3):100058, September 2022.

[99] Lori Sanfratello, Cheryl Aine, and Julia Stephen. Neuroimaging investigations of dorsal stream processing and effects of stimulus synchrony in schizophrenia. *Psychiatry Research: Neuroimaging*, 278:56–64, August 2018.

[100] Marie Chancel, Heather Iriye, and H. Henrik Ehrsson. Causal Inference of Body Ownership in the Posterior Parietal Cortex. *The Journal of Neuroscience*, 42(37):7131–7143, September 2022.

[101] Olivia Guest and Andrea E. Martin. On Logical Inference over Brains, Behaviour, and Artificial Neural Networks. *Computational Brain & Behavior*, 6(2):213–227, June 2023.

[102] James M. Gold, Rebecca L. Fuller, Benjamin M. Robinson, Elsie L. Braun, and Steven J. Luck. Impaired top–down control of visual search in schizophrenia. *Schizophrenia Research*, 94(1-3):148–155, August 2007.

[103] Tyler A Lesh, Tara A Niendam, Michael J Minzenberg, and Cameron S Carter. Cognitive Control Deficits in Schizophrenia: Mechanisms and Meaning. *Neuropsychopharmacology*, 36(1):316–338, January 2011.

[104] Cornelia Gröhn, Elin Norgren, and Lars Eriksson. A systematic review of the neural correlates of multisensory integration in schizophrenia. *Schizophrenia Research: Cognition*, 27:100219, March 2022.

[105] Lucy D Vanes. Reduced susceptibility to the sound-induced flash fusion illusion in schizophrenia. *Psychiatry Research*, page 8, 2016.

[106] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11:341–359, 1997.

[107] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Supplementary Material

A Multisensory causal inference network model

The network model used in this study is an extension of previous models developed by Cuppini and his collaborators [36, 45]. The model consists of three layers: two encode auditory and visual stimuli, separately, and connect to a multisensory layer where causal inference is computed. Each of these layers consists of 30 neurons arranged topologically to encode a 30° space. Hence, each neuron encodes 1° of space and neurons close to each other encode close spatial positions.

Each neuron will be noted with a superscript c indicating a specific cortical area (a, v or m for the auditory, visual or multisensory area respectively). Similarly, each neuron will hold a subscript j refereed to its spatial position within a given area.

Neurons in each layer have a sigmoid activation function and first-order dynamics:

$$\tau^c \frac{dy_j^c(t)}{dt} = -y_j^c(t) + F\left(u_j^c(t)\right), \qquad c = a, v, m \tag{4}$$

Here, $u(t)$ and $y(t)$ are used to represent the net input and output of a given neuron at time t. τ^c denotes the time constant of neurons belonging to a given area c. $F(u)$ represents the sigmoidal relationship:

$$F\left(u_j^c\right) = \frac{1}{1 + exp^{-s\left(u_j^c - \theta\right)}} \tag{5}$$

Here, s and θ denote the slope and the central position of the sigmoidal relationship respectively. Neurons in all regions differ only in their time constants, chosen to mimic faster sensory processing for stimuli in the auditory region compared to visual stimuli.

The net input of a neuron is the sum of an inside (i.e. within region) component (l_j^c) and an outside (i.e. extra-area) component ($o_j^c(t)$):

$$u_j^c(t) = l_j^c(t) + o_j^c(t) \tag{6}$$

The within region component l_j^c is defined as:

$$l_j^c\left(t\right) = \sum_k L_{jk}^c \cdot y_{jk}^c\left(t\right) \tag{7}$$

Here L_{jk}^c represents the strength of the lateral synapse from a presynaptic neuron at position k to a postsynaptic neuron at position j in the region c. y_{jk}^c is the activity of the presynaptic neuron at position k.

Such synapses are symmetrical and arranged according to a "Mexican hat" pattern (i.e. a central excitatory area surrounded by an inhibitory ring for each neuron, so that the entire layer generates excitation for spatially close stimuli and inhibition for distant stimuli):

$$L_{jk}^c = \begin{cases} L_{0ex}^c \cdot exp\left(-\frac{(D_{jk})^2}{2\cdot(\sigma_{ex}^c)^2}\right) - L_{0in}^c \cdot exp\left(-\frac{(D_{jk})^2}{2\cdot\left(\sigma_{in}^c\right)^2}\right), & D_{jk} \neq 0 \\ 0, & D_{jk} = 0 \end{cases} \tag{8}$$

L_{0ex}^c and L_{0in}^c denote the highest level of excitatory and inhibitory synaptic efficacy in the region c, respectively. D_{jk} indicate the distance between the pre-synaptic neuron and the post-synaptic neurons within a given area:

$$D_{jk} = \begin{cases} \left|j-k\right|, & \left|j-k\right| \leqslant N/2 \\ N - \left|j-k\right|, & \left|j-k\right| > N/2 \end{cases} \tag{9}$$

This defines a circular structure where each neuron receives the same number of lateral connections.

Importantly, the extra-area input is defined differently for unisensory and multisensory areas. The extra-area input for the unisensory areas includes a stimulus from the external world ($e_j^c(t)$), a cross-modal component coming from the other unisensory area ($c_j^c(t)$), a feedback component coming from the multisensory area ($b_j^c(t)$) and a noise component (n_j^c).

The stimulus from the external world is simulated as a 1-D Gaussian function to represent the uncertainty in the detection of external stimuli:

$$e_j^c(t) = E_0^c \cdot exp\left(-\frac{\left(d_j^c\right)^2}{2\left(\sigma^c\right)^2}\right) \tag{10}$$

Here, E_0^c denotes the strength of the stimulus, d_j^c the distance between neuron at position j and the stimulus at position p^c, and σ^c the degree of uncertainty in sensory detection. The distance d_j^c is defined as:

$$d_j^c = \begin{cases} |j - p^c|, & |j - p^c| \leqslant N/2 \\ N - |j - p^c|, & |j - p^c| > N/2 \end{cases} \tag{11}$$

Furthermore, the cross-modal input is defined as:

$$\begin{aligned} c_j^a(t) &= \textstyle\sum_{k=1}^{N} W_{jk}^{av} \cdot y_k^v(t - \Delta t_{cross}) \\ c_j^v(t) &= \textstyle\sum_{k=1}^{N} W_{jk}^{va} \cdot y_k^a(t - \Delta t_{cross}) \end{aligned} \tag{12}$$

Here Δt_{cross} represents the latency of cross-modal inputs between two unisensory regions. The synaptic weights are symmetrically defined ($W_0^{av} = W_0^{va}$ and $\sigma^{av} = \sigma^{va}$) by the Gaussian function:

$$W_{jk}^{cd} = W_0^{cd} \cdot exp\left(-\frac{\left(D_{jk}\right)^2}{2\left(\sigma^{cd}\right)^2}\right), cd = av\ or\ va \tag{13}$$

W_0 denotes the highest level of synaptic efficacy and D_{jk} is the distance between neuron at position j in the post-synaptic unisensory region and the neuron at position k in the pre-synaptic unisensory region. σ^{cd} defines the width of the cross-modal synapses.

Furthermore, the feedback input is defined as:

$$b_j^c(t) = \sum_{k=1}^{N} B_{jk}^{cm} \cdot y_k^c(t - \Delta t_{feed}) \tag{14}$$

Here Δt_{feed} represents the latency of feedback inputs between the multisensory and unisensory regions. The feedback synaptic weights are also symmetrically ($B_0^{am} = B_0^{vm}$ and $\sigma^{am} = \sigma^{vm}$) defined:

$$B_{jk}^{cm} = B_0^{cm} \cdot exp\left(-\frac{\left(D_{jk}\right)^2}{2\left(\sigma^{cm}\right)^2}\right), cm = am\ or\ vm \tag{15}$$

Here, B_0^{cm} denotes the highest level of synaptic efficacy and D_{jk} is the distance between neuron at position j in the post-synaptic unisensory region and the neuron at

position k in the pre-synaptic multisensory region. σ^{cd} defines the width of the feedback synapses.

The noise component n_j^c is extracted from a standard uniform distribution on the interval $[n_{max} + n_{max}]$. Here n_{max} is defined as the 40% of the strength of the external stimulus for each modality.

All these external sources are filtered by a second order differential equation to mimic the temporal dynamics of the stimuli in a cortex:

$$\begin{cases} \frac{d}{dt} o_j^c(t) = \delta_j^c(t) \\ \frac{d}{dt} \delta_j^c(t) = \frac{G^c}{\tau^c} \cdot \left[e_j^c(t) + c_j^c(t) + b_j^c(t) + n_j^c \right] - \frac{2 \cdot \delta_j^c(t)}{\tau^c} - \frac{o_j^c(t)}{(\tau^c)^2} \end{cases}, c = a, v \quad (16)$$

Here, G^c represents gain and τ^c the time constants of the dynamics.

In contrast, the extra-area input for the multisensory area comes only from feedforward synapses from the two unisensory areas:

$$i_j^m(t) = \sum_{k=1}^{N} W_{jk}^{ma} \cdot y_k^a(t - \Delta t_{feed}) + W_{jk}^{mv} \cdot y_k^v(t - \Delta t_{feed}) \quad (17)$$

Here Δt_{feed} represents the latency of feedforward inputs between the unisensory and multisensory regions. W_{jk}^{ma} and W_{jk}^{mv} are the synapses connecting the pre-synaptic neuron at position k in a given unisensory area and the post-synaptic neuron at position j in the multisensory area.

The weights of these feedforward synapses are symmetrically defined as:

$$W_{jk}^{mc} = W_0^{mc} \cdot exp\left(-\frac{(D_{jk})^2}{2(\sigma^{mc})^2} \right), c = a, v \quad (18)$$

Here W_0^{mc} denotes the highest value of synaptic efficacy, D_{jk} the distance between the multisensory neuron at position j and the unisensory neuron at position k, and σ^{mc} the width of the feedforward synapses.

These external sources are also filtered by a second order differential equation:

$$\begin{cases} \frac{d}{dt} o_j^m(t) = \delta_j^m(t) \\ \frac{d}{dt} \delta_j^m(t) = \frac{G^m}{\tau^m} \cdot \left[i_j^m(t) \right] - \frac{2 \cdot \delta_j^m(t)}{\tau^m} - \frac{o_j^m(t)}{(\tau^m)^2} \end{cases} \quad (19)$$

Here, G^m represents gain and τ^m the time constants of the dynamics in the multisensory neurons.

B Values of L-SPQ model parameters

Neurons			
N = 30	θ = 20	s = 0.3	τ = 1
External stimuli			
E_0^a = 2.425	σ^a = 4°	E_0^v = 1.375	σ^v = 32°
Input dynamics			
$G^{a,v,m}$ = e^1	τ^a = 6.71	τ^v = 23.56	τ^m=117.5
Δt_{cross} = 16 ms	Δt_{feed} = 95 ms		
Lateral synapses			
L_{0ex}^c = 1.6	L_{0in}^c = 1.8	σ_{ex}^c = 3°	σ_{in}^c = 24°
Inter-areal synapses			
W_0^{mc} = 2.461	σ^{mc} = 0.5°	B_0^{cm} = 0.242	σ^{cm} = 0.5°
$W_0^{av,va}$ = 0.013	$\sigma^{av,va}$ = 5°		

TABLE 1

Values of the parameters in the L-SPQ model. This parameterisation remained fixed throughout the simulations computed to identify the H-SPQ model.

C Fitting procedure

The probability of perceiving two flashes in a single trial $P(2f)$ was defined as the product of the two peak values above .15 read from visual neurons. We used a fitting procedure with the cost function defined by Equation 20.

$$Cost = \sum_{i=1}^{N} \left(\frac{P(2f)_i^{data} - P(2f)_i^{model}}{P(2f)_i^{data}} \right)^2 \tag{20}$$

Here, $P(2f)_i^{data}$ and $P(2f)_i^{model}$ denote the P (2f) measured in the *i*th stimulus onset

asynchrony (SOA). N represents the number of SOAs measured (i.e. 15 in the empirical study).

This cost function was minimised by the implementation of the differential evolution algorithm [106] available in the SciPy v1.11.4 library for the Python programming language [107].

$P(2f)^{model}$ is composed of the 15 $P(2f)$ generated by the model at each SOA defined in the experimental task. Similarly, $P(2f)^{data}$ is composed of 15 proportion values taken from the mean sigmoid fit obtained out of the experimental data of a given group. We assumed that the percentage of perceived second illusory flash reported in the original study [24] is equivalent to the probability of perceiving two flashes in each condition.

C.1 L-SPQ Fitting

We fitted the model to the L-SPQ group data. Here, W_0^{mc}, $W_0^{av,va}$, B_0^{cm}, τ^a, τ^v and τ^m (see Equations 18, 13, 15 16 and 19) varied freely to represent the group proneness to the illusion due to stable anatomical factors (i.e. immutable throughout the duration of the experiment). All other parameters (see Table 1) remained fixed and set to the values published by [36, 45]. The bounds given to the optimisation algorithm to find $W_0^{av,va}$, B_0^{cm}, W_0^{mc}, τ^a, τ^v and τ^m were (0.001, 0.25), (0.001, 1.2), (1, 6), (6.5, 15), (9, 60) and (6.5, 120) respectively. The results of the fit to the data of the L-SPQ group are presented in Table 1.

C.2 H-SPQ Fitting

We fitted the L-SPQ model to the H-SPQ data. The bounds given to the optimisation algorithm to find L_{0ex}^c and $W_0^{av,va}$ were (1.5, 1.7) and (0, 0.06) respectively. These limits were chosen because the response profile generated by values outside of this range ceases to resemble a sigmoid curve, which is characteristic of the average response profile found in the empirical study [24].