Balancing Prior Knowledge and Sensory Data in a Predictive Coding Model: Insights into Coherent Motion Detection in Schizophrenia

Elnaz Nemati[1], David B. Grayden[1,2], Anthony N. Burkitt[1, 2], and Parvin Zarei Eskikand[1,2*]

[1]Department of Biomedical Engineering, The University of Melbourne, Victoria, Australia
[2]Graeme Clark Institute, The University of Melbourne, Victoria, Australia

*pzarei@unimelb.edu.au

Abstract

This study introduces a biologically plausible computational model based on the predictive coding algorithm, providing insights into motion detection processes and potential deficiencies in schizophrenia. The model decomposes motion structures into individual and shared sources, highlighting a critical role of surround suppression in detecting global motion. This biologically plausible model sheds light on how the brain extracts the structure of motion and comprehends shared or coherent motion within the visual field. The results obtained from random dot stimuli underscore the delicate balance between sensory data and prior knowledge in coherent motion detection. Model testing across varying noise levels reveals increasing convergence time with higher noise, aligning with psychophysical experiments that show an increase in the response duration with an increase in the level of noise. The model suggests that an excessive emphasis on prior knowledge extends the convergence time in motion detection. Conversely, for faster convergence, the model requires a certain level of prior knowledge to prevent excessive disturbance due to noise. These findings contribute potential explanations for motion detection deficiencies observed in schizophrenia.
Keywords: Predictive Coding; Motion Detection; Schizophrenia; Surround Suppression.

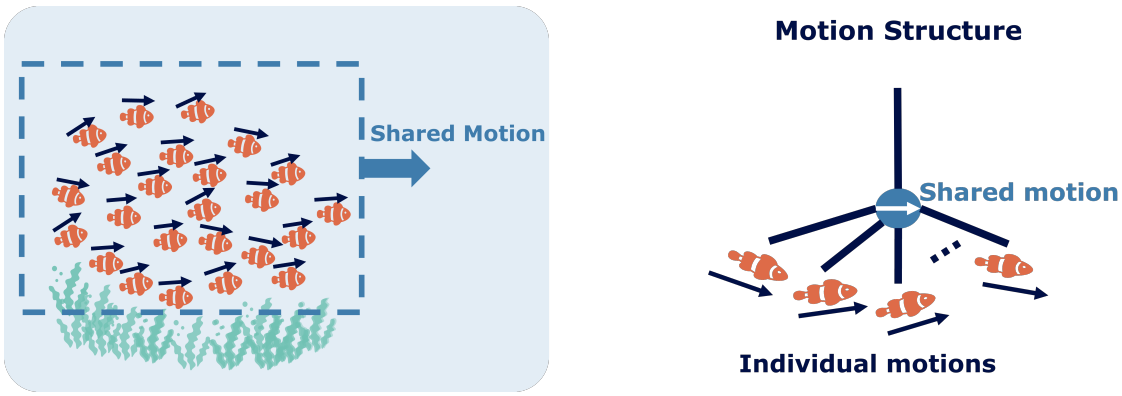
1 Introduction

In a world where visual information is constantly changing and can be ambiguous, efficiently understanding our environment necessitates that the brain can discern from the visual input received what motion is occurring in the environment, including both self-motion and the motion

of other objects. Motion detection is a complex perceptual process that involves the integration of prior knowledge, namely past visual experiences, with sensory data, namely the current visual information. This process is not performed in isolation. Our visual system takes into account our self-motion and the relative motion and interaction of neighboring objects (Born, 2000; Braddick, 1993), which play crucial roles in how motion stimuli are identified and interpreted. To achieve this, the neural circuits in the visual cortex learn their connectivity structure through synaptic plasticity during development. Neurons in the visual pathway have specific receptive fields, each attuned to a particular area of the visual field (Hubel and Wiesel, 1968). The activity of the neurons in the visual pathway is strongly influenced by the motion in the surround of their receptive fields. The presence of nearby stimuli shapes how coherent motion is detected and comprehended (Braddick, 1993; Born, 2000; Tadin et al., 2003). The complex microcircuits of the brain that are learned through this process allow the seamless integration of past and present motion information, enabling effective navigation in the environment.

1.1 Exploring Hierarchical Motion Perception: From Decomposition Principles to Predictive Coding Models

Achieving motion perception involves the brain using statistical relationships in the velocities of objects (Bill et al., 2022) and breaking down the motions of objects into underlying motion sources. These motion sources may not have a one-to-one relationship with objects. For example, for a school of fish, the movement of the fish can be broken down into both shared motion of the school and the individual motion of each fish, as illustrated in Fig. 1. Although it may seem abstract in concept, shared motion significantly enhances our perception and understanding of such a visual scene that involves many individual components that display coherent motion. By identifying the underlying motion structure, interpretation of the scene is enhanced, facilitating navigation and prediction tasks.



(a) (b)

Figure 1: Illustration of visual motion structure separation in complex scenes. (a) A scene of a school of fish, all moving collectively to the right (indicated by the pale blue arrow), with each fish exhibiting its unique motion (indicated by a navy blue arrow above each fish). (b) The scene's motion can be decomposed into shared motion (averaged over all fish) and individual motion components, relative to the shared motion.

The exploration of hierarchical motion perception has significantly progressed since the pi-

oneering studies by Johansson (1950), which were instrumental in exploring the decomposition of motion through displays of moving dots. These studies uncovered how simple motion of individual components in the visual scene can evoke complex perceptual experiences. This work supports the concept that the visual system uses vector analysis within a scene, distinguishing between the shared motion of objects and their individual motions. Furthermore, subsequent studies by Johansson (1973) on biological motion found that subtracting the shared motion of the components from the visual input allows the observer to perceive the relative motions of individual parts. This iterative subtraction method enabled a deeper understanding of the multilevel motion structure in visual scenes, laying the groundwork for our current understanding of hierarchical motion perception.

The resulting vector analysis theory in hierarchical motion perception, which breaks down object motion into common and relative components, is insightful but computationally limited. It assumes the presence of known motion components and structure but does not inherently detect these elements from sensory data, thus presenting a gap in its application to real-world motion analysis (Shum and Wolford, 1983). This issue is amplified in complex scenes with multiple possible vector analyses. Several theories have emerged to clarify ambiguities in motion interpretation. The "minimum principle" (Restle, 1979) posits a preference for interpretations that are the most straightforward, in line with Bayesian methods, that are applicable mainly in ideal, noise-free conditions with specific parametric motions. In contrast, the "adjacency principle" (Gogel, 1974) focuses on the cues from relative motion between closely located points. Additionally, the "belongingness principle" (DiVita and Rock, 1997) emphasizes how the coplanarity and reference frames of objects affect motion perception.

Recently, normative models based on Bayesian inference over tree structures have provided insights into more intricate aspects of human motion structure perception. Gershman et al. (2016) introduced a Bayesian framework for understanding how humans perceive motion structures through probabilistic inference over hierarchical motion structures, known as "motion trees". Subsequently, Yang et al. (2021) proposed that normative Bayesian models over tree structures could explain crucial aspects of motion perception. These studies primarily focused on modeling motion integration concerning perceptual outcomes, leaving unaddressed the complex challenge of real-time motion parsing while concurrently inferring scene structure. To tackle this challenge, Bill et al. (2022) proposed an innovative approach, formulating visual motion perception as online hierarchical inference within a generative model. This continuous-time model effectively explains various aspects of human motion perception, shedding light on the complex interplay between motion parsing and structure inference in real-time perception.

Motion detection heavily relies on past motion information. In the study of motion perception, the role of past information cues and prior experience has been traditionally overlooked, presenting a significant gap in the understanding of motion processing. Hierarchical predictive coding offers a promising approach to bridge this gap. This framework posits that the brain constructs a multi-level model, where high-level information influences the processing of sensory data at lower levels (Rao and Ballard, 1999; Millidge et al., 2021). In the context of motion

perception, this implies that the brain integrates current visual stimuli with past experiences to predict and interpret motion patterns. By employing predictive coding at lower levels of this hierarchy, the brain can process motion efficiently and rapidly, enhancing the understanding of complex motion structures. Despite the theoretical appeal and potential applicability of predictive coding in this domain, there is a scarcity of such models in the literature.

This study aims to illuminate the brain's adaptive mechanisms in refining motion processing through neural connectivity and networks. The central hypothesis of this research is grounded in the concept of "predictive coding" (Rao and Ballard, 1999), where the activity in the higher levels of the hierarchy determines (or"predicts") the activity in the lower levels. Although classical predictive coding is predictive in this hierarchical fashion, it has also been postulated to be utilized by the brain in the temporal domain to forecast future events (Millidge et al., 2023). We test the classical predictive coding model with different stimuli and show that the model successfully extracts the global motion of the stimuli and replicates the perception observations recorded in psychophysical experiments (Johansson, 1973; Duncker, 1929).

1.2 Motion detection deficiency in schizophrenia

Schizophrenia impacts around 24 million individuals globally and is recognized as one of the most debilitating and economically devastating disorders, as reported by the World Health Organization (Institute of Health Metrics and Evaluation, 2021). Individuals diagnosed with schizophrenia have been found to exhibit impairments in tasks related to motion processing, specifically in tasks such as velocity discrimination, where patients seek to identify which stimulus is moving at a faster pace (Chen et al., 1999; Kim et al., 2013). In an fMRI study, it was observed that the Middle Temporal (MT) area in the visual cortex, which is known to be highly specialized for motion detection, is significantly less active in patients with schizophrenia compared to control subjects when conducting motion detection tasks for direction discrimination and speed discrimination (Chen et al., 2008).

A frequently used type of stimulus to investigate the neural circuitry of motion detection is random dots. In the context of schizophrenia, Random Dot Kinematograms (RDKs) prove to be very informative, as they capture the ability of the nervous system to integrate the local motion of individual dots to estimate global motion under different levels of noise. A psychophysical study showed that individuals with schizophrenia require a higher level of coherence in random dots (i.e., less noise) to estimate the global motion of the dots compared to healthy controls (Kim et al., 2013). An essential consideration in the study of neuronal circuitry for motion detection is the observation that patients with schizophrenia exhibit weaker surround suppression compared to their healthy counterparts (Tibber et al., 2013).

The model we present here enables exploration into the underlying neuronal circuitry associated with the observed motion detection deficiency in schizophrenia. This model synthesizes various factors, notably highlighting the impact of weaker surround suppression in subjects

with a motion detection deficiency. Moreover, in this study, for the first time, we aim to investigate the necessity of a balance between sensory input and prior knowledge in detecting the global motion of stimuli, and how it relates to the observed motion detection deficiency in schizophrenia.

2 Materials and Methods

2.1 Model Framework Overview

To formalise the process of estimating motion sources for an object, we consider motion dynamics where the velocity of an object, denoted as v_i, is influenced by a set of M hidden motion sources. Each of these sources, represented as r_j for the jth source, contributes to the object's motion to varying degrees. This variability in influence is captured by the weight $u_{i,j}$, which quantifies the degree to which each motion source affects the velocity of object i. Specifically, the velocity of an object can be expressed as the sum of the products of the weights and the corresponding motion source contributions,

$$\hat{v}_i = \sum_{j=1}^{M} u_{i,j} r_j, \tag{1}$$

where $\hat{v}_i$ represents the estimated velocity of object i. The weight $u_{i,j}$ indicates the influence of the j-th motion source on object i, and these weights are organized into the matrix $\mathbf{U}$. The motion sources r_j are the components of the vector $\mathbf{r}$, which contains the values of all motion sources. The weights $u_{i,j}$ vary from 0 to 1, with 0 indicating no impact and higher values indicating a stronger influence on the object's motion. It is crucial to note that the value of each motion source r_j is consistent for all objects; only the weight $u_{i,j}$ changes to reflect the total velocity contribution from each motion source to the object. This framework provides a structured approach for dissecting and analyzing the complex interactions between objects and their motion sources, facilitating a deeper understanding of motion dynamics.

To explain the complexities of motion sources and their structural underpinnings, we introduce a methodology that leverages the classical predictive coding framework (Rao and Ballard, 1999), further enhanced by incorporating a surround suppression mechanism. This enhancement aims to more accurately depict the interactions between historical motion data and the effects exerted by the neighbouring objects. Predictive coding, in this context, serves to estimate motion information by evaluating the neuronal activity within the receptive fields of neurons, a process that essentially combines neuronal activity with synaptic weights. Synaptic weights here are indicative of the motion's strength (e.g., the influence of the respective motion source on the overall object's motion), while neuronal activity is associated with identifying the source of the motion. The neuronal activity indicates the [perceived] direction of motion within the receptive field of a neuron. The synaptic weights indicate an object's motion strength,

which is defined as the influence of that object's motion on the overall motion observed in the scene.

A pivotal extension to our model is the incorporation of surround suppression, a neurobiological mechanism that adjusts neural excitability based on the motion context provided by adjacent stimuli (Tsui et al., 2010). This regulatory mechanism fine-tunes the initial motion predictions derived from predictive coding, where the extent of refinement is graphically denoted by a triangle in Fig. 2. Initial motion predictions are generated through the predictive coding paradigm, establishing these estimates as the product of neural activities and synaptic weights **Ur**. The activation of surround suppression subsequently gates these motion strength predictions.

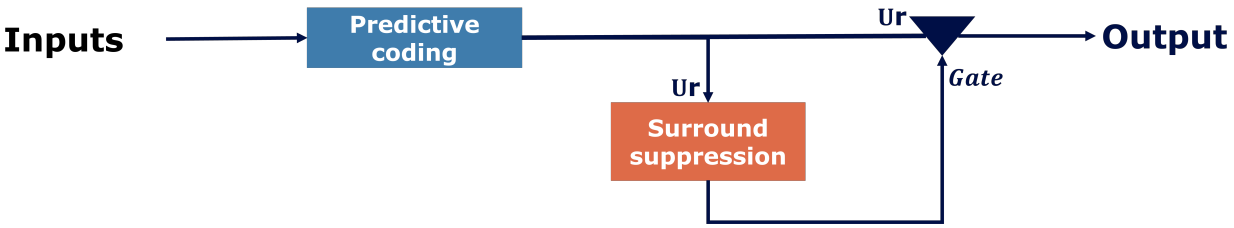


Figure 2: The proposed Motion Estimation Method. The estimated motions generated by the predictive coding, which is the activity of the neurons **r** multiplied by the synaptic weights **U**, marked on the arrows, are processed by the surround suppression mechanism. Depending on the activation of surround suppression, the estimated motions by the predictive coding are gated and symbolised by a triangle.

Crucially, in instances where the initial motion estimation for an object fails to encapsulate the entirety of its motion, our model employs subsequent iterations of predictive coding. This recursive process ensures the comprehensive capture of motion dynamics, persisting until the motion estimations holistically represent the object's motion.

Our study systematically examines the two-dimensional spatial configurations of objects in visual stimuli, employing a predictive coding model enhanced with the surround suppression module. This novel approach allows for the identification and quantification of motion sources. The model considers the influence of adjacent objects' movements through a surround suppression mechanism, which activates to suppress estimated motion when an object's movement significantly deviates from its surroundings, as suggested by Eskikand et al. (2016). Conversely, this mechanism remains inactive when there is alignment, preserving the integrity of the motion estimation.

Incorporating this approach, we are able to extract both the global shared motion and individual motion sources, thereby providing a detailed dissection of motion perception. This methodology, which marries the classical framework of predictive coding (Rao and Ballard, 1999) with a surround suppression mechanism, not only enhances the model's ability to capture the nuanced interplay between historical motion data and the effects of proximal stimuli but also enables the differentiation between shared and unique motion dynamics, as elaborated in subsequent sections.

2.2 Classical Predictive Coding

The predictive coding model provides a compelling framework for understanding brain function, likening it to an algorithm that predicts future events based on sensory data. It is akin to the brain continuously making educated guesses about incoming sensory information, then adjusting these predictions as new data comes in. This model is organized in a multilevel structure, Fig. 3a, each playing a unique role in processing information.

Central to the predictive coding model are two types of units within each level: error units and activity units, indicated by blue circles and orange rectangle in Fig. 3b, respectively. Error units function as the brain's mismatch identifiers, comparing its predictions against sensory inputs to pinpoint discrepancies, termed "prediction errors", and relaying these errors to subsequent levels in the hierarchy. In contrast, activity units are tasked with creating predictions from sensory data and previous prediction errors, and transmitting them to the preceding level in the hierarchy.

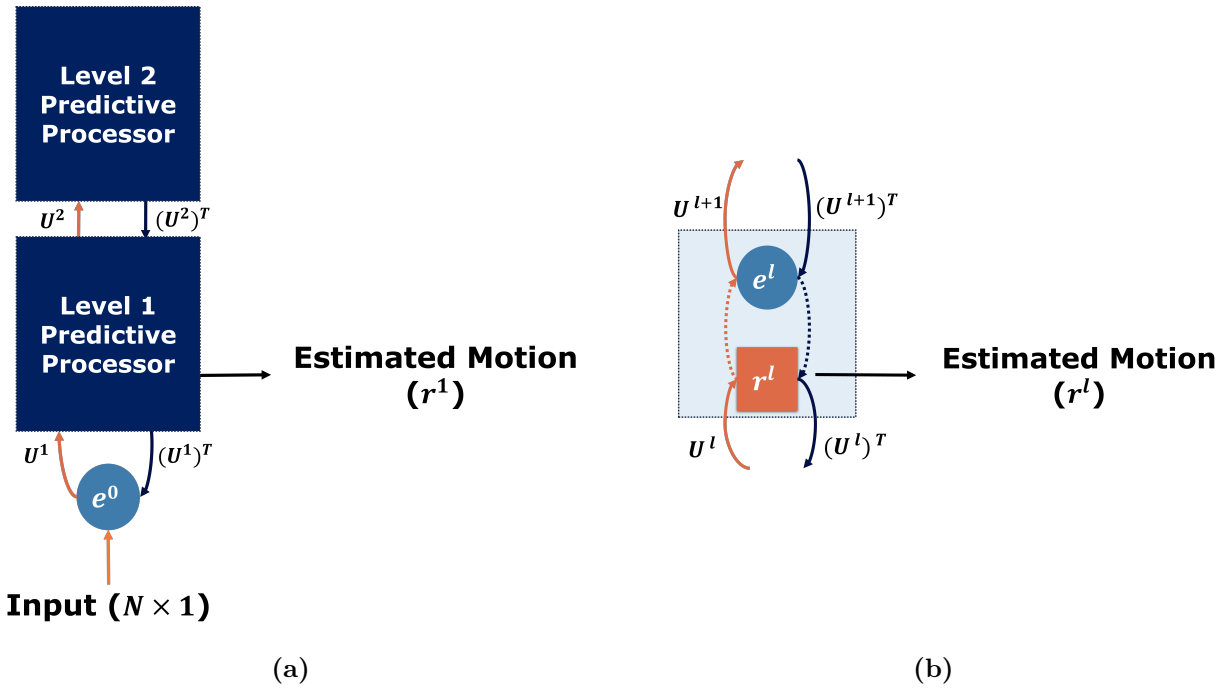


Figure 3: Classical Predictive Coding for motion processing. (a) Two-level classical predictive coding. (b) One level of a predictive processor. Both subplots illustrate the network's outputs with black arrows. Excitatory and inhibitory connections are highlighted in orange and blue lines, respectively. Error units are depicted as blue circles, while activity units are shown as orange rectangles. Synaptic weights, marked with '$\mathbf{U}^l$', are displayed on the arrows, where l indicates the layer in the hierarchy. Dotted orange and blue arrows represent internal excitatory and inhibitory connections. Here, l is the level index, and N is the number of input objects. Adapted from Rao and Ballard (1999).

The objective of the predictive coding model is to minimize the prediction error at each level of the hierarchy within the network. This is typically achieved by implementing gradient

descent of the prediction errors to adjust connection strengths within the network accordingly. The objective function, E, is a sum-of-squares error,

$$E = \frac{1}{2}\sum_{l}(\mathbf{e}^l)^2 = \frac{1}{2}\sum_{l}(\mathbf{r}^{l-1} - \mathbf{U}^l\mathbf{r}^l)^2, \tag{2}$$

where $\mathbf{e}^l$ denotes the prediction error at level l, $\mathbf{r}^l$ is a vector containing the output of the activity units, denoted as the prediction, and $\mathbf{U}^l$ denotes the matrix of synaptic weights that determine the prediction strength. The model treats $\mathbf{r}^0$ as the direct sensory input, iterating through levels to minimize E.

Consequently, the predictions r^l are updated over time within each level according to

$$\frac{\mathrm{d}\mathbf{r}^l}{\mathrm{d}t} = -\gamma\frac{\partial E}{\partial \mathbf{r}^l} = -\gamma\left((\mathbf{U}^l)^T\mathbf{e}^{l-1} - \mathbf{e}^l\right) = -\gamma\left((\mathbf{U}^l)^T\left(\mathbf{r}^{l-1} - \mathbf{U}^l\mathbf{r}^l\right) - (\mathbf{r}^l - \mathbf{U}^{l+1}\mathbf{r}^{l+1})\right). \tag{3}$$

The prediction, $\mathbf{r}^l$, within each level l is influenced by two factors: (1) the feed-forward input as the prediction error, $\mathbf{e}^{l-1}$, from the preceding level, adjusted through the synaptic weight, $\mathbf{U}^l$, depicted by the solid orange line in Fig. 3b, and (2) the difference between the current prediction, $\mathbf{r}^l$, and the feedback from higher levels, $\mathbf{U}^{l+1}\mathbf{r}^{l+1}$, illustrated by the dotted blue line in Fig. 3b. The update rate, γ, determines the degree to which these inputs affect the update of the current prediction, $\mathbf{r}^l$, ensuring a balance is struck between incorporating new data and maintaining prior predictions. This equation demonstrates how the network methodically enhances predictions by applying both feedback and the propagation of errors and adjustments, highlighting the model's capability for adaptive learning.

Additionally, the synaptic weights are updated to reflect the learning process, also according to gradient descent,

$$\frac{\mathrm{d}\mathbf{U}^l}{\mathrm{d}t} = -\eta\frac{\partial L}{\partial \mathbf{U}^l} = -\eta\left(\mathbf{e}^{l-1}\mathbf{r}^l\right), \tag{4}$$

where η is the learning rate. This equation embodies the Hebbian rule (Hebb, 1949), illustrating that the weight update is proportional to the product of the post-synaptic activity, $\mathbf{r}^l$, and the pre-synaptic activity, $\mathbf{e}^{l-1}$. This iterative process continues until the model converges to an optimal set of weights, indicated by a minimum prediction error.

A two-level classical predictive coding algorithm was used in the current study. The dynamics of the activity unit in the predictive coding are updated by

$$\frac{\mathrm{d}\mathbf{r}^1}{\mathrm{d}t} = -\gamma\left((\mathbf{U}^1)^T(x - \mathbf{U}^1\mathbf{r}^1) - (\mathbf{r}^1 - \mathbf{U}^2\mathbf{r}^2)\right), \tag{5}$$

$$\frac{\mathrm{d}\mathbf{r}^2}{\mathrm{d}t} = -\gamma\left((\mathbf{U}^2)^T(\mathbf{r}^1 - U^2\mathbf{r}^2)\right), \tag{6}$$

where $\mathbf{r}^l$ denotes the output of the activity unit and x denotes the input data, specifically the positions of objects within the environment. The variable $\mathbf{U}^l$ represents the synaptic weight that modulates the influence of activity units across the network levels, with their subscripts

1 and 2 denoting the first and second level indices, respectively. The update process for $\mathbf{r}^1$ and $\mathbf{r}^2$ is hierarchical: the update of $\mathbf{r}^1$ incorporates the current estimates of $\mathbf{r}^2$ along with the synaptic weights $\mathbf{U}^1$ and $\mathbf{U}^2$, and vice versa for the update of $\mathbf{r}^2$ the current estimates of $\mathbf{r}^1$, $\mathbf{U}^1$, and $\mathbf{U}^2$ are considered. Note that synaptic weights are updated at a time scale that is slower than the activities are updated.

To draw a parallel with kinematics, consider the dynamic equation, $x(t) = -\frac{1}{2}at^2 + vt + x_0$, which relates an object's position $x(t)$ over time t to its acceleration a and velocity v, with x_0 being the initial position. In this analogy, $\mathbf{r}^2$, representing an estimate of the system's acceleration, aligns with the acceleration term a, while $\mathbf{r}^1$, representing an estimate of velocity, parallels the velocity term v. Our model utilizes the predictive coding framework to model each motion source, where $\mathbf{r}^1$ in each predictive coding unit corresponds to the motion source, and the synaptic weight $\mathbf{U}^1$ reflects the influence of each motion source on the object.

At each time step, the estimated motion derived from the predictive coding framework is forwarded to the surround suppression mechanism for refinement. A detailed description of the surround suppression mechanism operation within the context of our predictive coding framework is given in the following section.

2.3 Surround Suppression Mechanism

At the core of the proposed model, surround suppression holds a pivotal role, as illustrated in Fig. 4. This mechanism is grounded in the principle that the response of motion-sensitive neurons to a stimulus within the central receptive field is modulated by concurrent motion stimuli in the surrounding areas, a concept previously outlined by Born (2000). We postulate that the influence of the surround is also affected by the historical activity within this surrounding suppression mechanism.

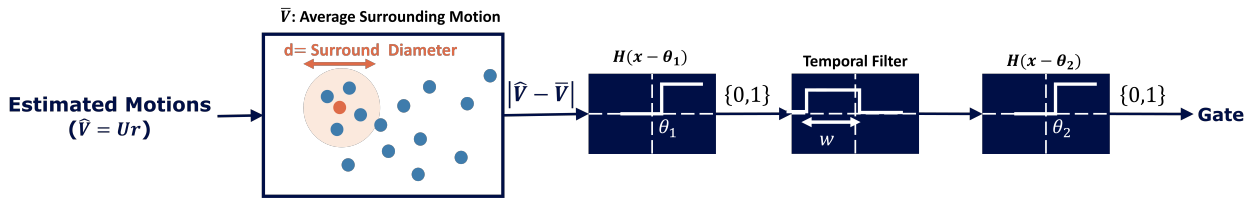


Figure 4: Surround Suppression mechanism. This diagram illustrates the process of comparing motion estimation at the center of the receptive field, denoted by $\hat{v} = Ur$, with the average motion in the surrounding area, denoted by $\bar{V}$, at each temporal step. This comparison is conducted using a Heaviside function with the threshold parameter θ_1. The result is then integrated with the historical average motion within a temporal window of width w. This integrated value is again subjected to a comparison, this time passing through a Heaviside function with a threshold parameter θ_2. The output is binary, indicating the activation state of the surround suppression mechanism."

The surround suppression mechanism remains inactive when the individual object's motion is aligned with the motion within the surround. In contrast, if there is a misalignment between the surround and the object's motion estimate, the surround suppression mechanism becomes active. This computation also incorporates weighted historical activity to account for temporal

dynamics. When the surround suppression mechanism is activated, it suppresses the activity of the neuron. Conversely, when the mechanism is inactive, the object's estimate in the center of the receptive field will not be suppressed, resulting in distinguishing its motion from the surrounding activity.

In implementing the surround suppression activation function, the model does a comparison between the motion estimation of an object, $\hat{\mathbf{v}} = \mathbf{Ur}$, and the average motion observed within the object's surrounding, which is assumed to be circular with diameter d. This average motion, is defined as

$$\bar{v_i} = \frac{1}{Q} \sum_{o \in q_i} \sum_{j} u_{o,j} r_j \tag{7}$$

where q_i represents the set of indices corresponding to objects within the circular area surrounding object i. The variable Q denotes the total number of these surrounding objects. The double summation calculates the combined influence of all motion sources (indexed by j) on each surrounding object o, thereby estimating their average motion in the surround.

The model takes into account not only the current motion but also the history of movement within the surround. It calculates a historical average of surround suppression activation, denoted as $\bar{f}_w$, over a specified time window w. This historical average serves to incorporate the influence of past suppression activation into the current assessment of motion.

The overall behavior of the surround suppression module, which takes into account both the current and historical motion data, is governed by

$$f(\hat{\mathbf{v}}) = H\left(H\left(|\hat{\mathbf{v}} - \bar{\mathbf{v}}| - \theta_1\right) - \bar{f}_w(\hat{\mathbf{v}}) - \theta_2\right) \ , \tag{8}$$

where $H()$ is the Heaviside step function, and θ_1 and θ_2 are threshold parameters that regulate the immediate and historical factors, respectively, determining the activation state of surround suppression based on the current motion estimate's divergence from the average surround motion and its historical context. Threshold θ_1 evaluates the similarity between the motion of an individual object, $\hat{v}_i$, and the average motion of its surrounding objects, $\bar{v}_i$. A secondary threshold, θ_2, is then utilized to evaluate the impact of this historical average against the present activation of the suppression function. The application of this threshold essentially determines how much the past activation will modify the current response of the surround suppression mechanism.

In this study, we focus on the effects of parameters θ_1 and w on the model's ability to detect coherent motion. θ_1 is crucial for processing immediate motion within the model's receptive field, whereas w relates to the duration of the temporal window used for evaluation. Conversely, θ_2 is not the subject of our investigation as it pertains to the historical impact of motion data. This parameter, θ_2, is associated with how previously observed movements affect the current analysis of motion, which is indirectly studied through the temporal window duration w. In other words, while θ_2 governs the integration of past motion information, impacting how the

model discerns between relevant and irrelevant historical movements, our research specifically examines the immediate processing governed by θ_1 and the temporal scope defined by w.

The parameter θ_1 is pivotal as it sets the threshold for surround suppression, dictating how the motion within the centre of the receptive field is compared with the collective motion in the surrounding visual field. A larger value of θ_1 represents an increased threshold for surround suppression, which means that the difference between the estimated motion and the average estimated motion in the surround needs to be greater for it to be influential.

The parameter w delineates the span of the historical window utilized by the surround suppression mechanism. It determines the extent to which past information is integrated into the current processing. Employing a broader window allows the model to leverage a more extensive history of information, which aids in mitigating the influence of transient noise and in fostering a more stable detection of coherent motion.

Our objective is to scrutinize how the incorporation of historical motion data affects the model's capacity to discern coherent motion amidst the ongoing sensory input. By fine-tuning the balance between immediate sensory information and accumulated historical data, we aim to enhance the model's ability to reliably detect coherent motion patterns.

2.4 Stimuli

The algorithm is evaluated using simulation results based upon three different psychophysical experiments: (I) the Johansson experiment (Johansson, 1973), (II) the Duncker wheel (Duncker, 1929), and (III) the Random Dot Kinematogram (Kim and Shadlen, 1999). Each experiment is selected for its unique contribution to studying motion perception and processing.

In the Johansson experiment (experiment I) (Johansson, 1973), illustrated in Row I of Fig. 5, three vertically aligned dots move backwards and forwards across a display. The two outer dots move horizontally, while the middle dot moves diagonally while maintaining collinearity with the outer dots (Fig. 5a). Observers tend to perceive the middle dot as moving vertically, oscillating between the two horizontally moving outer dots, despite its diagonal trajectory (Fig. 5b). This illusion is attributed to the synchronized horizontal motion of all three dots, where the brain interprets the movement based on the relative positions and synchronized movements of the dots.

The Duncker wheel experiment (experiment II) (Duncker, 1929) is a classic study in motion perception that features two dots placed on an invisible wheel, as illustrated in Fig. 5, Row II. One dot is positioned at the hub of the wheel and moves linearly (typically left to right), while the other dot follows a complex spiral path along the rim, resulting from the wheel's rotational and lateral movement (Fig. 5c). Observers perceive the motion of these dots differently: the dot at the hub is seen moving linearly, consistent with its actual motion, but the dot on the rim, despite its spiral path, is perceived as moving in a simple circular motion around the hub dot

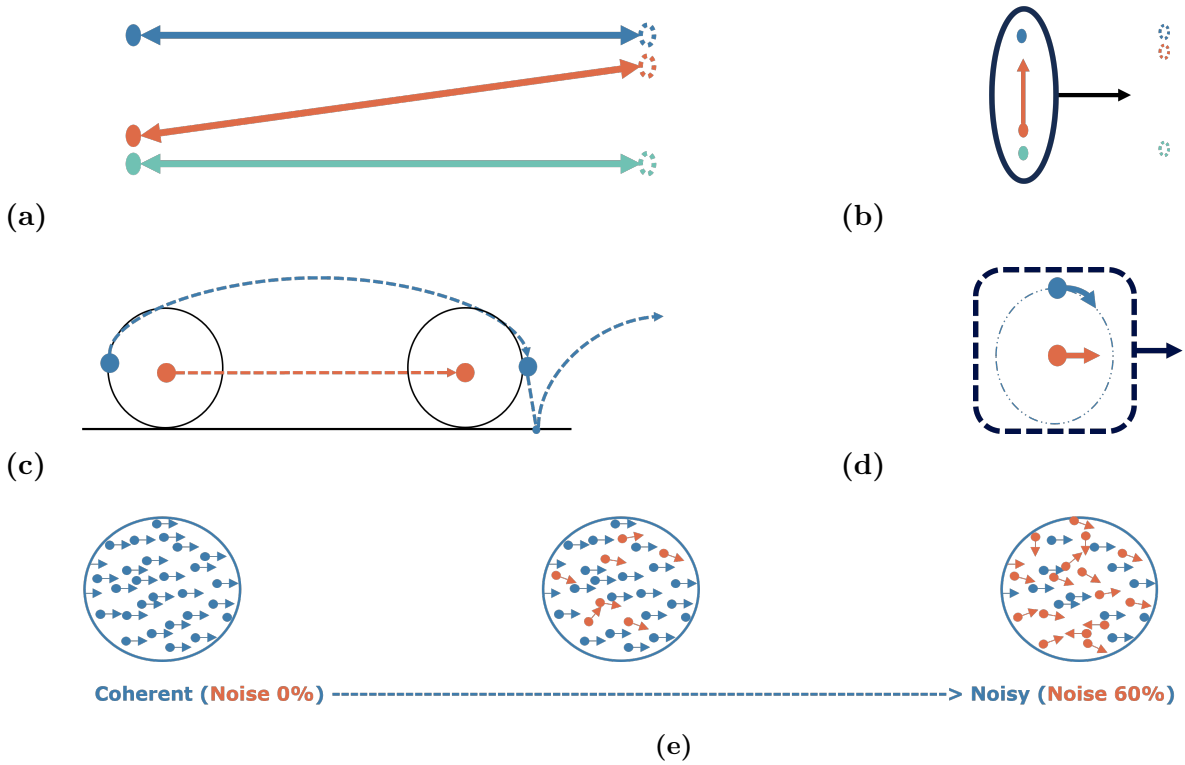


Figure 5: Experimental Paradigms Overview. Row I: A schematic view of Johansson's experiment with three dots. (a) Motion vectors for each of the dots, and (b) the perceived motion of the group (denoted by the oval containing the three dots moving rightwards) and the central dot (moving upwards). Adapted from Gershman et al. (2016). Row II: A schematic view of Duncker wheel's experiment with three dots. (c) Motion vectors for each of the hub- and rim-dots, and (d) the perceived motion of the group (denoted by the dashed square containing both dots moving rightwards) and the rim-dot (moving in a circular motion). Adapted from Gershman et al. (2016). Row III: Random Dot Kinematogram. (e) The difficulty of the experiment escalates from left to right with an increasing number of noise dots, depicted in orange (although the dot colors are not different in the experiment). The blue dots represent signal dots that have coherent motion towards the right.

(Fig. 5d). This phenomenon demonstrates our tendency to break down complex movements into recognizable components, such as interpreting the spiral motion here of the rim-dot as a combination of translation and rotation when the additional reference point of the hub dot is provided.

The Random Dot Kinematogram (RDK) (experiment III) is a useful tool for studying motion perception, particularly in relation to schizophrenia (Kim and Shadlen, 1999). In a RDK, numerous small dots move randomly on a screen, as illustrated in Fig. 5e, Row III. A subset of dots (blue) moves coherently in the same direction among the remaining dots (orange) that move randomly, and the observer is asked to identify the coherent direction. For instance, in a 95% coherent motion display, 95% of the dots (signal dots) move in a designated direction from one frame to the next, while the remaining 5% (noise dots) move randomly. The perception of global motion direction becomes more accessible with higher coherence levels. This experimental setup is crucial for gaining insights into how the visual system processes and distinguishes motion information from visual noise. In the context of our study, we introduce the concept of "time to convergence", which is defined as the number of discrete time steps, each with a duration of Δt, required for our model to accurately identify the coherent motion direction within

the Random Dot Kinematogram (RDK). Coherent motion is considered accurately identified when the motion strength of the coherent motion stabilizes, showing little or no change, with variations less than 0.05 units over subsequent time steps. This criterion ensures that the model has reliably converged to a stable representation of the coherent motion direction.

The parameters of the stimuli are consolidated in Table 1. The behavior of the model simulations of these psychophysical experiments in elucidating aspects of human visual motion perception is contained in the following section.

Table 1: Parameters of stimuli used in the study

Stimulus	Parameter	Value
Johansson Experiment	Frequency	0.5 Hz
	Horizontal Amplitude	$2\sqrt{0.3}$
	Vertical Amplitude	$\cos(45°)2\sqrt{0.3}$
	Time Constant	1/60 sec
Duncker Wheel	Wheel Radius (R)	10 unit
	Horizontal Speed	0.5 unit/sec
	Angular Speed	0.1 deg/sec
	Time Constant	0.05 sec
Random Dot Kinematogram	Number of Dots	1000
	Dot Density	500 dot/unit
	Coherent Velocity	0.01 unit/sec (horizontal motion)
	Random Velocity	0.01 unit/sec (vertical motion)
	Time Constant	0.05 sec

3 Results

We evaluated the performance of the model in identifying the motion structure and extracting coherent motion in three different experiments. Additionally, we investigated the effect of prior knowledge on the model's performance with motion data and the influence of the surround suppression mechanism on detecting coherent motion.

3.1 The model successfully identified the motion structure in classical motion patterns

Fig. 6 presents the outcomes of classical motion perception experiments modeled through our predictive coding framework. The model accurately captured the motion structures in the Johansson (1973) and Duncker (1929) experiments (I & II), differentiating between shared and individual motion types.

Psychophysics experiments of Johansson motion in experiment I (see Fig. 6a) show that humans perceive this stimulus as shared horizontal motion, with the central dots oscillating vertically Johansson (1973). Therefore, we expect the model to extract equal (synchronous)

shared motion across all three dots and individual motion for the central dot. The outcome of the model was connectivity weights **U** that represent the strength of the contribution of each motion source (motion strength). Fig. 6b shows the estimated motion strengths for the shared motion and the individual motions. The individual motions for blue and green dots converge to zero over time. However, the individual motion of the central dot (orange) remains pronounced. The shared motion of all three objects converged to the same values, close to 1. This demonstrates the model's ability to distinguish between shared motion, where all dots synchronize, and individual motion, which decreases for the outer dots but remains pronounced for the central dot. The observed oscillation in motion strength is due to the sinusoidal motion of objects, which moves back and forth from left to right and then right to left. In this experiment, shared motion refers to the horizontal movement of all dots, causing them to oscillate back and forth collectively. Individual motion, on the other hand, is characterized by vertical movement. For the blue and green dots, the vertical movement is zero, whereas for the orange dot, the vertical movement is non-zero, as indicated by the motion strength.

The model's application to the Duncker experiment in Experiment II (Fig. 6c) provided further validation of its capability to represent complex motion patterns. As shown in Fig. 6d, the shared horizontal motion of the dots was captured, while the individual motion of the dot on the rim (blue) was distinguished by its significant individual motion strength. Conversely, the other horizontally moving (orange) dot exhibited an individual motion strength that approached zero, effectively simulating the perception of a rolling wheel. In this experiment, shared motion refers to the horizontal movement of both objects, causing them to move horizontally to the left. Individual motion can be interpreted as radial motion. The radial motion converges to zero for the horizontally moving dot, while for the blue dot, it remains non-zero. The pronounced oscillation observed in the shared motion of the radial ball, resulting from the constant value of the motion source, effectively represents the motion strength for the radial movements of the blue dots.

The Fig. 7 displays outcomes from a Random Dot Kinematogram (RDK) in experiment III, which delineates the discernment of individual versus shared motion. Noise dots (dots with noncoherent motion) exhibit higher and more variable individual motion strengths compared to signal dots (orange line). However, the high number of noise dots prevents their individual motion strength from reaching 1. The mean individual motion strength (10.4 units/ms) is higher for noise dots than for signal dots (3.25 units/ms), with noise dots displaying significant fluctuations and signal dots maintaining a relatively stable, lower value. This highlights the unpredictable and incoherent movements of noise dots compared to the coherent movement of signal dots. In contrast, the shared motion graph shows that signal dots quickly reach and maintain a high shared motion strength close to 1, while noise dots exhibit almost no shared motion, indicating that the coherent motion stands out amidst the random background. As the model stabilizes, the individual motion strengths of signal dots diminish to zero, marking the successful detection of coherent motion.

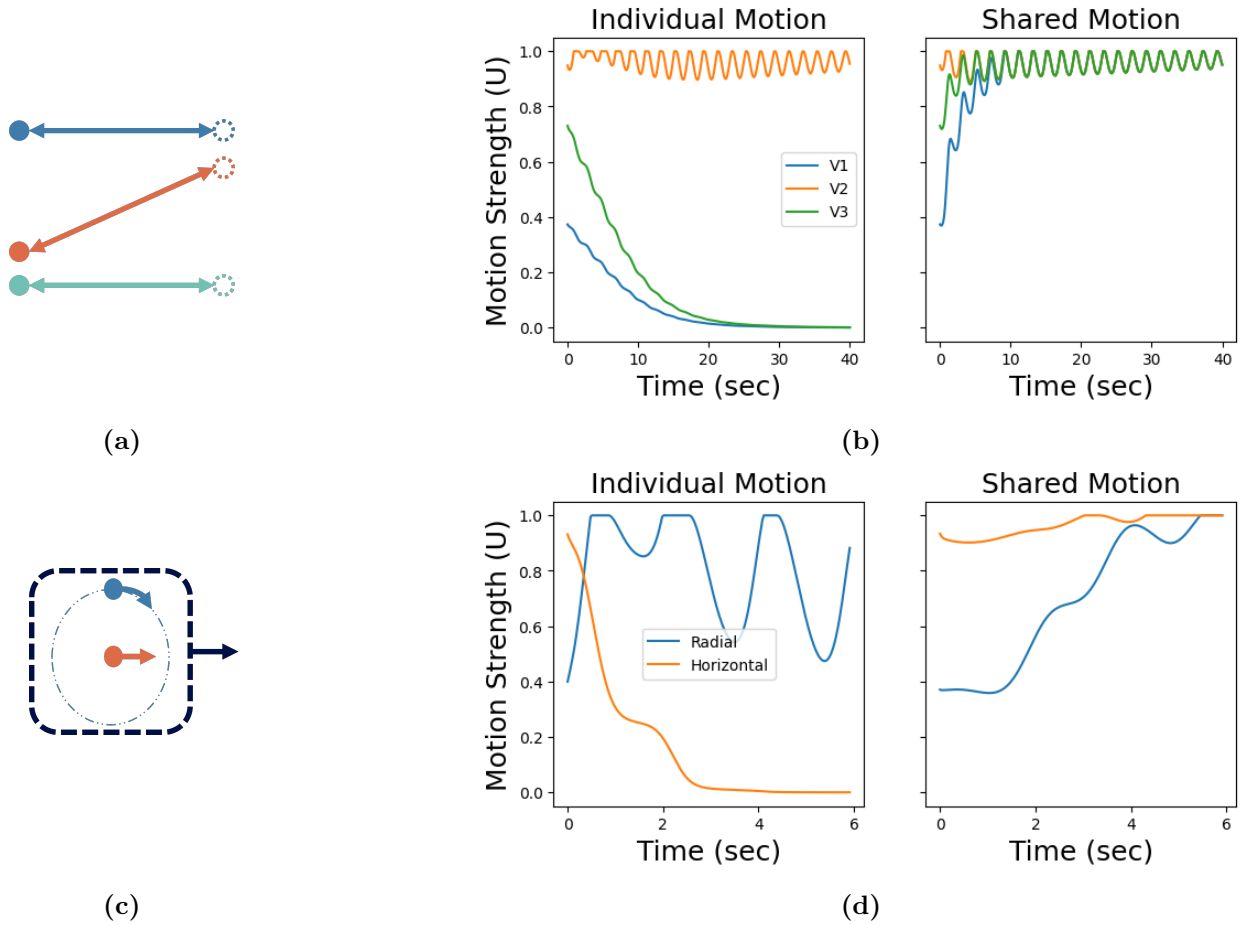


Figure 6: The model successfully identified the motion structure in experiments I and II. (a) A schematic view of Johansson's experiment I (Johansson, 1973) with three dots. (b) Model estimates of motion strength: shared and individual, horizontal and vertical motions, respectively. The individual motion strengths reach zero for the two horizontal moving dots (blue and gree), but is close to maximum for the middle dot (orange).(c) The Duncker wheel experiment II (Duncker, 1929) with a central dot (orange) and a dot on a circle (blue). (d) The model recognises shared motion (horizontal motion) and an individual element (rotational motion) of the rotating dot. The plots are colour coded, matching the colours of the objects with their corresponding plot lines.

3.2 The Balance between sensory data and prior knowledge in coherent motion detection

The model's performance in detecting coherent motion in the RDK task was evaluated by measuring the time to convergence. Variations in the time to convergence of the model as noise levels in the stimuli change is illustrated in Fig. 8, along with the impact of prior knowledge on the model's performance. Fig. 8a indicates a strong relationship between noise proportion and convergence time, with increasing noise leading to prolonged convergence duration. We conducted this experiment for window durations of 0, 10, 30, 50, 70, and 90 ms, with 150 trials per window duration.

Fig. 8a depicts the influence of noise proportion on convergence time across two historical windows with durations of 50 ms and 30 ms, with the longer window yielding longer convergence times compared to the shorter window. The shaded areas represent the standard deviation across the 150 trials.

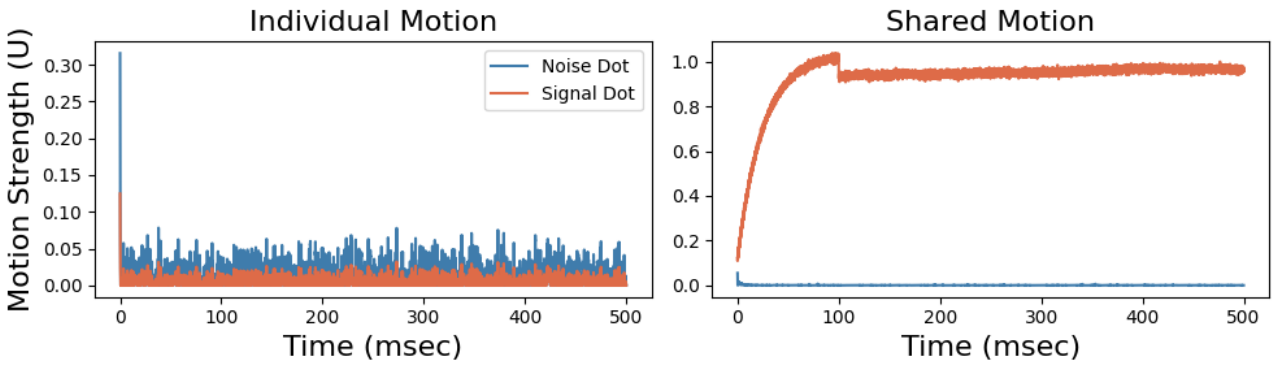
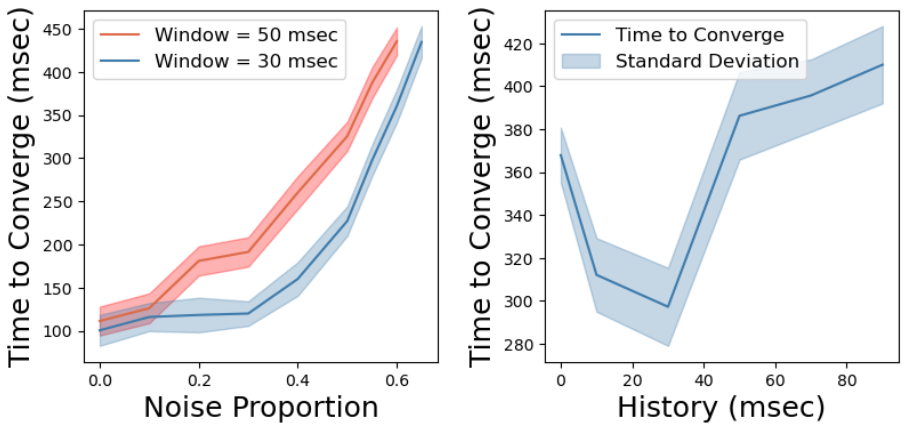


Figure 7: Individual and shared motion in a random dot kinematogram, experiment III. Blue lines represent the motion strengths of noise dots, while the orange lines depict the motion strengths of signal dots with coherent motion.

Note the reduced scale axis of the individual motion plot, used to highlight the low motion strengths of the dots.

The relationship between the historical window duration and convergence time is shown in Fig. 8b for a noise proportion of 0.6. There was an initial decrease in convergence time with increasing historical window duration up to a certain point (30 ms), after which the convergence time began to rise. This suggests an optimal range of historical information that facilitates the detection of the coherent motion. Beyond this optimal range, an excess of historical data appears to hinder the model's performance.



(a) **(b)**

Figure 8: Effects of noise and historical window duration on the time to convergence of model (experiment III). (a) The time to converge as a function of noise proportion for two different historical window durations: 50 msec and 30 msec.(b) Convergence time's dependence on the historical window duration. The shaded areas around the lines represent the standard deviations across 150 trials for each window duration.

3.3 Surround suppression and the detection of coherent motion

The performance of the model was evaluated by modifying the the influence of surround suppression on the detection of coherent motion in the RDK task (experiment III). Specifically, we evaluated the effect of modifying the threshold in the surround suppression mechanism. A higher threshold value implies that the motion in the centre and surround needs to be signifi-

cantly different for the surround suppression to be activated.

Fig. 9 summarizes the model's performance for the RDK task. The axes correspond to the threshold of surround suppression mechanism and noise proportion. The color gradient indicates the duration of convergence time, with cooler colors showing shorter times and warmer colors indicating longer times. The figure illustrates that, as the noise level increased, the time to converge also increased, suggesting that the model required more time to process the coherent motion direction in the presence of higher noise. Additionally, the influence of the surround suppression threshold on convergence time is evident, with lower thresholds corresponding to longer convergence times. Increasing the surround suppression threshold to a sufficiently high level improves the performance of the model in noisy conditions and prevents excessive changes in surround suppression due to noise.

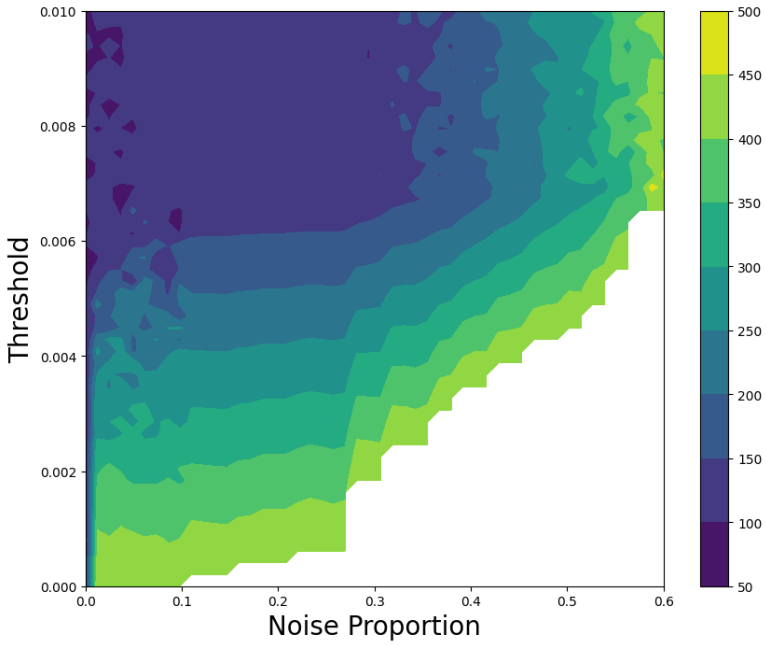


Figure 9: Effects of noise level and surround suppression threshold on model convergence time in the RDK task, experiment III. The color indicates the convergence time of the model in msec, with cooler colors representing shorter times and warmer colors indicating longer times.

4 Discussion

In this study, the results of a biologically plausible computational model for coherent motion detection based on the predictive coding algorithm are presented. The model not only demonstrates its capability to extract the structure of motion in the visual field by decomposing the motion into individual and shared motion sources, but it also reveals some interesting findings that could be related to motion detection deficiencies in schizophrenia.

4.1 Decomposing the Motion Structure into Shared and Individual Motion Sources

The results obtained here from testing the model with various stimuli underscore the significance of incorporating a surround suppression mechanism in modeling the process of motion detection. While classical predictive coding offers some amount of surround suppression, it inadequately captures some characteristics of extra-classical receptive field effects necessary for accurately interpreting the complex interplay of object motion with its neighbors. This includes the integrative and antagonistic effects observed in neurophysiological studies, where the motion in the surround affects the activity of the neurons (Braddick, 1993; Zarei Eskikand et al., 2020; Eskikand et al., 2023). This influence is either antagonistic, where the activity of the neurons increases when the motion in the center and surround are in different directions, or integrative, where the activity of the neurons is suppressed when the motion in the center and the surround are in different directions (Braddick, 1993; Tadin et al., 2003, 2019; Eskikand et al., 2023; Born, 2000; Huang et al., 2007). By incorporating this integrative effect, our method offers a more nuanced and comprehensive model for analysing motion interactions within dynamic scenes. This inclusion is crucial for effectively decomposing the motion structure and extracting shared motion within the visual field. The surround suppression mechanism assesses motion within both the center and the surround of the receptive field, checking for consistency in the visual field. If motion in the center and the surround occurs in different directions, it suppresses neuronal activity (Born, 2000; Huang et al., 2007; Zarei Eskikand et al., 2020). Previous studies have shown that the center-surround mechanism plays a pivotal role in figure-ground segregation to distinguishing an object from its background, edge detection, and discerning the movement of two overlapping or nearby objects in the receptive field (Braddick, 1993; Tadin et al., 2003, 2019; Eskikand et al., 2023). The model presented here underscores the importance of this process in detecting global (shared) motion in the visual field. This is evident through the implementation of the center-surround mechanism in the model based on the predictive coding algorithm.

Tests were conducted on the model using three distinct types of stimuli known for their significant contributions to the study of global motion detection in the visual cortex: (I) the Johansson experiment, (II) Dunker wheel, and (III) random dot kinematogram. The Johansson experiment I shows how the brain integrates information from multiple points into a coherent perception of global motion. The Dunker wheel experiment II demonstrates the tendency of our brain to break down multiple complex movements into recognizable, coherent motion, such as that of a rotating wheel in this experiment. The random dot kinematogram experiment III allows motion coherence to be manipulated by changing the level of noise in order to study how the brain integrates or segregates motion signals from different spatial locations. The model developed here exhibits proficiency in decomposing motions within the visual field into shared and individual motions of objects across all of these stimuli, replicating human brain perception in the detection of global motion (see Fig. 6b and Fig. 6d). The implementation of a model based on the predictive coding algorithm also illustrates how the brain predicts

an object's future position by considering current and past states of motion, aligning with kinematic principles.

4.2 Balancing sensory data and prior knowledge in the detection of coherent motion and the relationship to schizophrenia.

The modeling results presented here show that increasing the level of noise in the random dot kinematogram experiment III leads to an increase in the convergence time for the model to detect the global motions of the dots (see Fig. 8a). This simulates the psychophysics experiment conducted by Spencer et al. (2013), where they showed that participants require a longer duration of the stimuli in order to detect global motion for larger levels of noise. In this experiment, they also compared the performance of healthy controls with patients with schizophrenia in the detection of coherent motions of random dots. They observed that patients with schizophrenia require a longer duration compared to healthy counterparts to detect the global motions of random dots across different levels of noise. Using the model developed here, we simulated this observed deficiency in the detection of coherent motion that they observed in patients with schizophrenia. Our model shows that increasing the influence of prior knowledge in the motion detection process will have a significant impact on the convergence time of the model in detecting motion. Increasing the duration of the window for utilizing past information in motion detection was found to prolong the convergence time of the model, as shown in Fig. 8a. This simulation mirrors the differences observed in the required duration of stimuli during the detection of coherent motion in patients with schizophrenia compared to the healthy controls.

To analyze this phenomenon further, we investigated the performance of the model by varying the duration of the window for utilizing past information from 0 to 100 ms. The results revealed an intriguing finding, namely that increasing the window duration to a minimum threshold of 30 ms optimizes the model's performance, resulting in a reduced time to convergence. However, an increase in the window duration beyond this 30 ms threshold gradually lengthens the convergence time, ultimately leading to a point where the model may fail to converge, as illustrated in Fig. 8b. This is very interesting as it highlights the importance of the balance between prior knowledge and incoming sensory data in the detection of coherent motion. It shows that the model should not overly rely on past information (i.e., a very large window duration) to detect coherent motion. On the other hand, for optimal performance the model requires some past information to prevent it from excessively responding to noise. Having very little influence from past information (i.e., a very small window duration) will result in a longer convergence of the model to detect coherent motion, as it is overly disturbed by the noise in the stimuli.

These results could explain the motion detection deficiency observed in patients with schizophrenia. There are studies that have shown that the distortion in perception and beliefs among individuals with schizophrenia stems from the accurate encoding of prior knowledge, rendering them less responsive to feedback (i.e., under-learning) (Corlett et al., 2019; Horga and

Abi-Dargham, 2019; Nassar et al., 2021). On the other hand, there are theories suggesting that these abnormalities in motion perception could stem from interpreting prediction errors as significant changes, leading to an exaggerated response to noisy information and rapid adaptation in a changing environment compared to controls (i.e., over-learning) (Kaplan et al., 2016; Stephan et al., 2016). A recent study has demonstrated the coexistence of these conflicting theories, revealing that the learning process in these patients is not stationary and depends on the statistical context. The applied task exhibited both over-learning and under-learning in different situations (Nassar et al., 2021). The model presented here suggests that the motion detection deficiency observed in patients with schizophrenia could result from the same mechanism observed in either overly relying on past information (under-learning) or excessively updating with respect to feedback (over-learning).

4.3 Surround suppression and the detection of coherent motion

The results presented in this study demonstrate the impact of surround suppression on the detection of coherent motion. We observed that increasing the threshold for surround suppression improves the performance of the model in detecting coherent motion. Decreasing the threshold for surround suppression suppresses the activity of neurons, consequently leading to a longer convergence time for the model. This suggests an experiment that could be tested in patients with schizophrenia compared to control cases, focusing on the surround suppression mechanism in the detection of coherent motion. This could provide a new perspective on the underlying neural circuitry of the surround suppression mechanism and how it is affected in patients with schizophrenia.

References

Bill, J., Gershman, S. J., and Drugowitsch, J. (2022). Visual motion perception as online hierarchical inference. *Nature Communications*, 13(1):7403. Number: 1 Publisher: Nature Publishing Group.

Born, R. T. (2000). Center-surround interactions in the middle temporal visual area of the owl monkey. *Journal of Neurophysiology*, 84(5):2658–2669.

Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*, 16(7):263–268.

Chen, Y., Grossman, E. D., Bidwell, L. C., Yurgelun-Todd, D., Gruber, S. A., Levy, D. L., Nakayama, K., and Holzman, P. S. (2008). Differential activation patterns of occipital and prefrontal cortices during motion processing: Evidence from normal and schizophrenic brains. *Cognitive, affective & behavioral neuroscience*, 8(3):293. Publisher: NIH Public Access.

Chen, Y., Palafox, G. P., Nakayama, K., Levy, D. L., Matthysse, S., and Holzman, P. S. (1999). Motion perception in schizophrenia. *Archives of General Psychiatry*, 56(2):149–154.

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., and Powers, A. R. (2019). Hallucinations and Strong Priors. *Trends in Cognitive Sciences*, 23(2):114–127.

DiVita, J. C. and Rock, I. (1997). A belongingness principle of motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, 23(5):1343. Publisher: American Psychological Association.

Duncker, K. (1929). Über induzierte Bewegung. Ein Beitrag zur Theorie optisch wahrgenommener Bewegung. [Concerning induced movement. A contribution to the theory of visually perceived movement.]. *Psychologische Forschung*, 12:180–259. Place: Germany Publisher: Springer.

Eskikand, P. Z., Grayden, D. B., Kameneva, T., Burkitt, A. N., and Ibbotson, M. R. (2023). Understanding visual processing of motion: completing the picture using experimentally driven computational models of MT. *Reviews in the Neurosciences*. Publisher: De Gruyter.

Eskikand, P. Z., Kameneva, T., Ibbotson, M. R., Burkitt, A. N., and Grayden, D. B. (2016). A Possible Role for End-Stopped V1 Neurons in the Perception of Motion: A Computational Model. *PLOS ONE*, 11(10):e0164813. Publisher: Public Library of Science.

Gershman, S. J., Tenenbaum, J. B., and Jäkel, F. (2016). Discovering hierarchical motion structure. *Vision Research*, 126:232–241.

Gogel, W. C. (1974). Relative motion and the adjacency principle. *Quarterly Journal of Experimental Psychology*, 26(3):425–437. Publisher: SAGE Publications Sage UK: London, England.

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. Wiley, Oxford, England. Pages: xix, 335.

Horga, G. and Abi-Dargham, A. (2019). An integrative framework for perceptual disturbances in psychosis. *Nature Reviews Neuroscience*, 20(12):763–778. Number: 12 Publisher: Nature Publishing Group.

Huang, X., Albright, T. D., and Stoner, G. R. (2007). Adaptive Surround Modulation in Cortical Area MT. *Neuron*, 53(5):761–770.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243. Publisher: Wiley Online Library.

Institute of Health Metrics and Evaluation (2021). Global health data exchange (GHDx). Publisher: Institute of Health Metrics and Evaluation Seattla, WA, USA.

Johansson, G. (1950). *Configurations in event perception; an experimental study*. Configurations in event perception; an experimental study. Almquist & Wiksells, Oxford, England. Pages: 226.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211.

Kaplan, C. M., Saha, D., Molina, J. L., Hockeimer, W. D., Postell, E. M., Apud, J. A., Weinberger, D. R., and Tan, H. Y. (2016). Estimating changing contexts in schizophrenia. *Brain: A Journal of Neurology*, 139(Pt 7):2082–2095.

Kim, J., Norton, D., McBain, R., Ongur, D., and Chen, Y. (2013). Deficient biological motion perception in schizophrenia: results from a motion noise paradigm. *Frontiers in Psychology*, 4:391.

Kim, J. N. and Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience*, 2(2):176–185.

Millidge, B., Seth, A., and Buckley, C. L. (2021). Predictive coding: a theoretical and experimental review. *arXiv preprint arXiv:2107.12979*.

Millidge, B., Tang, M., Osanlouy, M., Harper, N. S., and Bogacz, R. (2023). Predictive coding networks for temporal prediction. *bioRxiv*, pages 2023–05. Publisher: Cold Spring Harbor Laboratory.

Nassar, M. R., Waltz, J. A., Albrecht, M. A., Gold, J. M., and Frank, M. J. (2021). All or nothing belief updating in patients with schizophrenia reduces precision and flexibility of beliefs. *Brain*, 144(3):1013–1029.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87. Publisher: Nature Publishing Group.

Restle, F. (1979). Coding theory of the perception of motion configurations. *Psychological Review*, 86(1):1. Publisher: American Psychological Association.

Shum, K. H. and Wolford, G. L. (1983). A quantitative study of perceptual vector analysis. *Perception & Psychophysics*, 34(1):17–24.

Spencer, J., Sekuler, A., Bennett, P., and Christensen, B. (2013). Contribution of coherent motion to the perception of biological motion among persons with Schizophrenia. *Frontiers in Psychology*, 4.

Stephan, K. E., Diaconescu, A. O., and Iglesias, S. (2016). Bayesian inference, dysconnectivity and neuromodulation in schizophrenia. *Brain*, 139(Pt 7):1874–1876.

Tadin, D., Lappin, J. S., Gilroy, L. A., and Blake, R. (2003). Perceptual consequences of centre-surround antagonism in visual motion processing. *Nature*, 424(6946):312–315.

Tadin, D., Park, W. J., Dieter, K. C., Melnick, M. D., Lappin, J. S., and Blake, R. (2019). Spatial suppression promotes rapid figure-ground segmentation of moving objects. *Nature Communications*, 10(1):2732. Number: 1 Publisher: Nature Publishing Group.

Tibber, M., Anderson, E., Bobin, T., Antonova, E., Seabright, A., Wright, B., Carlin, P., Shergill, S., and Dakin, S. (2013). Visual Surround Suppression in Schizophrenia. *Frontiers in Psychology*, 4.

Tsui, J. M. G., Hunter, J. N., Born, R. T., and Pack, C. C. (2010). The Role of V1 Surround Suppression in MT Motion Integration. *Journal of Neurophysiology*, 103(6):3123–3138.

Yang, S., Bill, J., Drugowitsch, J., and Gershman, S. J. (2021). Human visual motion perception shows hallmarks of Bayesian structural inference. *Scientific Reports*, 11:3714.

Zarei Eskikand, P., Kameneva, T., Burkitt, A. N., Grayden, D. B., and Ibbotson, M. R. (2020). Adaptive Surround Modulation of MT Neurons: A Computational Model. *Frontiers in Neural Circuits*, 14.