Increased excitation enhances the sound-induced flash illusion by impairing multisensory causal inference in the schizophrenia spectrum

Renato Paredes[a,b], Francesca Ferri[c], Vincenzo Romei[d], Peggy Seriès[e,*]

[a] *Department of Psychology, Pontifical Catholic University of Peru, Av. Universitaria 1801, San Miguel, 15088, Lima, Peru*
[b] *Instituto de Investigaciones Psicológicas, Facultad de Psicología, Universidad Nacional de Córdoba-CONICET, Enf. Gordillo esq. Enrique Barros, 3er piso, Cdad. Univ. UNC, X5000, Córdoba, Argentina*
[c] *Department of Neuroscience, Imaging and Clinical Sciences, University of Chieti-Pescara, Via dei Vestini 31, Chieti, 66100, Italy*
[d] *Consciousness Group, Centre for Studies and Research in Cognitive Neuroscience, Department of Psychology "Renzo Canestrari", University of Bologna, Via Zamboni 33, Bologna, 40126, Italy*
[e] *Institute for Adaptive and Neural Computation, University of Edinburgh, 10 Crichton Street, EH8 9AB, Edinburgh, UK*

Abstract

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. 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 recurrent excitation, cross-modal and feedback weights, and synaptic density on the visual illusory responses of the network. Using quantitative fitting to empirical data, we found that an increase in the

*Corresponding author
Email address: pseries@inf.ed.ac.uk (Peggy Seriès)

weights of the recurrent excitatory 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 excitation increases the probability of inferring a common cause from the stimuli. 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

Individuals with schizophrenia show a disturbed sense of self, with impairments in somatosensory processing such as tactile sensitivity (Chang and Lenzenweger, 2001, 2004, 2005) and proprioception (Michael and Park, 2016; Thakkar et al., 2011). They also have deficits in higher-order processes essential for a coherent sense of self, like body ownership (Rossetti et al., 2020; Costantini et al., 2020; Zopf et al., 2021; He et al., 2022), bodily-self boundary (Park et al., 2009; Holt et al., 2015; Di Cosmo et al., 2018, 2021), and self-recognition (Ferroni et al., 2019; Sandsten et al., 2020). These differences often stem from disruptions in multisensory integration, the neural process that combines sensory information to form a unified perceptual experience (Stein and Stanford, 2008; Stein et al., 2010). Effective interaction with the world depends on integrating vestibular, proprioceptive, and interoceptive information with external sensory data, making multisensory integration crucial for self-coherence and consciousness (Noel et al., 2018b; Serino, 2019).

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 (Zvyagintsev et al., 2017), the McGurk effect (Martin et al., 2013) and the double flash illusion (Haß et al., 2017). These differences have also been observed in high schizotypal individuals for the McGurk (Muller et al., 2021) and the double flash illusion (Ferri et al., 2018). Despite the amount of empirical evidence available, there is as yet no theoretical understanding of why multisensory integration is altered 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 (Ferri et al., 2017; Zhou et al., 2018; Dalal et al., 2021; Di Cosmo et al., 2021). 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 (Zhou et al., 2018; Dalal et al., 2021; Di Cosmo et al., 2021; Fotia et al., 2021).

In the double flash illusion (DFI) paradigm, schizophrenia patients more frequently perceived an illusory flash and had a broader temporal window of integration (TWI) for auditory and visual stimuli. This is likely due to the increased cross-modal influence of the second auditory input onto the representation of the visual input, which is then perceived to be repeated, even at longer sound onset asynchronies (Haß et al., 2017; Cooke et al., 2019). Similarly, high-schizotypy individuals showed an enlarged TWI and greater susceptibility to the illusion than those with low schizotypy, with both measures positively correlated (Ferri et al., 2018).

In general, audiovisual illusions have been consistently accounted for in healthy individuals with causal inference models of multisensory integration (Cao et al., 2019; Kayser and Shams, 2015; Magnotti et al., 2018; Mohl et al., 2020; Rohe et al., 2019). These models mainly use the Bayesian inference framework (Körding et al., 2007), which provides a high-level description (i.e. computational level of analysis according to Marr (2010)) of the computations carried out by the brain to integrate unisensory signals. To our knowledge, biologically plausible neural network models of multisensory causal inference (Cuppini et al., 2017; Fang et al., 2019; Rideaux et al., 2021) have not yet been used to explore the neural mechanisms behind the proneness to experience audiovisual illusions in individuals in the schizophrenia spectrum.

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 of how an enlarged TBW (Ferri et al., 2017; Zhou et al., 2018; Dalal et al., 2021; Di Cosmo et al., 2021) is related to failures in multisensory integration/segregation and causal inference in the schizophrenia spectrum (for a current debate, see Samaha and Romei (2024) and Schoffelen et al. (2024)).

Moreover, it is unclear how impairments in these neural processes are related to bodily self-aberrations (Costantini et al., 2020; Zopf et al., 2021; Holt et al., 2015; Di Cosmo et al., 2018; Sandsten et al., 2020) and symptoms of schizophrenia and schizotypal traits.

To this end, we build on recent advances in multisensory causal inference network models in healthy individuals (Cuppini et al., 2017) 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 the DFI (Hirst et al., 2020). In particular, we evaluate the hypothesis that an imbalance of E/I in sensory networks increases temporal binding windows (Ferri et al., 2017), which in turn alters causal inference.

2. Methods

2.1. Experimental task

Our computational model simulates the double flash illusion (DFI) experiment carried out in individuals with high schizotypy (Ferri et al., 2018) (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 Ferri et al. (2018) 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 (Cecere et al., 2015; Shams et al., 2002; Cooke et al., 2019). 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) (Ferri et al., 2018).

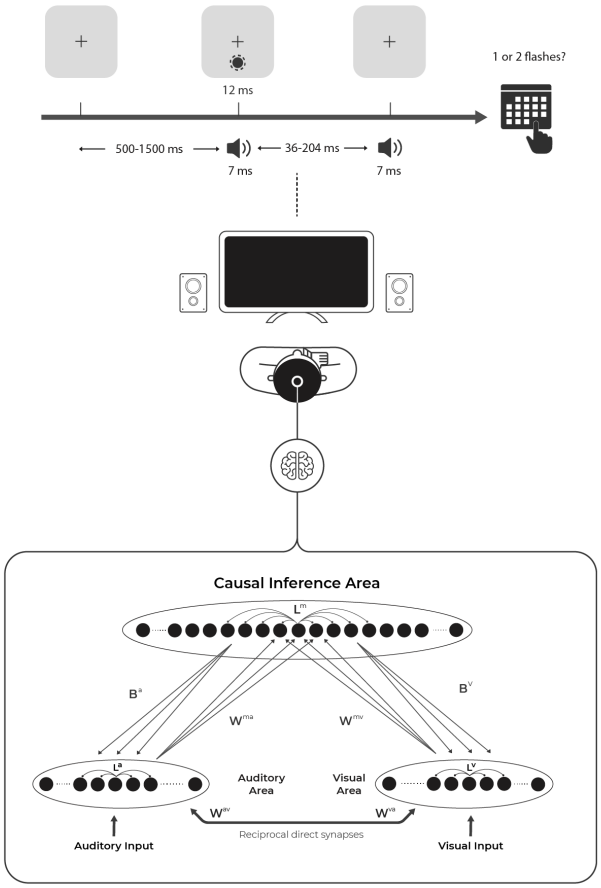

Figure 1: Double flash illusion experimental paradigm and Multisensory Causal Inference network model. In this experimental setup, two speakers and a monitor are placed in front of the participant. The black arrow indicates the evolution of the presentation of stimuli in time: one visual flash (lasting 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.

2.2. Multisensory Causal Inference network model of DFI

We model the DFI by extending the audiovisual network model proposed by Cuppini et al. (2014). Following Cuppini et al. (2017), the main modification is the incorporation of a multisensory causal inference layer with feedforward and feedback connectivity to unisensory layers. This multisensory layer implements a mechanism for causal inference by determining the likelihood that the presented sensory stimuli (flash and beeps) originate from a common source versus independent sources. 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° 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). Our network uses a "minimal" spatial representation, extending beyond the stimulus to incorporate recurrent connectivity and to be able to capture spatiotemporal relationships.

Our model includes cross-modal connectivity, as supported by modelling studies on audio-visual integration (Cuppini et al., 2014, 2017; Ursino et al., 2019) and empirical evidence obtained from research on primates (Eckert et al., 2008) and humans (Werner and Noppeney, 2010; Gurtubay-Antolin et al., 2021). 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 Information for an extensive mathematical description of the model).

Following Rohe and Noppeney (2015b), our network model incorporates two consecutive hierarchical levels of multisensory processing. In the unisensory areas, the model computes the spatiotemporal location of external stim-

uli (solving the so-called "implicit"[1] causal inference problem). In the multisensory area, the model evaluates whether stimuli originate from a single source (solving the "explicit" causal inference problem), mimicking the responses of neurons in parietal-temporal association cortices. The network also features feedback connectivity, enabling iterative refinement of causal inferences across the network (Rohe et al., 2019; Cao et al., 2019).

Our model does not account for any noise in the representation of the inputs or neural activities. We did not conduct an exhaustive comparison of all conceivable network architectures that could potentially explain the DFI, as such an analysis lies beyond the scope of this study. To test whether the causal inference area was necessary in the model, we simulated an architecture without it and found that it could not account for response performances (see dashed lines in Figure 4A). An extensive description of the model and values of the parameters can be found in the Supplementary Information.

2.3. Experiment simulation in the network

In simulating audiovisual stimuli in the DFI, the network receives Gaussian 1-D inputs centred in the encoded space. Visual inputs last 12 ms, auditory 7 ms. Visual and first auditory stimuli start 16 ms after trial onset. The second auditory stimulus starts at 15 delays ranging from 36 to 204 ms after the first stimulus, with the network running for 600 ms for each delay. Given the deterministic nature of the model, each experimental condition is only simulated once.

We model DFI as a multisensory integration process with early and late cross-modal interactions in distinct cortical hierarchies (Hirst et al., 2020; Keil, 2020). 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 (Shams et al., 2005; Raij et al., 2010), and functionally supported by TMS and EEG evidence (Romei et al., 2007; Cappe et al., 2010; Romei et al., 2012). Next, a delayed interaction arises from the input of feedforward-feedback connectivity with higher cortical regions (e.g. angular gyrus and the superior temporal sulcus) observed in the 130-600 ms post-stimulus range (Mishra et al., 2007; Balz et al., 2016a; Keil and Senkowski, 2018; Keil,

[1]Explicit and implicit causal inference refers to tasks in which participants have to directly evaluate the causal relationship of the stimuli (unity judgement) or indirectly by providing estimates of their spatiotemporal location (Acerbi et al., 2018).

2020). Critically, brain activity consistent with Bayesian Causal Inference[2] has been observed in anterior parietal areas in the 200-400 ms range (Rohe et al., 2019).

To approximate these temporal dynamics in the network model, we consider a temporal delay of 16 ms for cross-modal interactions (as in Cuppini et al. (2014)) and 95 ms for feedforward and feedback interactions. We established these temporal delays to model the empirical observations that cross-modal delays (typically around 35-90 ms) are considerably lower than feedforward-feedback delays (typically around 130-600 ms) (Keil and Senkowski, 2018; Keil, 2020). Moreover, following Cuppini et al. (2014), 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 (Ferri et al., 2018) can be simulated by a principled readout of the unisensory 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 Cuppini et al. (2017), 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 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].

To reproduce the sigmoid fit of the average responses of the low schizotypy group (Ferri et al., 2018), 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 Information 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

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

[3]When more than two visual peaks exceed the threshold, the average of all possible product combinations of these peaks is calculated.

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 in time (lower SOA), the higher the combined feedforward signal, as a consequence of cross-modal neural input.

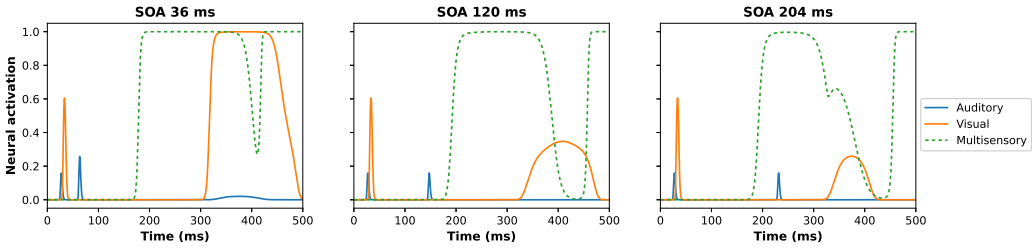

Figure 2: Base model (L-SPQ) dynamics. The network receives a 12 ms visual input and a 7 ms auditory input simultaneously, followed by a second auditory input at 15 different stimulus onset asynchronies (SOAs) from 36 to 204 ms. The figure shows the activity of auditory (blue line), visual (orange line) and multisensory (dashed green line) neurons. The auditory neuron shows two peaks of activity generated by the auditory stimuli received by the network. The visual neuron shows two peaks of activity: the first generated by the visual stimulus received by the network, and the second as a result of neural feedback. The multisensory neuron shows one activity peak at lower SOAs (36 ms) and two distinct peaks at larger SOAs (120 and 204 ms). This pattern results from feedforward input from auditory and visual neurons, where closer auditory stimuli increase the combined signal due to cross-modal neural interactions. Consequently, the greater and more sustained the activity of the multisensory neuron, the stronger the feedback input to the unisensory areas, resulting in an elevated peak of illusory visual activity. By design, feedforward input coming from unisensory activity is delayed by 95 ms before reaching the multisensory area.

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 that the visual input and

the second auditory input were caused by a single cause (peak activity values below this threshold were assigned a probability of zero). The probability of inferring a single cause from 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).

2.4. Modelling the influence of high schizotypy in the network

We evaluated the potential consequences of four observed impairments in the SCZ spectrum on the sound-induced flash illusion: excitation/inhibition (E/I) imbalance (Jardri et al., 2016; Leptourgos et al., 2022), early cross-modal processing deficits (Magnée et al., 2009; Balz et al., 2016b), failures in top-down signalling (Sterzer et al., 2018), and synaptic density decrease (Ellison-Wright and Bullmore, 2009).

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:

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

$$c = a, v, m$$

L^c_{jk} denotes the weight of the synapse from the pre-synaptic neuron at position k to post-synaptic neuron at position j within the area c (auditory, visual or multisensory). D^c_{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.

2.4.2. Cross-modal processing deficits

We modelled early cross-modal processing deficits by increasing or decreasing the weights of cross-modal connectivity. This was implemented by

manipulating the value of the parameter W_0^{cd} in both sensory modalities, defined in the equation 2:

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

W_0^{cd} 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 ($W_0^{av} = W_0^{va}$ and $\sigma^{av} = \sigma^{va}$) by the Gaussian function described above.

2.4.3. 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_0^c, defined in equation 3:

$$B_{jk}^c = B_0^c \cdot exp\left(-\frac{\left(D_{jk}^c\right)^2}{2 \cdot (\sigma^c)^2}\right) \tag{3}$$

Here, D_{jk}^c is the distance between the multisensory neuron at position j and the unisensory neuron at position k. B_0^c denote the value of the feedback when D_{jk}^c is equal to zero, representing the highest synaptic efficacy. σ^c represents the width of the feedback synapses.

2.4.4. Decrease of synaptic density

Following Hoffman and Dobscha (1989), Hoffman and McGlashan (2006) and Paredes et al. (2022), 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 the strength of lateral excitatory (L_{0ex}), feedback (B_0^c) and cross-modal (W_0^{cd}) weights, and feed-

forward ($\rho_{W^{mc}}$) 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 an increase in recurrent excitation weights enlarges the temporal window of illusion. Changes in feedback weights or pruning feedforward synapses had a weaker effect.

Similarly, we found that an increase in lateral excitatory 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 the pruning of feedforward 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 excitatory weights or pruning feedforward synapses increases the average probability of inferring a single common cause to the stimuli across all SOAs (see Figure 3C). Changes in cross-modal connectivity did not affect causal inference and are excluded from the figure.

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 (Ferri et al., 2018), we considered the Root Mean Square Error (RMSE) adjusted by the number of free parameters, as in Paredes et al. (2022).

The model that best fit the data from the H-SPQ group is presented in Figure 4A. Our findings indicate that increased lateral excitation in unisensory and multisensory areas matches H-SPQ responses in the DFI task (adj RMSE = 2.55). We discarded models with increased cross-modal (adj RMSE = 7.43) or feedback weights (adj RMSE = 5.89), or feedforward pruning (adj RMSE = 7.94). The model responses stay within participants' standard error range, except for SOAs under 75 ms (see Supplementary Information). The TWI from our L-SPQ and H-SPQ models approximate the observed data, as shown in Figure 4B.

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 (Ferri et al., 2018). For this purpose, an existing double flash

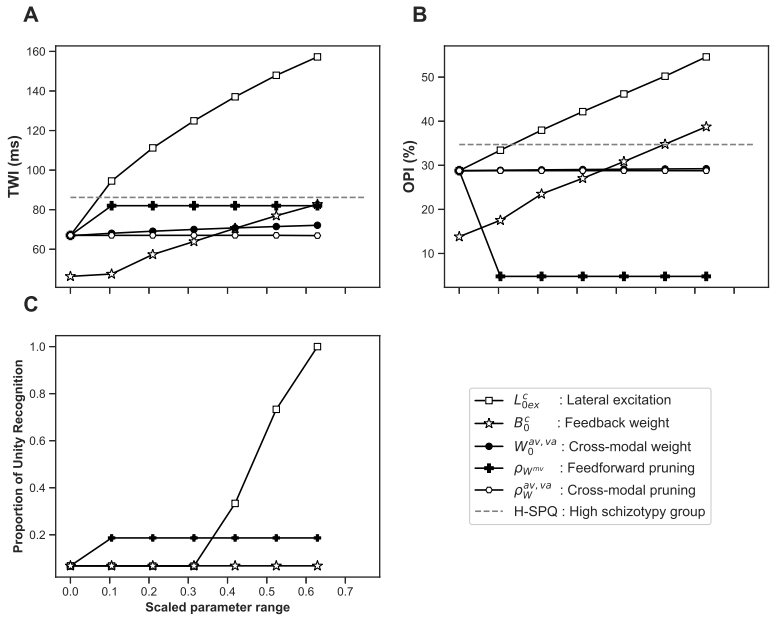

Figure 3: Effects of potential neural impairments in the temporal window of illusion, overall proneness to the illusion and causal inference in the network model. **Panels A**, **B** and **C** show the effect of systematic variation of parameters in the range at which they produce sensible responses 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 (Ferri et al., 2018). **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}). **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 excitatory (L^c_{0ex}) or feedback weights (B^c_0). **Panel C** indicates that both an increase in lateral excitation (L^c_{0ex}) weights or pruning feedforward synapses ($\rho_{W^{mc}}$) increases the overall probability of inferring a common cause to the presented stimuli.

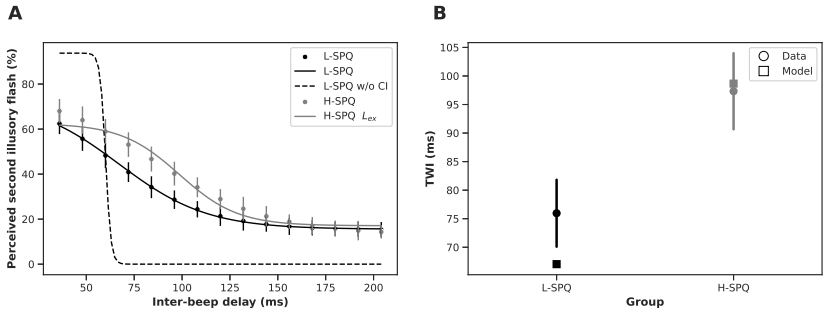

Figure 4: L-SPQ and H-SPQ network models identified by the fitting procedure. **Panel A** shows the DFI task responses generated by the L-SPQ and H-SPQ network models. The solid lines depict the sigmoid fit obtained from the data generated by the models. The dots represent the discretisation of the sigmoid fit obtained from the data collected in the experiment (Ferri et al., 2018). The quantitative fitting of the L-SPQ model (black) using L^c_{0ex} (grey) is sufficient to reproduce the H-SPQ data. The black dashed lines represent an alternative L-SPQ model without the causal inference area. **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 Ferri et al. (2018).

illusion audiovisual network model (Cuppini et al., 2014) was adapted to incorporate a multisensory layer with feedforward and feedback connectivity to unisensory layers, allowing iterative causal inference across the network (Rohe and Noppeney, 2015a; Rohe et al., 2019). 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 examined how E/I imbalance, cross-modal processing deficits, top-down signalling failures, and synaptic decrease affect illusory responses. These neural impairments align with NMDA receptor and GABA neuron dysfunctions, along with heightened D_2 receptor activity, crucial in schizophrenia models (Sterzer et al., 2018; Lanillos et al., 2020; Leptourgos et al., 2022). Analysing the experimental data (Section 3.2), we identified a neural mechanism related to a higher susceptibility to the double flash illusion in high schizotypy individuals: increased recurrent excitation in unisensory and multisensory neurons. Simulations (Section 3.1) show that this mechanism expands the illusion's temporal window (TWI) and raise overall proneness (OPI), also increasing the probability that the network infers a common cause from stimuli.

4.1. An E/I imbalance account of reduced temporal discrimination

Our results are broadly in line with magnetic resonance spectroscopy (MRS) studies that consistently show alterations in glutamatergic and GABAergic neurotransmission in the schizophrenia spectrum (Reddy-Thootkur et al., 2022). 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 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 (Ferri et al., 2017).

Nevertheless, our model reproduces the E/I balance perturbation but does not distinguish between specific impairments in excitation or inhibition (e.g. failures in GABAergic or glutamatergic receptors). The model could be expanded to account for differences in inhibition by adopting a modelling approach based on spiking neurons and the direct manipulation of NMDA and GABA input currents (e.g. Abbasi et al. (2023)). This would enable the possibility link our results 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 (Schormans and Allman, 2023). Here, wild-type rats received Gabazine (a GABA-A receptor antagonist) in the V2L cortex (an area known to process audiovisual stimuli) and evaluated during audiovisual stimulation and the Temporal Order Judgement (TOJ) task. 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.

Future work could also extend the model to better describe neural dynamics and explore how changes in E/I balance influence oscillatory patterns in multisensory integration areas (Fotia et al., 2021). Electrophysiological evidence has consistently shown a reduction in the power of gamma band activity in patients with SCZ (Leicht et al., 2010; Lenz et al., 2011; Leicht et al., 2015; Keil et al., 2016). 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 (Balz et al., 2016b). Such an approach could enable the computational understanding of how the concentration of GABA in STS is correlated with the

perception of DFI and influences the oscillatory activity of the gamma band (Balz et al., 2016a).

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 (Ferri et al., 2018; Haß et al., 2017). In line with Paredes et al. (2022), 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) (Fotia et al., 2021) and related spatiotemporal discrimination paradigms (Tseng et al., 2015; Ferri et al., 2016; Zhou et al., 2018; Dalal et al., 2021; Di Cosmo et al., 2021).

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 expand previous network modelling of audiovisual integration showing that cross-modal synapses encode prior probabilities of the co-occurrence of audiovisual stimuli in spatial localisation tasks (Ursino et al., 2017, 2019). This view aligns with Bayesian Causal Inference modelling of the Simultaneity Judgement task, indicating that temporal binding windows are influenced by an a priori bias to bind sensory information (Magnotti et al., 2013; Noel et al., 2018a; Chancel et al., 2022).

Future research could aim to link our results to Bayesian accounts of schizophrenia (Friston et al., 2016; Jardri et al., 2016; Sterzer et al., 2018; Tarasi et al., 2022a,b, 2023). A recent Bayesian model of the sound-induced flash illusion suggests that wider temporal binding windows and higher illusion rates would be a consequence of reduced precision in unisensory modalities (Zhu et al., 2024). However, it is unclear how our findings on increased excitation in recurrent connectivity within the multisensory integration network align with Bayesian accounts of psychotic symptoms: strong priors (Teufel et al., 2015; Powers et al., 2017; Cassidy et al., 2018); weaker or variable priors (Nazimek et al., 2012; Samaha and Romei, 2024; Schmack et al., 2015, 2017; Jardri et al., 2017; Valton et al., 2019; Fletcher and Teufel, 2022; Goodwin et al., 2023); stronger priors at higher sensory processing levels with weaker priors at lower levels (Yang et al., 2016; Petrovic and Sterzer, 2023); or circular inference in multisensory integration (Leptourgos et al., 2022). We

encourage direct evaluation of the activity of the unisensory and multisensory areas during sound-induced flash illusion in patients with schizophrenia to identify whether aberrations in top-down or bottom-up processing drive the increased proneness to 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 (French and DeAngelis, 2020). We showed at the network level how an increase in excitation in recurrent connectivity potentially increases the overall probability of inferring a common cause from sensory stimuli (see Figure 3C).

Speculatively, our predictions about multisensory causal inference impairments due to increased excitation could be related to bodily self-aberrations observed in the SCZ spectrum (Klaver and Dijkerman, 2016; Michael and Park, 2016; Sandsten et al., 2020; Di Cosmo et al., 2021). 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 (Rossetti et al., 2020; Costantini et al., 2020; Zopf et al., 2021; He et al., 2022).

We tentatively suggest that an increase in excitatory weights of recurrent synapses within multisensory integration networks increases susceptibility to RHI and bodily perceptual symptoms observed in the SCZ spectrum (Laurin et al., 2021; He et al., 2022; Torregrossa and Park, 2022). We encourage further research into the neurobiology of causal inference in SCZ during the RHI and other multisensory illusions (Klaver and Dijkerman, 2016; Michael and Park, 2016; Ferroni et al., 2019; Sandsten et al., 2020) to better understand bodily self-disturbances in patients with SCZ.

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 (Balz et al., 2016a,b; Fotia et al., 2021). Moreover, our network does not accurately reproduce the responses observed at lower inter-beep delays (see Supplementary Information). At best, the network model presented here serves as a tool for theoretical reasoning that can be incrementally improved for biological realism (Guest and Martin, 2023).

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 (Keil, 2020) and specific impairments in selective attention and cognitive control have been observed in the schizophrenia spectrum (Gold et al., 2007; Lesh et al., 2011). Furthermore, neuroimaging evidence suggests a critical role for fronto-temporal regions in multisensory integration impairments observed in SCZ (Gröhn et al., 2022; Leptourgos et al., 2022).

The experimental evidence directly evaluating sound-induced flash illusions in the SCZ spectrum has only begun to accumulate (Vanes et al., 2016; Haß et al., 2017; Balz et al., 2016b; Ferri et al., 2018). This study focuses on temporal binding without ruling out other factors contributing to temporal distortions in schizophrenia, like context integration gain (Cohen et al., 1999) or prediction delay mechanisms (Whitford et al., 2012; Okimura et al., 2023). Our findings are based on modelling a study with healthy individuals without evaluating causal inference (Ferri et al., 2018). Various experimental paradigms explore sound-induced flash illusions with different stimuli and spatiotemporal manipulations (Hirst et al., 2020; Keil, 2020). These paradigms remain underexplored in the SCZ spectrum, complicating the understanding of illusion proneness in SCZ (Vanes et al., 2016). Further studies with diverse stimuli and clear causal inference and spatiotemporal sensitivity measurements in the SCZ spectrum are needed (Klaver and Dijkerman, 2016; Michael and Park, 2016; Ferroni et al., 2019; Sandsten et al., 2020).

5. 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).

6. CRediT authorship contribution statement

Renato Paredes: Conceptualisation, Methodology, Software, Formal Analysis, Writing - Original Draft. **Francesca Ferri:** Conceptualisation, Investigation, Writing - Review and Editing. **Vincenzo Romei:** Conceptualisation, Investigation, Writing - Review and Editing. **Peggy Seriès:** Conceptualisation, Methodology, Supervision, Writing - Reviewing and Editing.

7. Declaration of competing interest

None.

8. Availability of data and materials

The model was implemented using the Scikit-NeuroMSI framework for modelling multisensory integration (Paredes et al., 2023). The code used to produce the simulations presented in this manuscript can be found at: https://github.com/renatoparedes/SCZ_DFI_model.

9. Acknowledgements

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

References

Abbasi, S., Wolff, A., Çatal, Y., Northoff, G., 2023. Increased noise relates to abnormal excitation-inhibition balance in schizophrenia: a combined empirical and computational study. Cereb. Cortex 33, 10477–10491. doi:10.1093/cercor/bhad297.

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. doi:10.1371/journal.pcbi.1006110.

Balz, J., Keil, J., Roa Romero, Y., Mekle, R., Schubert, F., Aydin, S., Ittermann, B., Gallinat, J., Senkowski, D., 2016a. GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. Neuroimage 125, 724–730. doi:10.1016/j.neuroimage.2015.10.087.

Balz, J., Roa Romero, Y., Keil, J., Krebber, M., Niedeggen, M., Gallinat, J., Senkowski, D., 2016b. Beta/Gamma Oscillations and Event-Related Potentials Indicate Aberrant Multisensory Processing in Schizophrenia. Front. Psychol. 7. doi:10.3389/fpsyg.2016.01896.

Cao, Y., Summerfield, C., Park, H., Giordano, B.L., Kayser, C., 2019. Causal Inference in the Multisensory Brain. Neuron 102, 1076–1087.e8. doi:10.1016/j.neuron.2019.03.043.

Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. J. Neurosci. 30, 12572–12580. doi:10.1523/JNEUROSCI.1099-10.2010.

Cassidy, C.M., Balsam, P.D., Weinstein, J.J., Rosengard, R.J., Slifstein, M., Daw, N.D., Abi-Dargham, A., Horga, G., 2018. A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. Curr. Biol. 28, 503–514.e4. doi:10.1016/j.cub.2017.12.059.

Cecere, R., Rees, G., Romei, V., 2015. Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. Curr. Biol. 25, 231–235. doi:10.1016/j.cub.2014.11.034.

Chancel, M., Ehrsson, H.H., Ma, W.J., 2022. Uncertainty-based inference of a common cause for body ownership. eLife 11, e77221. doi:10.7554/eLife.77221.

Chang, B.P., Lenzenweger, M.F., 2001. Somatosensory processing in the biological relatives of schizophrenia patients: a signal detection analysis of two-point discrimination. J. Abnorm. Psychol. 110, 433–442. doi:10.1037//0021-843x.110.3.433.

Chang, B.P., Lenzenweger, M.F., 2004. Investigating Graphesthesia Task Performance in the Biological Relatives of Schizophrenia Patients. Schizophr. Bull. 30, 327–334. doi:10.1093/oxfordjournals.schbul.a007082.

Chang, B.P., Lenzenweger, M.F., 2005. Somatosensory processing and schizophrenia liability: proprioception, exteroceptive sensitivity, and graphesthesia performance in the biological relatives of schizophrenia patients. J. Abnorm. Psychol. 114, 85–95. doi:10.1037/0021-843X.114.1.85.

Cohen, J.D., Barch, D.M., Carter, C., Servan-Schreiber, D., 1999. Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. J. Abnorm. Psychol. 108, 120. doi:10.1037//0021-843x.108.1.120.

Cooke, J., Poch, C., Gillmeister, H., Costantini, M., Romei, V., 2019. Oscillatory properties of functional connections between sensory areas mediate cross-modal illusory perception. J. Neurosci. 39, 5711–5718.

Costantini, M., Salone, A., Martinotti, G., Fiori, F., Fotia, F., Di Giannantonio, M., Ferri, F., 2020. Body representations and basic symptoms in schizophrenia. Schizophr. Res. 222, 267–273. doi:10.1016/j.schres.2020.05.038.

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G., Ursino, M., 2014. A neurocomputational analysis of the sound-induced flash illusion. Neuroimage 92, 248–266. doi:10.1016/j.neuroimage.2014.02.001.

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. doi:10.1111/ejn.13725.

Dalal, T.C., Muller, A.M., Stevenson, R.A., 2021. The relationship between multisensory temporal processing and schizotypal traits. Multisens. Res. 34, 511–529. doi:10.1163/22134808-bja10044.

Di Cosmo, G., Costantini, M., Ambrosini, E., Salone, A., Martinotti, G., Corbo, M., Di Giannantonio, M., Ferri, F., 2021. Body-environment integration: Temporal processing of tactile and auditory inputs along the schizophrenia continuum. J. Psychiatr. Res. 134, 208–214. doi:10.1016/j.jpsychires.2020.12.034.

Di Cosmo, G., Costantini, M., Salone, A., Martinotti, G., Di Iorio, G., Di Giannantonio, M., Ferri, F., 2018. Peripersonal space boundary in schizotypy and schizophrenia. Schizophr. Res. 197, 589–590. doi:10.1016/j.schres.2017.12.003.

Eckert, M.A., Kamdar, N.V., Chang, C.E., Beckmann, C.F., Greicius, M.D., Menon, V., 2008. A cross-modal system linking primary auditory and visual cortices: Evidence from intrinsic fMRI connectivity analysis. Hum. Brain Mapp. 29, 848–857. doi:10.1002/hbm.20560.

Ellison-Wright, I., Bullmore, E., 2009. Meta-analysis of diffusion tensor imaging studies in schizophrenia. Schizophr. Res. 108, 3–10. doi:10.1016/j.schres.2008.11.021.

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

Ferri, F., Ambrosini, E., Costantini, M., 2016. Spatiotemporal processing of somatosensory stimuli in schizotypy. Sci. Rep. 6, 38735. doi:10.1038/srep38735.

Ferri, F., Nikolova, Y.S., Perrucci, M.G., Costantini, M., Ferretti, A., Gatta, V., Huang, Z., Edden, R.A.E., Yue, Q., D'Aurora, M., Sibille, E., Stuppia, L., Romani, G.L., Northoff, G., 2017. A Neural "Tuning Curve" for Multisensory Experience and Cognitive-Perceptual Schizotypy. Schizophr. Bull. 43, 801–813. doi:10.1093/schbul/sbw174.

Ferri, F., Venskus, A., Fotia, F., Cooke, J., Romei, V., 2018. Higher proneness to multisensory illusions is driven by reduced temporal sensitivity in people with high schizotypal traits. Conscious. Cogn. 65, 263–270. doi:10.1016/j.concog.2018.09.006.

Ferroni, F., Ardizzi, M., Sestito, M., Lucarini, V., Daniel, B., Paraboschi, F., Tonna, M., Marchesi, C., Gallese, V., 2019. Shared multisensory experience affects Others' boundary: The enfacement illusion in schizophrenia. Schizophrenia Research 206, 225–235. doi:10.1016/j.schres.2018.11.018.

Fletcher, P.C., Teufel, C.R., 2022. The Changing Weight of Expectation: How Shifting Priors Underpin Variability in Hallucination Frequency. Biol. Psychiatr. 92, 752–753. doi:10.1016/j.biopsych.2022.08.027.

Fotia, F., Cooke, J., Van Dam, L., Ferri, F., Romei, V., 2021. The temporal sensitivity to the tactile-induced double flash illusion mediates the impact of beta oscillations on schizotypal personality traits. Conscious. Cogn. 91, 103121. doi:10.1016/j.concog.2021.103121.

French, R.L., DeAngelis, G.C., 2020. Multisensory neural processing: from cue integration to causal inference. Curr. Opin. Physiol. 16, 8–13. doi:10.1016/j.cophys.2020.04.004.

Friston, K., Brown, H.R., Siemerkus, J., Stephan, K.E., 2016. The dysconnection hypothesis (2016). Schizophr. Res. 176, 83–94. doi:10.1016/j.schres.2016.07.014.

Gold, J.M., Fuller, R.L., Robinson, B.M., Braun, E.L., Luck, S.J., 2007. Impaired top–down control of visual search in schizophrenia. Schizophr. Res. 94, 148–155. doi:10.1016/j.schres.2007.04.023.

Goodwin, I., Kugel, J., Hester, R., Garrido, M.I., 2023. Bayesian accounts of perceptual decisions in the nonclinical continuum of psychosis: Greater imprecision in both top-down and bottom-up processes. PLOS Comput. Biol. 19, e1011670. doi:10.1371/journal.pcbi.1011670.

Gröhn, C., Norgren, E., Eriksson, L., 2022. A systematic review of the neural correlates of multisensory integration in schizophrenia. Schizophr. Res. Cogn. 27, 100219. doi:10.1016/j.scog.2021.100219.

Guest, O., Martin, A.E., 2023. On Logical Inference over Brains, Behaviour, and Artificial Neural Networks. Comput. Brain Behav. 6, 213–227. doi:10.1007/s42113-022-00166-x.

Gurtubay-Antolin, A., Battal, C., Maffei, C., Rezk, M., Mattioni, S., Jovicich, J., Collignon, O., 2021. Direct structural connections between auditory and visual motion-selective regions in humans. Journal of Neuroscience 41, 2393–2405.

Haß, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., Dillo, W., Oranje, B., Szycik, G.R., 2017. Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. Cogn. Neuropsychiatry 22, 145–158. doi:10.1080/13546805.2017.1287693.

He, J., Ren, H., Li, J., Dong, M., Dai, L., Li, Z., Miao, Y., Li, Y., Tan, P., Gu, L., Chen, X., Tang, J., 2022. Deficits in Sense of Body Ownership, Sensory Processing, and Temporal Perception in Schizophrenia Patients With/Without Auditory Verbal Hallucinations. Front. Neurosci. 16, 831714. doi:10.3389/fnins.2022.831714.

Hirst, R.J., McGovern, D.P., Setti, A., Shams, L., Newell, F.N., 2020. What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. Neurosci. Biobehav. Rev. 118, 759–774. doi:10.1016/j.neubiorev.2020.09.006.

Hoffman, R.E., Dobscha, S.K., 1989. Cortical Pruning and the Development of Schizophrenia: A Computer Model. Schizophr. Bull. 15, 477–490. doi:10.1093/schbul/15.3.477.

Hoffman, R.E., McGlashan, T.H., 2006. Using a Speech Perception Neural Network Computer Simulation to Contrast Neuroanatomic versus Neuromodulatory Models of Auditory Hallucinations. Pharmacopsychiatry 39, 54–64. doi:10.1055/s-2006-931496.

Holt, D.J., Boeke, E.A., Coombs, G., DeCross, S.N., Cassidy, B.S., Stufflebeam, S., Rauch, S.L., Tootell, R.B., 2015. Abnormalities in personal space and parietal–frontal function in schizophrenia. NeuroImage Clin. 9, 233–243. doi:10.1016/j.nicl.2015.07.008.

Jardri, R., Duverne, S., Litvinova, A.S., Denève, S., 2017. Experimental evidence for circular inference in schizophrenia. Nat. Commun. 8, 14218. doi:10.1038/ncomms14218.

Jardri, R., Hugdahl, K., Hughes, M., Brunelin, J., Waters, F., Alderson-Day, B., Smailes, D., Sterzer, P., Corlett, P.R., Leptourgos, P., Debbané, M., Cachia, A., Denève, S., 2016. Are Hallucinations Due to an Imbalance Between Excitatory and Inhibitory Influences on the Brain? Schizophr. Bull. 42, 1124–1134. doi:10.1093/schbul/sbw075.

Kayser, C., Shams, L., 2015. Multisensory Causal Inference in the Brain. PLOS Biol. 13, e1002075. doi:10.1371/journal.pbio.1002075.

Keil, J., 2020. Double Flash Illusions: Current Findings and Future Directions. Front. Neurosci. 14, 298. doi:10.3389/fnins.2020.00298.

Keil, J., Roa Romero, Y., Balz, J., Henjes, M., Senkowski, D., 2016. Positive and Negative Symptoms in Schizophrenia Relate to Distinct Oscillatory Signatures of Sensory Gating. Front. Hum. Neurosci. 10. doi:10.3389/fnhum.2016.00104.

Keil, J., Senkowski, D., 2018. Neural Oscillations Orchestrate Multisensory Processing. Neuroscientist 24, 609–626. doi:10.1177/1073858418755352.

Klaver, M., Dijkerman, H.C., 2016. Bodily Experience in Schizophrenia: Factors Underlying a Disturbed Sense of Body Ownership. Front. Hum. Neurosci. 10. doi:10.3389/fnhum.2016.00305.

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. doi:10.1371/journal.pone.0000943.

Lanillos, P., Oliva, D., Philippsen, A., Yamashita, Y., Nagai, Y., Cheng, G., 2020. A review on neural network models of schizophrenia and autism spectrum disorder. Neural Netw. 122, 338–363. doi:10.1016/j.neunet.2019.10.014.

Laurin, A., Ramoz, N., Ameller, A., Dereux, A., Zajac, J., Bonjour, M., Tebeka, S., Le Strat, Y., Dubertret, C., 2021. Self-consciousness impairments in schizophrenia with and without first rank symptoms using the moving rubber hand illusion. Conscious. Cogn. 93, 103154. doi:10.1016/j.concog.2021.103154.

Leicht, G., Andreou, C., Polomac, N., Lanig, C., Schöttle, D., Lambert, M., Mulert, C., 2015. Reduced auditory evoked gamma band response and cognitive processing deficits in first episode schizophrenia. World J. Biol. Psychiatry 16, 387–397. doi:10.3109/15622975.2015.1017605.

Leicht, G., Kirsch, V., Giegling, I., Karch, S., Hantschk, I., Möller, H.J., Pogarell, O., Hegerl, U., Rujescu, D., Mulert, C., 2010. Reduced Early Auditory Evoked Gamma-Band Response in Patients with Schizophrenia. Biol. Psychiatry 67, 224–231. doi:10.1016/j.biopsych.2009.07.033.

Lenz, D., Fischer, S., Schadow, J., Bogerts, B., Herrmann, C.S., 2011. Altered evoked gamma-band responses as a neurophysiological marker of schizophrenia? Int. J. Psychophysiol. 79, 25–31. doi:10.1016/j.ijpsycho.2010.08.002.

Leptourgos, P., Bouttier, V., Denève, S., Jardri, R., 2022. From hallucinations to synaesthesia: A circular inference account of unimodal and multimodal erroneous percepts in clinical and drug-induced psychosis. Neurosci. Biobehav. Rev. 135, 104593. doi:10.1016/j.neubiorev.2022.104593.

Lesh, T.A., Niendam, T.A., Minzenberg, M.J., Carter, C.S., 2011. Cognitive Control Deficits in Schizophrenia: Mechanisms and Meaning. Neuropsychopharmacol. 36, 316–338. doi:10.1038/npp.2010.156.

Magnée, M.J., Oranje, B., van Engeland, H., Kahn, R.S., Kemner, C., 2009. Cross-sensory gating in schizophrenia and autism spectrum disorder: Eeg

evidence for impaired brain connectivity? Neuropsychologia 47, 1728–1732.

Magnotti, J.F., Ma, W.J., Beauchamp, M.S., 2013. Causal inference of asynchronous audiovisual speech. Front. Psychol. 4. doi: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. doi:10.1038/s41598-018-36772-8.

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

Martin, B., Giersch, A., Huron, C., van Wassenhove, V., 2013. Temporal event structure and timing in schizophrenia: Preserved binding in a longer "now". Neuropsychologia. 51, 358–371. doi:10.1016/j.neuropsychologia.2012.07.002.

Michael, J., Park, S., 2016. Anomalous bodily experiences and perceived social isolation in schizophrenia: An extension of the Social Deafferentation Hypothesis. Schizophr. Res. 176, 392–397. doi:10.1016/j.schres.2016.06.013.

Mishra, J., Martinez, A., Sejnowski, T.J., Hillyard, S.A., 2007. Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. J. Neurosci. 27, 4120–4131. doi:10.1523/JNEUROSCI.4912-06.2007.

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. doi:10.1152/jn.00046.2020.

Muller, A.M., Dalal, T.C., Stevenson, R.A., 2021. Schizotypal personality traits and multisensory integration: An investigation using the McGurk effect. Acta Psychol. 218, 103354. doi:10.1016/j.actpsy.2021.103354.

Nazimek, J., Hunter, M., Woodruff, P., 2012. Auditory hallucinations: Expectation–perception model. Med. Hypotheses 78, 802–810. doi:10.1016/j.mehy.2012.03.014.

Noel, J., Stevenson, R.A., Wallace, M.T., 2018a. Atypical audiovisual temporal function in autism and schizophrenia: similar phenotype, different cause. Eur. J. Neurosci. 47, 1230–1241. doi:10.1111/ejn.13911.

Noel, J.P., Blanke, O., Serino, A., 2018b. From multisensory integration in peripersonal space to bodily self-consciousness: from statistical regularities to statistical inference: Multisensory integration and self-consciousness. Ann. N. Y. Acad. Sci. 1426, 146–165. doi:10.1111/nyas.13867.

Okimura, T., Maeda, T., Mimura, M., Yamashita, Y., 2023. Aberrant sense of agency induced by delayed prediction signals in schizophrenia: a computational modeling study. Schizophrenia 9, 72. doi:10.1038/s41537-023-00403-7.

Paredes, R., Ferri, F., Seriès, P., 2022. Influence of E/I balance and pruning in peri-personal space differences in schizophrenia: A computational approach. Schizophr. Res. 248, 368–377. doi:10.1016/j.schres.2021.06.026.

Paredes, R., Seriès, P., Cabral, J., 2023. Scikit-NeuroMSI: a Python framework for multisensory integration modelling. IX Congreso de Matemática Aplicada, Computacional e Industrial 9, 545–548.

Park, S.H., Ku, J., Kim, J.J., Jang, H.J., Kim, S.Y., Kim, S.H., Kim, C.H., Lee, H., Kim, I.Y., Kim, S.I., 2009. Increased personal space of patients with schizophrenia in a virtual social environment. Psychiatry Res. 169, 197–202. doi:10.1016/j.psychres.2008.06.039.

Petrovic, P., Sterzer, P., 2023. Resolving the Delusion Paradox. Schizophr. Bull. 49, 1425–1436. doi:10.1093/schbul/sbad084.

Powers, A.R., Mathys, C., Corlett, P.R., 2017. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Sci. 357, 596–600. doi:10.1126/science.aan3458.

Raij, T., Ahveninen, J., Lin, F.H., Witzel, T., Jääskeläinen, I.P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hämäläinen, M.,

Belliveau, J.W., 2010. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices: Onset timing of audiovisual processing for simple stimuli. Eur. J. Neurosci. 31, 1772–1782. doi:10.1111/j.1460-9568.2010.07213.x.

Reddy-Thootkur, M., Kraguljac, N.V., Lahti, A.C., 2022. The role of glutamate and GABA in cognitive dysfunction in schizophrenia and mood disorders – A systematic review of magnetic resonance spectroscopy studies. Schizophr. Res. 249, 74–84. doi:10.1016/j.schres.2020.02.001.

Rideaux, R., Storrs, K.R., Maiello, G., Welchman, A.E., 2021. How multisensory neurons solve causal inference. Proc. Natl. Acad. Sci. 118, e2106235118. doi:10.1073/pnas.2106235118.

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

Rohe, T., Noppeney, U., 2015a. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. PLoS Biol. 13, e1002073. doi:10.1371/journal.pbio.1002073.

Rohe, T., Noppeney, U., 2015b. Sensory reliability shapes perceptual inference via two mechanisms. Journal of Vision 15, 22. doi:10.1167/15.5.22.

Romei, V., Gross, J., Thut, G., 2012. Sounds reset rhythms of visual cortex and corresponding human visual perception. Curr. Biol. 22, 807–813. doi:10.1016/j.cub.2012.03.025.

Romei, V., Murray, M.M., Merabet, L.B., Thut, G., 2007. Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. J. Neurosci. 27, 11465–11472. doi:10.1523/JNEUROSCI.2827-07.2007.

Rossetti, I., Romano, D., Florio, V., Doria, S., Nisticò, V., Conca, A., Mencacci, C., Maravita, A., 2020. Defective Embodiment of Alien Hand Uncovers Altered Sensorimotor Integration in Schizophrenia. Schizophr. Bull. 46, 294–302. doi:10.1093/schbul/sbz050.

Samaha, J., Romei, V., 2024. Alpha-band frequency and temporal windows in perception: a review and living meta-analysis

of 27 experiments (and counting). J. Cognn. Neurosci. , 1–15doi:10.1162/jocn$_{a0}$2069.$Sandsten, K.E., Nordgaard, J., Kjaer, T.W., Gallese, V., Ardizzi, M., recognitioninpatientswithschizophrenia. Schizophr.Res.218, 116 - -123. doi : 10.1016/j.schres.2020.01.022.$

Schmack, K., Gòmez-Carrillo De Castro, A., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J.D., Heinz, A., Petrovic, P., Sterzer, P., 2013. Delusions and the Role of Beliefs in Perceptual Inference. J. Neurosci. 33, 13701–13712. doi:10.1523/JNEUROSCI.1778-13.2013.

Schmack, K., Rothkirch, M., Priller, J., Sterzer, P., 2017. Enhanced predictive signalling in schizophrenia. Hum. Brain Mapp. 38, 1767–1779. doi:10.1002/hbm.23480.

Schmack, K., Schnack, A., Priller, J., Sterzer, P., 2015. Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. Schizophr. Res. Cogn. 2, 72–77. doi:10.1016/j.scog.2015.03.005.

Schoffelen, J.M., Pesci, U.G., Noppeney, U., 2024. Alpha oscillations and temporal binding windows in perception—a critical review and best practice guidelines. J. Cognn. Neurosci. , 1–36doi:10.1038/s41562-022-01294-x.

Schormans, A.L., Allman, B.L., 2023. An imbalance of excitation and inhibition in the multisensory cortex impairs the temporal acuity of audiovisual processing and perception. Cereb. Cortex 33, 9937–9953. doi:10.1093/cercor/bhad256.

Serino, A., 2019. Peripersonal space (PPS) as a multisensory interface between the individual and the environment, defining the space of the self. Neurosci. Biobehav. Rev. 99, 138–159. doi:10.1016/j.neubiorev.2019.01.016.

Shams, L., Iwaki, S., Chawla, A., Bhattacharya, J., 2005. Early modulation of visual cortex by sound: an MEG study. Neurosci. Lett. 378, 76–81. doi:10.1016/j.neulet.2004.12.035.

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

Stein, B.E., Burr, D., Constantinidis, C., Laurienti, P.J., Meredith, M.A., Perrault, T.J., Ramachandran, R., Röder, B., Rowland, B.A., Sathian, K.,

Schroeder, C.E., Shams, L., Stanford, T.R., Wallace, M.T., Yu, L., Lewkowicz, D.J., 2010. Semantic confusion regarding the development of multisensory integration: a practical solution. Eur. J. Neurosci. 31, 1713–1720. doi:10.1111/j.1460-9568.2010.07206.x.

Stein, B.E., Stanford, T.R., 2008. Multisensory integration: current issues from the perspective of the single neuron. Nat. Rev. Neurosci. 9, 255–266. doi:10.1038/nrn2331.

Sterzer, P., Adams, R.A., Fletcher, P., Frith, C., Lawrie, S.M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., Corlett, P.R., 2018. The Predictive Coding Account of Psychosis. Biol. Psychiatry 84, 634–643. doi:10.1016/j.biopsych.2018.05.015.

Tarasi, L., Martelli, M.E., Bortoletto, M., di Pellegrino, G., Romei, V., 2023. Neural signatures of predictive strategies track individuals along the autism-schizophrenia continuum. Schizophr. Bull. 49, 1294–1304. doi:10.1093/schbul/sbad105.

Tarasi, L., di Pellegrino, G., Romei, V., 2022a. Are you an empiricist or a believer? neural signatures of predictive strategies in humans. Prog. Neurobiol. 219, 102367.

Tarasi, L., Trajkovic, J., Diciotti, S., di Pellegrino, G., Ferri, F., Ursino, M., Romei, V., 2022b. Predictive waves in the autism-schizophrenia continuum: a novel biobehavioral model. Neurosci. Biobehav. Rev. 132, 1–22.

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P.R., Goodyer, I.M., Fletcher, P.C., 2015. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc. Natl. Acad. Sci. 112, 13401–13406. doi:10.1073/pnas.1503916112.

Thakkar, K.N., Nichols, H.S., McIntosh, L.G., Park, S., 2011. 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, e27089. doi:10.1371/journal.pone.0027089.

Torregrossa, L.J., Park, S., 2022. Body ownership across schizotypy dimensions: A rubber hand illusion experiment. Psychiatry Res. Commun. 2, 100058. doi:10.1016/j.psycom.2022.100058.

Tseng, H.H., Bossong, M.G., Modinos, G., Chen, K.M., McGuire, P., Allen, P., 2015. A systematic review of multisensory cognitive–affective integration in schizophrenia. Neurosci. Biobehav. Rev. 55, 444–452. doi:10.1016/j.neubiorev.2015.04.019.

Ursino, M., Crisafulli, A., di Pellegrino, G., Magosso, E., Cuppini, C., 2017. Development of a Bayesian Estimator for Audio-Visual Integration: A Neurocomputational Study. Front. Comput. Neurosci. 11, 89. doi:10.3389/fncom.2017.00089.

Ursino, M., Cuppini, C., Magosso, E., Beierholm, U., Shams, L., 2019. Explaining the Effect of Likelihood Manipulation and Prior Through a Neural Network of the Audiovisual Perception of Space. Multisens. Res. 32, 111–144. doi:10.1163/22134808-20191324.

Valton, V., Karvelis, P., Richards, K.L., Seitz, A.R., Lawrie, S.M., Seriès, P., 2019. Acquisition of visual priors and induced hallucinations in chronic schizophrenia. Brain 142, 2523–2537. doi:10.1093/brain/awz171.

Vanes, L.D., White, T.P., Wigton, R.L., Joyce, D., Collier, T., Shergill, S.S., 2016. Reduced susceptibility to the sound-induced flash fusion illusion in schizophrenia. Psychiatry Res. 245, 58–65. doi:10.1016/j.psychres.2016.08.016.

Werner, S., Noppeney, U., 2010. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. 30, 2662–2675. doi:10.1523/JNEUROSCI.5091-09.2010.

Whitford, T.J., Ford, J.M., Mathalon, D.H., Kubicki, M., Shenton, M.E., 2012. Schizophrenia, myelination, and delayed corollary discharges: a hypothesis. Schizophr. Bull. 38, 486–494. doi:10.1093/schbul/sbq105.

Yang, G.J., Murray, J.D., Wang, X.J., Glahn, D.C., Pearlson, G.D., Repovs, G., Krystal, J.H., Anticevic, A., 2016. Functional hierarchy underlies preferential connectivity disturbances in schizophrenia. Proc. Natl. Acad. Sci. 113. doi:10.1073/pnas.1508436113.

Zhou, H.y., Cai, X.l., Weigl, M., Bang, P., Cheung, E.F., Chan, R.C., 2018. Multisensory temporal binding window in autism spectrum disorders and schizophrenia spectrum disorders: A systematic review and meta-analysis. Neurosci. Biobehav. Rev. 86, 66–76. doi:10.1016/j.neubiorev.2017.12.013.

Zhu, H., Beierholm, U., Shams, L., 2024. The overlooked role of unisensory precision in multisensory research. Curr. Biol. 34. doi:10.1016/j.cub.2024.01.057.

Zopf, R., Boulton, K., Langdon, R., Rich, A.N., 2021. Perception of visual-tactile asynchrony, bodily perceptual aberrations, and bodily illusions in schizophrenia. Schizophr. Res. 228, 534–540. doi:10.1016/j.schres.2020.11.038.

Zvyagintsev, M., Parisi, C., Mathiak, K., 2017. Temporal processing deficit leads to impaired multisensory binding in schizophrenia. Cogn. Neuropsychiatry 22, 361–372. doi:10.1080/13546805.2017.1331160.

Increased excitation enhances the sound-induced flash illusion by impairing multisensory causal inference in the schizophrenia spectrum - Supplementary Information

1 Participants

We modelled the data collected in the study conducted by Ferri et al. (2018). One hundred and ninety-six adult volunteers were screened for schizotypal traits using the Schizotypal Personality Questionnaire (SPQ) (Raine, 1991). The questionnaire was administered online through Qualtrics. Participants were categorized into quintiles based on their SPQ scores. 64 participants participants falling within the first and fifth quintile were selected: 32 with low schizotypy scores (range: 5-16) and 32 with high schizotypy scores (range: 36-68). All participants had normal or corrected vision and hearing and no history of substance abuse or other psychiatric disorders. Age and gender were matched between the two groups. Demographic details are presented in Table 1.

Three participants in the low schizotypy group and four in the high schizotypy group were excluded from the analysis due to unreliable responses. This resulted in a final sample of 57 participants: 29 in the low schizotypy group and 28 in the high schizotypy group. All participants provided written informed consent before participating in the study.

Trait	L-SPQ	H-SPQ
SPQ Score Range	5-16	36-68
No. of Females	25	25
Mean Age (SD)	29 (9)	26 (7)
British White	41%	44%
Other White	50%	50%
Asian/Indian	9%	6%

Table 1: Demographics

2 Temporal Window of Illusion

Following Ferri et al. (2018), to determine the Temporal Window of Illusion (TWI), we analysed the proportion of trials in which participants reported perceiving the DFI (i.e., "two" responses) at each Stimulus Onset Asynchrony (SOA). For each participant, we plotted the proportion of DFI reports against the SOA and fitted a psychometric sigmoid function to these data points:

$$y = a + \frac{b}{1 + \exp\left(-\frac{(x-c)}{d}\right)} \tag{1}$$

where:

y : Proportion of DFI reports
x : SOA
a : Upper asymptote
b : Lower asymptote
c : Inflection point (TWI)
d : Slope

The estimated SOA (ms) corresponding to the inflection point (c) of the sigmoid function was defined as the individual's TWI, representing the temporal window within which the illusion was maximally experienced.

3 Multisensory causal inference network model

The network model used in this study is an extension of previous models developed by Cuppini and his collaborators (Cuppini et al., 2014, 2017). 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:

$$\tau^c \frac{dy_j^c(t)}{dt} = -y_j^c(t) + F\left(u_j^c(t)\right), \qquad c = a, v, m \tag{2}$$

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:

$$F\left(u_j^c\right) = \frac{1}{1 + exp^{-s\left(u_j^c - \theta\right)}} \tag{3}$$

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 ($o_j^c(t)$):

$$u_j^c\left(t\right) = l_j^c\left(t\right) + o_j^c\left(t\right) \tag{4}$$

The within region component l_j^c is defined as:

$$l_j^c\left(t\right) = \sum_k L_{jk}^c \cdot y_{jk}^c\left(t\right) \tag{5}$$

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):

$$L_{jk}^c = \begin{cases} L_{0ex}^c \cdot exp\left(-\frac{(D_{jk})^2}{2\cdot(\sigma_{ex}^c)^2}\right) - L_{0in}^c \cdot exp\left(-\frac{(D_{jk})^2}{2\cdot\left(\sigma_{in}^c\right)^2}\right), & D_{jk} \neq 0 \\ 0, & D_{jk} = 0 \end{cases} \tag{6}$$

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:

$$D_{jk} = \begin{cases} |j-k|, & |j-k| \leq N/2 \\ N - |j-k|, & |j-k| > N/2 \end{cases} \tag{7}$$

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:

$$e_j^c(t) = E_0^c \cdot exp\left(-\frac{\left(d_j^c\right)^2}{2\left(\sigma^c\right)^2}\right) \tag{8}$$

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:

$$d_j^c = \begin{cases} |j - p^c|, & |j - p^c| \leq N/2 \\ N - |j - p^c|, & |j - p^c| > N/2 \end{cases} \tag{9}$$

Furthermore, the cross-modal input is defined as:

$$\begin{array}{c} c_j^a(t) = \sum_{k=1}^{N} W_{jk}^{av} \cdot y_k^v(t - \Delta t_{cross}) \\ c_j^v(t) = \sum_{k=1}^{N} W_{jk}^{va} \cdot y_k^a(t - \Delta t_{cross}) \end{array} \tag{10}$$

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:

$$W_{jk}^{cd} = W_0^{cd} \cdot exp\left(-\frac{\left(D_{jk}\right)^2}{2\left(\sigma^{cd}\right)^2}\right), cd = av \ or \ va \tag{11}$$

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:

$$b_j^c(t) = \sum_{k=1}^{N} B_{jk}^{cm} \cdot y_k^c(t - \Delta t_{feed}) \tag{12}$$

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:

$$B_{jk}^{cm} = B_0^{cm} \cdot exp\left(-\frac{\left(D_{jk}\right)^2}{2\left(\sigma^{cm}\right)^2}\right), cm = am \ or \ vm \tag{13}$$

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:

$$\begin{cases} \frac{d}{dt} o_j^c(t) = \delta_j^c(t) \\ \frac{d}{dt} \delta_j^c(t) = \frac{G^c}{\tau^c} \cdot \left[e_j^c(t) + c_j^c(t) + b_j^c(t) + n_j^c \right] - \frac{2 \cdot \delta_j^c(t)}{\tau^c} - \frac{o_j^c(t)}{(\tau^c)^2} \end{cases} , c = a, v \quad (14)$$

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:

$$i_j^m(t) = \sum_{k=1}^{N} W_{jk}^{ma} \cdot y_k^a(t - \Delta t_{feed}) + W_{jk}^{mv} \cdot y_k^v(t - \Delta t_{feed}) \quad (15)$$

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:

$$W_{jk}^{mc} = W_0^{mc} \cdot exp\left(-\frac{(D_{jk})^2}{2(\sigma^{mc})^2} \right), c = a, v \quad (16)$$

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:

$$\begin{cases} \frac{d}{dt} o_j^m(t) = \delta_j^m(t) \\ \frac{d}{dt} \delta_j^m(t) = \frac{G^m}{\tau^m} \cdot \left[i_j^m(t) \right] - \frac{2 \cdot \delta_j^m(t)}{\tau^m} - \frac{o_j^m(t)}{(\tau^m)^2} \end{cases} \quad (17)$$

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

4 Values of L-SPQ model parameters

	Neurons		
$N = 30$	$\theta = 20$	$s = 0.3$	$\tau = 1$
	External stimuli		
$E_0^a = 2.325$	$\sigma^a = 32°$	$E_0^v = 1.45$	$\sigma^v = 4°$
	Input dynamics		
$G^{a,v,m} = e^1$	$\tau^a = 6.56$	$\tau^v = 9.19$	τ^m=120
$\Delta t_{cross} = 16$ ms	$\Delta t_{feed} = 95$ ms		
	Lateral synapses		
$L_{0ex}^c = 0.5$	$L_{0in}^c = 0.4$	$\sigma_{ex}^c = 3°$	$\sigma_{in}^c = 24°$
	Inter-areal synapses		
$W_0^{mc} = 3.892$	$\sigma^{mc} = 0.5°$	$B_0^{cm} = 0.623$	$\sigma^{cm} = 0.5°$
$W_0^{av,va} = 0.001$	$\sigma^{av,va} = 5°$		

Table 2: Values of the parameters in the L-SPQ model. This parameterisation remained fixed throughout the simulations computed to identify the H-SPQ model.

5 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 3).

In this parameter exploration, we examined seven uniformly distributed values within the operational range of each aforementioned parameter. The operational range is characterized by the parameter variations for which the base model (L-SPQ) demonstrates responses that align with those observed in human participants during the behavioral experiment (Ferri et al., 2018). The bounds considered for the parameter exploration for L_{0ex}, B_0^c, W_0^{cd}, $\rho_{W^{mc}}$, and $\rho_{W^{cd}}$ were (0.5, 0.65), (0.24, 0.25), (0.01, 0.055), (0, 0.35), and (0, 0.013), respectively.

6 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 18.

$$Cost = \sum_{i=1}^{N} \left(\frac{P(2f)_i^{data} - P(2f)_i^{model}}{P(2f)_i^{data}} \right)^2 \tag{18}$$

Here, $P(2f)_i^{data}$ and $P(2f)_i^{model}$ denote the P (2f) measured in the ith 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 (Storn and Price, 1997) available in the SciPy v1.11.4 library for the Python programming language (Virtanen et al., 2020).

$P(2f)^{model}$ 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 Ferri et al. (2018) is equivalent to the probability of perceiving two flashes in each condition.

6.1 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 16, 11, 13 14 and 17) 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 2) remained fixed and set to the values published by Cuppini et al. (2014, 2017). The bounds given to the optimisation algorithm to find τ^a, τ^v, τ^m, $W_0^{av,va}$, B_0^{cm} and W_0^{mc}, were (6, 20), (6.25, 60), (6, 120), (0.001, .05), (0.001, 1) and (0.001, 15) respectively. The results of the fit to the data of the L-SPQ group are presented in Table 2.

We also fitted an alternative L-SPQ model without the causal inference area. We maintained the same base parameterisation, and used the optimisation algorithm to find τ^a, τ^v and $W_0^{av,va}$. The bounds given were (6, 20), (6.25, 60), (6, 120) and (0.001, .05). This latter model failed to reproduce responses to SOAs beyond 60 ms, as shown in Figure 4A.

6.2 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 (Ferri et al., 2018).

It is observed that the model may not accurately reproduce the DFI responses observed at very short inter-beep intervals (below 75 ms). This discrepancy may be attributed, at least in part, to the fixed assumptions pertaining to the temporal delays of cross-modal and feedforward-feedback inputs incorporated within the model. Additionally, it is essential to acknowledge that the literature indicates substantial variability in DFI responses at very short inter-beep intervals across different individuals (Hirst et al., 2020). We recognize that further refinement of the model's parameters and architecture may be requisite to more precisely capture the complexities of DFI perception over the entire spectrum of inter-beep intervals.

References

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G., Ursino, M., 2014. A neurocomputational analysis of the sound-induced flash illusion. Neuroimage 92, 248–266. doi:10.1016/j.neuroimage.2014.02.001.

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. doi:10.1111/ejn.13725.

Ferri, F., Venskus, A., Fotia, F., Cooke, J., Romei, V., 2018. Higher proneness to multisensory illusions is driven by reduced temporal sensitivity in people with high schizotypal traits. Conscious. Cogn. 65, 263–270. doi:10.1016/j.concog.2018.09.006.

Hirst, R.J., McGovern, D.P., Setti, A., Shams, L., Newell, F.N., 2020. What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. Neurosci. Biobehav. Rev. 118, 759–774. doi:10.1016/j.neubiorev.2020.09.006.

Raine, A., 1991. The spq: a scale for the assessment of schizotypal personality based on dsm-iii-r criteria. Schizophrenia bulletin 17, 555–564.

Storn, R., Price, K., 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization 11, 341–359.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al., 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods 17, 261–272.