Neural mechanisms of visual motion extrapolation

William Turner[1,2*], Charlie Sexton[2], and Hinze Hogendoorn[1,2]

[1]Queensland University of Technology, Brisbane, Australia, 4059

[2]The University of Melbourne, Melbourne, Australia, 3010

[*]Corresponding Author

Email: w6.turner@qut.edu.au

Address: O Block, D Wing, Room 611, QUT Kelvin Grove, Brisbane, Australia, 4059

Conflict of interest: The authors declare no competing interests.

Abstract

Because neural processing takes time, the brain only has delayed access to sensory information. When localising moving objects this is problematic, as an object will have moved on by the time its position has been determined. Here, we consider predictive motion extrapolation as a fundamental delay-compensation strategy. From a population-coding perspective, we outline how extrapolation can be achieved by a forwards shift in the population-level activity distribution. We identify general mechanisms underlying such shifts, involving various asymmetries which facilitate the targeted 'enhancement' and/or 'dampening' of population-level activity. We classify these on the basis of their potential implementation (intra- vs inter-regional processes) and consider specific examples in different visual regions. We consider how motion extrapolation can be achieved during inter-regional signaling, and how asymmetric connectivity patterns which support extrapolation can emerge spontaneously from local synaptic learning rules. Finally, we consider how more abstract 'model-based' predictive strategies might be implemented. Overall, we present an integrative framework for understanding how the brain determines the real-time position of moving objects, despite neural delays.

Keywords:

Predictive processing; motion extrapolation; visual processing; neural delays

1. Introduction

Light falling on the photoreceptors in our eyes triggers waves of chemical and electrical activity in the nervous system. These carry information across the broad network of brain regions that makes up our 'visual system' to be processed and made sense of. Given the neural paths that must be travelled and the sense-making computations that are performed, the delay between light first arriving at the retina and the brain forming high-level perceptual representations spans many tens to hundreds of milliseconds. This raises a puzzling question: if visual processing inevitably incurs delays, how are we able to localize and accurately interact with moving objects?

Consider a professional tennis player facing a ball travelling at 180 km/h. Assuming it takes them 80 milliseconds to process positional information, in this time the ball will already have moved a further 4 metres. So how, then, are they able to hit it? Or, in more evolutionarily-relevant terms, how are hunters able to take down speeding prey? In both these scenarios, by the time the location of the moving object has been determined, it will no longer be there. The fact that such behaviours are nevertheless possible indicates that the brain must have developed ways of compensating for its own processing delays when localizing objects.

This review considers *predictive motion extrapolation* as a fundamental delay-compensation strategy employed by the brain. Broadly put, this involves using information about an object's past trajectory to infer its probable present position. In the past two and a half decades, the existence of motion extrapolation mechanisms in the visual system, and their potential perceptual consequences, has been strongly debated (Hogendoorn, 2020; Nijhawan, 1994, 2008; Nijhawan & Wu, 2009). Here, we argue that there is now clear evidence of a variety of neural mechanisms in the visual system that serve to extrapolate the represented position of moving objects. Drawing together past research, and highlighting important recent

advances, we will present an integrative framework for understanding how these mechanisms allow the brain to encode objects, not where they were, but where they (probably) now are.

2. Scope of the present review

This review focusses on neural motion extrapolation mechanisms in the visual system. While related predictive mechanisms have been shown to exist outside the visual system (e.g., in the hippocampus & motor regions) these will not be covered in the present article (but see Stachenfeld et al., 2017; Wolpert, 1997; Wolpert & Flanagan, 2001). The perceptual consequences of the reviewed mechanisms will also not be considered in detail (but see Hogendoorn, 2020; Nijhawan, 2008). The consciously perceived location of an object depends on numerous underlying factors, and multiple (differing) position representations for a single object can coexist in the brain (see Goodale & Milner, 1992; Lisi & Cavanagh, 2015; Liu et al., 2019). We therefore restrict our focus to neural mechanisms that predictively modulate object-position representations (be they conscious or unconscious), without laying claim to how these mechanisms ultimately affect perceptual experience. To begin, we will briefly review the magnitude and cause of processing delays in the visual system, which necessitate the existence of compensatory extrapolation mechanisms.

3. A brief review of visual delays: their magnitude and cause

It takes time for the brain to process visual information. But how long exactly does this take, and why do delays exist in the first place?

Photons of light entering the eye first activate photoreceptors at the innermost layer of the retina. These photoreceptors, grouped into rods and cones, convert light into chemical energy through the process of phototransduction. This energy is then transmitted to retinal ganglion cells (RGCs) via bipolar cells, which synapse with photoreceptors. Modulatory

activity along this pathway is provided by horizontal cells, which form lateral connections across photoreceptors, and amacrine cells, which form lateral connections across bipolar cells and RGCs. Besides RGCs, all activity in the retina is transmitted in the form of graded potentials; only when this energy reaches RGCs are action potentials produced and transmitted to the brain via the optic nerve.

RGCs produce spikes at latencies of ~20 to 70 ms after the onset of a visual stimulus (salamanders and rabbits – Berry et al., 1999; cats – Bolz et al., 1982; goldfish – Johnston & Lagnado, 2015). As information passes from RGCs to the brain, and to the many different regions of the visual system, this initial delay is further compounded by transmission and integration delays. Transmission delays are relatively small and arise due to the time it takes for electrical signals to travel along the axon and across synapses. For instance, there can be as little as a ~1-2 ms transmission delay between the optic chiasm and the lateral geniculate nucleus (LGN; rhesus monkey – Schiller & Malpeli, 1978), and synaptic transmission similarly can be as fast as ~1-2 ms. However, neurons typically integrate multiple synaptic inputs before producing a response, such that integration delays are usually much more substantial. Spatial and temporal integration help to differentiate meaningful signals from background noise, at the cost of an 'activation delay' (Baldo & Caticha, 2005) as neurons typically only reach firing threshold after integrating many inputs. For this reason, the time it takes for a given signal to travel through the visual system depends on the intensity of the input, the level of preceding neural activity, and the degree of integration in the cortical route being travelled.

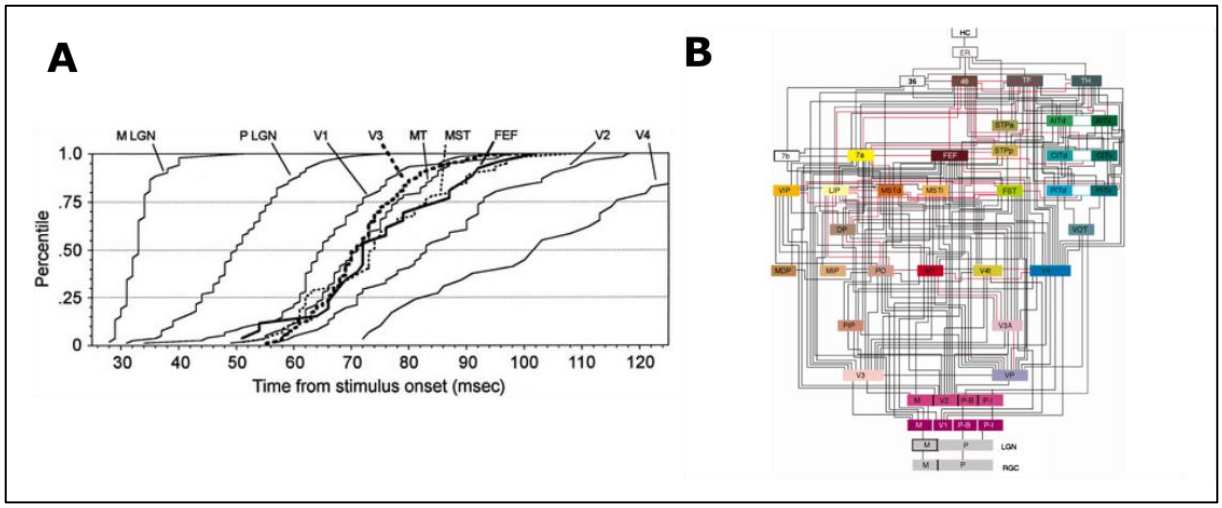

Figure 1. Neural delays across the visual hierarchy. A) Cumulative distributions of onset response latencies for cells in different regions of the macaque visual system — from Schmolesky et al. (1998) – *Journal of Neurophysiology*. **B)** Model diagram of the macaque visual system — from Felleman & Van Essen (1991) – *Cerebral Cortex*.

Figure 1A shows cumulative distributions of onset response latencies for neurons in select regions of the macaque visual system (Schmolesky et al., 1998). These data show that areas further along the visual hierarchy tend to have longer average response latencies, consistent with a compounding of delays. What is also apparent, however, is the high degree of variability both within and between regions. A number of cells in early regions, such as the LGN and V1, begin firing within just 20-40 ms, but many others, in both early and late regions, take between 50-100+ ms to begin responding. Similar variability has been observed in recordings from humans, with V1 latency estimates ranging from 26-100 ms (Inui & Kakigi, 2006). Considering the complex informational flow underlying visual processing (Figure 1B), it is apparent that there is no *single* fixed delay that must be compensated. Rather, variable delays accrue along each step of the way.

In and of themselves, delays are not always simply a nuisance. For example, in the (cat) LGN, distinct 'lagged' and non-lagged responses have been observed, suggesting that certain sub-populations of cells only becomes active after a fixed delay (note, this is irrespective of the upstream propagation delay; Dong & Atick, 1995; Saul & Humphrey, 1990). Crucially, delayed cells are thought to play a role in the temporal decorrelation of visual input (Dong & Atick, 1995), reminiscent of the spatial-decorrelation occurring in the retina (Atick & Redlich, 1992). More generally, many theoretical accounts of neural motion detection (e.g., Reichardt-type models, Hassenstein & Reichardt, 1956) explicitly employ transmission delays to enable forms of 'coincidence detection', registering motion in a given direction/velocity (see also Grimaldi & Perrinet, 2023; but see Heitmann & Ermentrout, 2020). Finally, more broadly again, delays (in combination with spike timing dependent plasticity) have been shown to play a crucial role in boosting the memory capacity of spiking neural networks (Izhikevich, 2006). However, as we outlined above, delays are problematic for time-sensitive interactions with dynamic environments, necessitating the existence of compensatory mechanisms, ideally on the local level.

Below, we consider specific examples of neural mechanisms, in different regions of the visual system, that compensate for delays by extrapolating the represented position of moving objects. To provide an integrative framework within which to understand these mechanisms, we will begin by teasing out a common thread running through them.

4. A general account of neural motion extrapolation

The visual system can be thought of as a network of interconnected neural populations set within distinct regions or 'layers' (retina, LGN, V1, etc.). Many of these layers are retinotopically organised, meaning that the retinal topology is broadly conserved. This is important as it means that smoothly moving stimuli sequentially activate neighboring cells,

and it allows for local modulatory processes to shape evoked activity. To extrapolate the position of a moving object within this network, really only one thing is required: at some point the encoded position of the object must shift from the originally stimulated retinotopic location to the probable present location of the stimulus, given its motion history and the intervening delay (Figure 2A). But how, in practice, can this be achieved?

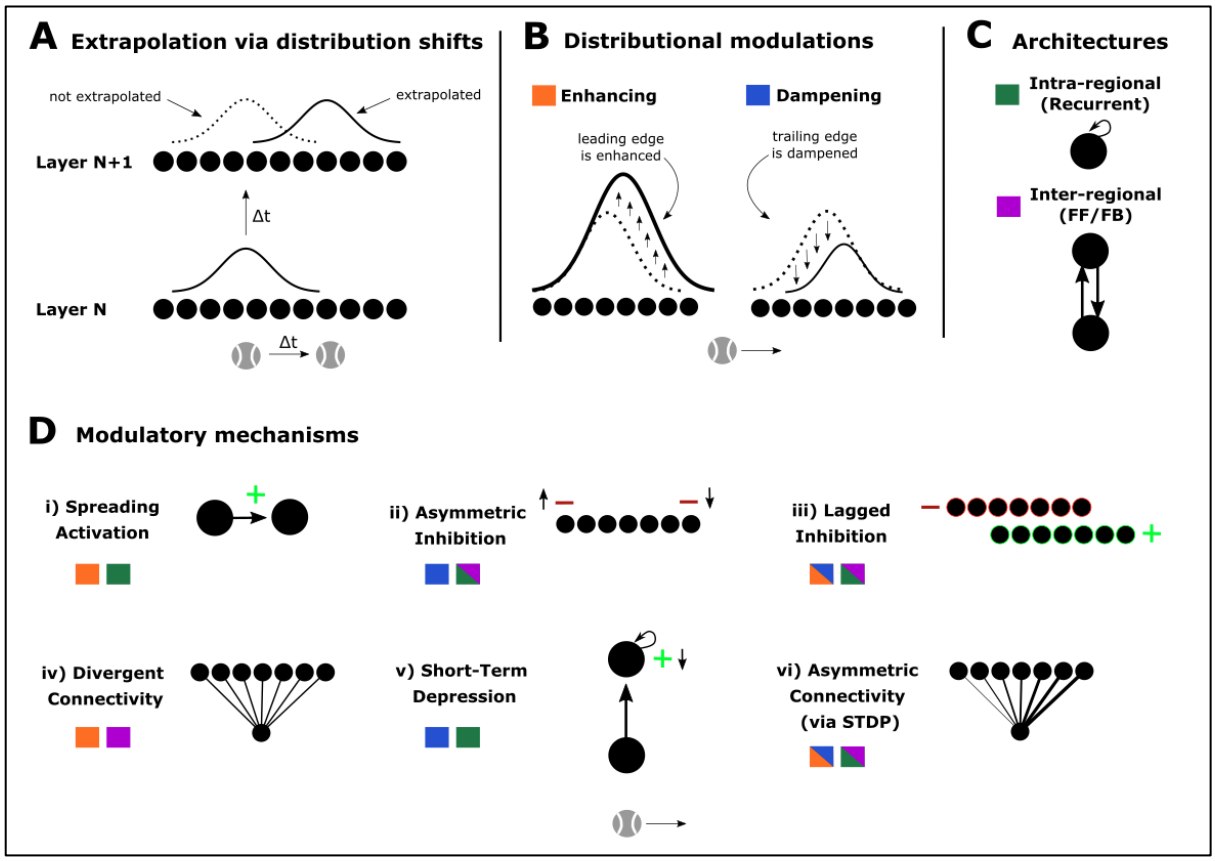

Figure 2. A general account of neural motion extrapolation. Motion extrapolation can be achieved by shifting the population-level distribution of evoked activity forwards along the stimulus' current motion trajectory. **A)** A moving stimulus (grey ball) generates a distribution of activity in layer N of the network. In the time it takes for this activity to be transmitted and registered in the next layer (N + 1), the ball continues moving. With no extrapolation, the distribution of activity evoked in layer N + 1 is not shifted to account for the intervening movement of the ball. With extrapolation, the distribution of activity shifts to account for this motion, encoding the current position of the ball despite the intervening delay. **B)** Neural extrapolation mechanisms can be classified on the basis of their modulatory effect(s). That is, whether they play a role in: 'enhancing' (orange) and/or 'dampening'

(blue) the leading/trailing edge of the evoked activity distribution. **C)** Mechanisms can also be classified according to whether they can be implemented via intra-regional (recurrent; green) and/or inter-regional (feedforward/feedback; purple) architectures. **D)** Illustrations of general neural mechanisms which support motion extrapolation. **i)** Spreading activation within a given region serves to enhance activity at the leading edge of the population-level distribution**. ii)** Spatially asymmetric inhibition (via intra- or inter-regional processes) serves to 'erase' the trailing edge of evoked activity. **iii)** Delayed inhibitory signals (within or between regions) can similarly erase the trailing edge of evoked activity. An enhancement of the leading edge also occurs due to an 'escape' from inhibition (relative to zero-lag inhibition; see Figure 5). **iv)** Divergent inter-regional connectivity facilitates responses at the leading edge of the activity distribution. **v)** Short-term synaptic depression (and related adaptation mechanisms, see Feuerriegel, 2023) dampens the trailing edge of evoked activity. **vi)** Asymmetric connectivity (either within or between regions) serves to 'diagonally transfer' activity (Nijhawan & Wu, 2009), simultaneously enhancing and dampening activity at the leading and trailing edges respectively.

Suppose that, in each retinotopically organised region of the visual system, the position of a moving object is encoded via some form of population code (Pouget et al., 2000). For example, individual neurons may code for specific regions of visual space, with the 'population position estimate' given by the average of the positions coded for by all active neurons, weighted by their activity strength (e.g., Erlhagen et al., 1999; Georgopoulos et al., 1986; see Section 6.2 for further discussion). From this viewpoint, a common thread can be found running through the neural mechanisms that we will review. They all involve an underlying *asymmetry* which shifts the population position estimate in the direction of stimulus motion. These mechanisms can be classified according to 1) the modulatory effect(s) they have on population-level activity – in particular, whether they play a role in enhancing activity at the leading edge of the population-level activity distribution, and/or dampening activity at the trailing edge (Figure 2B). These mechanisms can be further classified 2) on the basis of their

potential implementation (Figure 2C) – specifically, whether they can operate intra-regionally (recurrent processes) or inter-regionally (feedforward/feedback processes). The various mechanisms we will review (Figure 2D) all ultimately serve to shift the weight of evoked activity forwards along the trajectory of motion, extrapolating the population position estimate.

In the following sections, we will work our way along the visual stream, and consider specific examples of such mechanisms, at the neural circuit level. Beginning in the retina, we will consider how 'enhancing' and 'dampening' mechanisms predictively shift the encoded position of moving objects along their trajectory of motion. We will then consider related mechanisms in the brain. Next, we will consider how motion extrapolation can be achieved during inter-regional information transfer along feedback and feedforward pathways. We will highlight how the asymmetric connectivity patterns which best enable motion extrapolation can spontaneously emerge from local synaptic learning rules. Finally, we will broaden our focus and briefly consider how more abstract 'model-based' extrapolation strategies (e.g., recursive Bayesian estimation) might be neurally implemented.

5. Motion extrapolation in the retina

The first stage at which neural motion extrapolation can occur is the point at which the visual world is transformed into chemical and electrical energy: the retina. In this section, we consider how 'enhancing' and 'dampening' mechanisms within the retina facilitate motion extrapolation during initial visual processing. Drawing upon recent work, we will also consider the question of exactly *where* in the visual system motion extrapolation first begins.

5.1. Evidence of retinal extrapolation

Early work on retinal motion extrapolation compared the activity evoked by moving and static stimuli in (non-direction selective) retinal ganglion cells (RGCs) of salamander and

rabbit (Berry et al., 1999). Strikingly, the population-level activity evoked by moving stimuli was found to be shifted in the direction of motion, compared to activity evoked by static stimuli. When presented with a static stimulus, RGCs fired after an average delay of ~50ms. However, when exposed to a moving stimulus, these same cells began firing ~400 ms before the stimulus reached their receptive field centre – a phenomenon termed 'motion anticipation'. This led the peak of the population-level activity distribution to be roughly aligned with the leading edge of the stimulus, suggesting that rather than encoding the originally stimulated retinotopic location, RGCs were encoding the real-time position of the stimulus.

This seminal observation – that RGCs can anticipate the arrival of a moving stimulus – has been replicated several times, for a range of RGC subtypes and species (goldfish – Johnston & Lagnado, 2015; macaque – Liu et al., 2021; salamander and mouse – Schwartz et al., 2007; mouse – Trenholm et al., 2013). Most recently, it has been shown that both RGCs and upstream bipolar cells predictively encode a variety of different forms of motion (two and three point spatio-temporal correlations; Liu et al., 2021). The fact that bipolar cell activity predictively encodes motion information is particularly striking as it indicates that prediction must begin before or at the second synapse of the visual system.

But what are the specific neural mechanisms which enable this? Below, we will outline how the lateral spreading of activity between cells can prime responses ahead of a moving stimulus, driving anticipatory responding via an enhancement of the leading edge of evoked activity. We will then consider how inhibitory mechanisms refine these predictions by dampening the trailing edge of the population response.

5.2. Enhancing mechanisms: spreading activation

In many regions of the visual system, neighbouring cells are electrically 'coupled', meaning they share activity with one another laterally (Figure 3B). In the retina, the spreading

of activity between coupled cells has been shown to drive motion anticipation (Trenholm et al. 2013; Liu et al., 2021). On theoretical grounds, Liu et al. (2021) showed how coupling between bipolar cells, combined with a non-linear output function, could facilitate the encoding and transmission of predictive motion information. During visual motion, spatiotemporal correlations lead to the sequential stimulation of neighbouring bipolar cells. For cells that are electrically coupled, some current will spread out from an active cell to its neighbours. This causes a priming effect, leading to faster and greater depolarization in neighbouring cells that receive subsequent feed-forward excitation. Combined with an output thresholding mechanism, the overall effect is that coupled cells encode spatiotemporally correlated signals, while minimizing the encoding of uncorrelated inputs that fall below the output threshold.

Interestingly, photoreceptors have long been known to display direct electrical coupling, sometimes with as many as 200 other cells (Baylor et al., 1971; Copenhagen & Owen, 1976; DeVries et al., 2002). The basic building blocks for motion extrapolation are therefore present at the very start of the visual cascade. To our knowledge, however, the potential for anticipatory responding in photoreceptors has not been directly investigated. Bipolar cell outputs are therefore the earliest point at which predictive motion signaling has been observed (Liu et al. 2021).

While laterally spreading activity may be symmetric (i.e. a cell primes its neighbours omnidirectionally), in the context of motion this has asymmetric consequences for population-level activity (Liu et al., 2021). Specifically, for a moving object, spreading activation works as an extrapolation mechanism by selectively enhancing neural responses at the leading edge of the population-level response (Figure 3B). Under a population coding scheme, this causes a forward shift in the population position estimate. A relevant feature of this is that, when paired with a nonlinear output function, neurons that receive sub-threshold priming but not subsequent feed-forward input will not necessarily become active. This prevents unmet predictions (e.g.,

in the areas surrounding an object's trajectory) from being passed up the visual hierarchy. In fact, the requisite conjunction between feed-forward and spreading activity in theory means that the extrapolation of non-linear motion can be achieved. This is because it is ultimately the driven signal which determines where the benefit of sub-threshold priming is expressed.

For a given neural delay, the speed an object is moving dictates the extent of the spatial lag incurred – the faster an object moves, the greater the lag. An effective anticipatory response must therefore scale with stimulus velocity. To this end, Trenholm et al. (2013) showed how spreading activation can facilitate near-perfect 'lag-normalization' in electrically coupled (direction-selective) RGCs, across a range of stimulus velocities (Figure 3C). For uncoupled cells, they found that responses to a moving bar lagged as a function of its velocity. Conversely, coupled cells were found to begin responding when the bar reached an almost constant location on the retina relative to the cell, regardless of stimulus velocity. With a simple computational model, they showed that this remarkable 'lag-normalization' effect can be captured by the combined effects of velocity-dependence in the ganglion cell response (i.e., stronger responses to faster stimuli) and a gradual compounding of activity shared laterally between neighbouring cells (i.e., if a cell is initially primed, it in turn primes its neighbours more quickly, with this effect building up across successive neurons).

Spreading activation as a predictive mechanism is, however, incomplete. Firstly, the spatial extent of prediction that can be achieved via subthreshold spreading activation is limited. With non-linear thresholding, the system can only speed up to the moment that some afferent signal arrives and is further constrained by the spatial extent of local coupling. Secondly, spreading activation does little to limit activity at the trailing edge of evoked activity. As such, while coupled ganglion cells show lag-normalized response *onsets*, the timing of their peak response still displays a velocity-dependent lag (Trenholm et al., 2013, Figure 3C). To unambiguously encode the present location of a moving stimulus, the trailing edge of activity, which corresponds to the object's outdated prior location, should ideally be 'erased'. In the following section, we consider how various 'dampening' mechanisms can accomplish this.

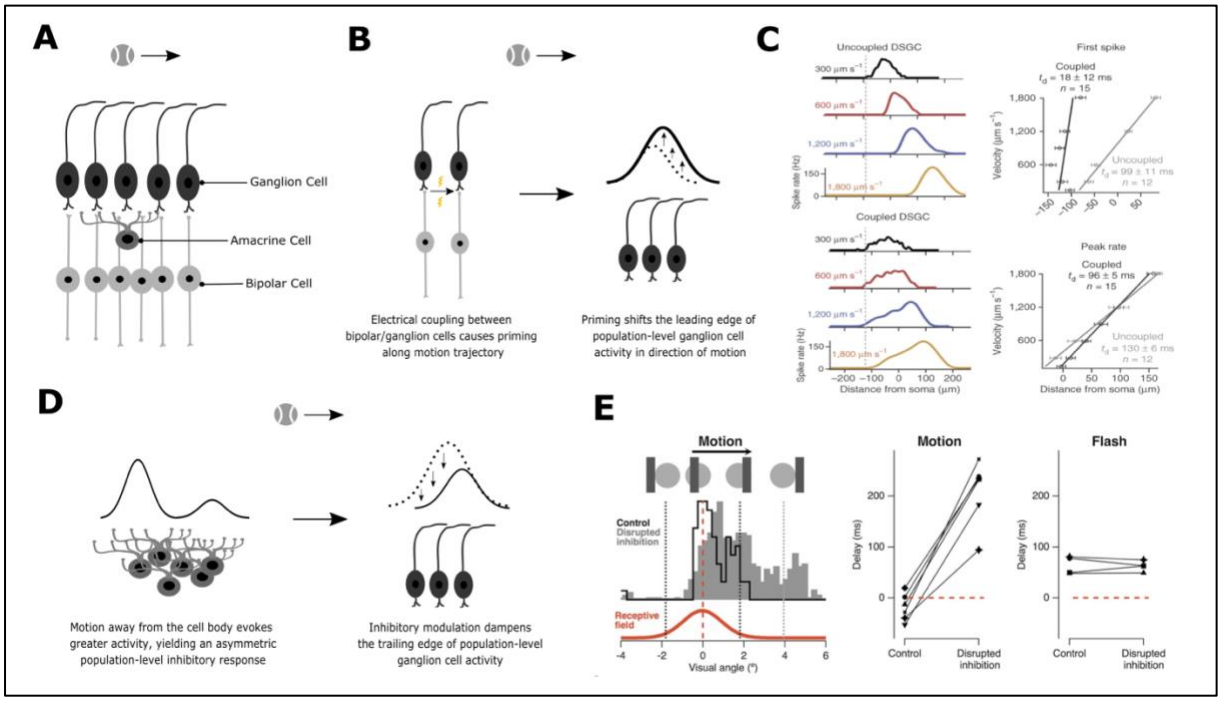

Figure 3. Retinal motion extrapolation. A) A subsection of the retinal network showing connections between bipolar, ganglion, and amacrine cells. **B)** Spreading activation as an excitatory extrapolation mechanism in the retina. Electrical coupling between bipolar/ganglion cells primes cells immediately along the current motion trajectory, causing a forwards shift in the leading edge of population-level ganglion cell activity. **C)** Empirical evidence of spreading activation via electrically coupled direction-

selective ganglion cells in mouse retina — from Trenholm et al. (2013) – *Nature Neuroscience*. The left panels show the activity profiles of uncoupled (top) and coupled (bottom) direction selective ganglion cells (DSGCs). Coupled cells display 'lag-normalization' in their response onsets and begin responding when a moving stimulus reaches a fixed distance from their soma, regardless of its speed. Conversely, the uncoupled cells display a speed dependent lag in the onset of firing (no delay compensation). The right panels show aggregate estimates of firing onset (left) and peak (bottom), where the dynamics of coupled and uncoupled cells can be directly compared. **D)** Amacrine cells as an inhibitory extrapolation mechanism in the retina. **E)** Empirical evidence of feed-forward inhibition as a driver of motion extrapolation in the goldfish retina — from Johnston & Lagnado (2015) – *eLife*. Peak firing occurs with ~0 delay relative to the arrival of the stimulus. However, when feedforward inhibition is disrupted, peak firing significantly lags the stimulus. Note that the onset of firing is, however, largely unaffected.

5.3. Dampening mechanisms: dynamic gain control

In their seminal investigation into retinal motion extrapolation, Berry et al. (1999) found that the degree of alignment between peak RGC activity and the stimulus was dependent on stimulus contrast. High contrast stimuli generated peaks of activity that roughly aligned with their leading edge, while low contrast stimuli generated peaks that lagged behind. To explain these observations, they invoked a contrast-dependent gain control mechanism. Under this account, if a cell receives strong input for an extended period of time its input 'gain' – that is, its responsiveness to further stimulation – is reduced. This dampens the cells response to the trailing edge of the stimulus, re-aligning population-level activity with the true position of the external stimulus. Strikingly, this mechanism breaks down at low image contrasts or very high movement speeds – where there is insufficient stimulation to trigger the gain control mechanism – at roughly the same points where human perceptual performance also becomes impaired (Berry et al., 1999).

Dynamic gain control features in many other models of retinal motion extrapolation (e.g., Chen et al., 2013; Trenholm et al., 2013). An important question has therefore been how dynamic gain control is actually implemented at the level of neural circuits. On theoretical grounds, it has been argued that feed-forward inhibition from starburst amacrine cells might play a role in modulating RGC gain (Nijhawan & Wu, 2009). This was based on the fact that different sections of their dendrites have different preferred motion directions, with direction-dependent activity profiles. Specifically, they produce stronger responses to motion away from their cell body than toward, leading them to asymmetrically inhibit RGCs in response to moving stimuli (Euler et al., 2002).

To illustrate this, consider a group of amacrine cells responding to a stimulus moving from left to right (Figure 3D). The population-level response of these neurons will be bi-modal, with a stronger response on the left side than the right. This is because cells on the left are exposed to motion that is predominately away from their cell bodies, generating stronger inhibitory activity, while cells on the right are exposed to motion which is predominantly towards their cell bodies, generating weaker activity. This asymmetry leads amacrine cells to more strongly inhibit RGC responses at the tail of the population-level response, shifting the weight of the activity distribution in the direction of motion.

Recently, it has been directly demonstrated that feed-forward inhibition from amacrine cells drives anticipatory responding (Johnston & Lagnado, 2015; Menz et al., 2020). Firstly, amacrine cell activity profiles have been shown to possess the spatiotemporal features required to implement dynamic gain control (Menz et al., 2020). More strikingly, however, the selective disruption of inhibitory inputs to single RGCs has been shown to abolish anticipatory activity (Johnston & Lagnado, 2015). Specifically, the blocking of feedforward inhibition shifts the peak firing rate of ganglion cells from being roughly aligned with the leading edge of the

stimulus to lagging by ~200 ms (although the onset of activity is largely unaffected, Figure 3E).

Passive interactions between inhibitory and excitatory signals within the dendrites of ganglion cells can also explain why inhibitory signals more strongly affect the trailing edge of stimulus-evoked activity (Johnston & Lagnado, 2015). The core idea here is rather simple. For inhibitory inputs to be most effective they must be located between excitatory input and the soma and must be activated before the distal excitatory synapse – allowing them to 'block' incoming excitatory signals from reaching the soma (Koch et al., 1983). To meet these conditions, two things are required. Firstly, inhibitory synapses must outnumber excitatory synapses, with both being randomly distributed. This will mean that, on average, for any excitatory synapse on the dendritic tree, there are a greater number of inhibitory synapses located between it and the cell body. Secondly, stimuli must be moving away from the ganglion cell body, meaning they activate the proximal inhibitory synapse before the more distal excitatory synapse (Johnston & Lagnado, 2015 see their Figure 6B). The latter condition is what causes inhibitory inputs to most strongly affect the trailing edge of the population response, as the stimulus will be moving away from the cell bodies of the ganglion cells underlying it.

The differing accounts for how asymmetric inhibitory profiles arise in the retina are not mutually exclusive. Rather they demonstrate how, viewed from two different perspectives — one population-level (Nijhawan & Wu, 2009) and one biophysical (Johnston & Lagnado, 2015) — the retinal circuitry possesses features which naturally facilitate motion extrapolation. By dampening activity at the trailing edge of the RGC activity distribution, this inhibitory modulation shifts the population position estimate forwards, predictively extrapolating an object's encoded position.

6. Motion extrapolation mechanism in the brain

Given the variety of mechanisms we have discussed, one might wonder whether all necessary extrapolation occurs in the retina, with delay-compensated representations sent to the brain to undergo further processing but not further extrapolation. This is not the case.

Even assuming perfect compensation can be achieved during retinal processing, delays will continue to accrue during subsequent processing. Normatively, this means that further compensation, and thus the involvement of additional extrapolation mechanisms, is required. Empirically, while extrapolation occurs in the retina, extrapolation can also occur without retinal involvement. We know this for two reasons. Firstly, people perceptually extrapolate the position of objects which move into their blind spot, where no photoreceptors are stimulated, and hence no retinal activity is generated (Maus & Nijhawan, 2008). Secondly, perceptual illusions driven by motion extrapolation have been shown to occur for 'cyclopean stimuli' (Nieman et al., 2006). These require the combining of signals from both eyes to be seen – a process which occurs in the cortex, not the retina. This suggests the existence of extrapolation mechanisms in cortical visual regions.

6.1. Evidence of extrapolation in the brain

As in the retina, neurons in the brain have been shown to respond more quickly to moving than to static stimuli, effectively 'anticipating' their arrival. This has been observed in cat LGN (Orban et al., 1985) and area 17 (Jancke et al., 2004; Orban et al., 1985), as well as macaque area V1 (Benvenuti et al., 2020; Guo et al., 2007; Subramaniyan et al., 2018) and V4 (Sundberg et al., 2006).

Early work was carried out by Jancke et al (2004) who compared the population activity evoked by moving and static stimuli in cat area 17 (analogous to primate V1). As in the retina, representations of moving stimuli were found to be shifted in the direction of motion, relative

to static stimuli (see Figure 4A). Similar observations were subsequently made in macaque area V1 (Benvenuti et al., 2020; Subramaniyan et al., 2018) and V4 (Sundberg et al., 2006), with anticipatory activity, in some cases, building up hundreds of milliseconds before an object reached a neuron's receptive field centre. This is reminiscent of (and indeed may partially reflect) retinal activity dynamics, where cells begin to fire hundreds of milliseconds before a moving stimulus reaches their receptive field centre (Berry et al., 1999).

Notably, the degree to which cortical position representations have been observed to shift during motion has varied significantly across studies/species. For example, Jancke et al (2004) observed a peak latency shift of just ~16 ms (in cat area 17), whereas Benvenuti et al. (2020) observed significantly larger shifts in both onset (~400 ms) and peak latencies (in macaque V1), mirroring observations from rabbit and salamander RGCs (Berry et al., 1999). At present, evaluating whether such differences reflect true inter-species differences in the degree to which position signals are extrapolated remains difficult, given the limited number of recordings from any one species, and the wide variety of experimental and analytical techniques adopted. As such, this remains an important avenue for future research.

Evidence of cortical extrapolation effects also comes from non-invasive human EEG recordings. For example, stimuli embedded in 'apparent motion sequences' – that is, chains of spatially and temporally separated flashes which generate the percept of a moving object – have been found to be processed more rapidly than when presented in isolation (Blom et al., 2020, 2021; Hogendoorn & Burkitt, 2018; Turner et al., 2023). Relatedly, position representations for smoothly moving stimuli have been shown to be activated substantially earlier than for unpredictable flashed stimuli (Johnson et al., 2023). These studies provide evidence that object-position representations in the human visual system are similarly predictively 'pre-activated' in motion contexts.

In the following sections, we consider specific examples of cortical extrapolation mechanisms. Following the structure of the previous section, we will first consider examples of asymmetric 'enhancing' and 'dampening' mechanisms. Then, we will shift focus to consider how motion extrapolation can occur as information is passed between neural regions.

6.2. Enhancing mechanisms in the cortex

In primate neurophysiology experiments, various techniques have been used to reveal the activity evoked by moving stimuli. Recently, voltage sensitive dye imaging was used to directly reveal spreading waves of activity within monkey V1, in response to both apparent (Chemla et al., 2019; see Figure 4B) and smooth motion (Benvenuti et al., 2020). For apparent motion (sequences of spatially and temporally separated flashes), travelling waves of activity have been observed to spread outwards omni-directionally after the presentation of a stimulus, facilitating the response to subsequent stimuli (Chemla et al., 2019). We note that this excitatory effect (as well as the corresponding inhibitory effects they observed, see below) can be modelled in a neural network with fully symmetric receptive fields, with asymmetries in excitation/inhibition arising at the population-level, in response to driven input from subsequent stimuli.

For smooth motion, neurons in V1 exhibit a slow build-up of activity before stimulus arrival, with the extent of this build up depending on the length of preceding motion (Benvenuti et al., 2020). Following this early build-up of activity, neural responses peak around the time the stimulus reaches the cells receptive field centre (see their Figure 3C&4C). This provides some emerging evidence that full delay-compensation may be achieved for early cortical position representations. Similar observations (albeit of only partial compensation) have also been made in cat area 17 (Jancke et al., 2004). This anticipatory effect can be accounted for by assuming that neurons ahead of a moving stimulus receive lateral input from their neighbours,

prior to the feed-forward driven signal (Jancke & Erlhagen, 2010). If these signals are recursively integrated along the motion trajectory, a gradual build-up in predictive activity occurs (as in the retinal lag-normalization model developed by Trenholm et al., 2013). Strong evidence that lateral connections generate the trajectory-dependent build-up of anticipatory activity is given by the fact that the dependence of anticipatory activity on stimulation duration disappears for inter-hemifield motion; when an object transitions between hemifields, neighbouring neurons are no longer stimulated (Benvenuti et al., 2020).

With human EEG recordings, it has been shown that the second stimulus in an apparent motion sequence is processed more quickly than the first (Blom et al., 2021). This is striking because until two stimuli have appeared it should be impossible to determine a motion trajectory, meaning it should be impossible to predict the location of the second stimulus. Again, this latency advantage can be accounted for by assuming there is an omni-directional travelling wave of activity, which spreads out from the initially stimulated retinotopic location (Jancke & Erlhagen, 2010). When this combines with feed-forward activity from the second stimulus, firing will occur more rapidly. As we noted above, this cortical 'spreading activation' mechanism is similar to the one found to contribute to 'lag-normalization' in the retina (Trenholm et al., 2013). However, the existence of long-range horizontal connections in early visual brain regions means that spreading activation can potentially contribute to more spatially extended forms of motion extrapolation (Allman et al., 1985; Blom et al., 2020; Bringuier et al., 1999; Jancke et al., 2004; Jancke & Erlhagen, 2010).

Generally speaking, travelling activity waves are a widespread phenomenon in visual cortex (Sato et al., 2012). Of particular relevance here are the results of two recent theoretical studies, which shed light on the potential role that travelling waves play in facilitating neural motion extrapolation (Benigno et al., 2023; Heitmann & Ermentrout, 2020). Firstly, Heitmann & Ermentrout (see also Xie & Giese, 2002), showed that stimulus-locked travelling waves can

form when the breadth of inhibitory tuning exceeds that of excitatory tuning (3:1), with the addition of a slight asymmetry in excitatory tuning. Compared to the symmetrical waves of spreading activation discussed above, these waves have the potential to more substantially compensate for neural processing delays. This is because the underlying asymmetry in neural tuning drives a simultaneous dampening of the trailing edge of evoked activity (see Figure 3C in Heitmann & Ermentrout, 2020). Relatedly, Benigno et al. (2023) trained a neural network to predict subsequent frames of video input. They found that after training, a few frames of input could trigger complex wave patterns within the network, driving accurate predictions of future frames. Crucially, shuffling the lateral/recurrent connections within the network abolished both the waves and the network's predictive accuracy, suggesting a crucial role for laterally spreading activity in visual prediction. Collectively, these studies further suggest that spreading activation within visual cortex, in the form of travelling waves, can naturally facilitate motion extrapolation.

One important question regards the degree to which ongoing motion extrapolation occurs following early cortical processing. Recently via human EEG recordings it was shown that position representations for smoothly moving stimuli are activated substantially earlier (~70 ms) than for static stimuli (Johnson et al., 2023). In line with Benvenuti et al. (2019; Figure 3C&4C), early representations (i.e. those formed during the initial feedforward sweep of processing) were found to align with the real-time position of the stimulus (Johnson et al., 2023; Figure 4D). This provides further evidence that early cortical position representations may be fully delay compensated. Crucially however, later position representations were progressively delayed suggesting that not all position information in the cortex is fully extrapolated.

Does this mean that extrapolation ceases to occur after early visual processing? One possibility is that no further extrapolation occurs in the visual system, with downstream regions

(e.g., in motor cortex) left to handle any remaining necessary delay-compensation. However, due to the spatial blurring of EEG, it is also possible that select position representations in the visual hierarchy continue to be extrapolated (e.g., those involved in time-sensitive motor behaviour) but that this is masked by a predominance of non-extrapolated representations involved in less time-sensitive processes. Indeed, simultaneously sampling a mixture of position representations (as is the case when examining late EEG responses) would result in a progressively lagged average position signal even if a select few representations are still being extrapolated. Ultimately, the question of for how long, and in which visual regions, extrapolation continues to occur, remains an avenue for future investigation. Testing this may require direct neural recording in higher visual areas, to examine whether certain sub-populations of cells (possibly those which project to motor areas and which play a role in representing the targets of speeded responses) continue to extrapolate position information.

Furthermore, the question of exactly how discrete position estimates are 'read out' from population-level activity (see, Deneve et al., 1999; Erlhagen et al., 1999; Groh et al., 1997; Lee et al., 1988) must be addressed. In this article, we have considered mechanisms which shift the weight of population activity forwards in some way. However, to fully appreciate the extent of the extrapolative effects these mechanisms are having, we must understand how more downstream brain areas 'read out' population codes from visual cortex. For example, does the activity peak encode object position (as some existing literature tends to implicitly assume) or are other distributional features (i.e. the centroid/vector average) used instead?

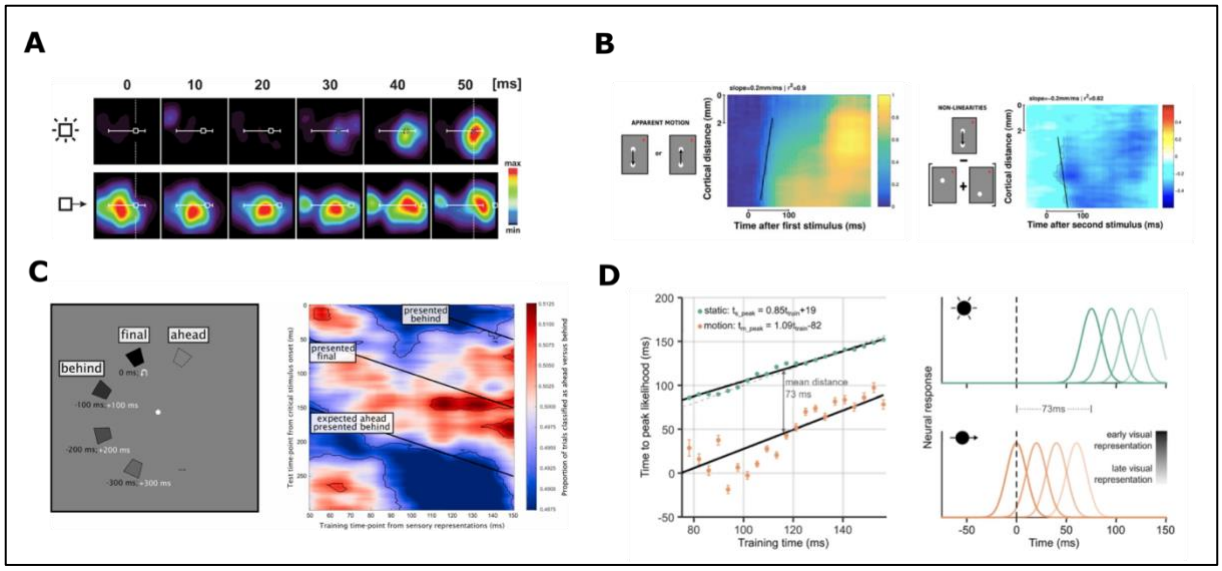

Figure 4. Motion extrapolation in the cortex. A) Comparison of the population-level activity evoked by moving and static stimuli in cat area 17 — from Jancke et al. (2004) – *Journal of Physiology*. Activity 50 ms after the onset of a flashed stimulus (top) vs activity 50 ms after a moving stimulus reached the same position (bottom). Activity for the moving stimulus is shifted in the direction of motion. **B)** Voltage sensitivity dye imaging (VSDI) of monkey visual cortex during the presentation of two stroke apparent motion (see inset stimulus diagrams) — from Chemla et al., (2019) – *Journal of Neuroscience*. The left sub-plot shows activity evoked by the apparent motion stimuli. Subthreshold excitatory activity spreads out from the initial flash, facilitating the response to the second stimulus and generating a directional wave of excitatory activity. However, the right sub-plot shows that this activity is sub-additive (i.e. less than would be expected if simply adding the activity generated by two separate flashes presented in isolation). Plotting the magnitude of this non-linearity reveals an inhibitory wave of activity in the direction opposite to the apparent motion. **C)** Decoding expected, but not presented, stimulus representations from human EEG recordings – from Blom et al., (2020) – *Proceedings of the National Academy of Science*. The left sub-plot shows an apparent motion stimulus moving along a circular trajectory. At 0 ms the stimulus reaches the 'final' location, after which it reverses direction. The right sub-plot shows a temporal generalization matrix showing the proportion of trials classified as immediately ahead or immediately behind the final position. After the final stimulus, there is a brief period in which classifier assignment favours the ahead position, before incoming information is processed and the reversal is encoded. **D)** Comparison of time-to-peak position likelihood for static and

moving stimuli, from human EEG recordings – from Johnson et al., (2023) – *eLife*. Peak position likelihood (i.e. when there is the strongest evidence for the stimulus being in its present location) consistently occurs ~70 ms earlier for moving compared to static stimuli. For early representations, there is almost perfect delay compensation, with delays gradually accruing once more for later position representations. The right sub-panel illustrates the degree of spatial compensation that the observed temporal shifts yield, for different cortical processing levels.

6.3. Dampening mechanisms in the cortex

For two-stroke apparent motion, a travelling wave of inhibitory activity has been observed moving backwards from the second stimulus to the first (Chemla et al., 2019; see Figure 4B). This reveals the existence of inhibitory processes in the cortex which dampen the trailing edge of evoked population-level activity. Earlier, we discussed the critical role that amacrine cells play in inhibition-driven motion extrapolation in the retina. However, these cells are not present in the cortex. So how are the asymmetric inhibitory profiles necessary for motion extrapolation generated?

Recall how passive interactions between excitatory and inhibitory signals in the dendrites of RGCs can selectively inhibit the trailing edge of stimulus evoked activity (Benigno et al., 2023; Heitmann & Ermentrout, 2020). This mechanism relies on general properties of synaptic arrangements, which likely apply to cells in many different brain areas. All that is needed is a proliferation of inhibitory synapses which are randomly and independently distributed relative to their excitatory counterparts. This ensures that for each excitatory synapse among the dendritic tree of a cortical neuron, there are a greater number of proximally located inhibitory synapses. Such an arrangement is crucial in optimizing the efficacy of inhibitory inputs (Koch et al., 1983). These conditions are almost certainly met in many regions of the visual cortex (Haider et al., 2013; Johnston & Lagnado, 2015), providing a general means for inhibitory gain control, and thus motion extrapolation, to be implemented.

Theoretically speaking, short-term synaptic depression (a reduction in synaptic efficacy due to the depletion of neurotransmitters; Abbott et al., 1997; Zucker, 1989) can have functionally similar effects to the retinal 'gain control' mechanisms discussed earlier. In particular, it has been shown that short-term depression of synapses in an attractor network model can support zero-lag tracking of a continuously moving stimulus (Fung et al., 2012). With sufficiently strong depression, the neural position can even predictively lead the stimulus by a fixed degree. More broadly, these mechanisms (retinal gain control, short-term synaptic depression) can be considered as special cases of the broader class of 'neural adaptation' mechanisms (see Feuerriegel, 2023). These all have the potential to dampen the trailing edge of stimulus evoked activity, and thus act as generic drivers of neural motion extrapolation.

Finally, another generic way of dampening the trailing edge of evoked activity, is if excitatory and inhibitory representations are transmitted at different speeds (Figure 5, Barlow, 1981). If a fast excitatory position signal is followed by a slow inhibitory signal, then the trailing edge of the evoked activity distribution will be erased by the lagging inhibition. This is similar to the inhibitory mechanisms in the retina. However, given the delay in inhibition, excitatory responses at the leading edge of evoked activity may avoid inhibition altogether, leading to a relative enhancement of activity and an even greater shift in the population position estimate. While this effect is most easily understood in retinotopic contexts (e.g., Figure 5), it theoretically enables the extrapolation of representations encoded in the distributed activity of non-contiguous cells. As long as inhibition lags excitation for given cell, this will shift the evolving population-level activity along its current representational trajectory. As such, despite the structural differences between the retina and early visual brain regions, there are multiple general mechanisms in the cortex capable of producing the asymmetric inhibitory profiles required for motion extrapolation.

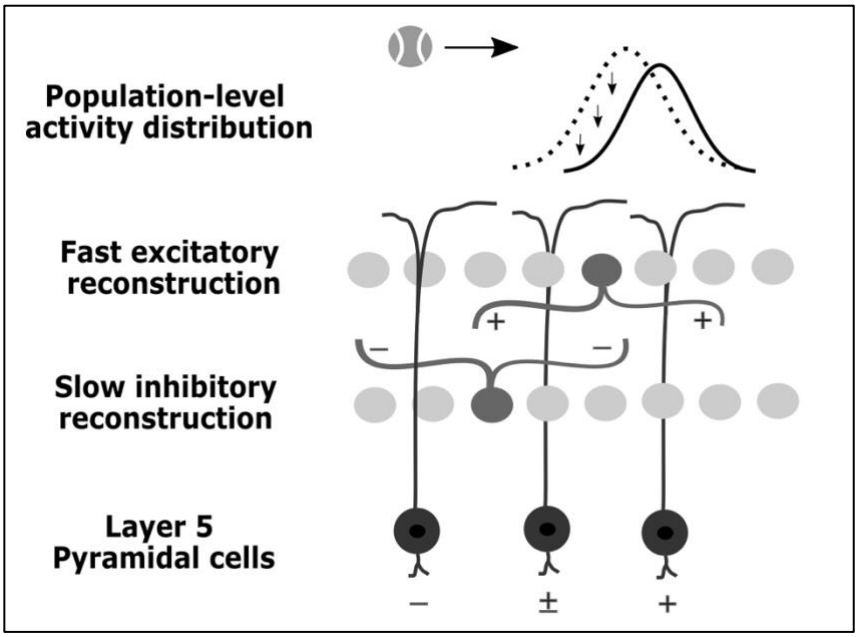

Figure 5. Delayed inhibition as a generic extrapolation mechanism. A hypothetical scheme for predicting the future position of a moving object — adapted from Barlow (1981). A delayed inhibitory signal spatially lags behind a fast excitatory signal, resulting in an asymmetric inhibitory profile. This serves to 'erase' the trailing edge of the population level activity distribution, shifting the population position estimate forwards along the current trajectory of motion. This may also lead to an increase in the leading edge of the activity distribution, relative to when there is zero-lag inhibition, since the initial excitatory activity 'escapes' inhibition.

7. Inter-regional extrapolation mechanisms

So far our focus has predominantly been on extrapolation mechanisms operating *within* a neural region (i.e. recurrent processes). However, the way in which information is transferred *between* regions can also drive motion extrapolation. In this section, we consider how inter-regional feedback mechanisms might extend the spatial range of motion extrapolation. We will also examine how specific patterns of feed-forward connectivity can facilitate responses to moving stimuli in downstream cortical regions. We will highlight recent research showing how

simple synaptic learning can drive the spontaneous emergence of asymmetric connectivity, allowing for efficient motion extrapolation during inter-regional information transfer.

7.1. Feedback connectivity

In apparent motion sequences there is a latency advantage for the third stimulus in the sequence, relative to the second (Blom et al., 2021). One might argue that this is simply due to a compounding of laterally spreading activity (e.g., Trenholm et al, 2015). However, it is also possible that this reflects the existence of more 'high-level' predictive mechanisms, which are employed once a motion trajectory can be extracted, and precise position predictions can be made. Given the large distances between stimuli in apparent motion designs, this likely requires the involvement of motion processing areas with large receptive fields. Consistent with this, numerous studies have suggested that feedback from area MT to V1 may facilitate the generation of motion-based predictions (Matsuyoshi et al., 2007; Muckli et al., 2005; Sterzer et al., 2006; Vetter et al., 2015; Wibral et al., 2009). Specifically, there is evidence that disrupting MT activity with TMS diminishes perceptions of apparent motion (Matsuyoshi et al., 2007) and interferes with a detection advantage normally observed for targets presented in apparent motion traces (Vetter et al., 2015). There is also evidence that area MT is involved in the 'pre-play' of activity associated with spatiotemporally predictable stimulus sequences in V1 (Ekman et al., 2017). Following the presentation of the first stimulus in a learned motion sequence, a time-compressed wave of anticipatory activity in V1 representing the full sequence is observed, even when the rest of the sequence is not presented. Evidence for the involvement of MT in such pre-play comes from the fact that the amplitude of BOLD activity in this area correlates with the amplitude of anticipatory V1 activity.

Considering this feedback connection, one important question is whether feedback signals are ultimately inhibitory or excitatory. In other words, does inter-regional feedback

serve to boost activity at the leading edge, or diminish activity at the trailing edge of the evoked activity distribution? Classical accounts of predictive visual processing argue that feedback is predominantly inhibitory (Rao & Ballard, 1999). Consistent with this, there is evidence that feedback from MT may reduce V1 activity for predictable motion sequences (Alink et al., 2010). However, if feedback from MT drives pre-play in V1 (Ekman et al., 2017), this suggests that feedback must be excitatory, or must ultimately result in excitation. This is consistent with the fact that, in apparent motion sequences, representations of stimuli which are expected but not presented can nevertheless be decoded from EEG recordings (Blom et al., 2020; Blom et al., 2021). Similarly, computational simulations have suggested that even subthreshold excitatory feedback is sufficient to drive motion anticipation and extrapolation, even across periods of occlusion (Erlhagen, 2003). As such, there is reason to believe that both excitatory and inhibitory feedback might play a role in motion extrapolation in the cortex.

In addition to examining the effect of feedback connections (excitation/inhibition), future research may consider the relative breadth of these connections. For example, it may be important to consider whether feedback connections are more or less diffuse than forward or lateral connections, within a given cortical region. This is because, in existing models, relative differences in the breadth of excitatory and inhibitory tuning have been shown to play an important role in generating motion-extrapolation effects (Chemla et al., 2019; Heitmann & Ermentrout, 2020; Xie & Giese, 2002). To better understand the potential role that feedback connections play in neural motion extrapolation, it may be necessary for future studies to jointly consider the breadth and modulatory effect of such connections, within a given neural region.

7.2. Feedforward connectivity

Perhaps the simplest way to achieve motion extrapolation is via the presence of spatially symmetric divergent feedforward connections between cortical regions. In the same way that horizontal excitation ('spreading activation') within a region serves to prime neurons ahead of a moving stimulus, divergent feedforward excitation can prime neurons in a downstream region, ahead of the currently stimulated location. This effect was modelled by Baldo and Caticha (2005) in a network of leaky integrate-and-fire (LIF) neurons with location-tuning. In their model, divergent feedforward activity primes cells in the direction of motion, leading them to fire more quickly in response to subsequent input. This has the same effect as the spreading activation mechanisms detailed above, however the anticipatory input is received via laterally divergent feedforward connections rather than local horizontal connections.

Importantly, symmetric divergent connections alone will lead to omnidirectional priming (i.e. all retinotopically adjacent downstream neurons will receive 'anticipatory' input). As we noted above, in motion contexts omnidirectional priming can ultimately have asymmetric consequences on the population-level (i.e. it shifts the leading edge of activity forwards), particularly when combined with non-linear thresholding (Liu et al., 2021). As such, non-selective divergent connectivity alone can be a functional driver of motion extrapolation. However, predictive extrapolation is best achieved when activity is *selectively* transmitted to retinotopic areas ahead of the current trajectory of motion (Kaplan et al., 2013; Nijhawan & Wu, 2009). To achieve this, feedforward connectivity must be asymmetrically weighted, such that the most strongly weighted connections are with those downstream neurons that code for the future position of an object. In the following section, we review recent work showing how such connectivity can emerge spontaneously from local synaptic learning rules.

8. Unsupervised learning of asymmetric connectivity patterns

Here, we discuss the role that spike timing dependent plasticity (STDP) plays in the emergence of asymmetric connectivity profiles, enabling directionally-specific information transfer between cortical regions (Abbott & Blum, 1996; Burkitt & Hogendoorn, 2021; Fu et al., 2004; Lim & Choe, 2008; Rao & Sejnowski, 1999; Sexton et al., 2023). STDP is a form of long-term plasticity in which the relative timing of pre- and post-synaptic spikes determines whether a synapse is potentiated or depressed. The STDP learning rule dictates that presynaptic spikes occurring just before a postsynaptic spike lead to long-term potentiation, whereas those occurring just after lead to long-term depression (Bi & Poo, 1998; Markram et al., 1997). To visualize how STDP can generate spatially asymmetric patterns of connectivity, we can consider a single postsynaptic neuron that receives input from a set of direction-selective neurons (Figure 6A). Motion in the rightward direction would cause the leftwards presynaptic cells to fire before the postsynaptic cell, strengthening these connections. Conversely, rightward neurons would fire after the postsynaptic cell, leading to depression of these synapses. Importantly, because the neurons have direction selectivity, equal exposure to both directions (as would be expected during development) still yields an asymmetry – that is, even though STDP still occurs for the anti-preferred direction, its effect is less pronounced due to the lower level of activity (Fu et al., 2004). Consequently, STDP generates asymmetrically tuned connectivity between neurons, amplifying the extrapolatory effect already given by simple divergent connectivity (Baldo & Caticha, 2005).

Fu, Shen, & Gao (2004) provide physiological support for this theoretical mechanism. They examined the receptive field properties of neurons in cat V1, in response to drifting gratings which contained motion information in either the preferred or anti-preferred direction of the cell. They found that the receptive field peak shifted in the opposite direction to the drifting grating (Figure 6B). This is precisely the effect required to achieve motion extrapolation, as it enables neurons to fire in anticipation of a moving object, and shifts the

population-level position estimate forward along the motion trajectory. Additionally, they observed that motion in one direction selectively enhanced the opposing side of the receptive field, arguing that this phenomenon is consistent with an asymmetry in the spatial distribution of direction-selective inputs to the neuron. By modelling the proposed connections, they showed how STDP can naturally give rise to this asymmetry.

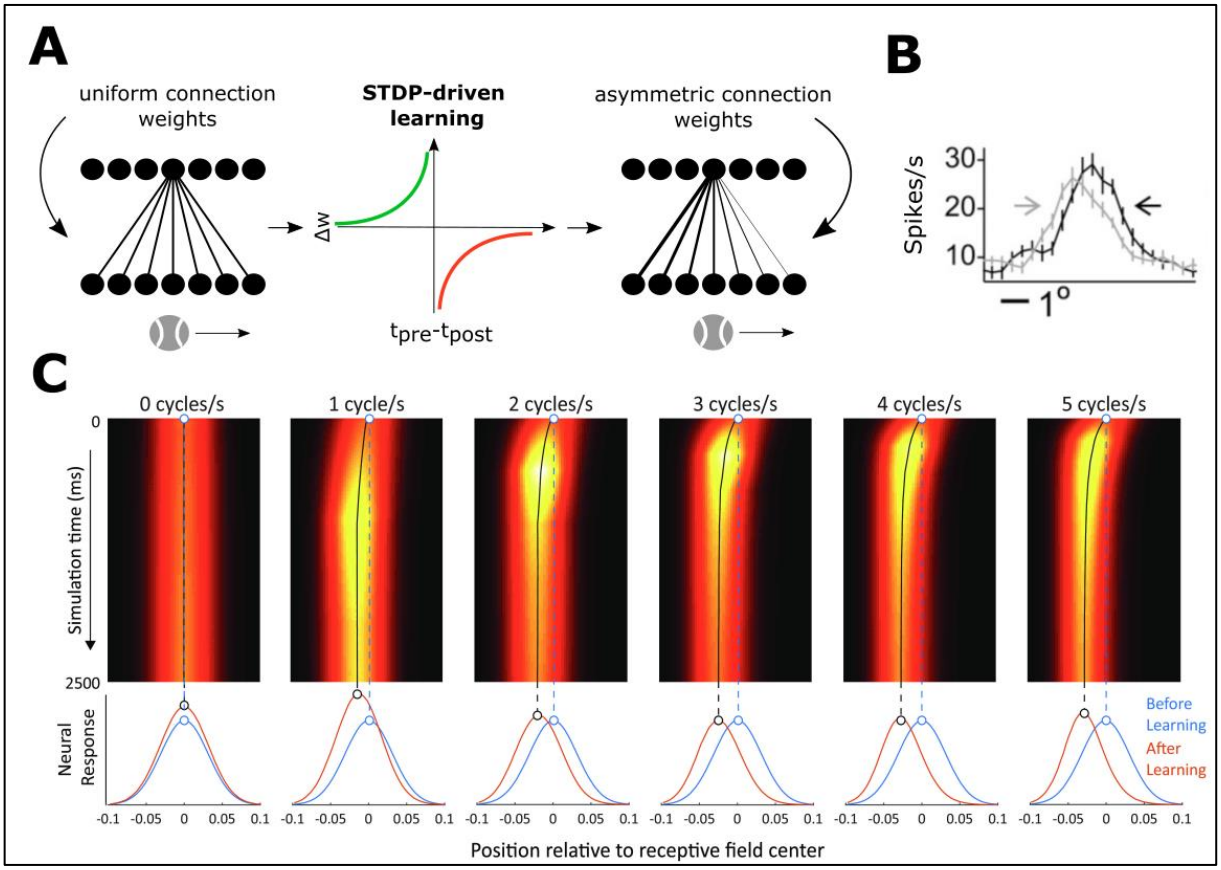

Figure 6. Unsupervised learning of asymmetric connectivity patterns A) Spike timing dependent plasticity (STDP) drives the learning of asymmetric connectivity patterns between layers of a neural network. Before learning, there are spatially uniform connection weights between neurons. The STDP learning rule dictates that presynaptic spikes occurring just before a postsynaptic spike (i.e. for positive t_{pre}-t_{post} values) lead to long-term potentiation (i.e. positive weight changes, Δw), whereas those occurring just after lead to long term depression. Simply allowing STDP-driven learning to unfold causes a spatial asymmetry in the connection weights for neurons in the higher layer, shifting the

neurons receptive field in the direction opposite to stimulus motion. This allows the higher level neuron to 'anticipate' the arrival of a moving stimulus. **B)** Empirical evidence of receptive field shifts in the direction opposite to stimulus motion – from Fu, Shen, & Gao (2004) – *Journal of Neuroscience*. The receptive field from a single neuron in cat visual cortex is shown, after being mapped with a grating stimulus containing either rightward (grey line) or leftward (black line) internal motion. The effective receptive field of the neuron is shifted in the direction opposite to stimulus motion. **C)** Average receptive field shifts before and after STDP-driven learning, across various stimulus speeds – from Burkitt & Hogendoorn (2021) – *Journal of Neuroscience*. The magnitude of the receptive field shift (displayed as a proportion of the entire visual space modelled) scales (non-linearly) with stimulus speed. In this study, an STDP time window spanning 200 ms was used, meaning that presynaptic spikes arriving within 100 ms before or after a postsynaptic spike would cause a change in the synaptic weight.

The role of STDP as a neural delay compensation mechanism was specifically considered by Burkitt and Hogendoorn (2021), who examined how STDP drives motion extrapolation across a range of stimulus velocities (Figure 6C). They simulated a two-layer, feedforward network comprised of several velocity-tuned subpopulations, with neural transmission delays along the feedforward connections. In this network, spikes are generated within an input layer of location- and velocity-tuned neurons in response to a moving stimulus. These then propagate to a second layer via divergent, feedforward excitatory connections. For connections that were initially spatially symmetric, simply allowing the dynamics of STDP to unfold caused the receptive fields of neurons in the second layer to shift in the direction opposite to motion. The magnitude of this shift depended on the velocity-tuning of the neural population, with greater velocities yielding greater receptive-field shifts.

Earlier we described how lateral priming effects can compound across neurons, such that neurons further along a motion trajectory receive greater priming. The same integrative effects of extrapolation may occur in the case of STDP along feedforward pathways. That is,

given that the visual system is comprised of multiple levels, receptive field shifts caused by STDP are likely to occur at several points throughout the visual hierarchy. This idea was recently addressed by looking at the accumulation of receptive field shifts in a multilayer network (Sexton et al., 2023). Sexton et al. (2023) showed that the overall capacity to extrapolate the position of a moving stimulus is greatly increased with the number of levels in the network, due to the accumulation of receptive field shifts at each higher layer. By looking at the distribution of spiking activity after network connection weights were trained with STDP, they showed that the degree of forward shift in the represented location of a moving object was in some cases sufficient to fully compensate for the delays embedded in the network (up to 100ms in a six-layer network). In terms of the temporal compensation achieved, comparing the magnitude of the receptive field shifts reported (~.001 to .028 cycles) to their corresponding velocities (0.1 to 1 cycles/s) yields estimates of temporal compensation on the order of 10 to 28ms per layer.

Given its prevalence among excitatory synapses in the visual cortex (Bi & Poo, 1998; Markram et al., 1997; Waters et al., 2008), STDP has the potential to act as a general delay-compensation mechanism, driving extrapolation during inter-regional information transfer (Note that because RGCs are the only neurons that produce action potentials in the retina, STDP is unlikely to be involved in retinal motion extrapolation). In addition to the potential role played by STDP in inter-regional connectivity, some authors have highlighted the potential for STDP to generate asymmetric intra-regional (recurrent) connectivity. For example, Lim and Choe (2008) modelled an array of orientation-tuned cells with bilateral (horizontal) excitatory connections. In response to a rotating stimulus, they found that STDP at lateral synapses led to an increase in excitation in the direction of motion, which caused a forward shift in the representation of the stimulus. Likewise, Rao and Sejnowski (1999) showed how STDP can produce predictive activity in a model network of recurrently connected neurons

with retinotopic inputs. In their network, an array of neurons receive feedforward input from a moving stimulus as well as from their horizontally connected neighbours. After allowing the recurrent connections to adjust their weights according to STDP, cells began to fire several milliseconds before arrival of the stimulus input. Finally, we described modelling work earlier in which asymmetries among local excitatory connections led to the generation of stimulus-locked travelling waves (Heitmann & Ermentrout, 2020). Although not specifically addressed, in principle it is possible that STDP could be involved in the formation of such asymmetries.

Looking forward, one open question is the degree to which STDP can dynamically compensate for specific neural delays, for example when there are different transmission delays between regions in a network. The work of Burkitt and Hogendoorn (2021) showed that STDP-driven extrapolation can scale across a range of stimulus velocities. However, in this work the degree of compensation achieved by STDP was fixed relative to the precise neural delay. In order to compensate for a specific delay, extrapolated representations generated at one level would have to be adjusted based on their correspondence with subsequent input at a lower level, in a manner similar to predictive coding (Rao & Ballard, 1999). This would entail 'calibrating' the degree of receptive field shift for a given delay between two visual areas, such that the downstream area is able to accurately represent the real-time position of an encoded object. This topic was addressed by Hogendoorn and Burkitt (2019), who extended the classical hierarchical predictive coding model to consider transmission delays between layers. They argued that prediction errors within each layer are minimized when some form of extrapolation is implemented along the feedforward and feedback pathways. This has the benefit of generating representations that are aligned across each hierarchical level at each moment in time. This type of hierarchical network structure would potentially be suitable to allow extrapolated representations to be calibrated to a specific neural transmission delay. We have described how STDP is a viable mechanism underlying extrapolation along feedforward

pathways, however the question of how such shifts could be calibrated, presumably via feedback pathways, remains an avenue for future research.

Finally, it is important to note an apparent contradiction between the theoretical findings described in this section and the empirical work described earlier. Specifically, whilst Sexton et al. (2023) found that STDP can (in theory) support motion extrapolation across multiple layers of processing, Johnson et al. (2023) found no clear evidence of ongoing extrapolation beyond the very earliest stages of cortical processing. Generally speaking, given the range of potential extrapolation mechanisms we have considered, the lack of evidence for sustained cortical extrapolation is puzzling. However, as noted earlier, a number of outstanding questions need to be addressed before we can confidently conclude that no ongoing extrapolation occurs. Specifically, neural recordings with increased spatial specificity are needed to examine whether select position signals (i.e. those involved in guiding speeded action) continue to be extrapolated, and the fundamental question of how discrete position estimates are actually 'read out' by higher cortical areas must be addressed (see Section 6.2).

9. The neural implementation of 'model-based' motion extrapolation

Up to this point we have considered extrapolation mechanisms that are essentially 'wired in' to the low-level structure of the visual system. Given this, it may be tempting to conclude that all extrapolation is achieved automatically, through hard-wired mechanisms, without the need for an internal model of the world (e.g., a model of the physical laws of motion). In this final section, we briefly caution against this view.

Many researchers have proposed that the brain ultimately employs 'model-based' predictive strategies – reliant upon internal generative models – to achieve accurate and efficient sensory processing (Erlhagen, 2003; Grush, 2005; Jiang & Rao, 2022; Khoei et al., 2017; Kwon et al., 2015; Mumford, 1992; Rao & Ballard, 1999). Theoretically speaking, these

mechanisms are extremely powerful, allowing for optimal inferences to be drawn from uncertain sources of information (under some assumptions). A classic example is recursive Bayesian estimation algorithms (e.g., Kalman filtering), in which uncertain sensory information is combined with predictions from an internal generative model, with their respective weights determined by their relative uncertainty. Computational models built around this general approach have been able to explain perceptual and neural data in object-localization tasks (Khoei et al., 2017; Kwon et al., 2015), in some cases explicitly accommodating for the effects of neural delays (Khoei et al., 2017; Perrinet et al., 2014). More generally, 'predictive coding' networks, in which hierarchical generative models learn the spatial (Rao & Ballard, 1999) or spatio-temporal (Jiang & Rao, 2022; Lotter et al., 2020) regularities of natural scenes, have been able to account for a range of neural and perceptual phenomena. The success of these models suggests that the brain may indeed rely on 'model-based' strategies during visual processing.

So how and where are these implemented? To answer this, it is important to recognize that the distinction we are drawing between hard-wired 'model-free' mechanisms and more abstract 'model-based' mechanisms, is in fact rather fuzzy. This is because the generative models core to 'model-based' strategies, must ultimately be coded into networks in a distributed fashion, via the modification of synaptic connection weights. As such, they are also in a sense 'wired in' to the structure of the network. To further complicate things, these networks typically *learn* their generative models via exposure to sets of images/videos, making it difficult to identify the underlying regularities that the model has learnt to exploit. Given this, it is possible that the neural mechanisms we have reviewed above, are actually constituent parts of a distributed 'model-based' processing strategy/strategies (e.g., Kaplan et al., 2013; Khoei et al., 2017).

There is, however, some evidence which suggests that more temporally extended predictive strategies are only implemented in high-level regions, with longer intrinsic timescales (Jiang & Rao, 2022; Murray et al., 2014; Runyan et al., 2017). For example, in the double-drift illusion, where there is a marked mismatch between the actual and perceived position of a moving object, the neural correlates of conscious perception appear to lie in anterior brain regions outside the visual system (Liu et al., 2019). Under such conditions, it has been argued that perceptual experience may be the end product of a model-based process, involving an internal model of motion dynamics (Kwon et al., 2015). The fact that in these cases the neural correlates of perception lie in frontal regions suggests that these higher areas may read out information from the visual system and, over relatively long time scales, combine it with internally generated predictions, to refine estimates of both past and future states (e.g., the 'timeline' model of perception; Grush, 2005; Hogendoorn, 2022; Jiang & Rao, 2022).

Broadly consistent with this, it has been suggested that early visual processing may 're-format' incoming sensory information, to facilitate subsequent linear extrapolation (Hénaff et al., 2019, 2021). To illustrate the core idea behind this line of research, consider a 16 x 16 pixel video. Each frame of this video can be represented as a point in a 256 dimensional 'pixel space', with each dimension representing the brightness of a specific pixel. As the video progresses, new points will appear, drawing out a 'trajectory' through this space. For naturalistic videos these trajectories are complex and curved, making it difficult to linearly extrapolate their evolution into the future. Strikingly however, empirical estimates of representational trajectories though neural and perceptual spaces (e.g., spaces made from considering the individual activity of neurons within a larger population) are much straighter (Hénaff et al., 2019, 2021). This suggests that visual processing serves to straighten the temporal trajectories of natural visual input, 're-formatting' it so as to facilitate linear extrapolation, potentially in higher cortical regions.

To sum up, the focus of this review has been on local neural mechanisms. However, it is important to highlight the rich concurrent stream of computational research which examines how more abstract predictive algorithms and more global information processing architectures enable the spatio-temporal regularities of visual input to be learnt. Bridging of the gap between these two levels (biological and computational), and understanding how these more abstract computational strategies can be instantiated in biologically-plausible networks, is an exciting avenue for future research.

10. Summary and conclusions

We have considered how the brain can predictively extrapolate the represented position of moving objects, (partially or fully) re-aligning internal representations with the current state of the outside world in the face of neural delays. We have identified a variety of mechanisms which facilitate extrapolatory shifts of population-level neural activity. These all involve some form of asymmetry which either enhances the leading edge and/or dampens the trailing edge of the evoked activity distribution. Working along the visual stream, we have considered specific examples, at the level of local neural circuits. In particular, we have shown how local (recurrent) excitatory and inhibitory processes serve to shift the represented position of a moving object along its current trajectory of motion. We have also considered how extrapolation can be achieved during inter-regional information transfer, and have highlighted how asymmetric connectivity patterns which support extrapolation can emerge spontaneously via spike-timing-dependent plasticity. Finally, we have considered how more abstract 'model-based' predictive strategies (e.g., recursive Bayesian estimation) might be neurally instantiated.

While the focus of this article has been on the visual system, it is important to recognize that neural delays are ubiquitous throughout the brain. As such, it is crucial that delays (and the mechanisms which serve to compensate for their deleterious effects) are accounted for in

models of neural processing. Moreover, consideration of delays can radically challenge intuitive (and widely held) views of neural processing. For example, the existence of heterogenous delays within our sensory systems means that the timing of neural events (and perceptual representations) is necessarily divorced from the timing of the external world. This renders an intuitive Newtonian view of sensory processing (in which the timing of sensory events is encoded as the timing of the neural activity representing those events) unviable (see Hogendoorn, 2022), and necessitates the internal representation(s) of time. Here, we have presented an integrative framework for understanding how the brain determines the real-time position of moving objects, despite neural delays. Looking forwards, further work remains to be done to fully understand the effect of delays on sensory processes and neural dynamics more generally.

References

Abbott, L. F., & Blum, K. I. (1996). Functional Significance of Long-Term Potentiation for Sequence Learning and Prediction. *Cerebral Cortex, 6*(3), 406–416. https://doi.org/10.1093/cercor/6.3.406

Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic Depression and Cortical Gain Control. *Science, 275*(5297), 221–224. https://doi.org/10.1126/science.275.5297.221

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus Predictability Reduces Responses in Primary Visual Cortex. *Journal of Neuroscience, 30*(8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus Specific Responses from Beyond the Classical Receptive Field: Neurophysiological Mechanisms for Local-Global Comparisons in Visual Neurons. *Annual Review of Neuroscience, 8*(1), 407–430. https://doi.org/10.1146/annurev.ne.08.030185.002203

Atick, J. J., & Redlich, A. N. (1992). What Does the Retina Know about Natural Scenes? *Neural Computation, 4*(2), 196–210. https://doi.org/10.1162/neco.1992.4.2.196

Baldo, M. V. C., & Caticha, N. (2005). Computational neurobiology of the flash-lag effect. *Vision Research, 45*(20), 2620–2630. https://doi.org/10.1016/j.visres.2005.04.014

Barlow, H. B. (1981). The Ferrier Lecture, 1980: Critical Limiting Factors in the Design of the Eye and Visual Cortex. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 212*(1186), 1–34.

Baylor, D. A., Fuortes, M. G. F., & O'Bryan, P. M. (1971). Lateral interaction between vertebrate photoreceptors. *Vision Research, 11*(10), 1195–1196. https://doi.org/10.1016/0042-6989(71)90134-9

Benigno, G. B., Budzinski, R. C., Davis, Z. W., Reynolds, J. H., & Muller, L. (2023). Waves traveling over a map of visual space can ignite short-term predictions of sensory input. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-39076-2

Benvenuti, G., Chemla, S., Boonman, A., Perrinet, L., Masson, G. S., & Chavane, F. (2020). Anticipatory responses along motion trajectories in awake monkey area V1. *bioRxiv*, 1–42.

Berry, M. J., Brivanlou, I. H., Jordan, T. A., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, *398*(6725), Article 6725. https://doi.org/10.1038/18678

Bi, G., & Poo, M. (1998). Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *Journal of Neuroscience*, *18*(24), 10464–10472. https://doi.org/10.1523/JNEUROSCI.18-24-10464.1998

Blom, T., Bode, S., & Hogendoorn, H. (2021). The time-course of prediction formation and revision in human visual motion processing. *Cortex*, *138*, 191–202. https://doi.org/10.1016/j.cortex.2021.02.008

Blom, T., Feuerriegel, D., Johnson, P., Bode, S., & Hogendoorn, H. (2020). Predictions drive neural representations of visual events ahead of incoming sensory information. *Proceedings of the National Academy of Sciences*, *117*(13), 7510–7515. https://doi.org/10.1073/pnas.1917777117

Bolz, J., Rosner, G., & Wässle, H. (1982). Response latency of brisk-sustained (X) and brisk-transient (Y) cells in the cat retina. *The Journal of Physiology*, *328*(1), 171–190. https://doi.org/10.1113/jphysiol.1982.sp014258

Bringuier, V., Chavane, F., Glaeser, L., & Frégnac, Y. (1999). Horizontal Propagation of Visual Activity in the Synaptic Integration Field of Area 17 Neurons. *Science*, *283*(5402), 695–699. https://doi.org/10.1126/science.283.5402.695

Burkitt, A. N., & Hogendoorn, H. (2021). Predictive Visual Motion Extrapolation Emerges Spontaneously and without Supervision at Each Layer of a Hierarchical Neural Network with Spike-Timing-Dependent Plasticity. *Journal of Neuroscience*, *41*(20), 4428–4438. https://doi.org/10.1523/JNEUROSCI.2017-20.2021

Chemla, S., Reynaud, A., Volo, M. di, Zerlaut, Y., Perrinet, L., Destexhe, A., & Chavane, F. (2019). Suppressive Traveling Waves Shape Representations of Illusory Motion in Primary Visual Cortex of Awake Primate. *Journal of Neuroscience*, *39*(22), 4282–4298. https://doi.org/10.1523/JNEUROSCI.2792-18.2019

Chen, E. Y., Marre, O., Fisher, C., Schwartz, G., Levy, J., Silveira, R. A. da, & Berry, M. J. (2013). Alert Response to Motion Onset in the Retina. *Journal of Neuroscience*, *33*(1), 120–132. https://doi.org/10.1523/JNEUROSCI.3749-12.2013

Copenhagen, D. R., & Owen, W. G. (1976). Coupling between rod photoreceptors in a vertebrate retina. *Nature*, *260*(5546), Article 5546. https://doi.org/10.1038/260057a0

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, *2*(8), Article 8. https://doi.org/10.1038/11205

DeVries, S. H., Qi, X., Smith, R., Makous, W., & Sterling, P. (2002). Electrical Coupling between Mammalian Cones. *Current Biology*, *12*(22), 1900–1907. https://doi.org/10.1016/S0960-9822(02)01261-7

Dong, D. W., & Atick, J. J. (1995). Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, *6*(2), 159. https://doi.org/10.1088/0954-898X/6/2/003

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms15276

Erlhagen, W. (2003). Internal models for visual perception. *Biological Cybernetics*, *88*(5), 409–417. https://doi.org/10.1007/s00422-002-0387-1

Erlhagen, W., Bastian, A., Jancke, D., Riehle, A., & Schöner, G. (1999). The distribution of neuronal population activation (DPA) as a tool to study interaction and integration in cortical representations. *Journal of Neuroscience Methods*, *94*(1), 53–66. https://doi.org/10.1016/S0165-0270(99)00125-9

Euler, T., Detwiler, P. B., & Denk, W. (2002). Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature*, *418*(6900), Article 6900. https://doi.org/10.1038/nature00931

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Feuerriegel, D. (2023). *Adaptation in the Visual System: Networked Fatigue or Suppressed Prediction Error Signalling?* PsyArXiv. https://doi.org/10.31234/osf.io/73hb4

Fu, Y.-X., Shen, Y., Gao, H., & Dan, Y. (2004). Asymmetry in Visual Cortical Circuits Underlying Motion-Induced Perceptual Mislocalization. *Journal of Neuroscience*, *24*(9), 2165–2171. https://doi.org/10.1523/JNEUROSCI.5145-03.2004

Fung, C., Wong, K., & Wu, S. (2012). Delay Compensation with Dynamical Synapses. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper_files/paper/2012/hash/85422afb467e9456013a2a51d4dff702-Abstract.html

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal Population Coding of Movement Direction. *Science*, *233*(4771), 1416–1419. https://doi.org/10.1126/science.3749885

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Grimaldi, A., & Perrinet, L. U. (2023). Learning heterogeneous delays in a layer of spiking neurons for fast motion detection. *Biological Cybernetics*, *117*(4), 373–387. https://doi.org/10.1007/s00422-023-00975-8

Groh, J. M., Born, R. T., & Newsome, W. T. (1997). How Is a Sensory Map Read Out? Effects of Microstimulation in Visual Area MT on Saccades and Smooth Pursuit Eye Movements. *Journal of Neuroscience*, *17*(11), 4312–4330. https://doi.org/10.1523/JNEUROSCI.17-11-04312.1997

Grush, R. (2005). Internal models and the construction of time: Generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions. *Journal of Neural Engineering*, *2*(3), S209. https://doi.org/10.1088/1741-2560/2/3/S05

Guo, K., Robertson, R. G., Pulgarin, M., Nevado, A., Panzeri, S., Thiele, A., & Young, M. P. (2007). Spatio-temporal prediction and inference by V1 neurons. *European Journal of Neuroscience*, *26*(4), 1045–1054. https://doi.org/10.1111/j.1460-9568.2007.05712.x

Haider, B., Häusser, M., & Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature*, *493*(7430), Article 7430. https://doi.org/10.1038/nature11665

Hassenstein, & Reichardt. (1956). Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Zeitschrift Für Naturforschung B*, *11*(9–10), 513-524.

Heitmann, S., & Ermentrout, G. B. (2020). Direction-selective motion discrimination by traveling waves in visual cortex. *PLOS Computational Biology*, *16*(9), e1008164. https://doi.org/10.1371/journal.pcbi.1008164

Hénaff, O. J., Bai, Y., Charlton, J. A., Nauhaus, I., Simoncelli, E. P., & Goris, R. L. T. (2021). Primary visual cortex straightens natural video trajectories. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-25939-z

Hénaff, O. J., Goris, R. L. T., & Simoncelli, E. P. (2019). Perceptual straightening of natural videos. *Nature Neuroscience*, *22*(6), Article 6. https://doi.org/10.1038/s41593-019-0377-4

Hogendoorn, H. (2020). Motion Extrapolation in Visual Processing: Lessons from 25 Years of Flash-Lag Debate. *Journal of Neuroscience*, *40*(30), 5698–5705. https://doi.org/10.1523/JNEUROSCI.0275-20.2020

Hogendoorn, H. (2022). Perception in real-time: Predicting the present, reconstructing the past. *Trends in Cognitive Sciences*, *26*(2), 128–141. https://doi.org/10.1016/j.tics.2021.11.003

Hogendoorn, H., & Burkitt, A. N. (2018). Predictive coding of visual object position ahead of moving objects revealed by time-resolved EEG decoding. *NeuroImage*, *171*, 55–61. https://doi.org/10.1016/j.neuroimage.2017.12.063

Inui, K., & Kakigi, R. (2006). Temporal Analysis of the Flow From V1 to the Extrastriate Cortex in Humans. *Journal of Neurophysiology*, *96*(2), 775–784. https://doi.org/10.1152/jn.00103.2006

Izhikevich, E. M. (2006). Polychronization: Computation with Spikes. *Neural Computation*, *18*(2), 245–282. https://doi.org/10.1162/089976606775093882

Jancke, D., & Erlhagen, W. (2010). Bridging the gap: A model of common neural mechanisms underlying the Fröhlich effect, the flash-lag effect, and the representational momentum effect. In *Space and time in perception and action* (pp. 422–440). Cambridge University Press. https://doi.org/10.1017/CBO9780511750540.025

Jancke, D., Erlhagen, W., Schöner, G., & Dinse, H. R. (2004). Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex. *The Journal of Physiology*, *556*(3), 971–982. https://doi.org/10.1113/jphysiol.2003.058941

Jiang, L. P., & Rao, R. P. N. (2022). *Dynamic Predictive Coding: A New Model of Hierarchical Sequence Learning and Prediction in the Cortex* (p. 2022.06.23.497415). bioRxiv. https://doi.org/10.1101/2022.06.23.497415

Johnson, P. A., Blom, T., & van Gaal, S. (2023). Position representations of moving objects align with real-time position in the early visual response. *eLife*.

Johnston, J., & Lagnado, L. (2015). General features of the retinal connectome determine the computation of motion anticipation. *eLife*, 1–19. https://doi.org/10.7554/eLife.06250

Kaplan, B., Lansner, A., Masson, G., & Perrinet, L. (2013). Anisotropic connectivity implements motion-based prediction in a spiking neural network. *Frontiers in Computational Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fncom.2013.00112

Khoei, M. A., Masson, G. S., & Perrinet, L. U. (2017). The Flash-Lag Effect as a Motion-Based Predictive Shift. *PLOS Computational Biology*, *13*(1), e1005068. https://doi.org/10.1371/journal.pcbi.1005068

Koch, C., Poggio, T., & Torre, V. (1983). Nonlinear interactions in a dendritic tree: Localization, timing, and role in information processing. *Proceedings of the National Academy of Sciences*, *80*(9), 2799–2802. https://doi.org/10.1073/pnas.80.9.2799

Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences*, *112*(26), 8142–8147. https://doi.org/10.1073/pnas.1500361112

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, *332*(6162), 357–360. https://doi.org/10.1038/332357a0

Lim, H., & Choe, Y. (2008). Extrapolative Delay Compensation Through Facilitating Synapses and Its Relation to the Flash-Lag Effect. *IEEE Transactions on Neural Networks*, *19*(10), 1678–1688. https://doi.org/10.1109/TNN.2008.2001002

Lisi, M., & Cavanagh, P. (2015). Dissociation between the Perceptual and Saccadic Localization of Moving Objects. *Current Biology*, *25*(19), 2535–2540. https://doi.org/10.1016/j.cub.2015.08.021

Liu, B., Hong, A., Rieke, F., & Manookin, M. B. (2021). Predictive encoding of motion begins in the primate retina. *Nature Neuroscience*, *24*, 1280–1290. https://doi.org/10.1038/s41593-021-00899-1

Liu, S., Yu, Q., Tse, P. U., & Cavanagh, P. (2019). Neural Correlates of the Conscious Perception of Visual Location Lie Outside Visual Cortex. *Current Biology*, *29*(23), 4036-4044.e4. https://doi.org/10.1016/j.cub.2019.10.033

Lotter, W., Kreiman, G., & Cox, D. (2020). A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nature Machine Intelligence*, *2*(4), 210–219.

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, *275*(5297), 213–215. https://doi.org/10.1126/science.275.5297.213

Matsuyoshi, D., Hirose, N., Mima, T., Fukuyama, H., & Osaka, N. (2007). Repetitive transcranial magnetic stimulation of human MT+ reduces apparent motion perception. *Neuroscience Letters*, *429*(2), 131–135. https://doi.org/10.1016/j.neulet.2007.10.002

Maus, G. W., & Nijhawan, R. (2008). Motion Extrapolation Into the Blind Spot. *Psychological Science*, *19*(11), 1087–1091.

Menz, M. D., Lee, D., & Baccus, S. A. (2020). Representations of the amacrine cell population underlying retinal motion anticipation Results. *bioRxiv*, *1*.

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary Visual Cortex Activity along the Apparent-Motion Trace Reflects Illusory Perception. *PLOS Biology*, *3*(8), e265. https://doi.org/10.1371/journal.pbio.0030265

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., & Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, *17*(12), Article 12. https://doi.org/10.1038/nn.3862

Nieman, D., Nijhawan, R., Khurana, B., & Shimojo, S. (2006). Cyclopean flash-lag illusion. *Vision Research*, *46*, 3909–3914. https://doi.org/10.1016/j.visres.2006.06.003

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, *370*(6487), Article 6487. https://doi.org/10.1038/370256b0

Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *The Behavioral and Brain Sciences*, *31*(2), 179–198; discussion 198-239. https://doi.org/10.1017/S0140525X08003804

Nijhawan, R., & Wu, S. (2009). Compensating Time Delays with Neural Predictions: Are Predictions Sensory or Motor? *Philosophical Transactions of the Royal Society A*, *367*(1891), 1063–1078.

Orban, G. A., Hoffmann, K. P., & Duysens, J. (1985). Velocity selectivity in the cat visual system. I. Responses of LGN cells to moving bar stimuli: A comparison with cortical areas 17 and 18. *Journal of Neurophysiology*, *54*(4), 1026–1049. https://doi.org/10.1152/jn.1985.54.4.1026

Perrinet, L. U., Adams, R. A., & Friston, K. J. (2014). Active inference, eye movements and oculomotor delays. *Biological Cybernetics*, *108*(6), 777–801. https://doi.org/10.1007/s00422-014-0620-8

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1*(2), Article 2. https://doi.org/10.1038/35039062

Rao, R. P. N., & 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.

Rao, R., & Sejnowski, T. J. (1999). Predictive Sequence Learning in Recurrent Neocortical Circuits. *Advances in Neural Information Processing Systems*, *12*. https://proceedings.neurips.cc/paper/1999/hash/b865367fc4c0845c0682bd466e6ebf4c-Abstract.html

Runyan, C. A., Piasini, E., Panzeri, S., & Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature*, *548*(7665), Article 7665. https://doi.org/10.1038/nature23020

Sato, T. K., Nauhaus, I., & Carandini, M. (2012). Traveling Waves in Visual Cortex. *Neuron*, *75*(2), 218–229. https://doi.org/10.1016/j.neuron.2012.06.029

Saul, A. B., & Humphrey, A. L. (1990). Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *Journal of Neurophysiology*, *64*(1), 206–224. https://doi.org/10.1152/jn.1990.64.1.206

Schiller, P. H., & Malpeli, J. G. (1978). Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey. *Journal of Neurophysiology*, *41*(3), 788–797. https://doi.org/10.1152/jn.1978.41.3.788

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal Timing Across the Macaque Visual System. *Journal of Neurophysiology*, *79*(6), 3272–3278. https://doi.org/10.1152/jn.1998.79.6.3272

Schwartz, G., Taylor, S., Fisher, C., Harris, R., & Berry, M. J. (2007). Synchronized Firing among Retinal Ganglion Cells Signals Motion Reversal. *Neuron*, *55*(6), 958–969. https://doi.org/10.1016/j.neuron.2007.07.042

Sexton, C. M., Burkitt, A. N., & Hogendoorn, H. (2023). Spike-timing dependent plasticity partially compensates for neural delays in a multi-layered network of motion-sensitive neurons. *PLOS Computational Biology*, *19*(9), e1011457. https://doi.org/10.1371/journal.pcbi.1011457

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), Article 11. https://doi.org/10.1038/nn.4650

Sterzer, P., Haynes, J., & Rees, G. (2006). Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT + / V5. *NeuroImage*, *32*, 1308–1316. https://doi.org/10.1016/j.neuroimage.2006.05.029

Subramaniyan, M., Ecker, A. S., Patel, S. S., Cotton, R. J., Bethge, M., Pitkow, X., Berens, P., & Tolias, A. S. (2018). Faster processing of moving compared with flashed bars in awake macaque V1 provides a neural correlate of the flash lag illusion. *Journal of Neurophysiology*, *120*(5), 2430–2452. https://doi.org/10.1152/jn.00792.2017

Sundberg, K. A., Fallah, M., & Reynolds, J. H. (2006). A Motion-Dependent Distortion of Retinotopy in Area V4. *Neuron*, *49*(3), 447–457. https://doi.org/10.1016/j.neuron.2005.12.023

Trenholm, S., Schwab, D. J., Balasubramanian, V., & Awatramani, G. B. (2013). Lag normalization in an electrically coupled neural network. *Nature Neuroscience*, *16*(2). https://doi.org/10.1038/nn.3308

Turner, W., Blom, T., & Hogendoorn, H. (2023). Visual Information Is Predictively Encoded in Occipital Alpha/Low-Beta Oscillations. *Journal of Neuroscience*, *43*(30), 5537–5545. https://doi.org/10.1523/JNEUROSCI.0135-23.2023

Vetter, P., Grosbras, M.-H., & Muckli, L. (2015). TMS Over V5 Disrupts Motion Prediction. *Cerebral Cortex*, *25*, 1052–1059. https://doi.org/10.1093/cercor/bht297

Waters, J., Nevian, T., Sakmann, B., & Helmchen, F. (2008). 4.39—Action Potentials in Dendrites and Spike-Timing-Dependent Plasticity. In J. H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 803–828). Academic Press. https://doi.org/10.1016/B978-012370509-9.00029-2

Wibral, M., Bledowski, C., Kohler, A., Singer, W., & Muckli, L. (2009). The Timing of Feedback to Early Visual Cortex in the Perception of Long-Range Apparent Motion. *Cerebral Cortex*, *19*(7), 1567–1582. https://doi.org/10.1093/cercor/bhn192

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209–216. https://doi.org/10.1016/S1364-6613(97)01070-X

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Xie, X., & Giese, M. A. (2002). Nonlinear dynamics of direction-selective recurrent neural media. *Physical Review E*, *65*(5), 051904. https://doi.org/10.1103/PhysRevE.65.051904

Zucker, R. S. (1989). Short-Term Synaptic Plasticity. *Annual Review of Neuroscience*, *12*(1), 13–31. https://doi.org/10.1146/annurev.ne.12.030189.000305

Acknowledgements

This work was supported by Australian Research Council Grants FT200100246, DP220101166 and DP180102268 awarded to Hinze Hogendoorn.