How nonlinear processing shapes natural stimulus encoding in the retina

Dissertation

for the award of the degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program
IMPRS Neurosciences
of the Georg-August University School of Science (GAUSS)

submitted by

Dimokratis Karamanlis
from Thessaloniki, Greece

Göttingen, 2021

Thesis Committee and Examination Board members

Prof. Dr. Tim Gollisch (1st referee)
Department of Ophthalmology, University Medical Center Göttingen

Prof. Dr. Alexander Gail (2nd referee)
Cognitive Neuroscience Laboratory, German Primate Center, Göttingen

Prof. Dr. Marion Silies
Institute of Developmental Biology and Neurobiology, Johannes-Gutenberg University Mainz

Further members of the Examination Board

Prof. Dr. Hansjörg Scherberger
Neurobiology Laboratory, German Primate Center, Göttingen

Prof. Dr. Tobias Moser
Institute for Auditory Neuroscience, University Medical Center Göttingen

Dr. Jan Clemens
European Neuroscience Institute, Göttingen

Date of oral examination: February 23, 2022

Wir müssen wissen. Wir werden wissen.

— DAVID HILBERT

CONTENTS

SUMMARY

Understanding natural vision is one of the fundamental goals of sensory neuroscience. To achieve such an understanding, we need to uncover how natural scenes are represented in early visual areas. What are the computations that lead to such neural representations? What cellular and molecular mechanisms are required to implement these computations? And what goals do these computations achieve in the context of a behaving animal? The only part of the visual system that may currently be amenable to such a complete understanding is the vertebrate retina, where there is increasing convergence of physiological and anatomical evidence that explain the transformation from visual stimulus to neural response.

Early theoretical work postulated that the main goal of the retina is to most efficiently represent natural scenes with the least amount of resources (e.g. action potentials). This perspective relied on the assumption that the retina acts as a parallel stream of linear filters, interfacing between the photoreceptors and the axons of retinal ganglion cells, which form the optic nerve. However, a linear picture may be inconsistent with the rich diversity of the—more than hundred—neuronal cell types in the retina, and the retinal circuit's ability to perform complex computations, such as distinguishing local from global motion. Perhaps most importantly, computational models based on linear filtering only partially capture the retinal output under naturalistic stimulation.

The shortcomings of this linear picture of the retina come as no surprise in the light of the multiple examples of nonlinear processing within the retinal circuit. The prime example is associated with signal transduction between bipolar and ganglion cells, which can be highly nonlinear. However, these nonlinearities are mostly studied in isolation with targeted artificial visual stimuli, and their significance for natural stimuli is still unclear. Natural scenes have different statistics from typically applied artificial stimuli and drive multiple retinal mechanisms in unison. Thus, it is still an open question how nonlinear spatial processing shows up in the retinal output under naturalistic stimulation. In this thesis I address this problem by asking two main questions: Is nonlinear processing necessary for explaining observed retinal responses to natural stimuli? And are computational models that include nonlinear processing sufficient to capture retinal responses to natural stimuli?

In Chapter 1, I introduce fundamental concepts about natural visual scenes and sensory encoding in the retina. I then examine different aspects of nonlinear spatial processing—the focus of this thesis—and review existing evidence about the role of retinal nonlinearities in the encoding of natural scenes.

In Chapter 2, I distinguish and summarize three perspectives under which vision researchers study the retinal encoding of natural scenes. The *normative* perspective asks about features of the retinal circuit that optimally represent natural scene statistics. The *coding* perspective seeks to understand which features of the natural visual world are conveyed by the spike trains of

ganglion cells. Finally, the *circuit* perspective is concerned with determining retinal mechanisms driven under naturalistic stimulation.

In Chapter 3, following the lines of the *circuit* perspective, I investigate the necessity of nonlinear spatial integration for natural image encoding in the mouse retina. I find differential sensitivity of ganglion cells to natural spatial structure. I then show that this sensitivity is a property of nonlinearities acting over the receptive field center, and I establish that different nonlinearity types can be found among different types of retinal ganglion cells.

In Chapter 4, to test the sufficiency of nonlinear integration, I develop models that can leverage nonlinear properties of the retinal circuit, and test their predictions with natural scene stimuli. I provide complete nonlinear receptive field descriptions of major cell types in the mouse retina, and I link my results to findings from the previous chapter. I then generate mouse-specific natural movies, and show that nonlinear models can capture the retinal output under stimuli with naturalistic temporal dynamics. Finally, using ideas from the *normative* perspective, I show that nonlinear processing alone is insufficient to explain pairwise response correlations between ganglion cells of the same type, and that retinal inhibition may be additionally required.

I summarize and discuss my findings in Chapter 5. Together, I show that nonlinear retinal processing is both necessary and sufficient for the encoding of natural scenes, extends functional descriptions of neuronal types beyond linear receptive fields, and can reduce the gap in predicting the retinal output to natural visual inputs.

1 GENERAL INTRODUCTION

Vision may be the most intuitive sense to humans. It is also through vision that we primarily communicate science: we use figures, diagrams and equations. It comes as no surprise that vision is the sensory modality that has been studied the most (Hutmacher 2019), from multiple angles and at different scales, ranging from Gestalt psychology to molecular biology. Detailed neurophysiological studies using artificial light patterns have allowed researchers to discover a number of principles underlying visual processing and have led to celebrated results, such as the first receptive field descriptions in the retina (Hartline 1938) or in the primary visual cortex (Hubel and Wiesel 1959). While these results form the conceptual backbone of modern computer vision (LeCun et al. 2010), they have yet to drive reliable biological applications, as exemplified by the limited visual restoration that current retinal prostheses achieve (Erickson-Davis and Korzybska 2021). Thus, despite the decades of research with a focus on precisely designed artificial visual stimuli, we may still be missing big parts of understanding vision in its *natural* settings (Felsen and Dan 2005).

Gestalt psychologists believe that perception may be irreducible, and that "the whole is more than the sum of parts"

In a behaving animal, natural vision drives many mechanisms in unison, and these mechanisms give rise to complex patterns of neural activity. Such complexity is particularly evident in the neocortex, where sensory encoding is multimodal (Ghazanfar and Schroeder 2006): for example, visual neurons may respond to animal behavior generated in response to auditory stimulation (Bimbard et al. 2021). The only part of the visual system that may be spared of such multimodal activity is the retina, whose output we are now comprehensibly cataloging in terms of anatomy (Bae et al. 2018), physiology (Baden et al. 2016), and transcriptomics (Rheaume et al. 2018). These detailed characterizations allow for a systematic treatment of questions that concern the visual system as a whole. How is information in the retinal output used in downstream visual areas such as the visual thalamus (Liang et al. 2018; Rompani et al. 2017) or the superior colliculus (Reinhard et al. 2019)? And is this information used in the same way under different behavioral states (Franke et al. 2021; Liang et al. 2020; Schröder et al. 2020)?

To completely address such questions we also need to understand the neural code of the retina under complex naturalistic stimulation. A way to do this is by building stimulus encoding models, that can predict the spiking output of the retina under arbitrary visual stimuli (Chichilnisky 2001). While somewhat successful general-purpose models of the retina exist (Keat et al. 2001; Pillow et al. 2008), they may fail to capture retinal responses under natural stimuli (Heitman et al. 2016), which drive many retinal mechanisms in parallel. A reason for this failure is that classical models may overlook that the retinal output is the result of elaborate interactions between neurons of the retinal circuit and that these interactions can be highly nonlinear (Gollisch 2013; Schwartz and Rieke 2011). And although we can capture nonlinear retinal processing under targeted artificial stimulation (Liu et al. 2017; Maheswaranathan et al. 2018a; Shah et al. 2020), it is still an open question whether nonlinear encoding models generalize under naturalistic

stimulation.

Understanding nonlinear retinal computation under naturalistic stimulation holds a great promise for understanding natural vision for the entirety of the visual system. In this chapter, we will first introduce the notion of natural vision from a statistical and a behavioral perspective. We will then review how scientists brought natural vision in the laboratory with simple artificial stimuli to systematically explore retinal function. These experiments generated major insights for the retina, which we will briefly discuss and then focus on nonlinear retinal processing. Finally, we will return to natural vision, and describe how we can link it with nonlinear processing, highlighting the topic of this thesis.

1.1 NATURAL VISION

Vision utilizes the differences in light reflected from object surfaces and may have evolved to help animals navigate and rapidly react to changing environments. Thus, the statistics of the natural visual environment may have led to specific retinal adaptations, or guided the design of the visual system as a whole. Below we discuss commonly studied aspects of those statistics and how they may change when an animal moves its eyes through space.

1.1.1 *Statistics of natural scenes*

In terms of spatial structure, natural scenes can be seen as a two-dimensional collection of brightness values, whose most characteristic property is their power spectrum. In particular, the power of spatial frequencies (f) in natural images scales with f^{-2} (Field 1987). Firstly, this power-law property endows images with self-similarity, i.e. the relationships between spatial scales are the same at any viewing distance. Thus, the same visual system can be used for objects that are close by or far away. Secondly, lower spatial frequencies dominate in natural images, and neighboring image pixels have correlated light intensities. The redundancy of such spatial correlations is exploited by image compression algorithms to reduce the size of saved files. It has been proposed that the early visual system specifically evolved to perform such a compression (Barlow 1961). This principle, called efficient coding (discussed in Chapter 2.3), argues that the retinal (or any early sensory) encoding should discard redundant information. The center-surround structure of retinal receptive fields can be directly derived from this principle (Atick and Redlich 1992). Instead of signaling every single pixel with a separate receptive field, the same information can be conveyed by using fewer receptive fields whose surrounding pixels are subtracted from the center ones.

A popular algorithm (JPEG compression) discards high-frequency components that are rare in natural images

Natural scenes show significant correlations also in the temporal domain (Dong and Atick 1995). A scene captured by a static camera shows very little change, except for when it's pointed at rapidly moving organisms. Even with object movement, the power spectrum of temporal frequencies follows a power-law (Dong and Atick 1995; Salisbury and Palmer 2016). The compression argument for temporal processing in the retina is similar to

the case of spatial structure: just encoding differences between successive frames, essentially visual contrast, can greatly improve the efficiency of retinal processing. Indeed, a lot of temporally-redundant information is discarded through visual adaptation in the very first layer of the visual system, i.e. by the rapid adaptation of the phototransduction cascade.

Although spatial and temporal correlations have been historically studied in relation to the retinal circuit, retinal encoding may be affected by other properties of natural scenes. In the spatial domain, natural images may also show textures, edges and gradients (Turiel and Parga 2000), which may require the reevaluation of the center-surround receptive field (Schwartz and Rieke 2011). In the temporal domain, light intensities may differ up to thousand-fold within the same scene (Frazor and Geisler 2006). The visual system is robust to such alterations of light intensity (Rieke and Rudd 2009), and nonlinear mechanisms may be required to support such a purpose (Grimaldi et al. 2019; Yedutenko et al. 2021).

1.1.2 *Viewing natural scenes*

In a behaving animal, it may be possible that the retinal encoding of a visual scene depends on the animal's inner state. Recent *in vivo* studies show systemic effects on the retinal output, such as the ones stemming from circadian rhythms (Hong et al. 2018) or arousal (Liang et al. 2020; Schröder et al. 2020). These effects are possibly mediated by the small number of retinopetal axons in the optic nerve, mainly carrying modulatory histaminergic and serotonergic projections (Gastinger et al. 2006). However, these phenomena may be rather slow compared to the scene changes caused by rapid gaze shifts.

The eyes are essentially motorized sensors, and gaze can be directed not only through the action of extraocular muscles, but also head movement (Land 2015). Freely-viewing primates show stereotyped gaze patterns comprised of a series of short fixations, separated by fast saccades. During fixations, the eyes stay mostly fixed, but still perform lower amplitude fixational eye movements, which can be further separated into few components (Rucci and Poletti 2015). Thus, the temporal dynamics of the scenes impinging on the retina are governed by the timescales relevant to the aforementioned eye movements. For the visual system overall, the exact timescales of eye movements may be of such significance that they are be compensated by head movements in the presence of extraocular muscle deficits (Gilchrist et al. 1997).

What is the purpose of eye movements? In the case of primate fovea, it may be clear that saccades happen for re-positioning attention and better resource allocation. However, the existence of saccadic eye movements seems to be a general pattern in vertebrates (Land 2019), as they are observed in other mammals (Collewijn 1970; Meyer et al. 2020) and even in fish (Easter 1975), and saccades may thus provide benefits independent from the existence of a fovea. For example, saccades can whiten the spatially correlated natural scenes in time, especially in the case low spatial frequencies (Mostofi et al. 2020), and thus provide the visual system with an independent mechanism for the decorrelation of the retinal output. Multiple roles have been

suggested for fixational eye movements (Martinez-Conde et al. 2004; Rucci and Poletti 2015), such as the prevention of visual fading, but their role in natural vision remains an open question. It is now becoming commonplace to include saccadic and fixational eye movements in the design of naturalistic visual stimuli for the retina (Heitman et al. 2016; Turner and Rieke 2016), and movement effects are already noticeable at the level of the retinal output (Roska and Werblin 2003; Segal et al. 2015).

1.2 VISUAL ENCODING AND THE RETINAL CIRCUIT

The retina is a part of the central nervous system in the back of the eye. Cajal (1893) was the first to describe the cellular components of the retinal structure (Figure 1.1A), which is highly conserved among all vertebrates (Baden et al. 2019). Light is captured by the photoreceptors, which transduce it into electrical activity and subsequently, synaptic release of glutamate. The neurotransmitter signal is then relayed to bipolar cells, which split the visual signal into parallel streams, and in turn excite ganglion cells. In between, two classes of interneurons, horizontal and amacrine cells modulate the signals. Horizontal cells are driven by and inhibit photoreceptors, and potentially bipolar cells (Behrens et al. 2021). While horizontal cell processing is limited in the outer plexiform layer, the results of this processing can be detected in the retinal output (Drinnenberg et al. 2018). Amacrine cells are driven mainly by bipolar cells, and in turn affect bipolar cell terminals, ganglion cells or other amacrine cells. While systematic physiological characterizations of ganglion and bipolar cells exist (Baden et al. 2016; Franke et al. 2017; Goetz et al. 2021), amacrine cells still remain a mystery (Masland 2012b). Because of their diverse morphology, connections, and functional roles (Yan et al. 2020), amacrine cells are thought to be the computational powerhouses of the retina.

How does the interplay of all these neuron types come together in the retinal output? The functional characterization of the retinal output can be achieved by probing the retina with structured visual stimuli and recording the responses of retinal ganglion cells. These experiments were first performed in either *ex vivo* preparations of the retina, or by direct recordings in anesthetized animals. Simple light flashes gave rise to the core concept of response polarity (Hartline 1938; Kuffler 1953), essentially the distinction of ganglion cells into ON- and OFF-types based on their sensitivity to light increments or decrements, respectively. This distinction arises from different glutamate receptors expressed in ON- and OFF-type bipolar cells (Nakajima et al. 1993), which either invert or retain the polarity of photoreceptor responses.

When presented within a limited spatial extent, light flashes led to another core concept of visual processing: the receptive field. The receptive field describes the region in visual space where a neuron responds to changes in light (Figure 1.1B), and was first coined for retinal ganglion cells (Hartline 1938; Kuffler 1953), following its original description in somatosensation (Sherrington 1906). Spots of light presented inside the receptive field can drive spiking responses, while the same spots presented further away will have no effect. The receptive field can be further subdivided into two con-

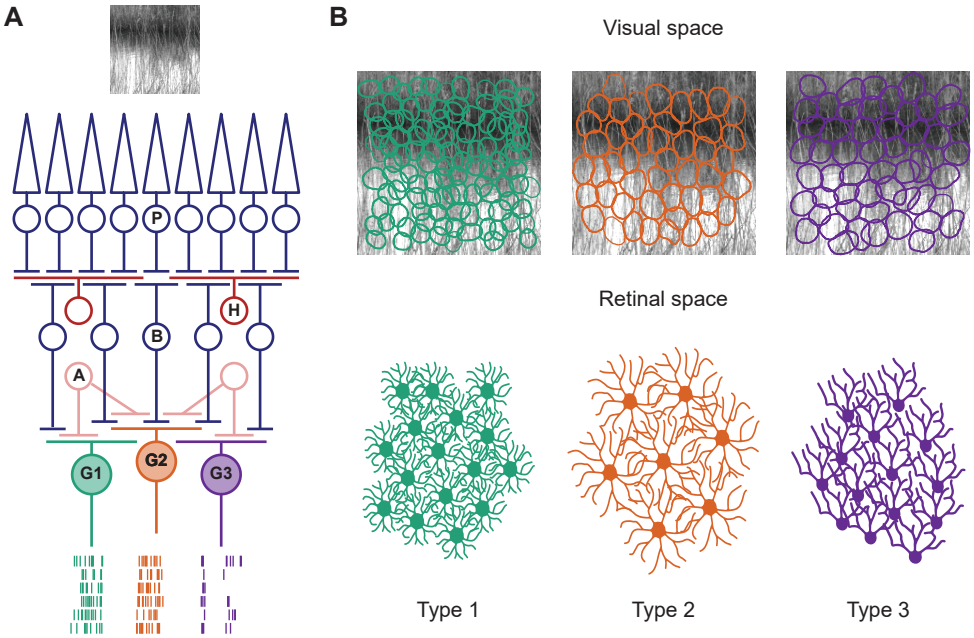

Figure 1.1: **The vertebrate retina splits visual information into multiple parallel outputs.** (**A**) Schematic structure of the vertebrate retina. Light from a natural scene is captured by the photoreceptors (P), which transduce the visual stimulus to an electrical signal. The signal is then passed on to the bipolar cells (B), which in turn excite different types of retinal ganglion cells (G1-3). Horizontal cells (H) modulate photoreceptor signals at the outer plexiform layer, while amacrine cells (A) receive bipolar cell inputs and modulate bipolar, ganglion and even other amacrine cells. The retinal output is sent to the brain in the form of spiking responses of different types of ganglion cells. (**B**) Receptive fields of neurons of the same type tile the visual scene, and different ganglion cell types send signals to the brain in parallel. The tiling of receptive fields in visual space matches the titling of ganglion cell dendritic fields in retinal space.

centric regions: center and surround (Barlow 1953; Kuffler 1953). Spots of the same size as the receptive field center drive strong responses, but larger spots will drive weaker responses, or none at all (Jacoby and Schwartz 2017; Zhang et al. 2012). The region where responses become weaker is the receptive field surround. ON-type ganglion cells have an OFF-type surround, and vice-versa. This means that an ON cell can either be driven by light increments in its center, or light decrements in its surround. Ganglion cell surround is mediated by lateral inhibition at two levels: through horizontal cells at the outer (Mangel 1991; McMahon et al. 2004) and through amacrine cells at the inner plexiform layer (Cook and McReynolds 1998; Johnson et al. 2018).

A common model of center-surround structure is the difference-of-Gaussians, which is also a feature enhancement algorithm

Computational models can summarize the core concepts of retinal processing and can generate response predictions under arbitrary visual stimuli. Perhaps the most successful encoding model in terms of its simplicity and component interpretability is the linear-nonlinear (LN) model (Figure 1.2A). The first part of this model consists of linear spatial filtering, which pools stimulus pixels with a set of weights corresponding to the spatial receptive field of the cell. The output of this pooling is typically followed by linear

temporal filtering, which captures the response kinetics of the receptive field center. In the nonlinear stage of the model, a static nonlinearity is typically included: this nonlinearity transforms the filter's output into a firing rate, which can directly be compared to neural responses. The nonlinearity can capture nonlinear effects in the ganglion cell such as the existence of a spiking threshold or firing rate saturation. The components of this LN model can be fit directly to retinal responses, a process typically done under stimulation with dense white noise (Chichilnisky 2001; Meister et al. 1994), however this is also possible under other stimuli, including natural movies (Heitman et al. 2016). Linear filters and nonlinearities are routinely used to characterize ganglion cells (Baden et al. 2016), as they offer a comprehensive summary of the cells' light responses. LN models and their variations have been quite successful in predicting the retinal output under simplified artificial stimuli, with no or coarse spatial structure (Keat et al. 2001; Pillow et al. 2008). However, it is an active research question whether these models suffice to capture ganglion cell responses under natural stimuli (Heitman et al. 2016; Turner and Rieke 2016).

1.2.1 *Retinal ganglion cell types*

The discussion about retinal coding of natural stimuli is complicated by the existence of multiple types of retinal ganglion cells. While bipolar and amacrine cells also show significant diversity, we focus on ganglion cells since they are the output channels of the retina. The estimate of retinal ganglion cell types in the mouse is now converging to more than 40 (Bae et al. 2018; Yan et al. 2020). Retinal ganglion cells of the same type tile visual space (Figure 1.1B). Each type is connected to particular regions in the brain (Martersteck et al. 2017; Morin and Studholme 2014), and is thought to convey a different picture of the visual world (Masland 2012a). A main target of ganglion cells is the lateral geniculate nucleus (LGN) of the thalamus, which leads retinal information to the visual cortex. However, the messages conveyed by the ganglion cells are not limited to image-forming functions. For example, ON direction-selective cells are connected to the accessory optic system and drive the optokinetic reflex (Krause et al. 2014), while intrinsically-photosensitive ganglion cells innervate the suprachiasmatic nucleus to entrain circadian rhythms (Berson et al. 2002), or the habenula for regulating mood (Fernandez et al. 2018).

The primate retina has a distinct composition of ganglion cell types. There are two major cell classes that constitute 50-80% of all ganglion cells, midget and parasol cells, both with corresponding ON and OFF types. Midget and parasol cells have been linked to the ventral and dorsal image-forming visual pathways, respectively, responsible for categorical and motion vision. Besides the two main classes, the so-called wide-field ganglion cells constitute the remaining types, and appear mainly in the retinal periphery. The function of many of those wide-field cells is still unknown. The total number of ganglion-cell types is around 18-20 (Masri et al. 2017; Peng et al. 2019), which is approximately half of the estimates for the mouse. This reduction of types in the primate from mammalian ancestors may reflect an

Ventral and dorsal pathways are sometimes called the 'what' and 'where' systems

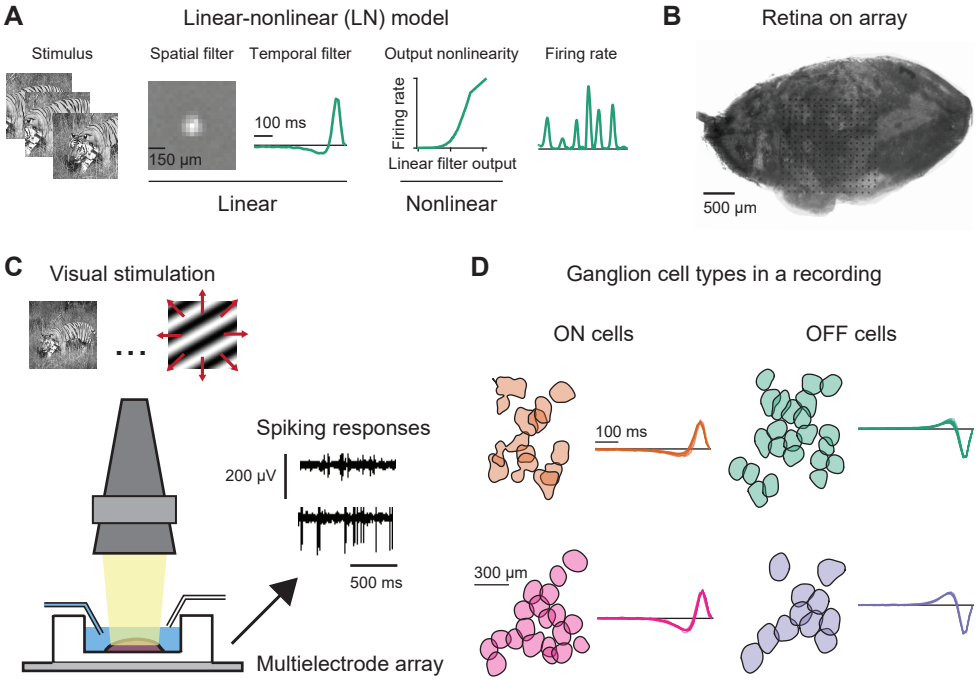

Figure 1.2: **Studying ganglion cell types in the retina with multi-electrode array recordings.** (**A**) The linear-nonlinear model generates a firing rate prediction for a single retinal ganglion cell under an arbitrary spatio-temporal visual stimulus. The stimulus is first convolved with a spatial and then a temporal filter, both of which linearly process the scene. The result of this operation is then transformed to spiking activity through the output nonlinearity. (**B**) Half of the mouse retina flat on top of a 252-electrode multi-electrode array (MEA). (**C**) Schematic setup of an MEA recording. Various visual stimuli are presented on the ex-vivo retina. The retina is continuously superfused with Ames medium, while extracellular voltage signals are captured from the array. (**D**) Different types of retinal ganglion cells can be distinguished in an MEA recording. Each type is typically summarized by the receptive field outlines of its members and their temporal filters, both components of the linear-nonlinear model. Note that all of the four types shown exhibit receptive field tiling.

increased need for access to detailed light intensity information from the highly developed ventral visual stream (Lindsey et al. 2019).

What roles do all of these types play in the context of the natural visual environment? A useful distinction here is the one of feature detectors and pixel encoders (Roska and Meister 2014). Under this picture, some cell types have developed to provide raw light intensity information to the brain for further processing, with the prime example of the midgets cells in the primate fovea, or similar types in the mouse (Johnson et al. 2018). Conversely, feature detectors or "bug-perceivers", first described in the frog retina (Lettvin et al. 1959), perform computations and provide an easy (e.g. linear) readout to downstream areas in order to efficiently guide behavior (Gollisch and Meister 2010). For example, different types of salamander ganglion cells can track small moving objects (Leonardo and Meister 2013) or represent global motion (Kühn and Gollisch 2019).

Ganglion cells of the same type are not always exact copies of one another. Continuous variation within cells of the same type can appear in different

parts of the nervous system (Cembrowski and Menon 2018; Cembrowski and Spruston 2019). Such inhomogeneity also exists in the retina, with the most illustrative example being the midget circuit in the fovea, which completely differs from its peripheral counterpart. Midget cells in the fovea receive input from as few as a single bipolar cell, which itself is connected to a single cone. Besides different convergence, midget cells in the fovea show slower light responses, and receive little inhibition compared to the periphery (Sinha et al. 2017). Regional specializations are also becoming apparent in the mouse retina (Heukamp et al. 2020; Sabbah et al. 2017) and are thought of as adaptations to the mouse's natural visual environment (Qiu et al. 2021).

The "private-line" midget circuit

1.2.2 *Capturing the retinal output with multi-electrode arrays*

To characterize stimulus-response relationships in the retina, the spiking responses of retinal ganglion cells are routinely recorded with extracellular electrodes. While nowadays brain studies can use high-density electrode arrays, such as the Neuropixels probe (Jun et al. 2017), the technology was already available for the retina two decades ago (Meister et al. 1994). The flat structure of the retinal tissue makes it perfect for a planar multi-electrode array (MEA; Figure 1.2B). Compared to single-electrode recordings, MEAs offer two main advantages. Firstly, their parallel throughput allows for the simultaneous characterization of light response properties in hundreds of ganglion cells—but also polyaxonal amacrine cells (Greschner et al. 2014; Kling et al. 2020). The mapping of receptive fields, for example, is typically parallelized though white noise stimulation (Wienbar and Schwartz 2018) which efficiently estimates the spatial filters of each ganglion cell. This is in stark contrast with single-cell recordings, which instead use flashed spots in the vicinity of the recording electrode. Secondly, MEAs allow for many-hour-long recordings because they are non-invasive to ganglion cells and compatible with high flow rates of extracellular medium (Figure 1.2B); thus MEA recordings offer a unique opportunity for fitting statistical models of retinal encoding that may require large amounts of data.

MEA recordings also come with a handful of challenges. The first challenge is technical, and involves the extraction of spiking events from extracellular voltage traces, and the attribution of those events to single neurons. This process is called spike-sorting and it involves the automated clustering of spiking events in different units, based on the electrical signature of those events on multiple electrodes. Despite a recent boom in the available algorithms (Lee et al. 2020; Pachitariu et al. 2016; Yger et al. 2018), automated spike-sorting may still sometimes fail to resolve all cells in the recording, which leads to spurious units that combine spikes from multiple neurons. To overcome this issue, manual curation of the identified units is typically required. The sorted units are used to study light response properties over the whole population of recorded ganglion cells (Carcieri et al. 2003; Reinhard and Münch 2021), but it is often illustrative to perform such analyses in a cell-type-specific manner (Field et al. 2007; Ravi et al. 2018; Rhoades et al. 2019). Another challenge associated with MEA recordings involves the identifiability of particular types of retinal ganglion cells. Ganglion cells with

small somata produce low-amplitude spikes that cannot be spike-sorted, and some cells may be completely silent for full-field stimuli (Jacoby and Schwartz 2017; Zhang et al. 2012), typically used in an MEA recording. Finally, understanding the mechanistic origin of stimulus encoding is limited compared to e.g. whole-cell patch clamp, where inhibitory and excitatory inputs to a ganglion cell can be readily distinguished.

Informed analysis methods can now overcome some of the shortcomings of MEA recordings. Starting from the example of the primate retina (Field et al. 2007; Rhoades et al. 2019), it is possible to group cell into multiple consistent types, as also exemplified in the rat (Ravi et al. 2018) retina. This can be achieved through clever light stimulation techniques in conjunction with modern clustering algorithms (Drinnenberg et al. 2018; Jouty et al. 2018). Although MEAs may fail to capture spiking responses of some cell types, receptive field mosaics of major cell types can be readily identified (Figure 1.2D). Non-overlapping mosaics are tell-tale signs of ganglion cell types, and populations of same-type ganglion cells can be used for both stimulus encoding and decoding studies (Brackbill et al. 2020; Roy et al. 2021). Finally, the growing availability of functional response datasets ranging from single-cell patch-clamp (Goetz et al. 2021) to calcium imaging (Baden et al. 2016), allows the matching of units to previously identified cell types based on responses to a standard set of visual stimuli (Román Rosón et al. 2019).

1.2.3 *The mouse retina as a model system*

The mouse (*Mus musculus*) retina is emerging as a model system (Huberman and Niell 2011) for studying the retinal output in a cell-type-specific manner. The anatomical (Bae et al. 2018) and functional (Baden et al. 2016; Goetz et al. 2021) catalogs of ganglion cell types, paired with genetic access to many of those types, allow the specific tracing of central projections of ganglion cell types (Martersteck et al. 2017), or assessing the effects of deactivating particular cell types in mouse behavior (Kim et al. 2020; Wang et al. 2021). Compared to primates, mice may rely less on their vision for navigating their environment. However, the organization of their visual system may be similar to primates, with candidate ventral and dorsal streams (Smith et al. 2017; Wang et al. 2011). There is also ample evidence (Seabrook et al. 2017) for various visually-guided behaviors in mice (Figure 1.3A). Examples include the avoidance of overhead predators (Yilmaz and Meister 2013), which may correspond to distinct behaviors (freezing or escaping) based on the exact parameters of the overhead visual stimulus (De Franceschi et al. 2016). Capturing moving prey also requires vision in multiple stages, starting from the initial approach to making the final capture (Hoy et al. 2016; Johnson et al. 2021; Michaiel et al. 2020). Finally, mice can innately discriminate depth, an ability usually tested with the "visual cliff" (Boone et al. 2021; Fox 1965).

Despite its lack of fovea, the mouse retina has its own interesting inhomogeneities (Figure 1.3B) that may ultimately affect the response properties of retinal ganglion cells. The most prominent one concerns the spectral sensitivity of cones: UV-sensitive S-cones dominate the ventral and green-sensitive M-cones the dorsal retina (Applebury et al. 2000; Szél et al. 1992), a

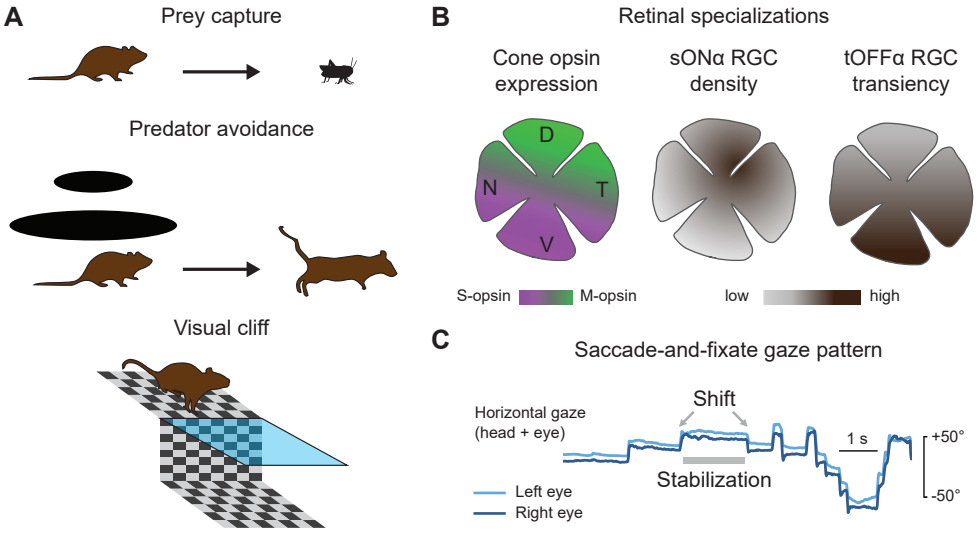

Figure 1.3: **The mouse as an animal model for vision science.** (**A**) Mice show consistent visual behaviors. Mice use vision to track, approach and capture moving prey (top), usually tested with moving crickets. In response to an expanding dark disk which emulates a predator, mice show stereotypical avoidance behavior (middle): they either escape, or freeze. When placed on a patterned cliff with a glass-covered deep side (visual cliff; bottom), mice prefer the shallow side, presumably due to height avoidance. (**B**) The expression of UV-sensitive S-opsin and green-sensitive M-opsin is inhomogeneously distributed across the surface of the mouse retina. Such inhomogeneities also show up in the retinal ganglion cell (RGC) layer: ON alpha RGCs show smaller dendritic trees and higher density in the dorsotemporal retina (middle), and OFF alpha transient RGCs become more sustained in the dorsal retina (right). (**C**) In the horizontal plane, mice show conjugate eye movements reminiscent of primate saccades, interleaved with stabilization periods of similar duration to fixations. Panel (C) is adapted from Meyer et al. (2020), licensed under CC BY 4.0.

distribution that may be linked to differences observed in the chromatic light response properties of ganglion cells (Szatko et al. 2020). Other gradients also exist in the light response properties of ganglion cells: sustained ON alpha cells have higher density and smaller receptive fields in the dorsotemporal retina (Bleckert et al. 2014), while transient OFF alpha ganglion cells become more sustained in the dorsal retina (Warwick et al. 2018). The exact arrangement of dorsal and ventral specializations is thought to be of particular importance, as mice show non-conjugate eye movements that align their gaze relative to the horizon (Meyer et al. 2020), thus keeping the ventral part of the retina facing overhead.

Mice also show conjugate eye movements in the horizontal plane (Meyer et al. 2020), which are similar to human saccades and can determine the temporal dynamics of the retinal image. These *saccade-and-fixate* patterns may allow mice to precisely track and capture prey (Michaiel et al. 2020). While mice have no fovea, a recent study suggested that a retinal area of increased ganglion cell coverage may serve prey capture (Holmgren et al. 2021). Both sustained ON and OFF alpha types show increased ganglion cell coverage in the same area (Bleckert et al. 2014), and it would be intriguing to compare whether the functional properties of cells in this area differ

compared to other areas, as is the case for midget cells in the primate fovea (Sinha et al. 2017).

1.3 NONLINEAR PROCESSING IN THE RETINA

In the retina, the extraction of complex visual features from natural scenes is thought to rely on nonlinear interactions between neurons upstream of retinal ganglion cells (Gollisch and Meister 2010). In general, nonlinear operations form the computational core of biological and artificial neural networks. The significance of stacking nonlinear operations stems from the theory of artificial neural networks (Hornik et al. 1989), because the application of consecutive nonlinearities to an input signal allows for non-trivial computations, such as the separation of points lying on two sides of a complex surface. Including such multi-stage nonlinearities in toy models of the retinal circuit can explain complex computations, from detecting dim light (Field and Rieke 2002) to separating local from global motion (Baccus et al. 2008; Ölveczky et al. 2003).

The universal approximation theorem states that nonlinear neural networks of arbitrary depth can approximate any continuous function

Because of their likely relevance for the encoding of natural stimuli, we will focus here on retinal nonlinearities that manifest over visual space. Such nonlinearities were first identified in the cat retina with experiments that introduced spatial structure within the receptive field center of ganglion cells (Enroth-Cugell and Robson 1966). A subset of ganglion cells, termed Y-cells, responded to the presentation of a visual grating that stimulated the receptive field with equal amounts of brightening and darkening: a net zero activation. Although cells that were silent were also found (X-cells), the existence of Y-cells contradicts the notion of the linear receptive field, which is included in LN-type models of retina. The Y-cell response is typically termed nonlinear spatial integration and is usually investigated with contrast-reversing gratings (Figure 1.4A). Y-cells have been reported in multiple species, such as salamanders (Bölinger and Gollisch 2012), mice (Borghuis et al. 2013; Carcieri et al. 2003) and primates (Crook et al. 2008a; Crook et al. 2008b; Freeman et al. 2015), including humans (Kling et al. 2020). Multiple ganglion cell types of the same species may behave as Y-cells (Crook et al. 2008a; Crook et al. 2008b; Krieger et al. 2017; Mani and Schwartz 2017), and this characterization also extends beyond the retina (Kaplan and Shapley 1982; Yeh et al. 2003). Nonlinear spatial integration persists for grating reversals of multiple scales, but disappears when bar widths become too narrow (Enroth-Cugell and Freeman 1987). This minimal scale of nonlinear spatial integration is much smaller than the receptive field centers of ganglion cells, indicating the phenomenon may originate in the ganglion cell inputs.

1.3.1 *Mechanisms underlying nonlinear spatial integration*

Where exactly do spatial nonlinearities upstream of the optic nerve originate from? All spiking neurons are essentially nonlinear, because of their spiking threshold stemming from voltage-gated sodium channels at the axon hillock. However, the inner network of the retina operates mostly on

graded potentials, which may be efficient over short distances (Sterling and Laughlin 2015). Demb et al. (2001) showed that nonlinear spatial integration comes about at the bipolar-to-ganglion cell interface and is probably related to nonlinear transformation between bipolar cell voltage and the consequent glutamate release (Borghuis et al. 2013). This explains why the spatial scale of nonlinear spatial integration typically matches the receptive field sizes of bipolar cells (Jacoby and Schwartz 2017; Mani and Schwartz 2017; Schwartz et al. 2012). For nonlinear spatial integration to manifest in the retinal output, the spatial distribution of the bipolar cell inputs must be relatively wide: overlapping bipolar cell receptive fields will encode the similar light intensities, making the nonlinearity effectively obsolete. This phenomenon is taken to the extreme in the primate fovea, where midget ganglion cells typically collect the output of a single bipolar cell. Nonlinear spatial integration depends also on the exact nonlinear transformation, as the synaptic release of glutamate may differ between different bipolar cell types: OFF cells show a more rectified glutamate response compared to ON cells (Borghuis et al. 2013). Besides ganglion cells, amacrine cells also receive bipolar cell inputs and thus may show nonlinear spatial integration (Kim et al. 2015; Murphy-Baum and Taylor 2015).

Other mechanisms have also been implicated in nonlinear spatial integration, and the nonlinearities upstream of retinal ganglion cells may be multi-stage. In the primate retina, smooth monostratified cells show an enlarged scale of nonlinear spatial integration which has been hypothesized to originate from active dendritic conductances (Rhoades et al. 2019). Cortical neurons also show active conductances upstream of their axon hillock, located at the dendrites. Nonlinear dendrites are being actively studied (Ujfalussy et al. 2015, 2018), and they are thought to increase the computational repertoire of the nervous system in general (Bicknell and Häusser 2021), as they add more nonlinear layers within each neuron. Another retinal mechanism for nonlinear spatial integration is located at the outer plexiform layer: some salamander bipolar cell types may be spatially nonlinear themselves (Schreyer and Gollisch 2021). Such a mechanism may grant ganglion cells sensitivity to scales below bipolar cell receptive fields, as is also observed in the mouse retina (Mani and Schwartz 2017).

The nonlinear transformation between bipolar and ganglion cells may be affected by the interaction of multiple mechanisms. A typical way to determine the nonlinearity in ganglion cell inputs is to measure the contrast-response function of the incoming excitation. Bipolar cell excitation can become more rectified in higher light intensities (Grimes et al. 2014), or when multiple adjacent bipolar cells are activated at the same time (Kuo et al. 2016). Inhibition can shift bipolar cell excitation to a more linear regime both through surround signals that depolarize the bipolar cell (Turner et al. 2018), or via crossover (from the ON to the OFF pathway and vice-versa) signals carried by narrow-field amacrine cells (Münch et al. 2009; Werblin 2010; Yu et al. 2021). It becomes thus evident that different spatial and temporal contexts, such as those provided by actively exploring a natural visual scene, may complicate the relevance of nonlinear spatial processing under naturalistic stimulation.

1.3.2 *Nonlinear subunit models capture nonlinear spatial integration*

How to capture nonlinear spatial integration in stimulus encoding models? Because linear receptive field models cannot capture nonlinear spatial integration, Victor and Shapley (1979) introduced nonlinear subunit models to explain the spatial sensitivity of cat Y-cells to contrast-reversing gratings (Enroth-Cugell and Freeman 1987). In subunit models (Figure 1.4B), the spatial receptive field of a ganglion cell is separated into functional subunits whose outputs are first nonlinearly combined, and then summed (Figure 1.4B). In the original descriptions, this nonlinearity was rectification: for an ON-type subunit, any darkening will cause a zero activation (Enroth-Cugell and Freeman 1987; Victor and Shapley 1979). In the retinal circuit, subunit receptive fields are thought to correspond to the receptive fields of bipolar cells providing input to the ganglion cell (Liu et al. 2017).

Variations of nonlinear subunit models have been used to explain various retinal responses to artificial stimuli. For example, nonlinear subunits are required to compute object-motion sensitivity (Baccus et al. 2008) and approach sensitivity (Münch et al. 2009), both elicited with complex motion stimuli. However, it has been challenging to fit general-purpose nonlinear models to ganglion cell responses. Compared to LN-type models that have a single spatial filter for the receptive field, subunit models require multiple filters equal to the number of subunits, and this number is not *a priori* known. Other parameters that need to be fit should describe the functional form of the subunit nonlinearities, and the weights under which subunits are pooled by the ganglion cell, which can be highly non-Gaussian (Schwartz et al. 2012). Recent advances in computational methods allow the estimation of some of the required subunit model parameters from neural responses (Liu et al. 2017; Maheswaranathan et al. 2018a; Shah et al. 2020). All of these methods are based on fitting ganglion cell responses under white-noise stimulation and focus on extracting the correct subunit layout by determining the correct number of subunits and their filters.

Because subunit models have filters that are localized in space, they may resemble computations across the whole hierarchy of the visual system. For example, the proposed circuit of orientation selectivity in the cat primary visual cortex includes nonlinear subunits located in the visual thalamus (Hubel and Wiesel 1962). To simplify parameter estimation, nonlinear subunit models in higher visual areas have been fit under the convolutional architecture, where subunits share the same filters which are applied over different locations in visual space (Vintch et al. 2015). A general-purpose nonlinear subunit model for retinal ganglion cells may greatly benefit by the convolutional architecture, as it resembles the structure of the retinal circuit. Many ganglion cells receive input just from a single bipolar cell type (Schwartz et al. 2012; Yu et al. 2018), and bipolar cell receptive fields also tile visual space. Thus, assuming that all subunit filters are the same—at least locally—for each ganglion cell may be a valid assumption that can aid model fitting. This approach has successfully been used to fit convolutional neural networks to retinal responses under natural stimuli (Goldin et al. 2021; Maheswaranathan et al. 2018b; McIntosh et al. 2016).

The first convolutional neural network, the "neocognitron", was inspired by the work of Hubel and Wiesel (Fukushima 1980)

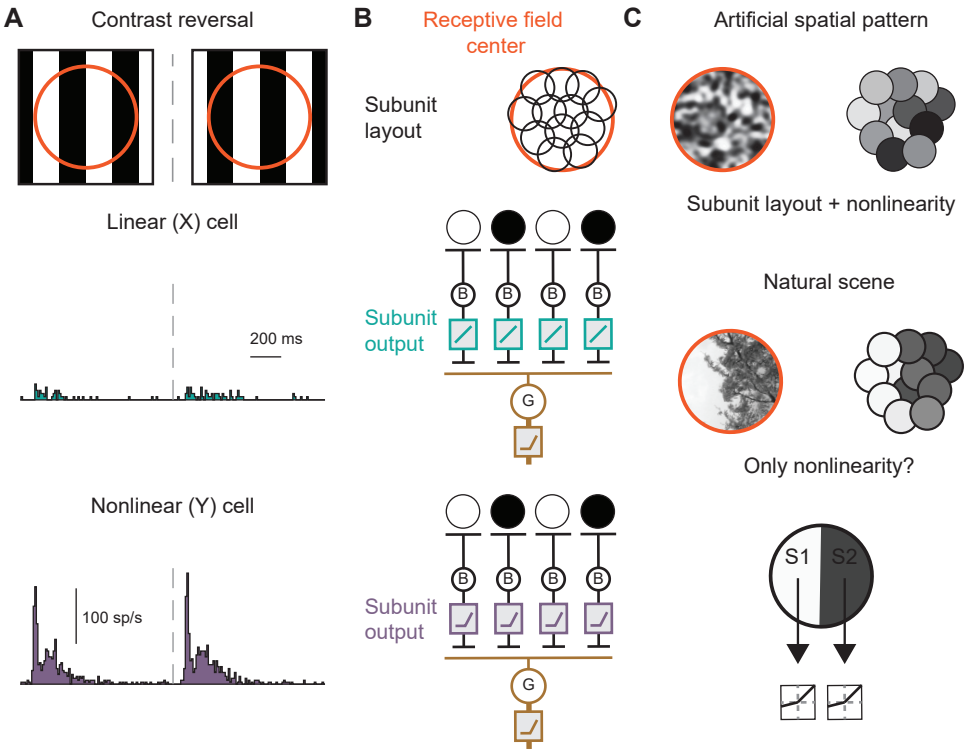

Figure 1.4: **Nonlinear spatial integration in retinal ganglion cells.** (**A**) A contrast-reversing grating will not drive responses in linear X-cells, because dark and bright regions inside the receptive field (orange) cancel each other out. Instead, Y-cells will respond to the grating because they integrate space nonlinearly. (**B**) The nonlinear subunit model can explain the Y-cell response. In a subunit model, the receptive field is partitioned into multiple functional subunits, whose outputs are first rectified and the summed (bottom). For a linear cell (middle), the subunit outputs are linear. (**C**) For an artificial spatial pattern with high frequency components (top), both the subunit layout and nonlinearity are expected to matter in shaping the response of the ganglion cell. For a natural scene (middle), where low frequency components dominate, nonlinearities in spatial integration are expected to matter the most, because any subunit layout that covers the receptive field would produce similar distributions of subunit outputs. Thus, it may be possible to study nonlinear spatial integration in natural scenes, just by identifying nonlinearities in integration of two contrast values (S1 and S2) inside the receptive field center.

1.4 NONLINEAR SPATIAL INTEGRATION AND NATURAL VISION

Is nonlinear spatial processing even relevant for the retinal encoding of natural stimuli? Usually spatial integration is assessed with reversing gratings of high spatial frequency that isolate the contributions of nonlinear excitation to the ganglion cell response (Figure 1.4A). However, lower spatial frequencies in natural scenes are more pronounced: spatial correlations dominate, with occasional sharp edges, gradients and textures (Turiel and Parga 2000). Furthermore, the timescales of natural vision are determined by saccade-and-fixate dynamics, which may strongly deviate from both reversing-gratings and white-noise stimulation. Thus, it may be the case that nonlinear processing is not at all driven under naturalistic stimulation or that nonlinear contributions to ganglion cell responses are actively suppressed (Yu et al. 2021).

Model-based approaches allow the study of nonlinear spatial integration with natural stimuli. If a given ganglion cell has a linear receptive field, it should respond the same way to both a natural image and a spot of light with same average light intensity. This approach was used to distinguish parasol cells in the macaque retina into linear ON cells and nonlinear OFF cells (Turner and Rieke 2016). Surprisingly, ON parasol cells showed nonlinear responses to classical reversing gratings, revealing an asymmetry between natural and artificial stimulus characterizations (Turner and Rieke 2016; Yu et al. 2021). Together, these initial results suggest that to understand nonlinear retinal processing, we need to study how different cell types respond to the spatial and temporal structure that the retina is exposed to when viewing natural scenes.

Do general-purpose nonlinear subunit models capture responses to naturalistic stimuli? Initial results suggest that, relative to classical LN models, nonlinear subunit models can indeed improve responses to natural stimuli (Liu et al. 2017; Shah et al. 2020). However, these studies used small numbers of retinal ganglion cells, and it is still unclear how to generalize response predictions over arbitrary cell types, that may differ substantially in how they integrate spatial contrast (Turner and Rieke 2016), even between spatially nonlinear types (Bölinger and Gollisch 2012). We hypothesize that the nonlinearity under which subunits are integrated largely determines responses to natural scenes: while the exact subunit layout might matter for detailed stimuli with high spatial frequencies (Schwartz et al. 2012), natural stimuli only have rough features and the spatial structure that they will most likely include is an edge, or a steep gradient, leading to a roughly homogeneous activation of subunits on one side of the receptive field (Figure 1.4C).

What are the functional consequences of a spatially nonlinear receptive field for the efficient coding of natural movies? Although center-surround receptive fields and biphasic temporal filters aid in the reduction of redundancies in the neural code of the retina, nonlinearities in the retinal output may be the major driver of retinal decorrelation (Pitkow and Meister 2012). It is thus intriguing to consider whether subunit nonlinearities, such as those existing in subunit models, may be the main contributor to such decorrelation, as it has been suggested through ganglion cell simulations (Maheswaranathan et al. 2018a).

1.5 THESIS OUTLINE

In this thesis, I investigate whether and how nonlinear spatial processing affects natural stimulus encoding in the retina. I first introduce the approaches under which retinal encoding of natural scenes is currently been studied (Chapter 2). Historically, the first approach was the *normative*, which seeks to understand which anatomical and physiological features of the retinal circuit can be predicted by natural scene statistics and first principles alone. Instead, the more modern *circuit* perspective aims at understanding retinal mechanisms relevant under naturalistic stimulation. Under the framework of the *circuit* perspective, I proceed to show that nonlinear spatial integration of ganglion cell inputs is relevant for natural images (Chapter 3). I also show that nonlinearities may differ between ganglion cells, and affect cell sensitivity to the spatial structure of natural scenes. To explain this differential sensitivity to spatial structure, I develop computational models of retinal ganglion cells and fit them to spiking responses from the mouse retina (Chapter 4). In particular, I extend currently existing nonlinear subunit models, by implementing the convolutional architecture under a stimulus that can strongly drive the subunit layout of retinal ganglion cells. These models explain the different types of spatial structure sensitivity I previously observed, identify key differences in the nonlinear receptive fields of common cell types, and can well capture responses to natural images and movies. Finally, I attempt to connect the *circuit* and *normative* perspectives, by investigating the functional consequences of a nonlinear receptive field for retinal decorrelation.

2 RETINAL ENCODING OF NATURAL SCENES

Attached manuscript. Invited review for *Annual Reviews in Vision Science*, currently in press. Reprinted below is the manuscript at its current version.

Authors:

Dimokratis Karamanlis[1,2,3], Helene Marianne Schreyer[1,2], Tim Gollisch[1,2]

1. University Medical Center Göttingen
2. Bernstein Center Göttingen
3. International Max Planck Research School for Neurosciences, Göttingen

Own contributions:

- Conceptualization (with TG)
- Visualization (with TG and HMS)
- Writing – original draft (with TG)
- Writing – reviewing and editing (with TG and HMS)

Retinal encoding of natural scenes

Dimokratis Karamanlis[1,2,3], Helene Marianne Schreyer[1,2], Tim Gollisch[1,2]

1. Department of Ophthalmology, University Medical Center Göttingen, 37073 Göttingen, Germany
2. Bernstein Center for Computational Neuroscience Göttingen, 37077 Göttingen, Germany
3. International Max Planck Research School for Neurosciences, 37077 Göttingen, Germany

Keywords: retina, natural stimuli, receptive field, neural code, circuit mechanisms, computational modeling

ABSTRACT

An ultimate goal in retina science is to understand how the neural circuit of the retina processes natural visual scenes. Yet, the vast majority of studies in laboratories have long been performed with simple, artificial visual stimuli, such as full-field illumination, spots of light, or gratings. The underlying assumption is that the features of the retina thus identified carry over to the more complex scenario of natural scenes. As the application of corresponding natural settings is becoming more commonplace in experimental investigations, this assumption is being put to the test and opportunities arise to discover processing features that are triggered by specific aspects of natural scenes. Here, we review how natural stimuli have been used to probe, refine, and complement knowledge accumulated under simplified stimuli, and we discuss challenges and opportunities along the way towards a comprehensive understanding of the encoding of natural scenes.

INTRODUCTION

The sense of vision has evolved to solve the challenges encountered by animals in the visual environment. Thus, natural visual scenes provide the most essential stimulus context under which the neurons and neural circuits of the visual system operate. And it is this mode of operation and the processes required to solve natural visual tasks that are of primary interest to vision researchers. The challenge is, of course, that natural scenes are complex with intricate structure on a wide range of spatial and temporal scales. Light levels and spatial contrast can vary drastically in different regions of the same image, object boundaries create sharp transitions in brightness and color information, whereas surface textures and illumination gradients provide smoothly varying components. Temporal dynamics add further complications in the form of moving objects and self-motion, including eye movements, as well as alterations in illumination conditions.

In the vertebrate visual system, the first neural stage that has to deal with the complex, multi-faceted structure of natural scenes is the retina, a thin sheet of neural tissue at the back of the eyeball. It is among the most investigated parts of the nervous system, and a great deal of anatomical and physiological detail has been revealed about its sensory transduction mechanisms, its immense diversity of neuron types, and their responses to light. This vast background knowledge and the immediacy with which the visual stimulus acts on this neural system make it a great target for studying the encoding of natural scenes and for evaluating how the visual system copes with different aspects of the natural environment.

Yet, even for the retina, taking the full complexity of natural scenes into account seems daunting. It is thus no surprise that much of vision research has been guided by a reductionist strategy of separating visual processing into smaller, manageable chunks. The separation may be into specific tasks (e.g., to detect a luminance change or a motion signal, to efficiently transmit information) or into specific stimulus features (temporal contrast, periodic spatial structure, uniform motion, color, etc.). This divide-and-conquer strategy has had undeniable success, from general characteristics of receptive fields to specific computations ascribed to individual ganglion cell types.

The hope is, of course, that the understanding of the stimulus encoding, of the visual tasks that are solved, and of the circuit mechanisms hold up also for natural stimuli. To turn hope into certainty, there is a growing interest in the retina field to re-evaluate classical concepts – originally derived with simple, artificial stimuli – under natural stimulation. Moreover, one may expect that challenging the retina with natural scenes may reveal response patterns or mechanisms beyond those that are observed with traditional artificial stimuli.

Here, we consider three ways in which natural stimuli have been used to analyze signal processing in the vertebrate retina and discuss the insights and challenges that have emerged in recent years (**Figure 1a**). Perhaps most famously, the *normative perspective* seeks to determine what retinal features are needed to best solve a specific visual task in the context of natural stimuli. From the *coding perspective*, natural stimuli are used to inspect the structure and information content of activity patterns in individual or populations of retinal neurons. Finally, the *circuit perspective* employs natural stimuli to ask which circuit mechanisms shape the neuronal responses under natural conditions. All three perspectives rely on appropriately designed computational models and selected natural stimuli, which brings us to the question of what choices there are to make when exploring how the retina encodes visual scenes.

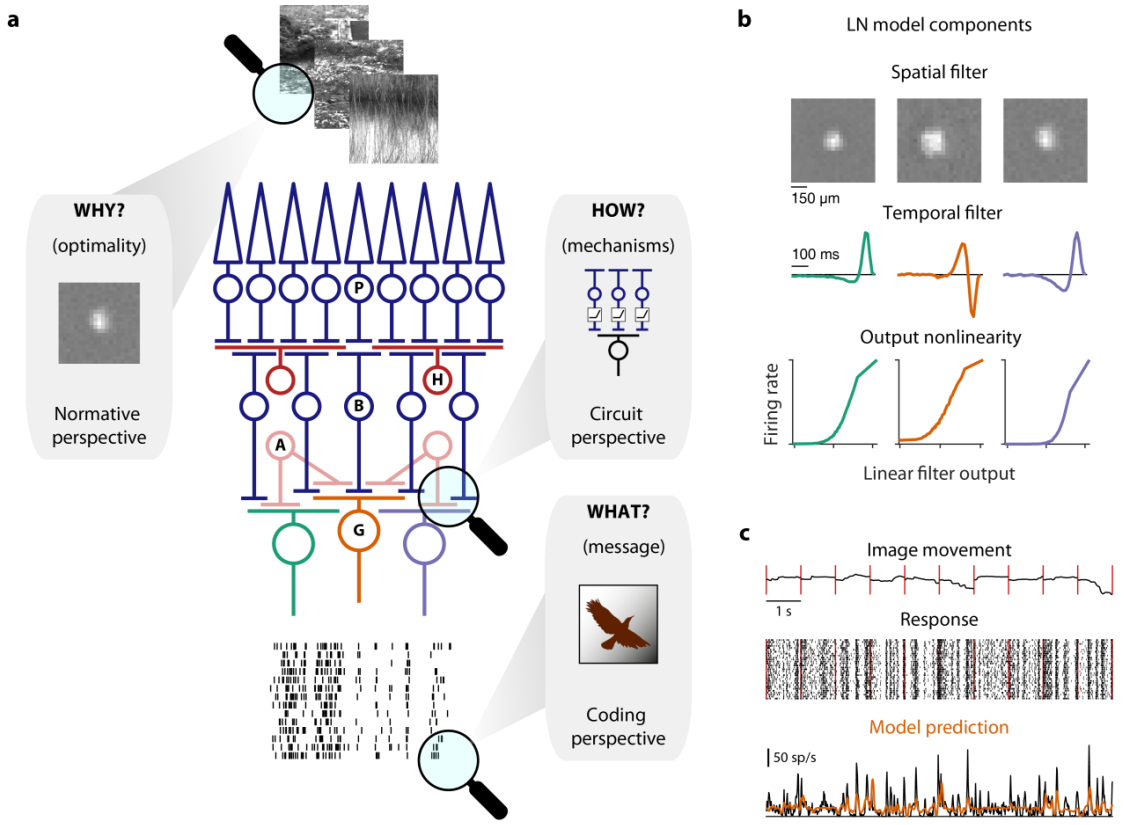

Figure 1. Three perspectives on retinal encoding of natural scenes. (a) Center: Schematic, simplified circuit of the retina with photoreceptors (P) and the excitatory bipolar (B) and ganglion (G) cells. Inhibition is provided by the horizontal (H) and amacrine (A) cells. Around: Schematic representation of the three perspectives under which natural stimuli are used in retinal research. (b) LN model components for three mouse ganglion cells of different types. (c) Spike raster (center) and corresponding firing rate (bottom) of the second cell in b to a naturalistic movie, constructed by shifting natural images on the retinal surface (top, red lines mark image switches). The LN model prediction is shown in orange.

FROM RECEPTIVE FIELDS TO COMPUTATIONAL MODELS OF RETINAL ENCODING

The retina's neural network is among the most intensely studied parts of the nervous system, and the basic layout and processing pathways are comparatively well understood (Cajal 1893; Masland 2001; Wässle 2004). Incoming light signals are captured by the photoreceptors, which pass on their signals via excitatory bipolar cells to the output neurons of the retina, the retinal ganglion cells (**Figure 1a**). In between, two classes of interneurons, horizontal cells and amacrine cells, provide feedforward, feedback, and lateral inhibition to gate and shape the signal transmission. Finally, the action potentials ("spikes") generated by the ganglion cells, travel down the cells' axons in the optic nerve to inform a multitude of brain regions about the natural visual world (Martersteck et al. 2017; Morin & Studholme 2014). Despite the long history of intense research, the retina has not ceased to surprise researchers with new insights about its complexity and signal processing capabilities. Recent years have seen a steady increase in the

number of identified cell types, which may now be converging to well above a hundred neuron types for the mouse retina (Yan et al. 2020). These include roughly 40 type of ganglion cells (Baden et al. 2016; Bae et al. 2018; Tran et al. 2019), each providing a separate information channel over the entire visual field, and a staggering diversity of more than 60 types of amacrine cells (Yan et al. 2020). The functional role of most of these individual cell types as well as of the diversity as a whole has remained unclear and is arguably among the most fundamental riddles of early-vision research. Investigating retinal processing in the context of complex, natural stimuli will likely be needed to reveal the full functional potential embedded in this diverse and interconnected circuit.

Early investigations of stimulus encoding by the retina were unimpeded by this daunting hidden complexity and made good headway with simple stimuli, such as light spots that were flashed onto different retinal locations (Barlow et al. 1957; Hartline 1938; Kuffler 1953). This led to the distinction of ON and OFF cells and the fundamental description of the center-surround receptive field, which still dominates the functional picture of the retina today. Ideas of more complex stimulus encoding then arose with the observation of ganglion cell sensitivity to specific visual patterns, such as small dark, moving spots for the "bug perceivers" described by Lettvin et al. (1959), and of nonlinear spatial integration, as revealed by frequency doubling in response to contrast-reversing gratings (Enroth-Cugell & Robson 1966; Victor & Shapley 1979).

These early findings were cast into computational models of retinal processing, which still form the backbone of current analyses of the encoding of natural scenes. The most fundamental concept here is to regard the receptive field of a ganglion cell as a spatial filter through which the cell processes the scene to determine the level of its activity (Rodieck 1965). Higher similarity between filter and visual stimulus corresponds to stronger activation. By supplementing the filter with a time course, one obtains a spatiotemporal filter (**Figure 1b**), which can be applied to dynamic stimuli. This basic filtering operation underlies the perhaps most widely used model of neural stimulus-response relationships, not only in the retina, the linear-nonlinear (LN) model. After the linear filtering (the "L" in the LN model), a nonlinear transformation (the "N") connects the neuron's activation to its spiking output, typically the firing rate or spike probability for a given image or brief time period in the case of spatiotemporal input, such as a natural movie (**Figure 1c**). In this way, the nonlinearity can capture effects of spiking threshold, response gain, and firing-rate saturation (Chichilnisky 2001). The LN model, together with its many extensions that are obtained, for example, by incorporating multiple parallel filters or feedback modules (McFarland et al. 2013; Pillow et al. 2008; Real et al. 2017) forms a powerful data analysis framework with a good compromise of biological interpretability and computational simplicity. Analyzing the encoding of natural scenes by the retina has relied heavily on this model framework by finding optimal model parameters (normative perspective), deciphering the messages that different response components carry about certain visual features (coding perspective), or identifying crucial model components and their biological counterparts (circuit perspective).

NORMATIVE PERSPECTIVE

An early focus for studying the encoding of natural scenes was the question in what sense the retina might be specifically adapted to handle natural stimuli. By assuming that evolution shaped the retina to optimally cope with the visual structure of the environment, one can investigate how the retina *should* operate to achieve this optimality. This *normative perspective* thereby seeks to answer *why* certain features of the retina are the way they are, i.e., what functional role they play in the encoding of natural stimuli. The typical working agenda has been to identify a plausible optimality criterion (or "objective function"), formalize retinal signal processing in a modeling framework, optimize the model parameters to maximize the objective function in the context of natural input signals, and compare the identified optimal parameters to the anatomical and physiological features of the retina. What makes the retina so amenable to this approach are the immediacy of stimulus encoding with no prior neural stage of preprocessing and the fairly good knowledge that we have about its basic circuitry, which allows for specific comparisons of theory and experiment (Balasubramanian & Sterling 2009).

A typical starting point for the normative approach is the idea of *efficient coding*, which posits that a sensory system should maximize the information contained in its output signals (Attneave 1954; Barlow 1961). For retinal encoding of natural scenes, this means that retinal ganglion cells should not simply encode raw light intensity in their receptive fields because neighboring locations in visual space tend to have similar intensities (Field 1987) and neighboring cells would thus spend their action potentials largely on redundant information. A maximally efficient code, on the other hand, should let different ganglion cells encode independent bits of information, leading to uncorrelated retinal activity. This *redundancy reduction* (Attneave 1954; Barlow 1961) from correlated natural scenes to uncorrelated neural activity was formalized in a spatial filter model of retinal processing (Atick & Redlich 1990, 1992) and found to yield filters with center-surround structure (**Figure 2a**), whose surround strength depends on the level of noise in the system (**Figure 2b**). The filters display a striking resemblance with recorded receptive fields, and the filtering matches psychophysical measurements of human contrast sensitivity under different light levels (Atick & Redlich 1992). An alternative, though closely related view of center-surround receptive fields comes from the idea of *predictive coding* (Srinivasan et al. 1982). Here, the surround is viewed as using the light-intensity correlations of natural images to predict the illumination in the center, so that the ganglion cell encodes only deviations from this prediction and thereby optimally uses its dynamic range.

The established connection between center-surround receptive fields and the statistics of natural stimuli is arguably one of the highlights in the convergence zone of theoretical and experimental neuroscience and continues to shape investigations of natural stimulus encoding. Many follow-up studies have confirmed, extended, and refined this view, with different model frameworks and optimization strategies (Jun et al. 2021; Karklin & Simoncelli 2011; Lindsey et al. 2019; Ocko et al. 2018; Roy et al. 2021). The retinal mechanisms of redundancy reduction, however, might go beyond center-surround filtering, as nonlinear thresholding of filtered signals provided the larger contribution to the decorrelation of salamander retinal ganglion cells (Pitkow & Meister 2012).

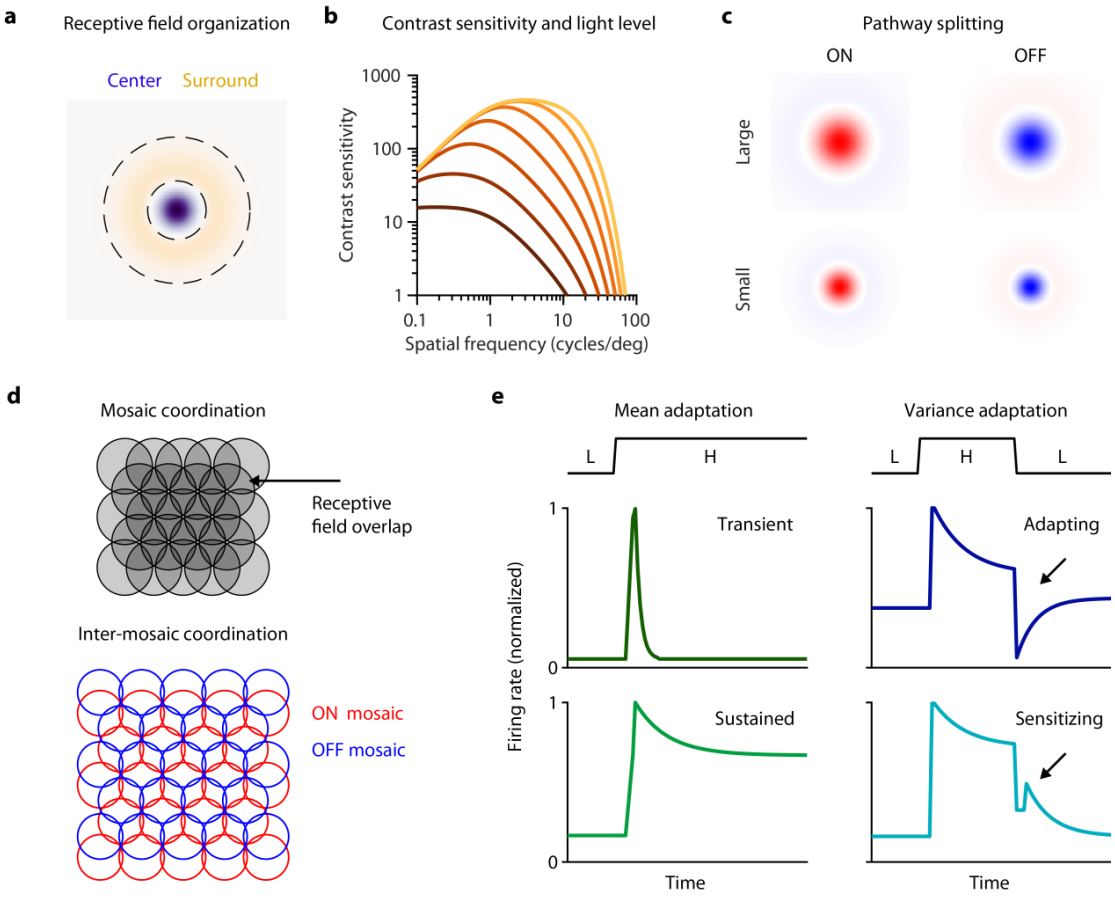

Figure 2. Normative perspective: Retinal features explained through efficient coding of natural scenes. (a) Schematic depiction of a center-surround receptive field. (b) Efficient-coding predictions of contrast sensitivity curves for different relative noise levels, corresponding to different background light levels (lighter shades denote lower noise/brighter light), following the calculations of Atick & Redlich (1992). The stronger bandpass filtering at brighter light results from a relatively stronger surround component. (c) Efficient coding can predict two types of pathway splitting: ON/OFF, and large/small, as depicted here by schematic receptive fields. (d) Top: Coordination of the receptive-field mosaic of a single ganglion cell type, illustrating receptive-field overlap. Bottom: Schematic depiction of anti-aligned ON and OFF-ganglion cell receptive-field mosaics, with receptive field overlap reduced for clarity. (e) Schematic response curves under steps from low (L) to high (H) mean luminance (left) and under steps between low and high contrast of flickering light (right). Different coding strategies predict transient or sustained response types for luminance steps and adapting or sensitizing response types for contrast steps, and arrows mark the response differences after returning to low contrast.

The principle of efficient coding in the context of natural scenes has also been used to explain the emergence of parallel pathways in the retina, such as the splitting of the neuronal population into ON and OFF cells (Gjorgjieva et al. 2014; Karklin & Simoncelli 2011) as well as into cells with small and large receptive fields (Ocko et al. 2018), matching, for example, the midget and parasol pathways of the primate retina (**Figure 2c**). Furthermore, OFF cells are often smaller and more densely packed than

corresponding types of ON cells, as observed in rat, macaque, and human retinas (Chichilnisky & Kalmar 2002; Kling et al. 2020; Ravi et al. 2018; Soto et al. 2020), which may reflect the predominance of OFF-type contrast in natural scenes (Ratliff et al. 2010). Populations of both ON and OFF receptive fields tightly tile visual space (**Figure 2d**, top), with an overlap that is found to be optimal for encoding natural images (Borghuis et al. 2008). Additionally, the receptive field midpoints of ON cells tend to show up in between those of OFF cells (**Figure 2d**, bottom), and the anti-alignment of the two mosaics contributes to optimal information transmission for natural images (Roy et al. 2021), though the optimality of this anti-aligned configuration appears to depend on the noise level in the system (Jun et al. 2021).

Natural stimuli contain correlations not only over space, but also over time, and the often bandpass temporal filtering of retinal ganglion cells, analogous to their center-surround receptive fields, can serve to reduce temporal redundancy (Dan et al. 1996; Pitkow & Meister 2012). Moreover, the natural visual environment is highly dynamic and entails large shifts in mean illumination and contrast (Rieke & Rudd 2009), for example, when saccades shift the gaze direction between regions of direct illumination and shade or between textures and object boundaries. The retina adapts to new light and contrast levels, and efficient coding theory can explain some of the observed phenomena of adaptation (Młynarski & Hermundstad 2021; Yedutenko et al. 2020). A model-based investigation of the decoding of dynamic natural stimuli (Młynarski & Hermundstad 2021) proposes that transient and sustained responses to step increases in luminance reflect optimal solutions to different decoding tasks (**Figure 2e**, left) and also captures the dependence of adaptation time scales on stimulus switching periods (Wark et al. 2009). In the same fashion, the model also suggests a similar task-specific optimality of cells that display either increased or decreased sensitivity after a switch to lower-contrast stimulation (**Figure 2e**, right), known as sensitization and adaptation, respectively (Appleby & Manookin 2019; Kastner & Baccus 2011).

Including not only luminance information but also chromatic statistics of natural images furthermore allows the derivation of color-opponent receptive fields (Atick et al. 1992; Doi et al. 2003), as observed in the midget pathway of the primate retina. For the mouse retina, on the other hand, it has been argued that chromatic encoding may require considerations beyond an efficient coding framework (Abballe & Asari 2021). The two cone types of the mouse retina, S-cones and M-cones, which are most sensitive to UV and green light, respectively, are inhomogeneously distributed across the retina, with S-cones dominating the ventral and M-cones the dorsal retina (Applebury et al. 2000; Szél et al. 1992). Though originally thought to be an adaptation to the prevalence of UV light coming from the sky and green light from the ground (Gouras & Ekesten 2004), it seems that the nonlinear responses of S-cones and their larger gain compared to M-cones make this configuration a near optimal detector for the natural distribution of achromatic contrast (Baden et al. 2013). Additionally, the spectral statistics of UV and green in a mouse's upper visual field (Qiu et al. 2021) may explain why green-UV color opponency is more pronounced in the ventral retina (Szatko et al. 2020). Similar task- and environment-specific adaptations of photoreceptor distributions and signals have been found for the zebrafish retina where UV-sensitive cones in a fovea-like spot on the retina support prey detection (Yoshimatsu et al. 2020) and signals from red- and green-sensitive cones already separate the chromatic content of natural daylight into achromatic and spectrally opponent components (Yoshimatsu et al. 2021).

Despite their undeniable successes, normative approaches often resemble post hoc explanations, deriving known features of the retinal circuitry under plausible, yet particular assumptions, such as the modeling framework, the objective function, and the metabolic cost or channel capacity (Sterling & Laughlin 2015). It is not always clear to what degree the derived optimal features remain robust to alterations in

these assumptions. The level of assumed noise, for example, can lead to fundamentally different optimal retinal architectures (Jun et al. 2021). Some experimental evidence may also question the redundancy reduction assumption of efficient coding, as retinal responses retain considerable redundancy under natural stimuli (Pitkow & Meister 2012; Puchalla et al. 2005).

CODING PERSPECTIVE

Studying the retina under the normative perspective typically starts with specific assumptions about what visual information is encoded, such as the spatial distribution of light levels across a scene, as well as about how this information is represented, e.g., by the firing rates of individual ganglion cells. Yet, what visual features retinal ganglion cells encode is also an empirical question, which needs to be viewed in the context of natural stimuli, giving rise to the *coding perspective*.

Before we dive into the topic of how the neural code of the retina is experimentally studied with natural scenes, let us note that the concept of a *neural code* (or relatedly of *neural representation* of sensory information) is riddled with difficulties of what it means to encode or represent certain information. For the neural activity X to encode or represent some visual feature Y, should we just expect X to correlate with Y? Or to somehow allow the (nonlinear) extraction of information about Y? Or to be causally related to behavioral responses to Y? Here, however, we put these important considerations aside and loosely speak of encoding when relevant information about Y is "readily available" and could "easily" (e.g., linearly) be read out from X by downstream circuitry, in line with the typical and practical use of "encoding" in the retina literature. In any case, we need to ask whether the encoding of a specific visual feature depends on the context of visual stimuli and whether findings obtained under simplified artificial stimuli need to be revised when considering natural scenes.

Direction-selective retinal ganglion cells, for example, are considered to encode the direction of visual motion because the cells respond strongly to spots of light or visual gratings that move in a particular direction (the preferred direction), but not for motion in the opposite (null-) direction (Barlow & Hill 1963). So a cell's activity apparently encodes motion direction, as the motion component along the preferred/null-axis can be fairly well determined from the evoked activity – at least in the context of the moving spots or gratings (**Figure 3a**). Natural stimuli, on the other hand, display much more complex combinations of spatial contrast and motion and may also activate direction-selective ganglion cells through light intensity changes that occur independently of motion or even through motion in the null-direction. This problem is partly mitigated by inhibition from the receptive field surround, which reduces light-intensity-related activity under natural stimulation (Im & Fried 2016). Nonetheless, despite the clear response properties of a direction-selective ganglion cell under simple motion stimuli, its encoding under natural scenes can be ambiguous. The relative activity of multiple direction-selective ganglion cells, however, can still robustly and unambiguously encode motion direction (**Figure 3b;** Kühn & Gollisch 2019). This suggests that a population code may help extract complex information from natural scenes, such as translatory and rotatory self-motion (Sabbah et al. 2017).

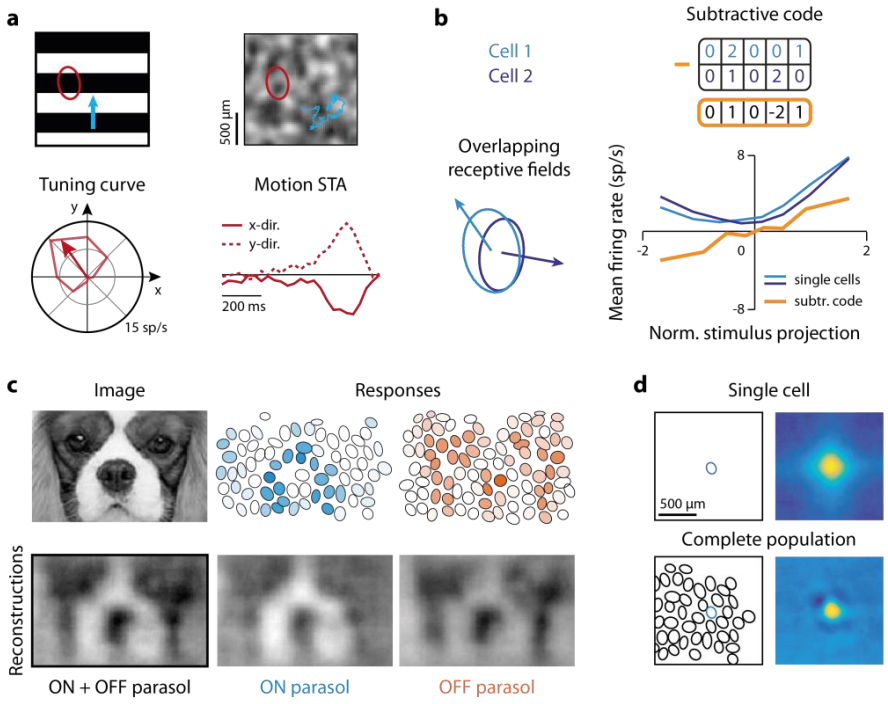

Figure 3. Coding perspective: The retina's messages about natural stimuli. (a) Comparing direction selectivity under drifting gratings (left) and textures moving similarly to fixational eye movements (right). The preferred direction under texture motion is obtained from the spike-triggered average of the motion steps in x and y direction (bottom right; data from a salamander direction-selective cell) and matches the preferred direction under drifting gratings for the sample cell (bottom left). (b) Ambiguity of direction encoding by single cells and resolution by a population code. Single direction-selective cells (blue) show elevated activity for motion along the preferred (positive stimulus projection) as well as the null direction (negative stimulus projection). The response difference (orange), on the other hand, depends monotonically on motion direction. Panels a-b adapted from Kühn & Gollisch (2019), licensed under CC BY 4.0. (c) Reconstruction of a natural image (top left) from responses of populations of ON and OFF parasol ganglion cells (top center and right; ellipses are receptive fields and saturation of color indicates elicited activity). Bottom row shows (left to right) reconstructions using both populations, using only ON, and using only OFF cells. (d) Optimal spatial reconstruction filters for a parasol ganglion cell taken in isolation (top) and from a joint decoding by a complete population (bottom). Panels c-d adapted from Brackbill et al. (2020), licensed under CC BY 4.0.

Even the simple notion that a ganglion cell encodes spatial contrast within its receptive field is worth scrutinizing, as the spatial correlations of natural stimuli lead to correlations of a cell's activity with light levels far outside the receptive field. Here, again, considering populations of neurons help clarify the matter, as reconstructions of natural images from spike counts of entire populations of primate ON and OFF parasol cells (Figure **3c,d**) works best when the contributions of individual cells actually reflects the cells' center-surround receptive field structure (Brackbill et al. 2020). Interestingly, decoding finer spatial details of natural images is possible by going beyond a linear decoder that considers individual filter-like contributions from each cell and incorporating nonlinear decoding techniques (Kim et al. 2021; Parthasarathy et al. 2017), suggesting that the temporal structure of ganglion cell spike patterns contains

information about natural images beyond what is available through the cells' trial-averaged firing rates (Kim et al. 2021).

An aspect of natural stimuli that may be particularly suited to reveal new aspects of stimulus encoding is the frequent occurrence of rapid shifts in gaze direction, either through saccadic eye movements or movements of the head or body (Land 2015). The rapid succession of fixated images, interspersed by fast, global motion signals might trigger modes of signal processing that remain untapped during artificial stimuli such as white noise or temporally isolated flashes. Indeed, simulated saccades can provoke strong ganglion cell activity or suppress it (Amthor et al. 2005; Idrees et al. 2020; Noda & Adey 1974; Roska & Werblin 2003). Yet, which aspects of the combination of pre- and post-saccadic image are encoded by ganglion cell activity across a saccade remains an open question for most ganglion cell types and is difficult to predict from response properties derived with isolated image presentations. For example, specific cells in the mouse retina displayed an unexpected sensitivity to recurring images across saccades (Krishnamoorthy et al. 2017).

Another important context for the encoding of a visual stimulus is given by the background on which it is presented. Different natural scenes provide different background, and a recent study found that the spatial sensitivity profile of salamander and mouse ganglion cells indeed depends on the background image (Goldin et al. 2021). ON-OFF ganglion cells may even invert their relative sensitivity to positive and negative contrast depending on the context-defining image. This context dependence of the preference for brightening or darkening can be viewed as an encoding of absolute contrast relative to the background image (Goldin et al. 2021).

Apart from asking what visual features are encoded by specific retinal ganglion cells, natural stimuli have also been used to study the structure of ganglion cell population activity in near-natural conditions. A particular focus has been to assess the prevalence and importance of correlated firing within ganglion cell populations (Nirenberg et al. 2001) and the statistical structure of multi-neuronal activity patterns (Ganmor et al. 2015; Schneidman et al. 2006). Relating multi-neuronal activity patterns to specific aspects of natural stimuli, however, has remained an open question, though certain stimulus features, which are part of natural viewing, may trigger synchronous activity, such as fixational eye movements (Greschner et al. 2002; Masquelier et al. 2016) and motion reversal (Schwartz et al. 2007).

CIRCUIT PERSPECTIVE

While the normative and coding perspectives aim at answering *why* the retina is designed the way it is and *what* is actually encoded by different patterns of ganglion cell activity, we also strive to understand *how* the retina actually implements the signal processing that underlies its coding schemes. Which elements of the neural circuitry are involved in a given encoding, which connections, which cellular properties and synaptic dynamics? These questions address the retina from a *circuit perspective* and have traditionally been targeted with simple, custom-made stimuli, aimed at isolating a particular phenomenon of interest. For example, flashed or flickering light spots can be used to investigate response kinetics, surround inhibition, as well as light and contrast adaptation. Focusing on these simple stimuli, however, leaves open two important questions: 1) Do the identified mechanisms act in the same way under natural stimuli, when stimulus statistics are different and when different types of mechanisms act simultaneously? 2) Do we miss specific mechanisms because they are not triggered under the commonly used artificial stimuli?

To cope with the complexity of naturalistic stimulation and still relate elicited responses to specific circuit features, the circuit perspective heavily relies on computational modeling and inference about the circuitry through parameter fitting. Typically, this involves cascade-type models, composed of mathematical primitives, such as filters, nonlinear transformations, and additive or multiplicative feedforward and feedback interactions. These components have been used to encapsulate various concepts from retinal physiology, such as the spatial receptive field (Brown et al. 2000; Chichilnisky 2001), adaptation to luminance and contrast (Jarsky et al. 2011; Ozuysal & Baccus 2012), spike generation (Weber & Pillow 2017), specific inhibitory interactions (Baccus et al. 2008; Geffen et al. 2007; Roska & Werblin 2003), or neuronal couplings (Meytlis et al. 2012; Pillow et al. 2008), and serve as an expedient tool to bridge artificial and natural stimulus scenarios. The models reside in a sweet spot of computational simplicity for fitting to experimental data combined with sufficient complexity to allow for biological interpretations. Enlarging this sweet spot through computational and experimental advances remains a hot and important topic for studying retinal function.

Spatial integration

A retinal mechanism that has recently received much attention because of its importance for the sensitivity to spatial contrast under natural stimuli is the pooling of bipolar cell inputs by ganglion cells. Simple models, such as the LN model, often consider the spatial integration over the excitatory bipolar cell inputs in the receptive field center to be linear, as reflected by the linear spatial filtering that forms the first model stage. Yet, many ganglion cells throughout various retinal model systems display nonlinear spatial integration under contrast-reversing spatial gratings (Bölinger & Gollisch 2012; Carcieri et al. 2003; Demb et al. 1999; Enroth-Cugell & Robson 1966; Petrusca et al. 2007), the stimulus that is classically used to investigate spatial stimulus integration (**Figure 4a**).

The lack of nonlinear spatial integration in common filter-based models has been hypothesized to contribute to the often unsatisfactory performance of these models under natural stimuli (Gollisch 2013; Heitman et al. 2016; Schwartz & Rieke 2011). It is not *per se* clear, however, that nonlinear signal transmission from bipolar to ganglion cells is actually relevant under natural stimuli. First, the gratings typically applied to identify these nonlinearities have high contrast and lack concomitant low spatial frequencies, thus making ganglion cell responses sensitive to even tiny nonlinear contributions. Second, due to the spatial correlations of natural images, the receptive field center may only rarely be filled with

patches of high spatial contrast, for which the nonlinearities matter most. And third, the typically periodic presentation of reversing gratings might drive adaptive mechanisms into a non-natural regime.

To test for the relevance of nonlinearities, several studies compare model predictions to ganglion cell responses under natural stimuli. Depending on cell type, accurate as well as inaccurate response predictions are observed (Cao et al. 2011; Heitman et al. 2016; Karamanlis & Gollisch 2021; Liu et al. 2017; Nirenberg & Pandarinath 2012; Turner & Rieke 2016). Additionally, subunit models with rectified input nonlinearities, representing the nonlinear signal transmission of bipolar cells as seen under artificial stimuli (Borghuis et al. 2013; Demb et al. 2001), can improve the prediction accuracy for many cells, highlighting the importance of nonlinear spatial integration also under natural stimuli (Liu et al. 2017; Shah et al. 2020; Turner & Rieke 2016). The cell-type specificity of nonlinear spatial integration under natural stimuli was also demonstrated by comparing ganglion cell responses under natural images and under homogenous spots of the same mean luminance (Turner & Rieke 2016) and by comparing responses under luminance-matched pairs of natural images (Karamanlis & Gollisch 2021; Liu & Gollisch 2021). Turner & Rieke (2016), for example, observed that OFF parasol cells were sensitive to spatial structure under natural stimuli, and their nonlinear spatial integration stems directly from their rectified excitatory inputs. ON parasol cells, on the other hand, were found to integrate inputs linearly under natural stimuli, despite their nonlinear spatial integration under reversing gratings (**Figure 4b**). This finding exemplifies that – at least for some cell types – insights from artificial stimuli may not easily generalize to natural stimuli.

Nonlinearities upstream of the bipolar-to-ganglion cell synapse may also contribute to natural scene encoding. Bipolar cell membrane potentials in the salamander retina show nonlinear contrast encoding and even nonlinear integration of photoreceptor signals under both artificial and natural stimuli (Schreyer & Gollisch 2021). And photoreceptors themselves can display pronounced nonlinear contrast encoding, especially when challenged with the broad range of light levels occurring in natural stimuli (Endeman & Kamermans 2010; Howlett et al. 2017), which effectively drives the phototransduction mechanism into its nonlinear regime (Clark et al. 2013; van Hateren 2005) and triggers adaptation mechanisms that contribute to photoreceptor nonlinearities (Angueyra et al. 2021; Howlett et al. 2017). In the mouse retina, nonlinear encoding of contrast in photoreceptors is particularly pronounced in S-cones, which may be a specific adaptation to the naturally occurring distribution of contrasts in the upper visual field (Baden et al. 2013). For the retinal encoding of natural scenes, the observed photoreceptor and bipolar cell nonlinearities may lead to sensitivity to spatial structure even below the scale of bipolar cell receptive fields.

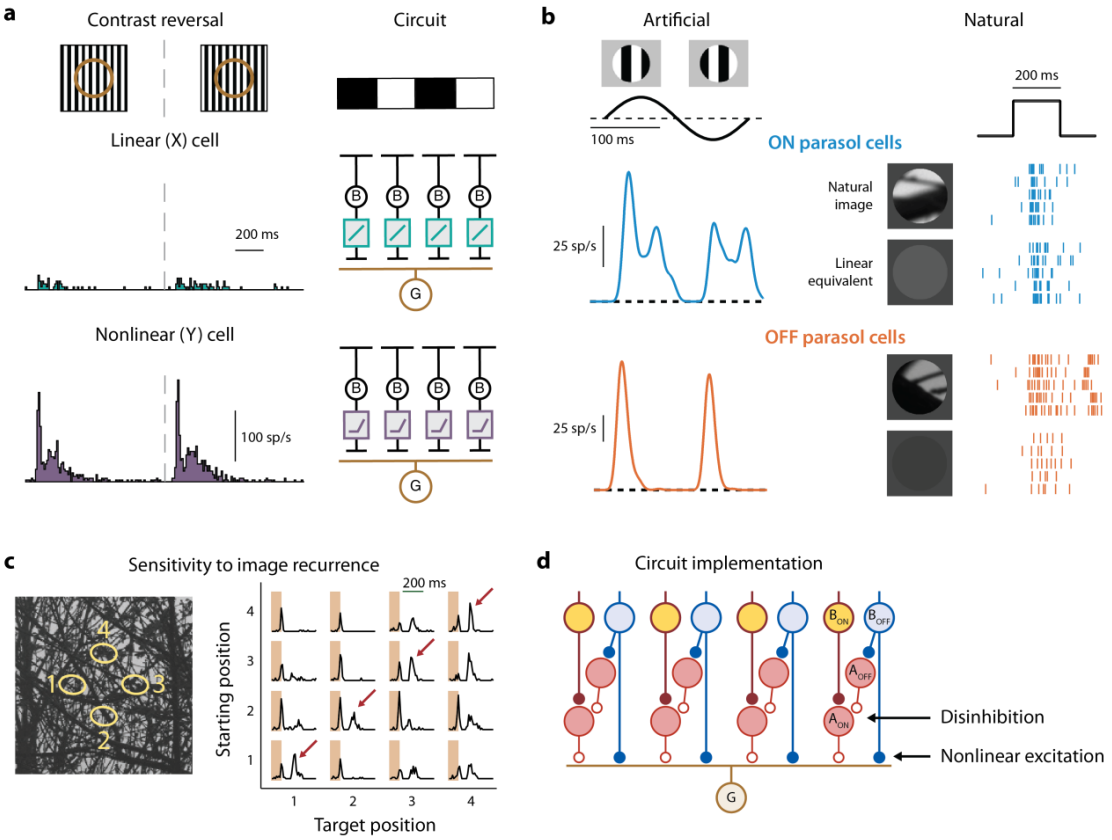

Figure 4. Circuit perspective: Neural circuitry involved in processing natural stimuli. (a) Responses of X and Y mouse retinal ganglion cells to reversing gratings (left) and corresponding circuit models (right) with linear and rectified transmission, respectively, from bipolar cells (B) to the ganglion cell (G). Adapted from Karamanlis & Gollisch (2021), licensed under CC BY 4.0. (b) Left: Responses of macaque ON and OFF parasol cells to reversing grating, showing frequency doubling indicative of nonlinear integration. Right: Responses to natural images and to homogenous spots of the same mean luminance (linear equivalent), indicating nonlinear spatial integration for OFF, but not ON cells. Adapted from Turner & Rieke (2016), Copyright (2016), with permission from Elsevier. (c) Responses of a mouse ganglion cell sensitive to image recurrence under saccade-like shifts of natural images. Left: Image with receptive field outlines for four fixation positions. Right: The cell's firing rate profiles for all transitions. The starting and target positions are the fixation positions before and the after the shift. The shaded regions mark the transition. For equal starting and target position, the image shifted to the central position and then returned. Red arrows mark firing rate peaks under recurring image position. (d) Circuit mechanism explaining image-recurrence sensitivity. The ganglion cell receives excitation from OFF bipolar cells and inhibition from ON amacrine cells. The ON amacrine cells are themselves inhibited by OFF amacrine cells, leading to disinhibition of the ganglion cell. Panels c-d adapted from Krishnamoorthy et al. (2017), licensed under CC BY 4.0.

Adaptation

The wide ranges of luminance and contrast levels encountered in natural scenes (Frazor & Geisler 2006) challenge the retina to continuously adjust its limited dynamic range (Rieke & Rudd 2009). Adaptation to mean luminance is thought to arise primarily from calcium-dependent feedback in the phototransduction

process (Matthews et al. 1988; Nakatani & Yau 1988) and is classically studied by presenting probe flashes on a steady illumination background. Contrast adaptation, on the other hand, is thought to first occur in bipolar cells or in the transmitter release at their synaptic terminals (Baccus & Meister 2002; Manookin & Demb 2006) and is often investigated under noise-like flicker of light intensity with sudden changes in variance. The distinct sites of luminance and contrast adaptation appear to fit the idea that the two should occur independently, as luminance and contrast levels in natural scenes were also found to be independent of each other (Mante et al. 2005). Yet, some components of luminance adaptation appear to arise downstream of photoreceptors at the bipolar cell output (Dunn et al. 2007; Jarsky et al. 2011), creating a mechanistic overlap of the two adaptation processes. Furthermore, under stimulation with the wide range and skewed distribution of natural light intensities (Endeman & Kamermans 2010; van Hateren et al. 2002), a contrast-adaptation-like mechanism can also be observed in goldfish photoreceptors, triggered by voltage-dependent hyperpolarization-activated ion channels (Howlett et al. 2017). In primate photoreceptors, using naturalistic light intensity sequences with saccade-fixation dynamics, Angueyra et al. (2021) observed adaptation components fast enough to adjust responses during fixation periods of few hundred milliseconds. This rapid adaptation has also important consequences for the balance of excitatory and inhibitory inputs to ganglion cells under natural stimuli (Yu et al. 2021), which appears to contribute to the discrepancy in spatial stimulus integration between natural stimuli and contrast-reversing gratings in ON parasol cells discussed above (Turner & Rieke 2016).

Inhibition

The perhaps least understood and most mysterious aspect of the retinal circuitry is the multitude and diversity of inhibitory interneurons, especially the more than 60 types of amacrine cells (Yan et al. 2020), which provide inhibition to ganglion cells, to bipolar cells, and to each other. These interactions likely play particular roles for natural scene encoding, which may therefore hold a key to solving the mystery of retinal inhibition. Inhibitory pathways are important, for example, in the context of gaze shifts, which provide global, coherent stimulation to the entire retina. Indeed, global saccade-like shifts of natural scenes activate retinal inhibition (Roska & Werblin 2003; Sivyer et al. 2019), which may contribute to saccadic suppression of visual perception. Furthermore, inhibition triggered by saccade-like shifts may not only suppress responses, but also shape post-saccadic activity through disinhibitory interactions (Geffen et al. 2007; Krishnamoorthy et al. 2017). For certain cells in the mouse retina (Krishnamoorthy et al. 2017), such disinhibition through serial connections of glycinergic and GABAergic amacrine cells leads to a particular sensitivity for the recurrence of spatial patterns across a saccade (**Figure 4c,d**), which may be relevant for correcting fixational drift through microsaccades.

Other inhibitory circuit features were identified via motion stimuli derived from natural scenarios, though the actual stimuli were simplified and schematized. The so-called object-motion-sensitive ganglion cells, for example, may help detect moving objects in the presence of eye-movement-induced background motion, owing to their specific sensitivity to relative motion (Ölveczky et al. 2003), whereas global motion signals are suppressed by inhibitory signals from polyaxonal amacrine cells in the surround (Baccus et al. 2008). Similarly, looming-sensitive ganglion cells in mouse were found to strongly respond to approaching, but not lateral motion (Münch et al. 2009). This stimulus selectivity has given rise to a circuit model with local cross-over inhibition, potentially from AII amacrine cells (Münch et al. 2009), but looming-sensitive excitatory input from glutamate-releasing amacrine cells has also been implicated as a circuit mechanism (Kim et al. 2020). Targeted ablation of these amacrine cell specifically diminishes

the defensive response of mice to overhead looming (Kim et al. 2020), thus suggesting a direct role of amacrine cell processing in visual behavior.

Activation of the receptive field surround is typically thought to provide response suppression. Yet, Turner et al. (2018) showed that the surround can also modulate the spatial integration properties of the receptive field center under natural stimuli. This is caused by a shift from nonlinear to more linear transmitter release at the bipolar cell synapse when surround signals depolarize the bipolar cell. To account for such surround effects that occur before input integration in the ganglion cell, circuit models with subunits that have their own antagonistic surrounds could help improve predictions of ganglion cell responses to natural stimuli (Enroth-Cugell & Freeman 1987; Turner et al. 2018).

Towards a complete model – Lego versus Michelangelo

Being able to predict ganglion cell responses to any natural stimulus by a model that represents the retina's neural circuitry and properties of its elements would be an ultimate goal of the circuit perspective. Achieving this goal will require an elaborate model containing the components discussed here and potentially more: integration over multiple bipolar cells, adaptation to light and contrast at the local and global scale, and inhibitory interactions that suppress or gate different information channels. Which road should we pursue towards such a complete model? Can we investigate the individual building blocks in isolation by custom-made stimuli and put them together like a set of Lego bricks? The potential of such an approach for predicting responses to natural stimuli has yet to be tested. An obvious challenge, however, comes from potential interactions between these building blocks. If, for example, the nonlinearity of spatial integration in the receptive field center depends on mean luminance (Grimes et al. 2014) or on activation of the surround (Turner et al. 2018) or if adaptation depends on the organization of a receptive field into subunits (Brown & Masland 2001; Garvert & Gollisch 2013; Khani & Gollisch 2017), studying these building blocks in unison is called for.

As an alternative, one may hope to set up a modeling framework that contains all the potential components and interactions and then extract the relevant elements and their parameters *en bloc*, like Michelangelo uncovering his David hidden in a single slab of marble. However, the challenge of optimizing the many parameters of such an intricate, nonlinear model seems daunting. While complex fits of cascade models designed to match retinal circuits are feasible (Maheswaranathan et al. 2018a; Real et al. 2017), they have yet to be attempted under natural stimuli. More abstract models in the style of artificial neural networks, on the other hand, have already been used to capture ganglion cell responses to natural scenes (Goldin et al. 2021; Maheswaranathan et al. 2018b; McIntosh et al. 2016). As an approach for studying retinal mechanisms in the context of natural stimuli, artificial neural networks may include circuit motifs that represent actual retinal elements (Turner et al. 2019), and it is possible to extract such motifs from networks trained on retinal responses to natural scenes (Tanaka et al. 2019).

CONCLUDING REMARKS

The topic of retinal encoding of natural scenes comprises many facets, including the questions why retinal signal processing is structured the way it is, what aspects of natural scenes are encoded by different retinal information channels, and how the retinal circuitry extracts and processes this information. Early insights about the relationship between ganglion cell receptive fields and natural scene statistics (Atick & Redlich 1992) still relied on findings from the retina made with simple, artificial stimuli. As technical limitations are overcome, probing the retina with actual naturalistic stimuli has turned from possible to routine over recent years. These experimental measurements are needed to sample the circuit in regimes that may have previously been unexplored and to test whether concepts derived under artificial stimulation still hold for naturalistic settings.

Analyzing retinal responses under natural stimuli will continue to heavily rely on computational models to cope with the complexity of the stimulus-response relationship. It will be necessary to explore model structures that can encompass the growing knowledge about relevant circuit mechanisms and remain computational tractable. Incorporating nonlinear input from bipolar cells, for example, is typically approached with subunit models, but may need to be extended to incorporate input from different types of bipolar cells, adaptive properties of bipolar cells, or surround and substructure of bipolar cell receptive fields. The inhibitory surround may require its own subunits (Takeshita & Gollisch 2014) and be supplemented with inhibitory effects beyond the classical receptive field surround, acting on large spatial scales (Baccus et al. 2008; Jadzinsky & Baccus 2015; Manookin et al. 2015) or locally shaping signal transmission in the receptive field center (Szatko et al. 2020). And considering chromatic image content brings in an additional stimulus dimensions with its own intricate interactions (Joesch & Meister 2016; Khani & Gollisch 2021; Szatko et al. 2020). The appropriate model structure is not only important for parameter fitting under the coding or circuit perspective, but also for defining the space of potential operations in which to search for optimal designs under the normative perspective. For example, rather than a single nonlinearity, normative approaches may profit by modeling retinal nonlinearities at multiple stages, which have been shown to display an information-boosting effect (Gutierrez et al. 2021).

Besides shaping the modeling framework, attention should be given to selecting the set of natural stimuli to be used in the investigations. For focusing on the encoding of spatial structure, flashing photographed images is often the approach of choice, and databases of calibrated and versatile natural images are readily available (Olmos & Kingdom 2004; Tkačik et al. 2011; van Hateren & van der Schaaf 1998). But flashed photographs, of course, lack natural temporal dynamics. Instead, movies of natural scenes can be used. Their temporal dynamics may, for example, be governed by the camera or environmental movement (Betsch et al. 2004; Salisbury & Palmer 2016) or emulate self-motion trajectories from saccades and fixational eye movements (Heitman et al. 2016; Roska & Werblin 2003; Segal et al. 2015; Turner & Rieke 2016).

Paying justice to all dynamics is a tremendous challenge, however, and it is unlikely that a one-size-fits-all natural stimulus exists. It seems likely that some retinal neurons, ganglion cells as well as amacrine cells, are triggered best by rare, but important visual features, which may not be frequent enough to substantially impact responses under generic natural scenarios. In the original characterization of the W3 ganglion cell in the mouse, for example, a striking feature was the cell's lack of spiking to what could be considered a generic natural stimulus, and the cell only revealed its functional characteristics when probed with specifically chosen isolated motion objects (Zhang et al. 2012). Other neurons may require

certain types of saccade-like shifts (Krishnamoorthy et al. 2017), combinations of object motion and fixational eye movements (Ölveczky et al. 2003), or small moving objects (Jacoby & Schwartz 2017) to reveal their full processing potential. Thus, a useful set of natural stimuli should strive to not only match the overall statistics of natural scenes, but also cover specific features and feature combinations that could be of particular importance.

Insight about the importance of visual stimulus features and the structure of natural visual stimuli in behaviorally important circumstances is now increasingly generated by studies of naturalistic behavior and concomitant sensory processing (Holmgren et al. 2021; Kim et al. 2020; Wallace et al. 2013). Fortunately, sharing data from such behavioral experiments, such as eye-movement traces appears to be common practice among vision science labs. An even greater benefit might be obtained if the use of natural stimuli could be more standardized by jointly casting available insight about natural scenes and visual behavior into a unified set of natural stimuli, covering different aspects of visual dynamics and visual tasks. This would allow different labs to probe different parts of the retinal circuitry – and perhaps also other visual areas of the brain – with the same natural stimuli for easier comparability and generalizability. This could be analogous to the versatile "chirp stimulus" (Baden et al. 2016), which is emerging as a standard artificial stimulus to characterize temporal response properties of visual neurons.

Let us finally also note that natural stimuli will not replace the application of custom-made simple, artificial stimuli. These still hold tremendous potential when teasing out specific aspects of signal processing while minimizing confounding effects from other dynamics in the system (Rust & Movshon 2005). Studying signal processing under constant mean light intensity or contrast, for example, has been a standard to minimize adaptation effects. In particular, hybrid stimuli that retain some aspects of natural stimuli while simplifying others have turned out very useful for identifying candidate retinal functions and related mechanisms. Examples also discussed in this text include natural time courses of light intensity and chromatic components with no spatial structure (Angueyra et al. 2021; Endeman & Kamermans 2010; Howlett et al. 2017; van Hateren et al. 2002), static presentations of natural photographic images (Brackbill et al. 2020; Cao et al. 2011; Karamanlis & Gollisch 2021; Turner & Rieke 2016; Turner et al. 2018), and eye-movement dynamics with simplified spatial patterns (Idrees et al. 2020; Krishnamoorthy et al. 2017; Kühn & Gollisch 2019; Ölveczky et al. 2003). Finding a good balance of naturalistic stimulus patterns and artificial simplifications and subsequently identifying the right natural stimuli for testing whether the hybrid-stimulus results generalize to truly natural scenarios will be among the critical steps towards a comprehensive understanding of retinal encoding of natural scenes.

SUMMARY POINTS

1. Investigating retinal processing with natural stimuli is becoming more commonplace to test findings from artificial stimulation and to search for new functional properties specific to certain natural scenarios.
2. Computational models of the retinal circuit provide important tools for handling the complexity of natural stimuli when studying retinal stimulus encoding.
3. Retinal encoding of natural scenes can be viewed from the normative perspective, asking why retinal features are the way they are, the coding perspective, asking what is encoded by activity patterns of different retinal ganglion cells, and the circuit perspective, asking how retinal processing shapes responses under natural scenes.
4. The normative perspective combines a traditional focus on explaining receptive field structure and new successes regarding cell-type diversity, adaptation, and arrangement of retinal mosaics.
5. Under the coding perspective, natural and hybrid stimuli are used to study the messages contained in the spiking output of the retina. Here, population codes may help establish robustness of coding concepts derived from simple stimuli against the additional variability and complexity under natural stimuli.
6. The circuit perspective focuses on investigating the mechanisms that shape retinal responses under natural scenes, re-evaluating the role of known circuit features and providing opportunities to discover new mechanisms, triggered by specific components of natural stimuli.

FUTURE ISSUES

1. How does the efficient coding hypothesis relate to complex signal processing beyond the single-filter models, including, for example, nonlinear integration, adaptation, or inhibition?
2. How can we probe and refine our understanding of the retina's neural code with natural stimuli? How do findings derived with artificial stimuli about the encoding of specific visual features, like object motion or looming, generalize to natural scenes?
3. How can we study inhibitory interactions with natural stimuli? How can we incorporate inhibition beyond surround suppression into computational models?
4. How can we combine different computational models of specific operations (adaptation, nonlinear signal pooling, inhibitory interactions) into a single, holistic model for processing of natural stimuli?
5. Is it feasible to set up and fit a holistic model of retinal processing to responses under natural scenes and retain biological interpretability?
6. Which natural or hybrid stimuli could serve as a standard stimulus set to probe diverse retinal functions and circuit operations in order to boost generalizability and comparability across laboratories and investigated cell types?

DISCLOSURE STATEMENT

The authors declare no competing interests.

ACKNOWLEDGEMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 724822), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project IDs 154113120 (SFB 889, project C01) and 432680300 (SFB 1456, project B05). D.K. was supported by a Boehringer Ingelheim Fonds fellowship.

LITERATURE CITED

Abballe L, Asari H. 2021. Natural Image Statistics for Mouse Vision. *bioRxiv*. 438953v3

Amthor FR, Tootle JS, Gawne TJ. 2005. Retinal ganglion cell coding in simulated active vision. *Vis. Neurosci.* 22(6):789–806

Angueyra JM, Baudin J, Schwartz GW, Rieke F. 2021. Multiple time scales of adaptation allow cones to encode the inputs created by visual exploration of natural scenes. *bioRxiv*. 431101v1

Applebury ML, Antoch MP, Baxter LC, Chun LLY, Falk JD, Farhangfar F, Kage K, Krzystolik MG, Lyass LA, Robbins JT. 2000. The murine cone photoreceptor: A single cone type expresses both S and M opsins with retinal spatial patterning. *Neuron*. 27(3):513–23

Appleby TR, Manookin MB. 2019. Neural sensitization improves encoding fidelity in the primate retina. *Nat. Commun.* 10(1):4017

Atick JJ, Li Z, Redlich AN. 1992. Understanding Retinal Color Coding from First Principles. *Neural Comput.* 4(4):559–72

Atick JJ, Redlich AN. 1990. Towards a Theory of Early Visual Processing. *Neural Comput.* 2(3):308–20

Atick JJ, Redlich AN. 1992. What Does the Retina Know about Natural Scenes? *Neural Comput.* 4(2):196–210

Attneave F. 1954. Some informational aspects of visual perception. *Psychol. Rev.* 61(3):183–93

Baccus SA, Meister M. 2002. Fast and Slow Contrast Adaptation in Retinal Circuitry. *Neuron*. 36(5):909–19

Baccus SA, Ölveczky BP, Manu M, Meister M. 2008. A retinal circuit that computes object motion. *J. Neurosci.* 28(27):6807–17

Baden T, Berens P, Franke K, Román Rosón M, Bethge M, Euler T. 2016. The functional diversity of retinal ganglion cells in the mouse. *Nature*. 529(7586):345–50

Baden T, Schubert T, Chang L, Wei T, Zaichuk M, Wissinger B, Euler T. 2013. A Tale of Two Retinal Domains: Near-Optimal Sampling of Achromatic Contrasts in Natural Scenes through Asymmetric Photoreceptor Distribution. *Neuron*. 80(5):1206–17

Bae JA, Mu S, Kim JS, Turner NL, Tartavull I, Kemnitz N, Jordan CS, Norton AD, Silversmith WM, Prentki R, Sorek M, David C, Jones DL, Bland D, Sterling ALR, Park J, Briggman KL, Seung HS. 2018. Digital Museum of Retinal Ganglion Cells with Dense Anatomy and Physiology. *Cell*. 173(5):1293-1306.e19

Balasubramanian V, Sterling P. 2009. Receptive fields and functional architecture in the retina. *J. Physiol.* 587(12):2753–67

Barlow H. 1961. Possible principles underlying the transformations of sensory messages. *Sens. Commun.* 6(2):57–58

Barlow HB, Fitzhugh R, Kuffler SW. 1957. Change of organization in the receptive fields of the cat's retina during dark adaptation. *J. Physiol.* 137(3):338–54

Barlow HB, Hill RM. 1963. Selective Sensitivity to Direction of Movement in Ganglion Cells of the Rabbit Retina. *Science*. 139(3553):412–14

Betsch BY, Einhäuser W, Körding KP, König P. 2004. The world from a cat's perspective--statistics of

natural videos. *Biol. Cybern.* 90(1):41–50

Bölinger D, Gollisch T. 2012. Closed-Loop Measurements of Iso-Response Stimuli Reveal Dynamic Nonlinear Stimulus Integration in the Retina. *Neuron*. 73(2):333–46

Borghuis BG, Marvin JS, Looger LL, Demb JB. 2013. Two-Photon Imaging of Nonlinear Glutamate Release Dynamics at Bipolar Cell Synapses in the Mouse Retina. *J. Neurosci.* 33(27):10972–85

Borghuis BG, Ratliff CP, Smith RG, Sterling P, Balasubramanian V. 2008. Design of a neuronal array. *J. Neurosci.* 28(12):3178–89

Brackbill N, Rhoades C, Kling A, Shah NP, Sher A, Litke AM, Chichilnisky E. 2020. Reconstruction of natural images from responses of primate retinal ganglion cells. *eLife*. 9:e58516

Brown SP, He S, Masland RH. 2000. Receptive field microstructure and dendritic geometry of retinal ganglion cells. *Neuron*. 27(2):371–83

Brown SP, Masland RH. 2001. Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells. *Nat. Neurosci.* 4(1):44–51

Cajal SR. 1893. La rétine des vertébrés. *La Cellule*. 9:119–257

Cao X, Merwine DK, Grzywacz NM. 2011. Dependence of the retinal Ganglion cell's responses on local textures of natural scenes. *J. Vis.* 11(6):11

Carcieri SM, Jacobs AL, Nirenberg S. 2003. Classification of Retinal Ganglion Cells: A Statistical Approach. *J. Neurophysiol.* 90(3):1704–13

Chichilnisky EJ. 2001. A simple white noise analysis of neuronal light responses. *Netw. Comput. Neural Syst.* 12(2):199–213

Chichilnisky EJ, Kalmar RS. 2002. Functional Asymmetries in ON and OFF Ganglion Cells of Primate Retina. *J. Neurosci.* 22(7):2737–47

Clark DA, Benichou R, Meister M, Azeredo da Silveira R. 2013. Dynamical Adaptation in Photoreceptors. *PLoS Comput. Biol.* 9(11):e1003289

Dan Y, Atick JJ, Reid RC. 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* 16(10):3351–62

Demb JB, Haarsma L, Freed MA, Sterling P. 1999. Functional circuitry of the retinal ganglion cell's nonlinear receptive field. *J. Neurosci.* 19(22):9756–67

Demb JB, Zaghloul K, Haarsma L, Sterling P. 2001. Bipolar Cells Contribute to Nonlinear Spatial Summation in the Brisk-Transient (Y) Ganglion Cell in Mammalian Retina. *J. Neurosci.* 21(19):7447–54

Doi E, Inui T, Lee TW, Wachtier T, Sejnowski TJ. 2003. Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Comput.* 15(2):397–417

Dunn FA, Lankheet MJ, Rieke F. 2007. Light adaptation in cone vision involves switching between receptor and post-receptor sites. *Nature*. 449(7162):603–6

Endeman D, Kamermans M. 2010. Cones perform a non-linear transformation on natural stimuli. *J. Physiol.* 588(3):435–46

Enroth-Cugell C, Freeman AW. 1987. The receptive-field spatial structure of cat retinal Y cells. *J.*

Physiol. 384(1):49–79

Enroth-Cugell C, Robson JG. 1966. The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol.* 187(3):517–52

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A.* 4(12):2379–94

Frazor RA, Geisler WS. 2006. Local luminance and contrast in natural images. *Vision Res.* 46(10):1585–98

Ganmor E, Segev R, Schneidman E. 2015. A thesaurus for a neural population code. *eLife*. 4:e06134

Garvert MM, Gollisch T. 2013. Local and global contrast adaptation in retinal ganglion cells. *Neuron.* 77(5):915–28

Geffen MN, de Vries SEJ, Meister M. 2007. Retinal Ganglion Cells Can Rapidly Change Polarity from Off to On. *PLoS Biol.* 5(3):e65

Gjorgjieva J, Sompolinsky H, Meister M. 2014. Benefits of pathway splitting in sensory coding. *J. Neurosci.* 34(36):12127–44

Goldin M, Lefebvre B, Virgili S, Ecker A, Mora T, Ferrari U, Marre O. 2021. Context-dependent selectivity to natural scenes in the retina. *bioRxiv*. 462157v1

Gollisch T. 2013. Features and functions of nonlinear spatial integration by retinal ganglion cells. *J. Physiol.* 107(5):338–48

Gouras P, Ekesten B. 2004. Why do mice have ultra-violet vision? *Exp. Eye Res.* 79(6):887–92

Greschner M, Bongard M, Rujan P, Ammermüller J. 2002. Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. *Nat. Neurosci.* 5(4):341–47

Grimes WN, Schwartz GW, Rieke F. 2014. The synaptic and circuit mechanisms underlying a change in spatial encoding in the retina. *Neuron*. 82(2):460–73

Gutierrez GJ, Rieke F, Shea-Brown ET. 2021. Nonlinear convergence boosts information coding in circuits with parallel outputs. *Proc. Natl. Acad. Sci.* 118(8):e1921882118

Hartline HK. 1938. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am. J. Physiol.* 121(2):400–415

Heitman A, Brackbill N, Greschner M, Sher A, Litke AM, Chichilnisky EJ. 2016. Testing pseudo-linear models of responses to natural scenes in primate retina. *bioRxiv*. 045336v2

Holmgren CD, Stahr P, Wallace DJ, Voit K, Matheson EJ, Sawinski J, Bassetto G, Kerr JN. 2021. Visual pursuit behavior in mice maintains the pursued prey on the retinal region with least optic flow. *eLife*. 10:e70838

Howlett MHC, Smith RG, Kamermans M. 2017. A novel mechanism of cone photoreceptor adaptation. *PLoS Biol.* 15(4):e2001210

Idrees S, Baumann MP, Franke F, Münch TA, Hafed ZM. 2020. Perceptual saccadic suppression starts in the retina. *Nat. Commun.* 11(1):1977

Im M, Fried SI. 2016. Directionally selective retinal ganglion cells suppress luminance responses during natural viewing. *Sci. Rep.* 6(1):35708

Jacoby J, Schwartz GW. 2017. Three Small-Receptive-Field Ganglion Cells in the Mouse Retina Are Distinctly Tuned to Size, Speed, and Object Motion. *J. Neurosci.* 37(3):610–25

Jadzinsky PD, Baccus SA. 2015. Synchronized amplification of local information transmission by peripheral retinal input. *eLife*. 4:e09266

Jarsky T, Cembrowski M, Logan SM, Kath WL, Riecke H, Demb JB, Singer JH. 2011. A Synaptic Mechanism for Retinal Adaptation to Luminance and Contrast. *J. Neurosci.* 31(30):11003–15

Joesch M, Meister M. 2016. A neuronal circuit for colour vision based on rod–cone opponency. *Nature*. 532(7598):236–39

Jun NY, Field GD, Pearson J. 2021. Scene statistics and noise determine the relative arrangement of receptive field mosaics. *Proc. Natl. Acad. Sci.* 118(39):e2105115118

Karamanlis D, Gollisch T. 2021. Nonlinear spatial integration underlies the diversity of retinal ganglion cell responses to natural images. *J. Neurosci.* 41(15):3479–98

Karklin Y, Simoncelli EP. 2011. Efficient coding of natural images with a population of noisy Linear-Nonlinear neurons. *Adv. Neural Inf. Process. Syst.* 24:999–1007

Kastner DB, Baccus SA. 2011. Coordinated dynamic encoding in the retina using opposing forms of plasticity. *Nat. Neurosci.* 14(10):1317–22

Khani MH, Gollisch T. 2017. Diversity in spatial scope of contrast adaptation among mouse retinal ganglion cells. *J. Neurophysiol.* 118(6):3024–43

Khani MH, Gollisch T. 2021. Linear and nonlinear chromatic integration in the mouse retina. *Nat. Commun.* 12(1):1900

Kim T, Shen N, Hsiang J-C, Johnson KP, Kerschensteiner D. 2020. Dendritic and parallel processing of visual threats in the retina control defensive responses. *Sci. Adv.* 6(47):eabc9920

Kim YJ, Brackbill N, Batty E, Lee J, Mitelut C, Tong W, Chichilnisky EJ, Paninski L. 2021. Nonlinear decoding of natural images from large-scale primate retinal ganglion recordings. *Neural Comput.* 33(7):1719–50

Kling A, Gogliettino AR, Shah NP, Wu EG, Brackbill N, Sher A, Litke AM, Silva RA, Chichilnisky EJ. 2020. Functional Organization of Midget and Parasol Ganglion Cells in the Human Retina. *bioRxiv*. 240762v1

Krishnamoorthy V, Weick M, Gollisch T. 2017. Sensitivity to image recurrence across eye-movement-like image transitions through local serial inhibition in the retina. *eLife*. 6:e22431

Kuffler SW. 1953. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* 16(1):37–68

Kühn NK, Gollisch T. 2019. Activity Correlations between Direction-Selective Retinal Ganglion Cells Synergistically Enhance Motion Decoding from Complex Visual Scenes. *Neuron*. 101(5):963-976.e7

Land MF. 2015. Eye movements of vertebrates and their relation to eye form and function. *J. Comp. Physiol. A*. 201(2):195–214

Lettvin J, Maturana H, McCulloch W, Pitts W. 1959. What the Frog's Eye Tells the Frog's Brain. *Proc. IRE*. 47(11):1940–51

Lindsey J, Ocko SA, Ganguli S, Deny S. 2019. A Unified Theory of Early Visual Representations from Retina to Cortex through Anatomically Constrained Deep CNNs. *arXiv*. 1901.00945v1

Liu JK, Gollisch T. 2021. Simple Model for Encoding Natural Images by Retinal Ganglion Cells with Nonlinear Spatial Integration. *bioRxiv*. 458067v1

Liu JK, Schreyer HM, Onken A, Rozenblit F, Khani MH, Krishnamoorthy V, Panzeri S, Gollisch T. 2017. Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nat. Commun.* 8(1):149

Maheswaranathan N, Kastner DB, Baccus SA, Ganguli S. 2018a. Inferring hidden structure in multilayered neural circuits. *PLOS Comput. Biol.* 14(8):e1006291

Maheswaranathan N, McIntosh LT, Tanaka H, Grant S, Kastner DB, Melander JB, Nayebi A, Brezovec L, Wang J, Ganguli S, Baccus SA. 2018b. The dynamic neural code of the retina for natural scenes. *bioRxiv*. 340943v5

Manookin MB, Demb JB. 2006. Presynaptic Mechanism for Slow Contrast Adaptation in Mammalian Retinal Ganglion Cells. *Neuron*. 50(3):453–64

Manookin MB, Puller C, Rieke F, Neitz J, Neitz M. 2015. Distinctive receptive field and physiological properties of a wide-field amacrine cell in the macaque monkey retina. *J. Neurophysiol.* 114(3):1606–16

Mante V, Frazor RA, Bonin V, Geisler WS, Carandini M. 2005. Independence of luminance and contrast in natural scenes and in the early visual system. *Nat. Neurosci.* 8(12):1690–97

Martersteck EM, Hirokawa KE, Evarts M, Bernard A, Duan X, Li Y, Ng L, Oh SW, Ouellette B, Royall JJ, Stoecklin M, Wang Q, Zeng H, Sanes JR, Harris JA. 2017. Diverse Central Projection Patterns of Retinal Ganglion Cells. *Cell Rep.* 18(8):2058–72

Masland RH. 2001. Neuronal diversity in the retina. *Curr. Opin. Neurobiol.* 11(4):431–36

Masquelier T, Portelli G, Kornprobst P. 2016. Microsaccades enable efficient synchrony-based coding in the retina: a simulation study. *Sci. Rep.* 6(1):24086

Matthews HR, Murphy RLW, Fain GL, Lamb TD. 1988. Photoreceptor light adaptation is mediated by cytoplasmic calcium concentration. *Nature*. 334(6177):67–69

McFarland JM, Cui Y, Butts DA. 2013. Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. *PLoS Comput. Biol.* 9(7):e1003143

McIntosh LT, Maheswaranathan N, Nayebi A, Ganguli S, Baccus SA. 2016. Deep Learning Models of the Retinal Response to Natural Scenes. *Adv. Neural Inf. Process. Syst.* 29:1369–77

Meytlis M, Nichols Z, Nirenberg S. 2012. Determining the role of correlated firing in large populations of neurons using white noise and natural scene stimuli. *Vision Res.* 70:44–53

Młynarski WF, Hermundstad AM. 2021. Efficient and adaptive sensory codes. *Nat. Neurosci.* 24(7):998–1009

Morin LP, Studholme KM. 2014. Retinofugal projections in the mouse. *J. Comp. Neurol.* 522(16):3733–53

Münch TA, Da Silveira RA, Siegert S, Viney TJ, Awatramani GB, Roska B. 2009. Approach sensitivity in the retina processed by a multifunctional neural circuit. *Nat. Neurosci.* 12(10):1308–16

Nakatani K, Yau KW. 1988. Calcium and light adaptation in retinal rods and cones. *Nature*. 334(6177):69–71

Nirenberg S, Carcieri SM, Jacobs AL, Latham PE. 2001. Retinal ganglion cells act largely as independent encoders. *Nature*. 411(6838):698–701

Nirenberg S, Pandarinath C. 2012. Retinal prosthetic strategy with the capacity to restore normal vision. *Proc. Natl. Acad. Sci.* 109(37):15012–17

Noda H, Adey WR. 1974. Excitability changes in cat lateral geniculate cells during saccadic eye movements. *Science*. 183(4124):543–45

Ocko S, Lindsey J, Ganguli S, Deny S. 2018. The emergence of multiple retinal cell types through efficient coding of natural movies. *Adv. Neural Inf. Process. Syst.* 31:9411–22

Olmos A, Kingdom FAA. 2004. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*. 33(12):1463–73

Ölveczky BP, Baccus SA, Meister M. 2003. Segregation of object and background motion in the retina. *Nature*. 423(6938):401–8

Ozuysal Y, Baccus SA. 2012. Linking the Computational Structure of Variance Adaptation to Biophysical Mechanisms. *Neuron.* 73(5):1002–15

Parthasarathy N, Batty E, Falcon W, Rutten T, Rajpal M, Chichilnisky EJ, Paninski L, Falcon W, Rutten T, Rajpal M, Chichilnisky EJ, Paninski L. 2017. Neural Networks for Efficient Bayesian Decoding of Natural Images from Retinal Neurons. *Adv. Neural Inf. Process. Syst.* 30:6434–45

Petrusca D, Grivich MI, Sher A, Field GD, Gauthier JL, Greschner M, Shlens J, Chichilnisky EJ, Litke AM. 2007. Identification and characterization of a Y-like primate retinal ganglion cell type. *J. Neurosci.* 27(41):11019–27

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*. 454(7207):995–99

Pitkow X, Meister M. 2012. Decorrelation and efficient coding by retinal ganglion cells. *Nat. Neurosci.* 15(4):628–35

Puchalla JL, Schneidman E, Harris RA, Berry MJ. 2005. Redundancy in the population code of the retina. *Neuron.* 46(3):493–504

Qiu Y, Zhao Z, Klindt D, Kautzky M, Szatko KP, Schaeffel F, Rifai K, Franke K, Busse L, Euler T. 2021. Natural environment statistics in the upper and lower visual field are reflected in mouse retinal specializations. *Curr. Biol.* 31(15):3233-3247.e6

Ratliff CP, Borghuis BG, Kao Y-H, Sterling P, Balasubramanian V. 2010. Retina is structured to process an excess of darkness in natural scenes. *Proc. Natl. Acad. Sci.* 107(40):17368–73

Ravi S, Ahn D, Greschner M, Chichilnisky EJ, Field GD. 2018. Pathway-Specific Asymmetries between ON and OFF Visual Signals. *J. Neurosci.* 38(45):9728–40

Real E, Asari H, Gollisch T, Meister M. 2017. Neural Circuit Inference from Function to Structure. *Curr. Biol.* 27(2):189–98

Rieke F, Rudd ME. 2009. The Challenges Natural Images Pose for Visual Adaptation. *Neuron.* 64(5):605–16

Rodieck RW. 1965. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Res.* 5(12):583–601

Roska B, Werblin F. 2003. Rapid global shifts in natural scenes block spiking in specific ganglion cell types. *Nat. Neurosci.* 6(6):600–608

Roy S, Jun NY, Davis EL, Pearson J, Field GD. 2021. Inter-mosaic coordination of retinal receptive fields. *Nature*. 592(7854):409–13

Rust NC, Movshon JA. 2005. In praise of artifice. *Nat. Neurosci.* 8(12):1647–50

Sabbah S, Gemmer JA, Bhatia-Lin A, Manoff G, Castro G, Siegel JK, Jeffery N, Berson DM. 2017. A retinal code for motion along the gravitational and body axes. *Nature*. 546(7659):492–97

Salisbury JM, Palmer SE. 2016. Optimal Prediction in the Retina and Natural Motion Statistics. *J. Stat. Phys.* 162(5):1309–23

Schneidman E, Berry MJ, Segev R, Bialek W. 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*. 440(7087):1007–12

Schreyer HM, Gollisch T. 2021. Nonlinear spatial integration in retinal bipolar cells shapes the encoding of artificial and natural stimuli. *Neuron*. 109(10):1692-1706.e8

Schwartz G, Rieke F. 2011. Nonlinear spatial encoding by retinal ganglion cells: when $1 + 1 \neq 2$. *J. Gen. Physiol.* 138(3):283–90

Schwartz G, Taylor S, Fisher C, Harris R, Berry MJ. 2007. Synchronized Firing among Retinal Ganglion Cells Signals Motion Reversal. *Neuron*. 55(6):958–69

Segal IY, Giladi C, Gedalin M, Rucci M, Ben-Tov M, Kushinsky Y, Mokeichev A, Segev R. 2015. Decorrelation of retinal response to natural scenes by fixational eye movements. *Proc. Natl. Acad. Sci. U. S. A.* 112(10):3110–15

Shah NP, Brackbill N, Rhoades C, Kling A, Goetz G, Litke AM, Sher A, Simoncelli EP, Chichilnisky E. 2020. Inference of nonlinear receptive field subunits with spike-triggered clustering. *eLife*. 9:e45743

Sivyer B, Tomlinson A, Taylor WR. 2019. Simulated Saccadic Stimuli Suppress ON-Type Direction-Selective Retinal Ganglion Cells via Glycinergic Inhibition. *J. Neurosci.* 39(22):4312–22

Soto F, Hsiang J, Rajagopal R, Piggott K, Harocopos GJ, Couch SM, Custer P, Morgan JL, Kerschensteiner D. 2020. Efficient Coding by Midget and Parasol Ganglion Cells in the Human Retina. *Neuron*. 107(4):656-666.e5

Srinivasan M V., Laughlin SB, Dubs A, Series L, Sciences B. 1982. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. London. Ser. B. Biol. Sci.* 216(1205):427–59

Sterling P, Laughlin S. 2015. *Principles of Neural Design*. The MIT Press

Szatko KP, Korympidou MM, Ran Y, Berens P, Dalkara D, Schubert T, Euler T, Franke K. 2020. Neural circuits in the mouse retina support color vision in the upper visual field. *Nat. Commun.* 11(1):3481

Szél, Röhlich P, Gaffé AR, Juliusson B, Aguirre G, Van Veen T. 1992. Unique topographic separation of two spectral classes of cones in the mouse retina. *J. Comp. Neurol.* 325(3):327–42

Takeshita D, Gollisch T. 2014. Nonlinear Spatial Integration in the Receptive Field Surround of Retinal Ganglion Cells. *J. Neurosci.* 34(22):7548–61

Tanaka H, Nayebi A, Maheswaranathan N, McIntosh L, Baccus SA, Ganguli S. 2019. From deep learning

to mechanistic understanding in neuroscience: The structure of retinal prediction. *Adv. Neural Inf. Process. Syst.* 32:8537–47

Tkačik G, Garrigan P, Ratliff C, Milčinski G, Klein JM, Seyfarth LH, Sterling P, Brainard DH, Balasubramanian V. 2011. Natural images from the birthplace of the human eye. *PLoS One*. 6(6):e20409

Tran NM, Shekhar K, Whitney IE, Jacobi A, Benhar I, Hong G, Yan W, Adiconis X, Arnold ME, Lee JM, Levin JZ, Lin D, Wang C, Lieber CM, Regev A, He Z, Sanes JR. 2019. Single-Cell Profiles of Retinal Ganglion Cells Differing in Resilience to Injury Reveal Neuroprotective Genes. *Neuron*. 104(6):1039-1055.e12

Turner MH, Rieke F. 2016. Synaptic Rectification Controls Nonlinear Spatial Integration of Natural Visual Inputs. *Neuron*. 90(6):1257–71

Turner MH, Sanchez Giraldo LG, Schwartz O, Rieke F. 2019. Stimulus- and goal-oriented frameworks for understanding natural vision. *Nat. Neurosci.* 22(1):15–24

Turner MH, Schwartz GW, Rieke F. 2018. Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. *eLife*. 7:252148

van Hateren H. 2005. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *J. Vis.* 5(4):331–47

van Hateren JH, Rüttiger L, Sun H, Lee BB. 2002. Processing of natural temporal stimuli by macaque retinal ganglion cells. *J. Neurosci.* 22(22):9945–60

van Hateren JH, van der Schaaf A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. B Biol. Sci.* 265(1394):359–66

Victor JD, Shapley RM. 1979. The nonlinear pathway of Y ganglion cells in the cat retina. *J. Gen. Physiol.* 74(6):671–89

Wallace DJ, Greenberg DS, Sawinski J, Rulla S, Notaro G, Kerr JND. 2013. Rats maintain an overhead binocular field at the expense of constant fusion. *Nature*. 498(7452):65–69

Wark B, Fairhall A, Rieke F. 2009. Timescales of Inference in Visual Adaptation. *Neuron*. 61(5):750–61

Wässle H. 2004. Parallel processing in the mammalian retina. *Nat. Rev. Neurosci.* 5(10):747–57

Weber AI, Pillow JW. 2017. Capturing the Dynamical Repertoire of Single Neurons with Generalized Linear Models. *Neural Comput.* 29(12):3260–89

Yan W, Laboulaye MA, Tran NM, Whitney IE, Benhar I, Sanes JR. 2020. Mouse Retinal Cell Atlas: Molecular Identification of over Sixty Amacrine Cell Types. *J. Neurosci.* 40(27):5177–95

Yedutenko M, Howlett MHC, Kamermans M. 2020. High Contrast Allows the Retina to Compute More Than Just Contrast. *Front. Cell. Neurosci.* 14:595193

Yoshimatsu T, Bartel P, Schröder C, Janiak FK, St-Pierre F, Berens P, Baden T. 2021. Ancestral circuits for vertebrate color vision emerge at the first retinal synapse. *Sci. Adv.* 7(42):eabj6815

Yoshimatsu T, Schröder C, Nevala NE, Berens P, Baden T. 2020. Fovea-like Photoreceptor Specializations Underlie Single UV Cone Driven Prey-Capture Behavior in Zebrafish. *Neuron*. 107(2):320-337.e6

Yu Z, Turner MH, Baudin J, Rieke F. 2021. Adaptation in cone photoreceptors contributes to an

unexpected insensitivity of On parasol retinal ganglion cells to spatial structure in natural images. *bioRxiv*. 450295v1

Zhang Y, Kim I-J, Sanes JR, Meister M. 2012. The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proc. Natl. Acad. Sci.* 109(36):E2391–98

3 NONLINEAR SPATIAL INTEGRATION UNDERLIES THE DIVERSITY OF RETINAL GANGLION CELL RESPONSES TO NATURAL IMAGES

Attached manuscript. Research article published in the *Journal of Neuroscience*. Reprinted below is the published manuscript, licensed under CC BY 4.0.

Authors:

Dimokratis Karamanlis[1,2,3], Tim Gollisch[1,2]

1. University Medical Center Göttingen
2. Bernstein Center Göttingen
3. International Max Planck Research School for Neurosciences, Göttingen

Own contributions:

- Conceptualization (with TG)
- Data curation
- Formal analysis
- Investigation
- Methodology
- Project administration (with TG)
- Software
- Validation (with TG)
- Visualization
- Writing – original draft
- Writing – reviewing and editing (with TG)

Systems/Circuits

Nonlinear Spatial Integration Underlies the Diversity of Retinal Ganglion Cell Responses to Natural Images

Dimokratis Karamanlis[1,2,3] and Tim Gollisch[1,2]

[1]Department of Ophthalmology, University Medical Center Göttingen, Göttingen, 37073, Germany, [2]Bernstein Center for Computational Neuroscience Göttingen, Göttingen, 37073, Germany, and [3]International Max Planck Research School for Neurosciences, Göttingen, 37077, Germany

How neurons encode natural stimuli is a fundamental question for sensory neuroscience. In the early visual system, standard encoding models assume that neurons linearly filter incoming stimuli through their receptive fields, but artificial stimuli, such as contrast-reversing gratings, often reveal nonlinear spatial processing. We investigated to what extent such nonlinear processing is relevant for the encoding of natural images in retinal ganglion cells in mice of either sex. We found that standard linear receptive field models yielded good predictions of responses to flashed natural images for a subset of cells but failed to capture the spiking activity for many others. Cells with poor model performance displayed pronounced sensitivity to fine spatial contrast and local signal rectification as the dominant nonlinearity. By contrast, sensitivity to high-frequency contrast-reversing gratings, a classical test for nonlinear spatial integration, was not a good predictor of model performance and thus did not capture the variability of nonlinear spatial integration under natural images. In addition, we also observed a class of nonlinear ganglion cells with inverse tuning for spatial contrast, responding more strongly to spatially homogeneous than to spatially structured stimuli. These findings highlight the diversity of receptive field nonlinearities as a crucial component for understanding early sensory encoding in the context of natural stimuli.

Key words: **natural visual stimuli; nonlinear spatial integration; receptive field; retinal ganglion cell**

Significance Statement

Experiments with artificial visual stimuli have revealed that many types of retinal ganglion cells pool spatial input signals nonlinearly. However, it is still unclear how relevant this nonlinear spatial integration is when the input signals are natural images. Here we analyze retinal responses to natural scenes in large populations of mouse ganglion cells. We show that nonlinear spatial integration strongly influences responses to natural images for some ganglion cells, but not for others. Cells with nonlinear spatial integration were sensitive to spatial structure inside their receptive fields, and a small group of cells displayed a surprising sensitivity to spatially homogeneous stimuli. Traditional analyses with contrast-reversing gratings did not predict this variability of nonlinear spatial integration under natural images.

Introduction

The natural visual world is communicated to the brain through an array of functionally distinct parallel channels that originate in the retina (Roska and Meister, 2014; Baden et al., 2016). A classical view of retinal function advocates that the retinal output channels, represented by types of retinal ganglion cells (RGCs), serve as linear filters for natural visual inputs (Atick and Redlich, 1990; Shapley, 2009). Recordings under artificial visual stimuli, such as contrast-reversing gratings (Enroth-Cugell and Robson, 1966; Demb et al., 1999; Petrusca et al., 2007; Krieger et al., 2017) or finely structured white noise (Freeman et al., 2015; Liu et al., 2017), however, have shown that several ganglion cell types have spatially nonlinear receptive fields (RFs). The nonlinearities arise in the RF center from the nonlinear integration of excitatory signals, which originate from presynaptic bipolar cells (Demb et al., 2001; Borghuis et al., 2013; Turner and Rieke, 2016). Furthermore, nonlinear RFs are proposed in circuit models of retinal computations that are thought to occur during natural vision (Gollisch and Meister, 2010), such as the distinction of object from background motion (Ölveczky et al., 2003; Baccus et al., 2008; Zhang et al., 2012).

Received Dec. 7, 2020; revised Feb. 5, 2021; accepted Feb. 9, 2021.

Author contributions: D.K. and T.G. designed research; D.K. performed research; D.K. analyzed data; D.K. wrote the first draft of the paper; D.K. and T.G. edited the paper.

The authors declare no competing financial interests.

This work was supported by Boehringer Ingelheim Fonds, the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant Agreement 724822), and Deutsche Forschungsgemeinschaft (German Research Foundation) Projects 154113120 (SFB 889, Project C01) and 432680300 (SFB 1456, Project B05). We thank Michael Weick for help in designing experiments and for discussion.

Correspondence should be addressed to Tim Gollisch at tim.gollisch@med.uni-goettingen.de.

https://doi.org/10.1523/JNEUROSCI.3075-20.2021

Together, these findings raise the question to what extent nonlinear RFs of different ganglion cell types play a role in natural vision. On the one hand, the spatial structure of natural stimuli is not as pronounced and rich in high spatial frequencies as in typical artificial stimuli used to detect nonlinear RFs because the light intensities of nearby regions in natural images are correlated (Burton and Moorhead, 1987). This leads to extensive areas of nearly homogeneous illumination, for which RF nonlinearities may play no role. On the other hand, object boundaries can induce pronounced changes of stimulus intensity over short distances (Turiel and Parga, 2000), and textures or illumination gradients may provide further structure within individual RFs. Despite the importance of evaluating stimulus encoding models under natural stimuli (Carandini et al., 2005; Felsen and Dan, 2005), only few studies have focused on whether the linear RF provides a good abstraction of RGCs for stimuli with natural spatial structure, and reported findings are mixed. Some studies support that linear RFs suffice to describe natural stimulus encoding in mouse and primate retina (Nirenberg and Pandarinath, 2012; Bomash et al., 2013), whereas others indicate that linear RFs may fail to predict natural scene responses in mammalian (Cao et al., 2011; Freeman et al., 2015; Heitman et al., 2016; Turner and Rieke, 2016; Shah et al., 2020) and salamander retinas (Liu et al., 2017; McIntosh et al., 2017).

In this work, we establish a connection of spatial RF nonlinearities to the encoding of natural images in RGCs. We do so in the mouse retina, in which spatial integration, as measured with artificial stimuli, appears to display a broad scope (Carcieri et al., 2003), with spatially linear (Krieger et al., 2017; Johnson et al., 2018) as well as strongly nonlinear cells (Zhang et al., 2012; Jacoby and Schwartz, 2017; Mani and Schwartz, 2017). We first show that linear RF models successfully predict responses to natural images for some ganglion cells and substantially fail for others. We then connect model failure to the characteristics of spatial nonlinearities in the RF center and analyze these under different stimulus layouts and for specific functional cell types.

Materials and Methods

Experimental design and statistical analysis. We used 13 retina pieces from 9 adult WT mice of either sex (6 C57BL/6J and 3 C57BL/6N; 7 male and 2 female), mostly between 8 and 12 weeks old (except for one 18- and one 26-week-old). All mice were housed in a 12 h light/dark cycle. Experimental procedures were in accordance with national and institutional guidelines and approved by the institutional animal care committee of the University Medical Center Göttingen, Germany. No statistical methods were used to predetermine sample size. Statistical tests and associated information (e.g., p values) are noted where appropriate in the text. For all statistical procedures, we used default MATLAB2019b functions.

Tissue preparation and electrophysiology. Mice were dark-adapted for at least 1 h before eye enucleation. After the animal had been killed, both eyes were removed and immersed in oxygenated (95% O_2-5% CO_2) Ames' medium (Sigma Millipore), supplemented with 22 mM $NaHCO_3$ (Merck Millipore) and 6 mM D-glucose (Carl Roth). We cut the globes along the ora serrata, removing the cornea, lens, and vitreous humor. In some experiments, the resulting eyecups were cut in half to allow two separate recordings. Before the start of each recording, we isolated retina pieces from the eyecups. We placed the pieces ganglion cell-side-down on planar multielectrode arrays (Multichannel Systems; 252 electrodes; 30 μm diameter, either 100 or 200 μm minimal electrode distance) with the help of a semipermeable membrane, stretched across a circular plastic holder (removed before the recording). The arrays were coated with poly-D-lysine (Merck Millipore). Throughout the recording, retinal pieces were continuously superfused with the oxygenated Ames solution flowing at ~250 ml/h. The bath solution was heated to a constant temperature of 34°C-35°C via an inline heater in the perfusion line and a heating element below the array. Dissection and mounting were performed under infrared light on a stereo-microscope equipped with night-vision goggles.

Extracellular voltage signals were amplified, bandpass filtered between 300 Hz and 5 kHz, and digitized at 10 kHz sampling rate. Spikes were detected by threshold crossings (4 SDs of the voltage trace), and spike waveforms were sorted offline into units with a custom-made IgorPro (WaveMetrics) routine based on Gaussian mixture models (Pouzat et al., 2002). We curated the routine's output and selected only well-separated units with clear refractory periods. Duplicate units were identified by temporal cross-correlations and removed. Finally, only units with stable electrical images (Litke et al., 2004) throughout the recording were considered for further analysis.

Visual stimulation. Visual stimuli were generated and controlled through custom-made software, based on Visual C++ and OpenGL. Different stimuli were presented sequentially to the retina through a gamma-corrected monochromatic white OLED monitor (eMagin) with 800×600 square pixels and 60 Hz refresh rate. The monitor image was projected through a telecentric lens (Edmund Optics) onto the photoreceptor layer of the retina, and each pixel's side measured 7.5 μm on the retina. All stimuli were presented on a background of low photopic light levels (2.5 or 3.5 mW/m^2, corresponding to 1500 or 1900 R*/rod/s), and their mean intensity was always equal to the background. We fine-tuned the focus of stimuli on the photoreceptor layer before the start of each experiment by visual monitoring through a light microscope and by inspection of spiking responses to contrast-reversing gratings with a bar width of 30 μm.

Linear RF identification. To estimate the RF of each cell, we used a spatiotemporal binary white-noise stimulus (100% contrast) consisting of a checkerboard layout with flickering squares (60 μm side). The update rate was either 30 or 60 Hz in different experiments. We measured the spatiotemporal RF by calculating the spike-triggered average (STA) over a 500 ms time window (Chichilnisky, 2001) and fitted a parametric model to the RF (Chichilnisky and Kalmar, 2002). The model was spatiotemporally separable and comprised a product of a spatial ($k_S(\boldsymbol{x})$) and a temporal component ($k_T(t)$).

The spatial component was modeled as a difference of Gaussians as follows:

$$k_S(\boldsymbol{x}) = N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) - A_S N\left(\boldsymbol{x}; \boldsymbol{\mu}, k^2\boldsymbol{\Sigma}\right)$$

where $N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$ is a two-dimensional Gaussian function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (describing the RF center's coordinates and shape), $A_s \in [0,1]$ captures the RF surround strength relative to the RF center, and $k \geq 1$ is a scaling factor for the surround's extent.

The temporal component was modeled as a difference of two low-pass filters as follows:

$$k_T(t) = p_1 \left(\frac{t}{\tau_1} e^{-\frac{t}{\tau_1}} + 1 \right)^n - p_2 \left(\frac{t}{\tau_2} e^{-\frac{t}{\tau_2}} + 1 \right)^n$$

with $t>0$ indicating the time before the spike and $p_1>0$, $p_2>0$, $\tau_1>0$, $\tau_2>0$, $n>0$ being free parameters.

We fitted the full parametric model ($k_S(\boldsymbol{x}) \cdot k_T(t)$) to the STA by minimizing the mean squared error using constrained nonlinear optimization. To get reasonable initial conditions, we first separately fitted the spatial component to the STA frame at which the element with the largest absolute value occurred and the temporal component to the time course of the same element. If the element was negative, the sign of the STA frame was inverted before the spatial component fit. We then seeded the obtained values of spatial and temporal fits as the initial parameters for the full spatiotemporal fit. The chosen initialization procedure is consistent with the positive center peak of k_S.

The diameter of the RF center was defined as the diameter of a circle with the same area as the 2σ (elliptical) boundary of the Gaussian center

profile (Baden et al., 2016). We also used the 2σ boundary for all RF center visualizations.

Natural image response predictions with a linear-nonlinear (LN) model. We selected natural images as stimuli from three sources: the van Hateren Natural Image Dataset (van Hateren and van der Schaaf, 1998), the McGill Calibrated Color Image Database (Olmos and Kingdom, 2004), and the Berkeley Segmentation Dataset (Arbeláez et al., 2011). The central square region of each image was resized to 512×512 pixels (400×400 pixels in a few experiments) by cropping (van Hateren and McGill images) or cropping and upsampling with nearest neighbor interpolation (Berkeley images). All color images (McGill and Berkeley databases) were converted to grayscale by weighted averaging over the color channels (Liu et al., 2017). We normalized the mean and SD of the pixel values for each image by appropriately shifting and scaling the values so that the mean pixel intensity was equal to the background and the SD was 40% of the mean intensity. Pixel values that then deviated from the mean by >100% in either direction were clipped to ensure that the maximal pixel values were within the physically available range of the display. Finally, all images were encoded at 8-bit color depth to match the range of our OLED monitor. The images were presented on top of a uniform gray background and centered on the multielectrode array, covering a region of 3.84×3.84 mm^2 on the retina (3×3 mm^2 for 400×400 pixels).

In every experiment, we used 300 natural images (100 from each database), except for one (200 images in total). Images were presented individually for 200 ms each, with an 800 ms interstimulus interval of homogeneous background illumination. We collected 10 trials for each image by consecutively presenting 10 different pseudo-randomly permuted sequences of all images. For each cell, we measured the response as the trial-averaged number of spikes over a 250 ms window following stimulus onset.

To compare a cell's responses to model predictions, we constructed an LN model (Chichilnisky, 2001), which generates average spike count responses $R_m \geq 0$ to natural image stimuli $\boldsymbol{s}_m$: $R_m = f(\boldsymbol{k}^T \cdot \boldsymbol{s}_m)$, where the vector $\boldsymbol{k}$ is a linear spatial filter, f is a nonlinear function, and m denotes the image index. For the analyses, all natural image stimuli were spatially clipped to the smallest square that could fit the 4σ boundary of the RF center, and their pixel intensity values were transformed to Weber contrast values, which constitute the elements of $\boldsymbol{s}_m$. For the linear filter ($\boldsymbol{k}$), we used the parametric spatial RF component (k_S) estimated from white noise, sampled at the center point of each pixel of the clipped natural image. The linear filter was normalized to a sum of unity of the absolute values of its elements.

In initial analyses, we also tried using the pixelwise spatial profile obtained from the reverse-correlation analysis as a spatial filter of the model. Overall, response predictions and model performances were similar to the ones under the parametric fit, yet often more noisy, owing to noise in the pixelwise filter estimate. All further analyses were therefore based on the parametric fit of the spatial filter.

Linear predictions (g) were estimated from the inner product of stimuli and the linear filter: $g_m = \boldsymbol{k}^T \cdot \boldsymbol{s}_m$. Because the linear filter is composed of mainly positive values, the sign of the linear prediction reflects the net contrast in the spatial RF. For the nonlinear part of the LN model (f), we used a bi-logistic nonlinearity of the following form:

$$f(g) = b + \frac{M_{ON} - b}{1 + e^{-r_{ON}(g - g_{ON})}} + \frac{M_{OFF} - b}{1 + e^{r_{OFF}(g - g_{OFF})}}$$

where $b \geq 0$, $M_{ON} \geq b$, $M_{OFF} \geq b$, $r_{ON}, \geq 0$, $r_{OFF} \geq 0$, and g_{ON}, g_{OFF} were free parameters that were fitted to data. To facilitate estimation of the parameters for both monotonic and U-shaped nonlinearities, we first fitted a single logistic nonlinearity $f(g) = b + (M - b)/[1 + e^{-r_o(g - g_o)}]$ to the data, and initialized the parameters of the bi-logistic nonlinearity to describe the dominant lobe (ON or OFF, determined by the sign of r_o). Such bi-logistic nonlinearities had been previously used to describe tuning curves in sensory neuroscience under the name "difference of sigmoids" (Fischer et al., 2009; Mel et al., 2018; Murgas et al., 2020).

To assess the prediction accuracy by the LN model, we applied a normalized correlation coefficient (CC_{norm}) as our model performance metric (Schoppe et al., 2016). This measure was used to account for differences in response reliability among cells, since we used a relatively small number of trials per image. Concretely, for M images and N trials per image, $R_{m,n}$ denoted the cell's response to image m for trial n, with $y_m = \sum_{n=1}^{N} R_{m,n}/N$ being the average response to a particular image, $\widehat{y_m}$ being the prediction for the same image, and y and $\widehat{y}$ denoting the corresponding distributions over images. CC_{norm} was then defined as follows:

$$CC_{norm} = \frac{\mathrm{Cov}(y, \widehat{y})}{\sqrt{\mathrm{Var}(\widehat{y}) \times SP}}$$

We calculated the necessary quantities as the sample covariance $\mathrm{Cov}(y, \widehat{y}) = \sum_{m=1}^{M} (y_m - y)(\widehat{y_m} - \widehat{y})/M$ and sample variance $\mathrm{Var}(\widehat{y}) = \sum_{m=1}^{M} (\widehat{y_m} - \widehat{y})^2/M$. SP denotes the signal power, a measure of signal-to-noise ratio, and is defined as follows:

$$SP = \frac{\mathrm{Var}\left(\sum_{n=1}^{N} R_{m,n}\right) - \sum_{n=1}^{N} \mathrm{Var}(R_{m,n})}{N(N-1)}$$

with the two variances in the numerator denoting sample variances over images.

We estimated LN model performance through 10-fold cross-validation. Briefly, the collection of average responses for all images was randomly split into 10 equally sized sets. Every set was used once as a test set for the full LN model, whose nonlinearity was fitted to the other 90% of image responses. For each cell, LN model performance was defined as the average CC_{norm} over all cross-validation sets. For all nonlinearity visualizations in the plots, we used the nonlinearity corresponding to the cross-validation set with the CC_{norm} value closest to the average.

Since CC_{norm} values are ill-defined for very low data reliability, we excluded cells whose responses for identical images were highly variable. For each cell, we therefore calculated the coefficient of determination (R^2) between responses averaged over even (r^e_m) and over odd trials (r^o_m), where $m = 1,...,M$ enumerates the images. Concretely, we used a symmetrized R^2, defined as follows:

$$R^2 = 1 - \frac{1}{2} \times \frac{\sum_{m=1}^{M} (r^e_m - r^o_m)^2}{\sum_{m=1}^{M} (r^o_m - \mu^o)^2} - \frac{1}{2} \times \frac{\sum_{m=1}^{M} (r^o_m - r^e_m)^2}{\sum_{m=1}^{M} (r^e_m - \mu^e)^2}$$

where μ^o and μ^e are the average odd and even trial responses over images. We excluded cells with $R^2 < 0.5$ from further analysis. We furthermore excluded cells that showed large response drift over the course of image presentations. In most cases, drift corresponded to a global scaling that approximately affected responses to all images proportionally. This is reflected in a high Pearson correlation over images between the average responses of the first five and last five trials; 94% of analyzed cells had a correlation coefficient of at least 0.7. Such global scaling of responses does not affect the analysis of differences in average responses. Thus, we excluded cells with coefficients <0.7 from further analyses. The two criteria (regarding reliability and drift) yielded 900 cells included in the analysis of 1209 recorded cells.

Calculation of spatial contrast (SC) sensitivity for natural images. We measured the SC of an image in the RF center of a given ganglion cell as the weighted SD of pixel contrast values inside the 2σ contour of the Gaussian center fit as follows:

$$SC = \sqrt{\frac{\sum_i w_i (p_i - \mu_w)^2}{\sum_i w_i}}$$

where the sums run over all pixels i within the 2σ contour, p_i is the pixel value, w_i is the pixel weight as given by the value at the pixel center of

the fitted RF center part, and μ_w is the weighted mean of the pixel values.

To obtain the SC sensitivity, we sorted the images according to their linear predictions in the cell's LN model and then grouped neighboring images into pairs, with each image belonging to a single pair, yielding 150 pairs per cell when 300 images had been applied. For each image pair, we calculated the SC difference and the trial-averaged response difference. To compare across cells, we normalized the response differences by the maximum response (over images) of the cell. We defined the SC sensitivity as the slope of the linear regression between the SC differences and the normalized response differences. Cells were defined as contrast-sensitive if they had a significant regression slope at the 5% significance level.

Assessment of spatial nonlinearity with contrast-reversing gratings. To compare our findings to classical analyses of spatial integration, we stimulated the retina with full-field square-wave gratings of 100% contrast. The contrast of the gratings was reversed every 1 s. The reversing gratings were presented sequentially from higher to lower spatial frequencies for 20-30 reversals each, and the whole sequence was repeated 2 times. Depending on the experiment, we sampled 5-8 spatial frequencies, with bar widths ranging from 15 to 240 μm. For each spatial frequency, we applied 1-4 equidistant spatial phases, with more phases for lower spatial frequencies (e.g., one for 15, two for 30, two for 60, four for 120, four for 240 μm bar width). In some of the recordings, we also included contrast reversals of homogeneous illumination (corresponding to a bar width of $\geq$6 mm). Between presentations of the different gratings, there was a gray screen at background intensity for 2 s. We constructed peristimulus time histograms (PSTHs) over one reversal period by binning ganglion cell spikes with 10 ms bins and averaging across reversals and repeats, leaving out the first reversal after a gray period. In some experiments, gratings were flashed for 200 ms, and presentations of reversed contrast were separated by an 800 ms gray screen at background intensity. For the subsequent analyses, PSTHs corresponding to one full reversal period were constructed by extracting the cell responses during the 200 ms grating flashes and concatenating the two PSTHs for the two spatial phases of the grating into a single 400 ms PSTH. For all analyses of responses to gratings, we excluded cells with unreliable responses by calculating R^2 values between average response vectors of even and odd trials, similar to the analysis of natural-image responses. We created the response vector of a single trial by concatenating single-trial PSTHs from all different spatial frequencies and phases. We only considered cells with $R^2 > 0.1$ for our population analyses. The criterion was satisfied by 890 of 1126 cells recorded for this stimulus.

To estimate the grating spatial scale for each cell, we extracted the peak firing rate in the PSTH (across time and spatial phases) for each bar width (Krieger et al., 2017). We then fitted a logistic function (compare Natural image response predictions with a linear-nonlinear (LN) model) to the relationship of peak firing rate versus bar width, and extracted the function's midpoint as an estimate of the spatial scale. The amplitudes of harmonics of the PSTH were calculated by temporal Fourier transforms for each combination of spatial frequency and phase (Hochstein and Shapley, 1976). From the PSTHs of all spatial scales and phases, we extracted the maximum amplitude F1 at the stimulus frequency as well as the maximum amplitude F2 at twice the stimulus frequency and defined the nonlinearity index as the ratio of F2 over F1. This definition is slightly different from other approaches, where the F2/F1 ratio is calculated for each spatial scale and phase separately, with the maximum being chosen as the nonlinearity index (Hochstein and Shapley, 1976; Carcieri et al., 2003; Petrusca et al., 2007). Our approach aimed at capturing the maximum mean-luminance-induced modulation in F1 and the maximum spatial-contrast-induced modulation in F2.

Assessment of spatial input nonlinearities with checkerboard flashes. To assess how local visual signals are transformed in nonlinear cells, we used a stimulus that had a checkerboard layout with square tiles of either 105 or 120 µm to the side. The tiles were alternatingly assigned to two sets (A and B) so that neighboring tiles were in different sets. For each individual stimulus presentation, each set of tiles was assigned an intensity s_A or s_B, respectively, expressed as the Weber contrast from background illumination. Similar to our presentation of natural images, these checkerboard stimuli were flashed for 200 ms with an interstimulus interval of 800 ms, during which background illumination was presented. The contrast pairs (s_A, s_B) were selected from a two-dimensional stimulus space, organized in polar coordinates, by using 24 equidistant angles, each with 10 equidistant radial contrast values ($\sqrt{{s_A}^2+{s_B}^2}$) between 3% and 100%, and presented in pseudorandom order. The set of all contrast pairs was presented to the retina 4 or 5 times, with a different pseudorandomly permuted sequence chosen each time. We calculated cell responses by counting the number of spikes for each ganglion cell over a 250 ms window following stimulus onset and averaging over trials. Iso-response contour lines were constructed from the cells' response profiles using MATLAB's contour function. To exclude cells with unreliable responses, we calculated R^2 values across the set of all contrast combinations between spike counts averaged over even and over odd trial numbers. We only considered cells with $R^2 > 0.1$ for our population analyses. This criterion was satisfied by 833 of 1204 cells for which the stimulus was recorded.

Rectification (RI) and convexity (CI) indices were calculated for a specific contrast level c ($c = 0.6$ for most analyses). To quantify rectification of nonpreferred contrasts (RI), we compared the responses r^{half} under stimulation with only one spatial input (e.g., $s_A = c$ and $s_B = 0$, corresponding to a stimulus on one of the four half-axes of the stimulus space) to the responses r^{oppos} under stimulation with this input and the other spatial input at opposite contrast ($s_A = c$ and $s_B = -c$). In cases with no direct response measurement for a particular required contrast pair, we estimated the response based on the measured responses to nearby contrast pairs, using natural neighbor interpolation, as implemented in MATLAB's scatteredInterpolant function. From all response measurements, we subtracted the background spike count, measured as the response to the (0, 0) pair, which was included as a regular stimulus in the sequence of contrast pairs. To use a single definition of RI for ON, OFF, and ON-OFF cells, we considered all four half-axes in the stimulus space (with either s_A or s_B at either positive or negative contrast) and computed a weighted average from the four r^{half} values as well as from the corresponding r^{oppos} values (there are only two r^{oppos} values that are each used twice) to define RI as their ratio as follows:

$$\mathrm{RI} = \sum_{i=1}^{4} w_i r_i^{oppos} \Big/ \sum_{i=1}^{4} w_i r_i^{half}$$

where the weights w_i are measures of sensitivity along each half-axis i. Concretely, we obtained w_i as the slope of a regression line, fitted to the contrast-response pairs along the corresponding half-axis.

Similarly, for quantifying integration of preferred contrasts (CI), we compared the r^{half} values to responses r^{same}, which were measured with the same contrast for the two stimulus components, $s_A = s_B = c/2$, corresponding to the spatially homogeneous stimulus that has the same linearly integrated contrast as the stimulus used to measure r^{half}. Again, we took all four half-axes into account for defining CI as follows:

$$\mathrm{CI} = 1 - \sum_{i=1}^{4} w_i r_i^{same} \Big/ \sum_{i=1}^{4} w_i r_i^{half}$$

We subtracted the ratio of r^{same} over r^{half} from unity so that CI = 0 corresponds to linear integration and CI > 0 to $r^{same} < r^{half}$, and thus a convex, outward-bulging shape of the iso-response contour line. We used RI and CI to formally define homogeneity-sensitive cells as cells with RI < 0 and CI < 0, corresponding to iso-response contour lines curving toward the origin.

To probe spatial integration in the RF center with minimal surround influence, we used local checkerboard flashes. The stimulus was similar to the one above, but with small patches of 2 × 2 tiles. Tiles here had a side length of 105 μm, and patches thus had a side length of 210 μm. To compare our results with the full-field version of the stimulus, patch tiles were placed to align with the tiles of the full-field stimulus. The local patches were flashed for 200 ms, with no interval between successive presentations. For each individual presentation, the applied patch locations

were randomly chosen to maximally fill the screen (typical number of locations = 44-61, median = 53) while ensuring a minimum center-to-center distance of three patch side lengths (630 µm) for simultaneously presented patches (see Fig. 5C, bottom left). The rest of the screen was kept at background illumination (see Fig. 5C, top). For each presented patch, the contrast combination (s_A, s_B) was selected randomly and independently from the contrast combinations at other, simultaneously displayed locations. We applied fewer contrast combinations than for the full-field version of the stimulus to ensure adequate numbers of trials for each contrast combination at each location. Specifically, we used 8 or 12 equidistant angles in the stimulus space, each with 5 or 6 equidistant radial values between either 20%-100% or 3%-100%.

We also used the weights w_i to calculate an index of relative sensitivity for the two types of tiles (A and B). For each cell, we selected the pair of half-axes (of either positive or negative s_A and s_B values) with the highest average weight. The relative sensitivity index was calculated as follows:

$$\frac{w_B - w_A}{|w_B| + |w_A|}$$

with w_A and w_B being the weights of the half-axes in the selected pair. An index of zero indicates a balanced sensitivity to both types of tiles.

For analysis, we selected for each ganglion cell the patch closest to its RF center and extracted the responses to flashes when this particular patch was used. We counted the number of spikes over a 250 ms window following presentation onset and again subtracted the background activity, which was here obtained by interpolation to the (0, 0) contrast pair. Response contour lines in stimulus space as well as rectification and convexity indices were calculated in the same way as for the full-field version of the checkerboard flashes. Similarly to the full-field stimulus, we calculated R^2 values between the average spike counts of even and odd trials with respect to all contrast combinations, and only considered cells with $R^2 > 0.1$ for our population analyses. Additionally, we required that cells had a relative sensitivity index for the Tiles A and B with an absolute value <0.5. Both criteria were satisfied by 289 of 564 cells.

Spatial scale estimation from blurred natural images. For recordings with blurred natural images, we selected either 30 or 40 images from our set of natural images. The images were blurred by convolution with a two-dimensional, spherically symmetric Gaussian function. We used different σ values of the Gaussian to implement different spatial scales of blurring, defined as the diameter of the 2σ Gaussian contour (Schwartz et al., 2012), to also match our RF center definition. Blurred and original images were presented in a pseudorandom sequence, similar to the presentation of the large set of natural images described above, collecting 10 trials for each image and blurring scale. Responses were again measured for each ganglion cell by counting the number of spikes over a 250 ms window following stimulus onset.

We calculated R^2 values (see Natural image response predictions with a linear-nonlinear (LN) model) between blurred and original spike counts for each scale. We also calculated an R^2 value for the original image responses by considering odd- and even-trial averages and assigned this value to a blurring scale of 0 μm. We then fitted logistic functions to the R^2 values with respect to the blurring scales. We defined the natural spatial scale for each ganglion cell as the midpoint of the fitted logistic function. Again, by requiring $R^2 > 0.1$ for odd- versus even-trial averages of the original images, we included 747 cells of 850 for which we had recorded the stimulus.

Detection of image-recurrence-sensitive (IRS) cells. We detected IRS cells as described previously (Krishnamoorthy et al., 2017). Briefly, we presented a square-wave grating of either 240 or 270 µm spatial period and 60% contrast in a sequence of 800-ms-long fixations, separated by 100 ms transitions. During a transition, the grating was shifted by approximately two spatial periods to land in one of four equidistant fixation positions (corresponding to four specific spatial phases of the grating). The sequence of the four fixation positions was randomly chosen so that all 16 possible transitions (between starting and target positions) appeared several times in the stimulus sequence.

IRS cells are described as cells that show a strong response peak after onset of the new fixation when the grating position is the same as before the transition, but not when it has reversed contrast across the transition. To detect this, as done previously (Khani and Gollisch, 2017; Krishnamoorthy et al., 2017), we measured the response after each of the 16 possible transitions by creating PSTHs with bins of 10 ms and extracting the maximal difference for successive time bins in the PSTH as a measure of response increase (maximal derivative of the PSTH) in the window from 50 to 200 ms after fixation onset. We compared for each target grating i the maximal derivative D_i^{rec} under image recurrence (when the starting grating was also i) to the maximal derivative D_i^{change} when the starting grating was contrast-reversed compared with grating i. We calculated a recurrence sensitivity index (RSI) as $\mathrm{RSI} = \frac{1}{4}\sum_{i=1}^{4}\left(D_i^{rec} - D_i^{change}\right)/\left(D_i^{rec} + D_i^{change}\right)$. Cells with RSI > 0.7 and an average peak firing rate of at least 50 Hz in the post-transition PSTHs of the four image recurrences were considered as IRS cells.

Detection of direction-selective (DS) and orientation-selective (OS) cells. To identify DS ganglion cells, we used drifting sinusoidal gratings of 100% contrast, 240 µm spatial period, and a temporal frequency of 0.6 Hz (Sabbah et al., 2017). The gratings were shown in a sequence of eight equidistant directions with four temporal periods per direction, separated by 5 s of background illumination. The sequence was repeated 4 or 5 times. For each angle (θ), we collected the average spike responses (r_θ) during the presentation of the grating (excluding the first period). We calculated a direction selectivity index (DSI) as the magnitude of the normalized complex sum $\sum_\theta r_\theta e^{i\theta} / \sum_\theta r_\theta$ (Mazurek et al., 2014). The preferred direction was obtained as the argument of the same sum.

We also used drifting square-wave gratings of 100% contrast, 225 µm spatial period, and a temporal frequency of 4 Hz to identify OS ganglion cells (Nath and Schwartz, 2016, 2017). The gratings were shown in a sequence of eight equidistant directions with 12 periods per direction, separated by 2 s of background illumination. The sequence was repeated 4 or 5 times. We calculated an orientation selectivity index (OSI) as the magnitude of the complex sum $\sum_\theta r_\theta e^{i2\theta} / \sum_\theta r_\theta$. The preferred orientation was obtained as the line perpendicular to half the argument of the same sum.

To calculate the statistical significance for both indices, we used a Monte Carlo permutation approach (Liu et al., 2017). For a given cell, we repeatedly shuffled the responses over all angles and trials 2000 times to obtain a distribution of DSI (or OSI) values under the null hypothesis that the firing rates are independent of the motion direction (or orientation). All cells with DSI > 0.25 (significant at 1% level) were considered as DS cells. Similarly, OS cells were identified as cells with OSI > 0.25 (significant at 1% level) that were not DS or IRS. We only included cells with a total mean firing rate >1 Hz during the presentation of the drifting gratings (Kühn and Gollisch, 2016).

OS cells were classified as either ON- or OFF-type based on the sign of the first peak (i.e., closest to zero) in the fitted temporal component $k_T(t)$. Here we disregarded a peak if its amplitude (unsigned) was <25% of the largest deflection.

Data and code availability. The spike-time data used in this study and sample code for stimulus reconstruction are available at https://gin.g-node.org/gollischlab/Karamanlis_Gollisch_2021_RGC_spiketrains_natural_image_encoding.

Results

Performance of LN models for predicting responses to natural images varies strongly among RGCs

Since our goal was to assess the role of spatial nonlinearities in the RF, we focused on stimuli that have natural spatial structure, but simplified temporal dynamics. We therefore stimulated the retina with briefly flashed achromatic natural images while recording the spiking activity of several hundred mouse RGCs with multielectrode arrays, to survey whether linear RF models

could capture the cells' responses. The images had been collected from three different databases (van Hateren and van der Schaaf, 1998; Olmos and Kingdom, 2004; Arbeláez et al., 2011) and were presented for 200 ms each, separated by 800 ms of background illumination (Fig. 1*A*). Flash duration was close to the typical fixation duration in "saccade-and-fixate" gaze patterns observed in freely moving mice (Meyer et al., 2020; Michaiel et al., 2020). To analyze ganglion cell responses in relation to the signal inside the RF, we determined the RFs (including center and surround) from responses to spatiotemporal white noise (Fig. 1*B*). Different cells sampled different parts of the images and displayed a variety of response patterns (Fig. 1*C*), with apparent sensitivity to positive or negative Weber contrast. Some ganglion cells responded to both stimulus onset and offset for some images (Fig. 1*D*), which may indicate ON-OFF-type RFs (Jacoby and Schwartz, 2017) or spatially nonlinear RFs (Mani and Schwartz, 2017). Furthermore, we observed both transient and sustained responses as well as response suppression (Fig. 1*C*, bottom left).

To test whether these diverse ganglion cell responses could originate from a spatially linear RF, we measured how well a simple linear RF model could reproduce such responses. To do so, we quantified a cell's response for each image by the average spike count over 250 ms following image onset. We then aimed at predicting this spike count with an LN model (Fig. 1*E*). The model's first stage is a linear spatial filter, which was estimated from the STA under white-noise stimulation by a parametric fit that contained a difference-of-Gaussians as the spatial component. The filter captured the location, size, shape, and relative surround contribution of the spatial RF. Applying the filter to the pixelwise Weber contrast values of a given image yielded a linear prediction: a single number that corresponded to the image's net Weber contrast as seen through the cell's RF. It quantifies how much the mean light level over the RF changed between background illumination and image presentation. The model then predicted the average spike count to the image by transforming this linear prediction with a parameterized nonlinear function, the model's nonlinearity. The nonlinearity was obtained by selecting a number of images (training set) and fitting a generic function to the relation between the linear predictions and the measured responses. The obtained nonlinearity was then used to compare predictions with actual responses for the remaining images (test set), using cross-validation to quantify prediction accuracy.

For cells that linearly integrate over space, the linear prediction of the LN model should be tightly coupled to the response strength, and the relationship between the two is

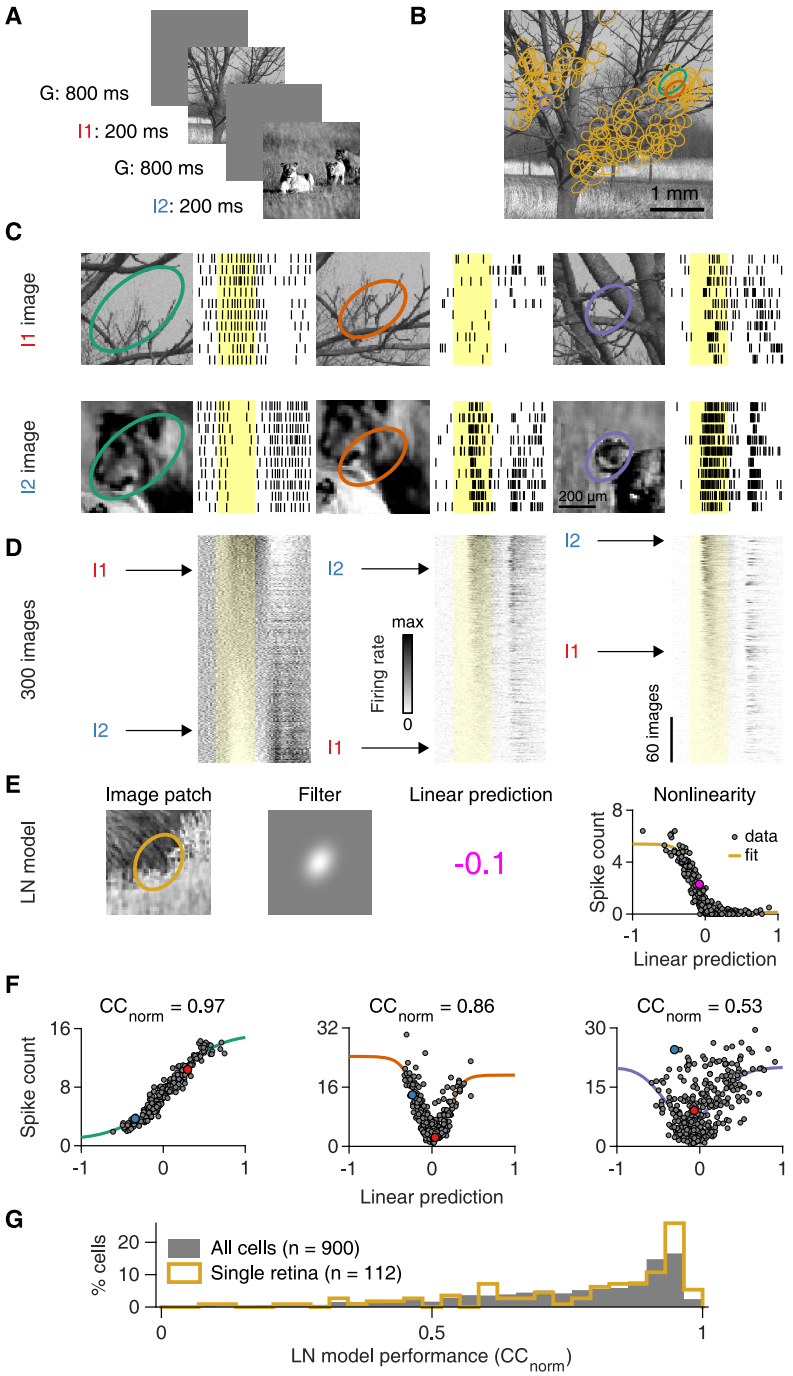

Figure 1. A spatially linear RF model often fails to predict natural image responses of RGCs. ***A***, Natural images were presented to the retina in a pseudorandom sequence for 200 ms with an interstimulus interval of 800 ms. ***B***, Sample natural image (I1). Overlaid ellipses (light orange) represent the outlines of RF centers (center parts of difference-of-Gaussians fits) of 130 RGCs from a single recording. The midline is RF-free because it contained the optic disk region. ***C***, Top, Raster plots with responses of three different ganglion cells to 10 presentations of image I1. Different RF outline colors correspond to different cells, also highlighted in ***B***. Bottom, Same as in top, but for presentations of another image (I2). Yellow-shaded areas correspond to the 200 ms image presentations. ***D***, PSTHs for 300 natural images, aligned to the raster plots of (***C***) and sorted by the average spike count during stimulus presentation. Rows corresponding to images I1 and I2 are marked. ***E***, The structure of an LN model that we used to predict average spike counts for natural images. The linear prediction is the inner product of the contrast values in the image patch and the filter. ***F***, Spiking nonlinearities fitted to observed spike counts for the three sample cells of ***C***. Data points for images I1 and I2 are highlighted. Top, The obtained normalized correlation coefficients (CC_{norm}). ***G***, LN model performance distribution for ganglion cells in a single retina preparation (light orange, same as in ***B***), and for all recorded cells (gray) from 13 preparations (9 animals).

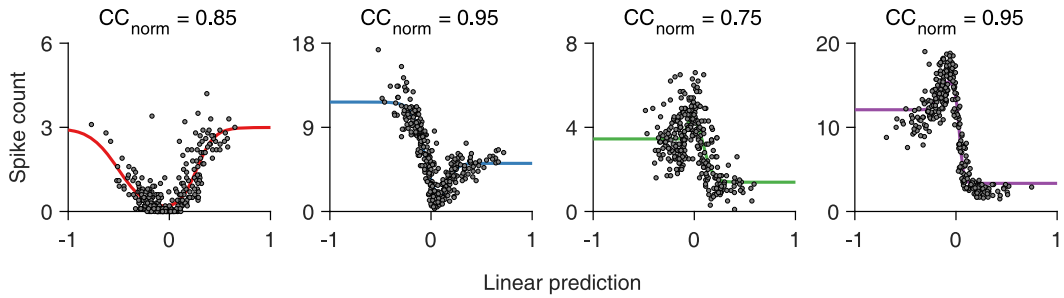

Figure 2. Examples of U- and bell-shaped nonlinearities. Four sample cells, with either ON-OFF-type nonlinearities (the two leftmost) or suppressed-by-contrast-type nonlinearities (the two on the right).

effectively a contrast-response function. We found cells, for example, for which the linear predictions displayed a clear, monotonic relationship to the responses, such as in Figure 1*E* (right, rank correlation between measured spike counts and corresponding model predictions: Spearman's $\rho = -0.88$, $n = 300$ images) or in Figure 1*F* (left, Spearman's $\rho = 0.97$, $n = 300$ images). As the spatial filter is defined to always have a positive central peak, an increasing monotonic relationship indicates a contrast-response function of an ON-type ganglion cell (Fig. 1*F*, left), whereas a decreasing one indicates a contrast-response function of an OFF-type ganglion cell (Fig. 1*E*). For such monotonic relationships, simple logistic nonlinearities provided good fits. Yet, we also found cells with a U-shaped relationship between linear predictions and responses (Fig. 1*F*, middle). To also capture such a nonmonotonic contrast-response function shape, we applied a bi-logistic nonlinearity, fitted to the contrast-response function of each cell. The bi-logistic functions captured nonmonotonic nonlinearities by combining an increasing and a decreasing logistic function, but also worked well for monotonic contrast-response relations, as the weight of one logistic component then naturally assumed a value near zero in the fit. Nonmonotonic contrast-response functions are expected to occur in the retina for ON-OFF (Burkhardt et al., 1998) or suppressed-by-contrast ganglion cells (Levick, 1967; Jacoby et al., 2015; Tien et al., 2015), and we indeed observed both cases as indicated by U- and bell-shaped functions, respectively (Fig. 2). The flexible parameterization of the nonlinearity allowed us to assess LN model performance and thus spatial nonlinearities for these cells in the same way as for pure ON and OFF cells. Finally, we found cells with no apparent relationship between linear predictions and responses (Fig. 1*F*, right, and Fig. 3*A*). For such cells, fits were poor because of the spread of data points, indicating that the LN model failed to predict responses to natural images.

How well the LN model captures the responses can be visually assessed by how tightly the data points cluster around the fitted nonlinearities and quantified by how strongly prediction and response are correlated. However, part of the deviation from the fit could result from noise in the response measure, as only 10 trials per image were available, rather than from an actual failure of the model. Thus, to quantify performance of the LN model, we computed a normalized correlation between response prediction and measured response, CC_{norm} (Schoppe et al., 2016), which takes the variability of responses across trials into account by assessing the model prediction relative to the reliability of the trial-averaged responses. Furthermore, we used cross-validation by averaging CC_{norm} over 10 different sets of held-out images not used to fit the nonlinearity.

Model performance varied considerably between cells. A sizeable proportion showed good model performance, indicated by a peak close to unity in the distribution of CC_{norm} values (Fig. 1*G*). On the other hand, we observed a broad tail of cells with low CC_{norm} values, indicating different degrees of model failure, both for individual retina pieces as well as for the entire population of recorded cells. Given the variability-adjusted measure of model performance and the flexibility of the applied nonlinearity, we hypothesized that the nonlinear part of the LN model was not the source of the observed diversity in natural image encoding. We therefore focused on investigating the relation between model performance and spatial signal integration.

Linear RF model performance correlates with SC sensitivity in the RF center

Figure 3*A* displays measured spike counts versus model predictions for a sample cell with low model performance. The model failure is apparent from the fact that the cell elicited widely different spike counts for images that yielded similar linear predictions of the model, corresponding to similar net contrast over the RF, and thus similar spike count predictions. The two images shown in Figure 3*B*, for example, had nearly identical linear predictions for the sample cell, but the cell clearly responded differently to the two images. These two images strikingly differed in their spatial structures inside the cell's RF center (Fig. 3*B*). We therefore quantified the spatial structure of each image within the center of a cell's RF by computing the spatial contrast (SC, see Materials and Methods), which measures the variability of image pixels inside the RF center.

To evaluate the impact of SC on the spike output for a given cell, we grouped the images into pairs of similar linear predictions by the cell's LN model. This allowed us to relate differences in spike count within a pair to differences in SC while minimizing confounding effects of mean light-level changes inside the RF. The analysis revealed that SC was systematically related to spike count for many cells, with more spikes elicited when SC was larger (Fig. 3*C*). Indeed, for the majority of cells (72%, $n = 651$ of 898 recorded cells), differences in SC and spike count were positively correlated, indicating that SC had a response-boosting effect beyond mean light level and that spatial integration was nonlinear.

Other cells (22%, 202 of 898) appeared insensitive to SC, as indicated by an approximately flat relationship between differences in SC and spike count and no significant correlation (Fig. 3*D*). This was expected as the LN model, which is based solely on mean light level in the RF, did provide an accurate description of spike counts for some RGCs.

Unexpectedly, however, we also found a small subset of cells (5%, 45 of 898) that responded vigorously to stimuli with spatially homogeneous illumination of preferred contrast, but displayed smaller spike counts for images with similar mean

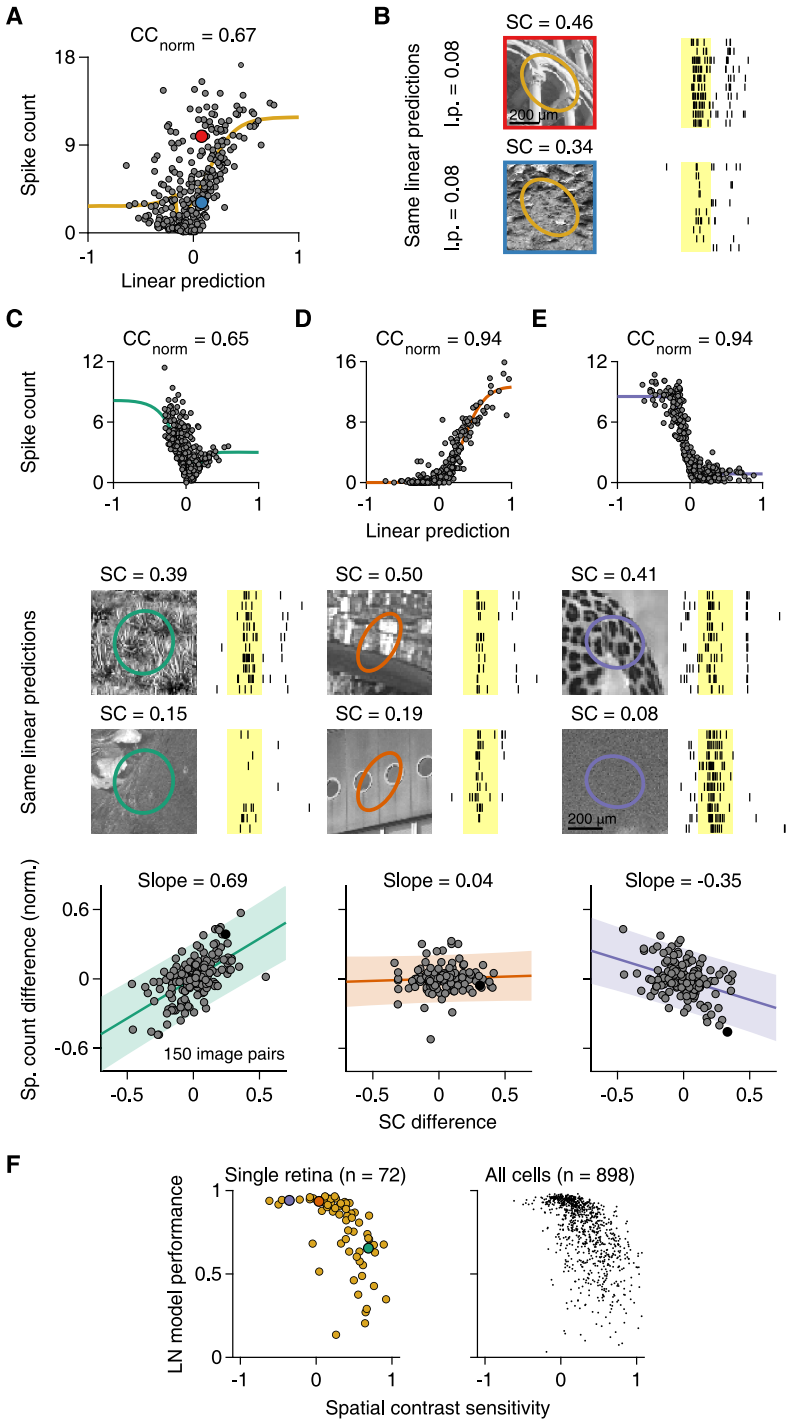

Figure 3. Sensitivity to natural SC is correlated with LN model performance. ***A***, Output nonlinearity fit for a cell with low LN model performance. Marked data points correspond to the responses shown in ***B***. ***B***, Different responses of the cell in ***A*** to natural images with approximately same linear predictions (l.p.), but different SC in the RF center. ***C***, Top, Output nonlinearity fit for another sample cell. Middle, Raster plots of the cell's responses to natural images with approximately same l.p., but with high (top) or low (bottom) SC in the RF center. Bottom, Relation of SC differences to average spike count differences for 150 pairs of natural images with similar l.p. in the RF center. Count differences are normalized to the maximum observed average spike count. Filled black data point represents the difference obtained from the pair of sample images above. Line and slope value correspond to least-squares estimate. Shaded area represents 95% confidence interval. ***D***, ***E***, Same as in ***C***, but for two other sample cells. ***B–E***, Shaded yellow areas represent the 200 ms image presentations. ***F***, Relation of LN model performance to SC sensitivity, defined as the slope of the relation between spike-count differences and SC differences, as in ***C–E***, for ganglion cells in a single preparation (left; Spearman's $\rho = -0.73$, $p < 10^{-3}$) and for all recorded cells (right; 13 retinas, 9 animals). ***C–E***, Data points for the sample cells are highlighted in the corresponding colors.

illumination and higher SC (Fig. 3*E*). Such inverse sensitivity to SC represents a different form of nonlinear spatial integration than the response-boosting effect of SC in the majority of cells and may be described as sensitivity to spatially homogeneous stimulation. However, despite the inverse sensitivity to SC, the response characteristics of these cells differ from those of suppressed-by-contrast cells because the preference for homogeneous stimuli does not extend to the temporal domain. Unlike for suppressed-by-contrast cells, temporal contrast at image onset can strongly activate the cells described here (Fig. 3*E*).

To assess whether sensitivity to SC was systematically related to LN model performance, we quantified the "SC sensitivity" of a given cell by the slope of the regression line between SC and response differences, normalized by the cell's maximum response. We found that SC sensitivity was indeed negatively correlated with LN model performance in individual experiments (e.g., Fig. 3*F*, left; median Spearman's $\rho = -0.60$, 10 of 13 had $p < 0.05$) as well as in the pooled data (Fig. 3*F*, right; Spearman's $\rho = -0.64$, $p < 10^{-3}$, $n = 898$ cells). Cells for which SC boosted activity (corresponding to large positive values of SC sensitivity) were generally not as well described by the LN model. This suggests that model performance is indeed limited by a systematic influence of SC on spike count in many cells. For cells with no detectable sensitivity to SC, on the other hand, model performance was generally good (CC_{norm} median = 0.91, $n = 202$). Also, the few cells with a suppressive effect of SC (negative SC sensitivity) showed fairly good LN model performance, despite the observed deviation from linear spatial integration.

Sensitivity to fine spatial gratings alone does not predict LN model performance

Sensitivity to spatial structure on a sub-RF scale is characteristic for nonlinear RFs. A classical test for nonlinear spatial integration is to stimulate the retina with full-field contrast-reversing gratings at different spatial scales and phases (Hochstein and Shapley, 1976; Demb et al., 1999). Applying such stimuli in our recordings, we found RGCs that clearly responded to the reversals of fine (30 µm bar width) gratings (e.g., Fig. 4*A*,*B*), revealing nonlinear spatial integration under reversing gratings, similar to previous measurements in the mouse retina with single-cell recordings (Schwartz et al., 2012; Tien et al., 2015; Krieger et al., 2017). These cells also responded to coarser gratings

←

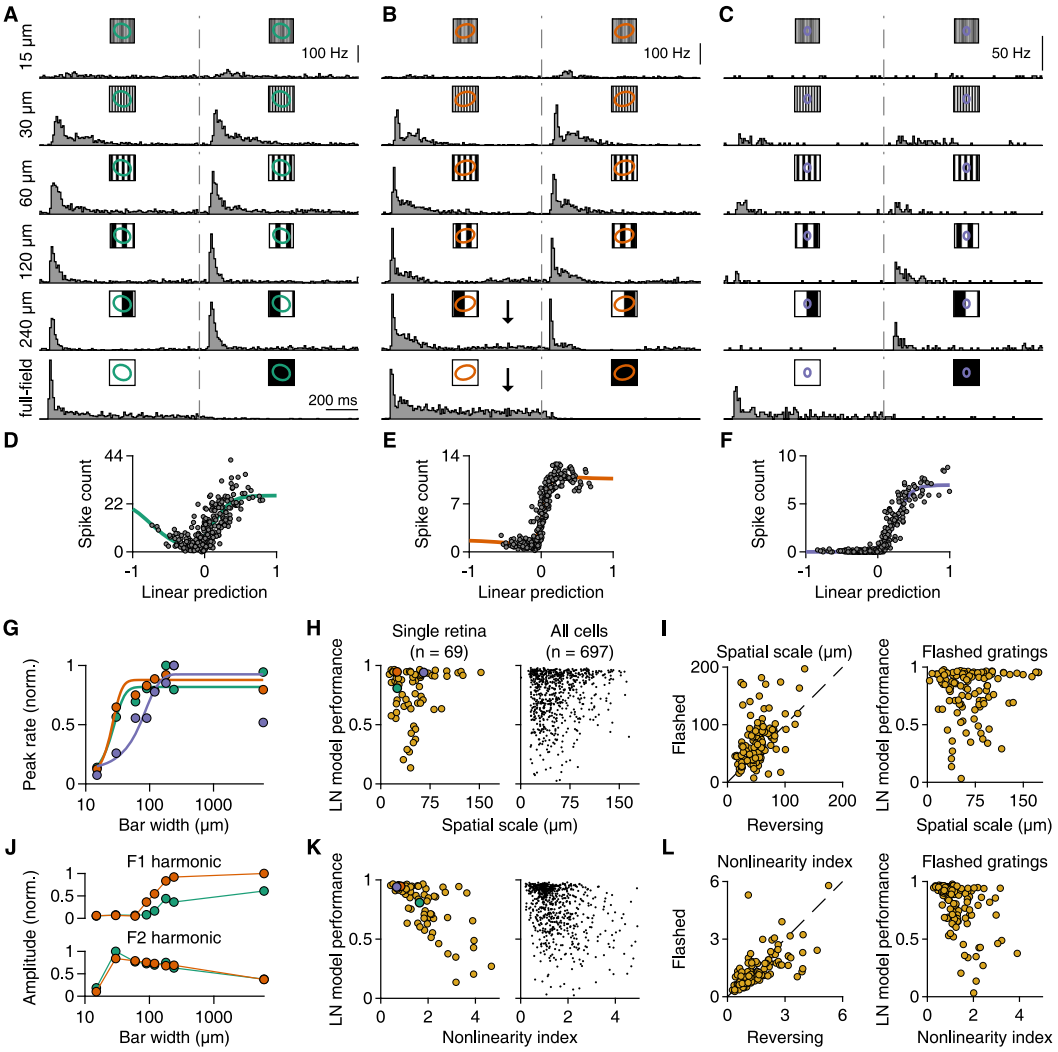

Figure 4. Relation of responses under reversing gratings to LN model performance. ***A–C***, PSTHs of three ganglion cells to full-field contrast-reversing gratings of six different bar widths, which are indicated to the left of the plots. For wider bars, the gratings were presented for multiple spatial phases, and the displayed PSTHs represent the phase with the smallest net brightness changes averaged over the RF. Dashed gray lines indicate the time of contrast reversal. ***B***, Arrows indicate the sustained component of the responses. ***D–F***, Relationship between linear prediction and average spike count of the same three cells for 300 natural images. Solid lines indicate fitted nonlinearities. ***G***, Relation of bar width to normalized peak firing rates (across time and spatial phases) for the cells in ***A–C***. Colors correspond to the RF colors of ***A–C***. Solid lines indicate logistic fits. ***H***, Relation of LN model performance to the spatial scale of each cell for a single retinal preparation (left, Spearman's $\rho = 0.01$, $p = 0.95$) and the total population (right) from 12 retinas (9 animals). Cells from ***A–C*** are highlighted. ***I***, Left, Comparison of spatial scales obtained from reversing or flashed gratings ($n = 126$). Right, Relationship between LN model performance for natural images and spatial scales from flashed gratings ($n = 126$). Data are from 3 experiments (2 animals). ***J***, Normalized F1 and F2 amplitudes (maximum over spatial phases) for the cells of ***A*** and ***B***. ***K***, Relation of LN model performance to the nonlinearity index of each cell for a single retinal preparation (left, Spearman's $\rho = -0.7$, $p < 10^{-3}$) and the total population (right) from 12 retinas (9 animals). The nonlinearity index is defined as the maximum F2 (across spatial frequencies) over the maximum F1 (across spatial frequencies) amplitude. Cells from ***A–C*** are highlighted. ***L***, Same as in ***I***, but for nonlinearity indices.

that split their RF centers in two halves. Other cells, however, barely responded to reversals of gratings for phases with zero net contrast across the RF, such as the sample cell of Figure 4*C*.

Interestingly, sensitivity to contrast reversals of gratings often seemed unrelated to LN model performance for natural images. One of the sample cells with clear responses to fine-scale gratings (Fig. 4*A*) had poor LN model performance for natural images (Fig. 4*D*), whereas the other (Fig. 4*B*) showed good model performance (Fig. 4*E*). For the third sample cell (Fig. 4*C*), model performance was good (Fig. 4*F*), consistent with the observed insensitivity to grating reversals, which suggests linear spatial integration.

In order to systematically compare the sensitivity to reversals of fine gratings with the LN model performance across multiple ganglion cells, we extracted two measures from a cell's responses to the reversing gratings. First, to assess the spatial scale at which a cell becomes sensitive to the grating as revealed by a sizeable response peak (Krieger et al., 2017), we examined the cell's peak firing rates for different grating bar widths, fitted this relationship with a logistic curve, and used the curve's midpoint as a measure of spatial scale (Fig. 4*G*,*H*). Second, to assess how nonlinear spatial integration contributes to the overall strength of the response at different spatial frequencies, we compared the response Fourier components for the stimulus frequency (F1) and for twice that frequency (F2, frequency-doubled component, corresponding to responses for both reversal directions). Large F2 amplitudes, compared with F1, are indicative of nonlinear spatial-integration effects (Hochstein and Shapley, 1976) at the

level of spike counts. We then computed a nonlinearity index as the ratio of the maximal F2 amplitude (over grating widths and phases) and the maximal F1 amplitude (see Materials and Methods).

This analysis showed that the spatial scale was rarely correlated to the LN model performance in individual experiments (median Spearman's $\rho = 0.15$, 2 of 12 experiments had $p < 0.05$; example in Fig. 4*H*, left); and for the entire dataset, this correlation was weak, albeit significant (Spearman's $\rho = 0.11$, $p = 0.002$, $n = 697$; Fig. 4*H*, right). Thus, sensitivity to reversals of high spatial frequency gratings, typically taken as a sign for nonlinear spatial integration, does not generally imply failure of the LN model. Indeed, many cells that start responding already for fairly fine spatial gratings (small spatial scale) showed remarkably good model performance as illustrated by the example in Figure 4*E*.

By contrast, the relative amplitudes of the F1 and F2 response components predicted model performance much better. The nonlinearity index that was computed from their ratio was negatively correlated to LN model performance under natural images both in single experiments (median Spearman's $\rho = -0.37$, 8 of 12 experiments had $p < 0.05$; example in Fig. 4*K*, left) as well as in the whole population (Spearman's $\rho = -0.34$, $p < 10^{-3}$, $n = 697$; Fig. 4*K*, right). Thus, the relative degree of nonlinear spatial integration as measured by the F2 response component is a better indicator of the importance of nonlinear spatial integration under natural images than the mere sensitivity to spatial gratings.

The responses of the sample cells in Figure 4 illustrate this difference between the sensitivity to fine spatial gratings and relative size of nonlinear response components. The cells of Figure 4*A*, *B* were both sensitive already to reversing gratings of bar widths of 30 µm (Fig. 4*G*), indicative of nonlinear RFs. Yet, although initial response peaks might be similar, leading to similar F2 response components for the two cells (Fig. 4*J*, bottom), responses for the second cell were more sustained with higher spike count when net-coverage of the RF with preferred contrast was larger (see Fig. 4*B*, arrows). Thus, the responses of this cell contain also a considerable linear component even for fairly fine spatial gratings, as reflected by a higher F1 response component (Fig. 4*J*, top). The resulting lower nonlinearity index matches the better performance of the LN model for this cell. Although the linear response component may not stand out in the response patterns under reversing gratings, it may dominate the spike count responses under natural images, which contain relatively larger mean luminance signals because of the abundance of power in low spatial frequencies. Thus, even cells with clear sensitivity to fine spatial gratings and a large F2 response component under reversals may display relatively good LN model performance.

While reversing gratings are a typical stimulus used to test for spatial nonlinearities, they differ from the flashed natural images not only in their spatial structure, but also in their temporal dynamics. This might contribute to the differences observed between responses to gratings and to images. To test this, we therefore also applied flash-like presentations of gratings in some of our recordings to provide a comparable stimulation time course as for the natural images. We found that results were quite similar to those obtained with contrast-reversing gratings and led to the same conclusions. In particular, spatial scales and nonlinearity indices were correlated between the two grating versions (Spearman's $\rho = 0.50$, $p < 10^{-3}$, for spatial scales, Fig. 4*I*, left; and $\rho = 0.74$, $p < 10^{-3}$ for nonlinearity indices, Fig. 4*L*, left). Furthermore, similar to reversing gratings, spatial scales from flashed gratings were not informative about LN model performance, displaying no significant correlation (Spearman's $\rho = 0.15$, $p = 0.11$, Fig. 4*I*, right), whereas nonlinearity indices were negatively correlated to LN model performance (Spearman's $\rho = -0.55$, $p < 10^{-3}$, Fig. 4*L*, right). This suggests that it was indeed the different spatial structure between natural images and gratings and not their temporal profiles that led to different nonlinear characteristics of some cells under these two stimulus types.

Although we found that LN model performance under natural images and the nonlinearity index from gratings are correlated, there is considerable remaining variability across cells around this relation, potentially stemming from drawbacks of the classical analysis with contrast-reversing gratings. First, for ON-OFF cells, the analysis cannot distinguish between nonlinear integration over space or over ON-type versus OFF-type inputs, as both phenomena can lead to large F2 components. Second, the analysis primarily detects that some rectification of nonpreferred contrasts exists (as effects of preferred and nonpreferred contrast do not cancel out), but is not fully determined by the degree of rectification and does not provide information about how contrast signals at different locations inside the RF are combined.

Responses to contrast combinations inside the RF reveal the components of natural SC sensitivity

To overcome the shortcomings of classical contrast-reversing grating stimulation and explore the relationship between SC sensitivity and LN model performance more systematically, we designed a stimulus that tests a range of contrast combinations by flashing checkerboards on the retina with different light intensities for the two sets of alternating checkerboard tiles. The idea is to independently stimulate two separate sets of spatial subunits within a cell's RF with different inputs (Bölinger and Gollisch, 2012; Takeshita and Gollisch, 2014). This allows comparing responses at different contrast levels of spatially homogeneous stimulation, stimulation of only one spatial stimulus component, or stimulation with opposite contrast of the two spatial components.

Concretely, we applied a batch of varied checkerboards (Fig. 5*A*, top), whose Contrasts A and B for the two sets of tiles, or spatial inputs, systematically covered the stimulus space of pairs of contrast values (Fig. 5*A*, bottom right) to explore a wide range of contrast combinations. To directly compare responses between artificial and natural stimuli, we flashed the contrast pairs for 200 ms each (the same duration as for the natural images) in a pseudorandom sequence, collecting 4 or 5 trials per pair. The subfields of the checkerboard spanned 105 or 120 μm to the side, approximately half of the average mouse RF center, to provide a strong, yet spatially structured stimulus inside the RF.

To visualize the responses for different contrast combinations, we extracted the average spike counts over 250 ms after stimulus onset, equivalent to the response measure under natural images, and displayed them as color maps over the stimulus space of contrast pairs (Fig. 5*B*, middle row). We then calculated iso-response contour lines (Fig. 5*B*, bottom row), which trace out those contrast pairs that led to the same response (here number of spikes). The shape of the iso-response contours can reveal whether stimulus integration is linear or nonlinear and is indicative of the type of subunit nonlinearity (Bölinger and Gollisch, 2012; Maheswaranathan et al., 2018). Notably, the contours are independent of any output nonlinearities that transform responses after stimulus integration has taken place, such as thresholding and saturation in the spike generation process

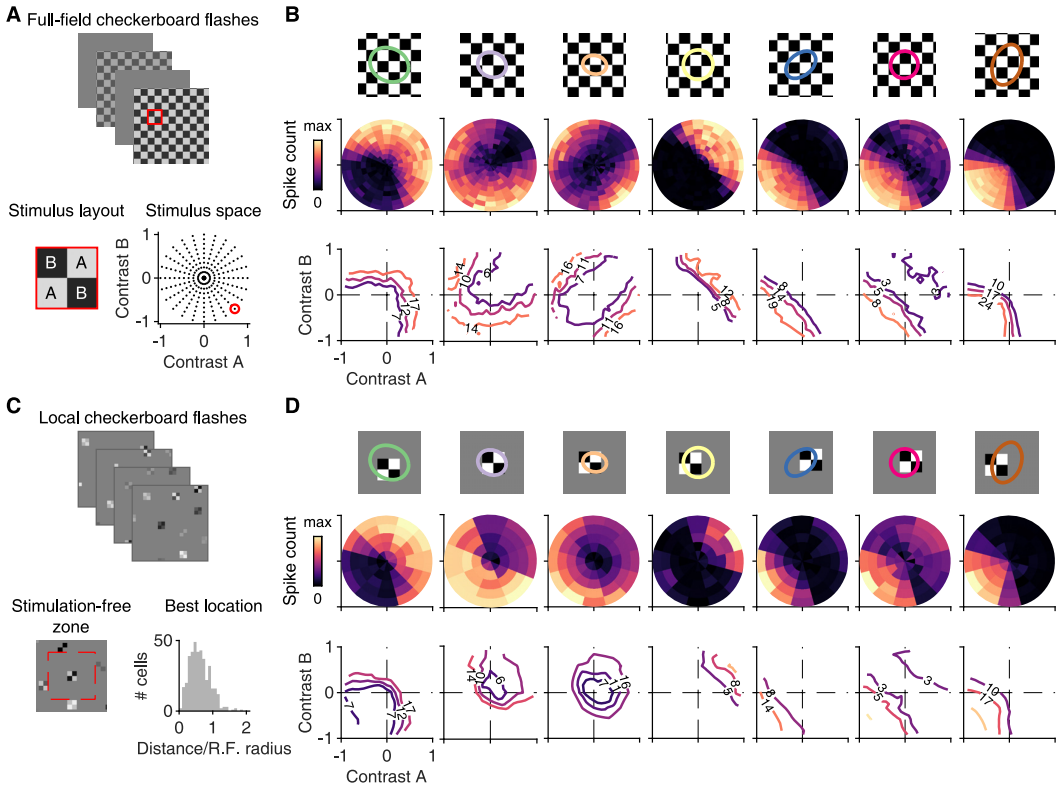

Figure 5. Stimulation with contrast combinations reveals nonlinearities in spatial input integration. ***A***, Depiction of the applied stimulus space, comprising flashed checkerboards with contrast combinations (A, B) sampled along different directions in stimulus space. Bottom right, Dots in the stimulus space represent all contrast combinations applied in the experiment. The contrast combination identified by the red circle corresponds to the example on the left, which shows a 2×2 cutout from the stimulus frame shown on top (marked by the red square). ***B***, Top, RF center outlines of seven sample ganglion cells, relative to the stimulus layout, with tiles for Contrast A and B shown in white and black, respectively. Middle, Color-coded average spike counts for all tested contrast pairs in the stimulus space. Bottom, Iso-response contours in the stimulus space for three selected spike counts (at 30%, 50%, and 70% of the maximum spike count), indicated by the number on the contour and the contour's color. Contour shapes are largely invariant to the selected response level. ***C***, Top, Example frames of the locally sparse stimulus. Bottom, Display of the region (dashed red line) that was excluded from further selection of stimulus locations around an already selected location (left). Distribution of distances from the RF midpoint to the center point of the closest grid square ($n = 404$), normalized by the RF radius. ***D***, Same as in ***B***, but for the locally sparse stimulus. Contour lines are shown for the same spike counts as in ***B***.

(Gollisch and Herz, 2012). Linear integration of the two inputs, for example, leads to straight contour lines, independent of any subsequent nonlinear transformation of the summed inputs. Curved iso-response contours, on the other hand, reflect nonlinear stimulus integration, and their shape can provide information about the type of nonlinearity, as discussed below.

The iso-response analysis under flashed checkerboards revealed a variety of spatial integration profiles among different RGCs. We found both ON and OFF varieties of nonlinear cells with contour lines curving convexly around the origin (Fig. 5*B*, Cells 1 and 2). This nonlinear signature may result from an expansive transformation of local signals, such as by a threshold-quadratic function (Bölinger and Gollisch, 2012) or by a sigmoid with high threshold (Maheswaranathan et al., 2018). Quadratic integration of inputs A and B, for example, leads to circular (or elliptic) parts of the contour lines, as $A^2 + B^2 = \text{const}$ is the circle equation. Furthermore, contours that run parallel to the axes in the quadrants where the stimulus components have opposite sign indicate a rectifying threshold, as one of the two input components can apparently vary without changing the response level.

We also found linear ON and OFF ganglion cells, as identified by their straight contour lines (Fig. 5*B*, Cells 4 and 5). Furthermore, our approach allowed us to visualize the spatial integration profiles of ON-OFF cells, and distinguish between spatially nonlinear and linear ON-OFF cells (Fig. 5*B*, Cells 3 and 6). Linear ON-OFF cells responded mostly to net-increases or decreases of light intensity, but not when the two contrast signals cancelled each other, leading to straight, parallel contour lines (Cell 6). On the other hand, nonlinear ON-OFF cells often had closed or nearly closed contour lines, corresponding to strong responses also for contrast combinations with opposing signs (Cell 3). Finally, we identified a unique nonlinear spatial integration profile in some cells, characterized by contour lines curving concavely away from the origin, coming closest to the origin on the diagonal of equal contrast for A and B (Fig. 5*B*, Cell 7). Such a profile indicates a particular preference to a spatially homogeneous change in light level, as a given response level can be reached with comparatively little contrast when both spatial components are stimulated in unison. We mainly found such profiles for OFF-type ganglion cells, but occasionally in ON-type cells as well (6 of 27 cells were ON-type). Cells with similar preference for homogeneous illumination of the RF have previously been observed in the salamander retina (Bölinger and Gollisch, 2012).

For comparison, we also devised a local version of checkerboard flashes to assess potential contributions of the RF surround to nonlinear spatial integration. Here, the display of each

contrast combination was spatially restricted to a patch of 2×2 tiles of the checkerboard, which roughly corresponds to typical RF center sizes. To nonetheless cover the entire recording area and obtain sufficient sampling of contrast combinations, multiple randomly chosen patches, obeying local sparsity (Hawrylycz et al., 2016; de Vries et al., 2020), were displayed simultaneously (Fig. 5*C*), and fewer contrast combinations were sampled compared with the full-field version of the stimulus. For further analysis, we selected for each cell the patch location closest to the RF center. This generally lay not further away than one RF radius (Fig. 5*C*, bottom right), indicating good overlap of the analyzed patch location with the RF center. Furthermore, the stimulus patch did not need to fill the RF center to trigger robust responses. Also, precise centering on the RF was not required to make the two stimulus components similarly effective. If, say, a tile of Component A was closer to the RF midpoint than the other three tiles and thus more effective in influencing the response, this was approximately balanced by the second tile of Component A being further away from the midpoint than the two tiles of Component B.

Using this local version of the flashed checkerboards, we found that spatial integration profiles, as captured by the shape of the contour lines in stimulus space, were qualitatively similar under local stimulation compared with full-field stimulation (Fig. 5*D*). This indicates that it is the nonlinear stimulus integration in the RF center that determines the shape of the contour lines. As the examples show, this shape can deviate from straight lines in different ways. Rectification of nonpreferred inputs, for example, becomes visible by how the contour line bends as it progresses from the quadrant in stimulus space that corresponds to preferred contrast for both stimulus components (top right quadrant for ON cells; bottom left for OFF cells) to the two neighboring quadrants that combine positive and negative contrast. In addition, there is also nonlinear integration of preferred contrast, which is visible in a nonlinear shape of the contour line inside the quadrant that corresponds to preferred contrast of both stimulus components.

To quantify these nonlinear signatures, we devised two corresponding indices as explained in Figure 6. We calculated a rectification index (RI, Fig. 6*A*) by comparing responses to flashes where both components had opposing, equal-magnitude contrast (here 60%; for comparison with different contrast levels, see Fig. 6*B*,*C*) with responses when only a single stimulus component was used (Molnar et al., 2009). Full rectification leads to equal responses for both configurations and an index of unity, whereas linear integration would make the opposing-contrast configuration effectively a null stimulus, resulting in no response and an index of zero. Similarly, we computed a convexity index (CI, Fig. 6*D*) by comparing responses from using just one spatial input at a specific contrast level (again 60%; compare Fig. 6*E*,*F*) with responses from using both inputs at half

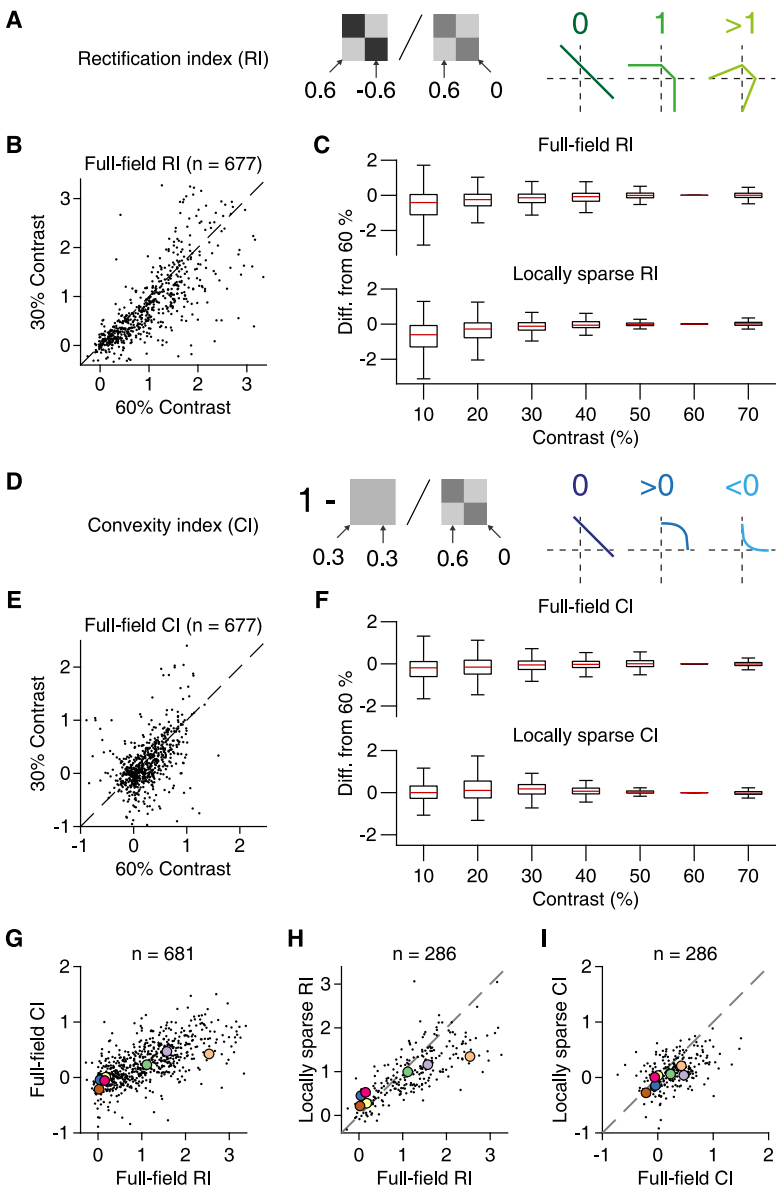

Figure 6. Rectification and convexity indices quantify iso-response contour shapes across contrast levels. ***A***, Schematic depiction of how rectification indices were computed from responses to different contrast combinations and how they relate to different shapes of contour lines. As indicated, the RI resulted from a ratio of responses when the tiles had opposing contrast (here 0.6 and −0.6) and when only one component was used (0.6 and 0). ***B***, Comparison of full-field rectification indices calculated for 30% and 60% contrast. ***C***, Rectification indices for different contrast levels and presentation modes (top: full-field, $n = 695$; bottom: local, $n = 289$). To compare with our default of 60%, values are shown as differences from the default. Each box represents the interquartile range (25th and 75th percentile), along with the median (red line). The whiskers extend to the maximal and minimal values, with outliers excluded (defined as data points >1.5 times the interquartile range away from the box). ***D***, Same as in ***A***, but for the CI. The CI was computed by comparing the responses when stimulation was homogeneous (here both contrast values at 0.3) and when only one component was used (contrast values 0.6 and 0) and subtracting the ratio of these responses from unity. ***E***, ***F***, Same as in ***B***, ***C***, but for the convexity indices. ***G***, Relation of full-field RI and CI in the pooled ganglion cell data from 13 retinas (9 animals). Colored dots correspond to the sample cells from Figure 5. ***H***, Relation of full-field and locally sparse rectification indices in the pooled ganglion cell data from 6 retinas (4 animals). Dashed line indicates the equality line. ***I***, Same as in ***H***, but for convexity indices.

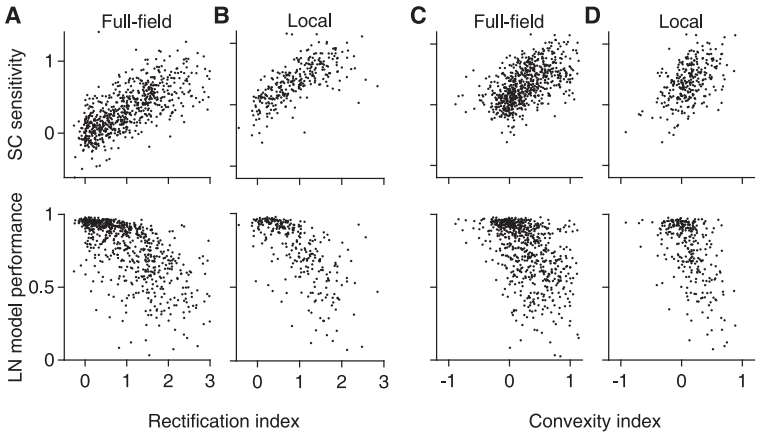

Figure 7. Spatial input nonlinearities correlate with SC sensitivity and model performance. ***A***, Relation of the SC sensitivity (top) and LN model performance (bottom), measured with natural images, to the full-field RI for all recorded cells ($n = 700$) from 13 retinas (9 animals). ***B***, Same as in ***A***, but using the data from the local checkerboard flash stimulus ($n = 289$) from 6 retinas (4 animals). ***C***, ***D***, Same as in ***A***, ***B***, but for the CI.

that contrast level. A CI of zero corresponds to linearity (equal responses for a single component at full contrast and for two components at half contrast), whereas values smaller or larger than zero correspond to increased or decreased preference for homogeneous stimuli, respectively. Over the population of all recorded cells, the two indices were correlated (for the full-field indices, see Fig. 6*G*; Spearman's $\rho = 0.68$, $p < 10^{-3}$, $n = 700$), indicating that the two nonlinearity components often coexist and may reflect the same mechanistic origin.

To systematically compare full-field and local spatial integration profiles, we compared RI and CI across the two conditions. Although both indices displayed a significant change between local and full-field stimulation (Wilcoxon signed-rank test, $p < 10^{-3}$ for both RI and CI, $n = 289$), the values were correlated between the two conditions (Spearman's $\rho = 0.82$, $p < 10^{-3}$ for the RI and Spearman's $\rho = 0.48$, $p < 10^{-3}$ for the CI), indicating that cells retained their relative characteristics of nonlinear spatial integration, in particular regarding rectification (Fig. 6*H*). One subtle change was that, for many cells with CI > 0 in the full-field condition, the index became smaller for the local stimulus (Fig. 6*I*), corresponding to a less outward-bulging shape of the contour line in the quadrant of preferred contrast (visible in the first two examples when comparing Fig. 5*B* and Fig. 5*D*). Thus, spatially homogeneous stimulation was less effective for these cells under full-field conditions compared with local stimulation because relatively larger contrast values were needed under homogeneous full-field stimulation to reach the activation level of the contour line. This may be explained by a linear component of spatial integration in the surround. A linear surround component would provide relatively more surround suppression under spatially homogeneous stimulation than a corresponding stimulation of only one spatial component and thereby decrease sensitivity to spatially homogeneous stimuli in the full-field condition.

How are the extracted components of nonlinear spatial integration related to responses under natural images? We first investigated the relationship to the SC sensitivity, as determined from the responses to natural images (compare Fig. 3*C–E*), and found that it was correlated with both the RI (Spearman's $\rho = 0.73$, $p < 10^{-3}$) and the CI (Spearman's $\rho = 0.59$, $p < 10^{-3}$), as obtained from full-field stimulation (Fig. 7*A*,*C*, top). Similar results were also found for the indices obtained from local stimulation (Fig. 7*B*,*D*, top; Spearman's $\rho = 0.75$, $p < 10^{-3}$ for rectification and $\rho = 0.49$, $p < 10^{-3}$ for convexity). In line with the analysis of SC sensitivity, LN model performance for natural images also displayed a clear dependence on the rectification indices (Fig. 7*A*,*B*, bottom) from both full-field (Spearman's $\rho = -0.71$, $p < 10^{-3}$) and local stimulation (Spearman's $\rho = -0.73$, $p < 10^{-3}$). This relationship was much more pronounced than that between LN model performance and the nonlinearity indices extracted from contrast-reversing gratings (compare Fig. 4*K*). The convexity indices from full-field (Spearman's $\rho = -0.56$, $p < 10^{-3}$) and local stimulation (Spearman's $\rho = -0.50$, $p < 10^{-3}$) were also correlated to LN model performance (Fig. 7*C*,*D*, bottom), but to a smaller extent than the rectification indices. We thus concluded that the degree of rectification of spatial inputs in the RF center is a primary factor that shapes ganglion cell responses to natural images and determines whether responses can be captured by the LN model.

The spatial scale of contrast sensitivity for natural images

We next asked on what spatial scale nonlinearities are relevant for encoding natural images. To do so, we compared responses under original natural images and blurred versions (Fig. 8), similar to previous analyses with white-noise patterns (Schwartz et al., 2012; Jacoby and Schwartz, 2017; Mani and Schwartz, 2017; Johnson et al., 2018). The blurring with a given spatial scale corresponds to low-pass filtering and removes fine spatial structure below this scale while keeping the mean intensity over larger regions approximately constant. At a blurring scale close to a cell's RF center diameter, blurring should diminish SC within the RF while keeping the mean light intensity approximately unchanged. Figure 8*A–C* compares responses to natural images and their blurred versions for three sample cells. At a scale of 240 μm, the blurring generally reduced responses for the first cell (Fig. 8*A*, middle; Wilcoxon signed-rank test, $n = 40$ images, $p < 10^{-3}$), but left responses for the second largely unaffected (Fig. 8*B*, middle; $p = 0.18$), and for the third cell even led to increased spike count (Fig. 8*C*, middle; $p = 0.02$).

To quantify the blurring effects for all cells, we calculated the mean response difference between the blurred and the original version of the images, normalized by the cell's maximum response over all images. This spike count difference was correlated to LN model performance for natural images (Fig. 8*D*) in both individual experiments (median Spearman's $\rho = 0.65$, 9 of 9 experiments had $p < 0.01$), and in the pooled population (Spearman's $\rho = 0.68$, $p < 10^{-3}$). Thus, cells that were more strongly affected by the blurring generally displayed worse LN model performance and had a stronger dependence of spike count on SC. This confirms the effect of spatial structure inside the RF for determining responses to natural images in particular ganglion cells.

When analyzing responses across different blurring scales, we observed that cells sensitive to SC reduced their spike counts already at scales smaller than their RF center (Fig. 8*A*, bottom). To quantify the spatial scale of blurring sensitivity for each cell, we measured the similarity between responses to original and blurred images by calculating the corresponding coefficient of

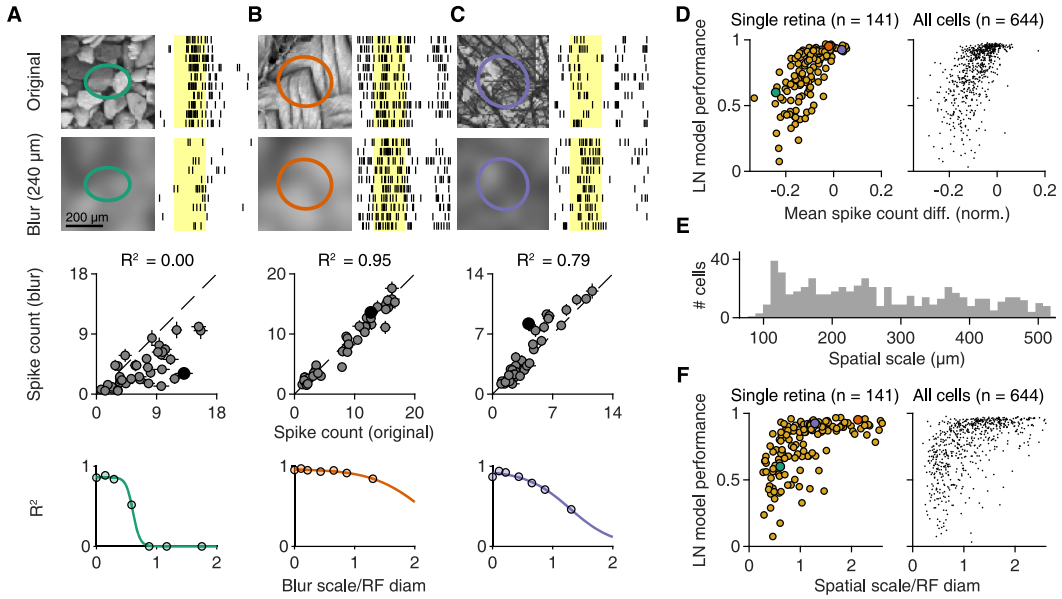

Figure 8. The scale of SC sensitivity for natural images. ***A–C***, Top, Responses of three sample cells to original natural images and their blurred versions for 10 trials. Images were blurred with a Gaussian function (scale $4\sigma = 240\ \mu$m). Shaded yellow areas represent the 200 ms image presentations. Middle, Relation of the average spike count for presentations of 40 natural images and their blurred counterparts for each of the three cells. Filled black dots correspond to the image pairs shown above. Error bars indicate SEM. Dashed line indicates the equality line. Bottom, Coefficient of determination (R^2) between original and blurred spike counts for different degrees of blurring, normalized by the RF diameter of each cell. Colored lines indicate logistic fits. ***D***, Relation between LN model performance for natural images and the mean (across images) spike count difference between blurred ($4\sigma = 240\ \mu$m) and original images for each cell from a single retina (left, Spearman's $\rho = 0.75$, $p < 10^{-3}$) and from the pooled ganglion cell population (right, Spearman's $\rho = 0.68$, $p < 10^{-3}$) from 9 retinas (6 animals). The differences were normalized to the maximum average spike count observed for each cell. Cells from ***A–C*** are highlighted. ***E***, The distribution of spatial scales across the pooled ganglion cell population ($n = 644$). The spatial scale was defined as the midpoint of fitted logistic functions (compare ***A–C***, bottom). ***F***, Relation of spatial scale, normalized by the RF diameter, to LN model performance for natural images for a single retina (left, Spearman's $\rho = 0.63$, $p < 10^{-3}$) and for the pooled ganglion cell population (right, Spearman's $\rho = 0.61$, $p < 10^{-3}$) from 9 retinas, 6 animals. Cells from ***A–C*** are highlighted.

determination (R^2), which is unity when responses with and without blurring are identical, and falls off toward zero as responses to blurred images deviate more and more from the original responses. Analogous to the analysis of contrast-reversing gratings, we fitted logistic functions to the decay of R^2 with blurring scale and defined the spatial scale as the midpoint of the logistic function. The obtained spatial scales ranged from 100 to 500 μm (Fig. 8*E*) and were only weakly correlated with the spatial scales measured with contrast-reversing gratings (Spearman's $\rho = 0.12$, $p = 0.007$). And unlike the spatial scale obtained from reversing gratings, the spatial scale from blurred images (normalized by the RF center diameter) was strongly related to LN model performance (Fig. 8*F*) in both individual experiments (median Spearman's $\rho = 0.60$, 9 of 9 experiments had $p < 0.01$) and in the pooled population (Spearman's $\rho = 0.61$, $p < 10^{-3}$).

SC sensitivity differs among RGC classes

The analyses so far have shown that the characteristics of spatial integration are consistent for individual ganglion cells across different stimulus conditions, including natural and artificial stimuli. We thus hypothesized that they reflect cell type-specific properties. To test this hypothesis, we looked at three readily identifiable cell classes, detected through a standard set of artificial stimuli.

First, we focused on IRS cells, which form a single functional cell type in the mouse retina and which correspond to transient OFF-α ganglion cells (Krishnamoorthy et al., 2017). We identified IRS cells by their characteristic response peaks to rapid shifts of a grating with no net displacement of the grating position (Fig. 9*A*). As expected, IRS cells were all OFF-type, with fast temporal filters and tiling RFs. For these cells, all our spatial integration measures displayed relatively narrow distributions. LN model performance for IRS cells was high (Fig. 9*D*, left; median = 0.94, $n = 29$), suggesting linear spatial integration. However, rather than showing no sensitivity to SC, the distribution of SC sensitivity for IRS cells was significantly shifted toward negative values (Fig. 9*D*, right; median = −0.11, $n = 29$, Wilcoxon sign-rank test, $p = 0.005$). This indicates that IRS cells had a particular preference for spatially homogeneous natural stimuli. Specifically, about half (14 of 29) of the IRS cells were inversely sensitive to SC of natural images, as identified by a significant negative slope comparing differences in SC and in spike count for image pairs with similar mean illumination (compare Fig. 3*C–E*). In terms of spatial integration measured by the checkerboard flashes, most IRS cells showed profiles, such as the one in Figure 5*B* (Cell 7), with low rectification (median = 0.12, $n = 28$) and slightly negative convexity indices, yet not significantly different from zero (median = −0.06, Wilcoxon sign-rank test, $p = 0.07$).

Second, we tested DS ganglion cells (Fig. 9*B*), detected through their responses to drifting gratings. DS cells had either ON- or OFF-type temporal filters (Fig. 9*B*, right top), with OFF-type filters likely corresponding to ON-OFF DS cells. DS cells with OFF-type filters typically showed U-shaped nonlinearities in LN models obtained from white-noise stimulation (data not shown) and responses under light-intensity steps to both increasing and decreasing intensity. DS cells generally displayed rather nonlinear spatial integration for natural images (Fig. 9*D*), with low LN model performance (median = 0.71, $n = 46$) and significant SC sensitivity (median = 0.41, $n = 46$, Wilcoxon sign-rank

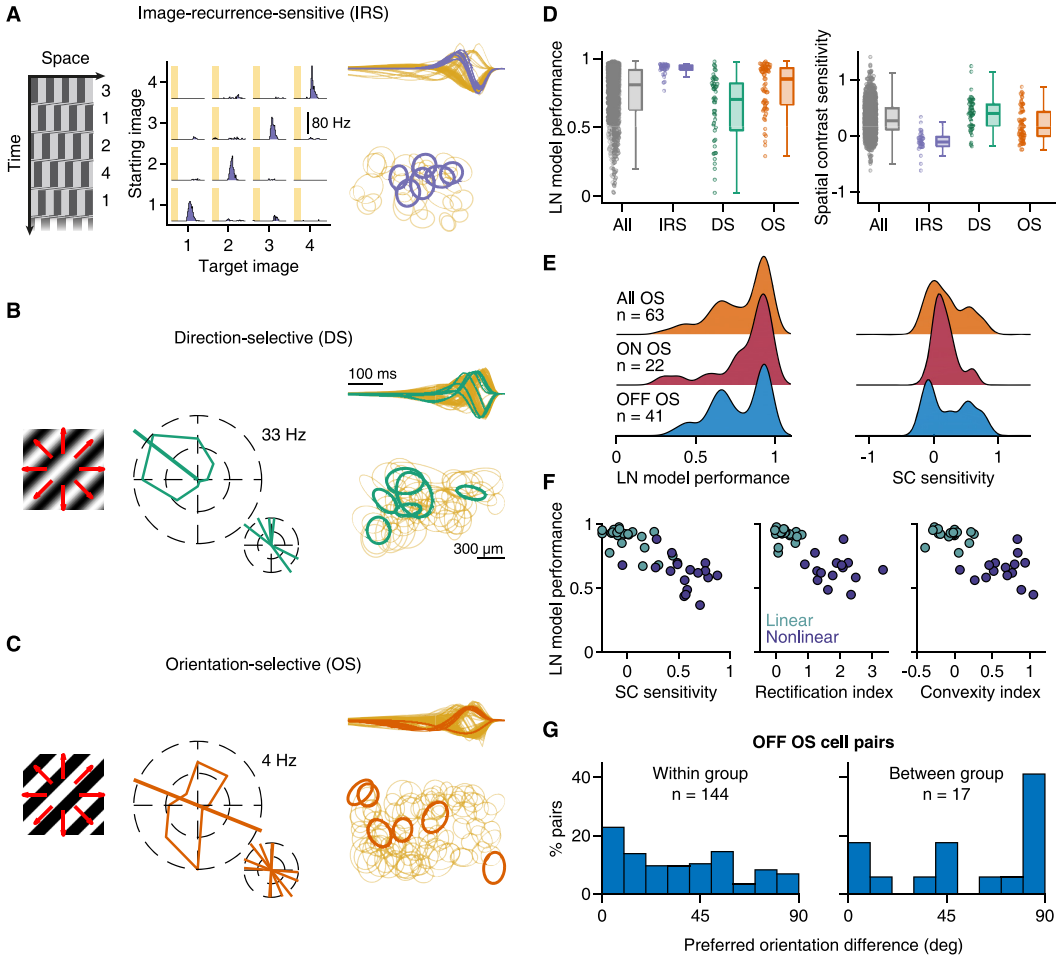

Figure 9. SC integration in three different functional cell classes. ***A***, IRS cells were identified with a sequence of saccadic-like grating movements (left), shown as a space-time plot with time going downward. Numbers indicate the four different fixation positions. PSTHs of a sample IRS cell for the 16 possible transitions (middle). Shaded areas represent the 100 ms transitions. Temporal filters (normalized to unit norm) of IRS cells in a single retina (right top), overlaid on top of the temporal filters of all recorded cells. Bottom, The RF centers of IRs cells (right bottom), overlaid on other nearby RFs. ***B***, DS ganglion cells were identified from their responses to slow drifting sinusoidal gratings (left). Tuning curve of a sample DS cell (middle), along with the preferred direction of all DS cells in a single recording. Temporal filters and RFs are shown as in ***A***. ***C***, OS ganglion cells were identified with drifting square-wave gratings (left), moving at higher speeds than in ***B***. Tuning curve of a sample OS cell (middle), along with the preferred orientations of all OS cells in a single recording. Temporal filters and RFs are shown as in ***A***. ***D***, Distributions of LN model performance (left) and SC sensitivity (right) under natural stimuli for IRS ($n = 29$), DS ($n = 46$), and OS ($n = 63$) cells from 13 retinas (9 animals). Gray represents all cells. ***E***, Distributions of LN model performance and SC sensitivity for ON and OFF OS cells compared with all OS ganglion cells. ***F***, OFF OS cells were assigned to two groups (linear and nonlinear) with *k*-means clustering. The features we used were the LN model performance, SC sensitivity, and full-field rectification as well as convexity indices. All four features were available for only $n = 36$ of 41 OFF OS cells, and clustering was performed with these cells only. For the other cells, responses to checkerboard flashes were not recorded, and the cells were assigned to the group whose cluster centroid was closest for the two available measures, LN model performance and SC sensitivity. ***G***, Distributions of differences in preferred orientation for pairs of OS cells that belonged to either the same group ("within group") or to different groups ("between group").

test, $p < 10^{-3}$). As expected for cells sensitive to SC, their full-field rectification indices were rather large (median = 1.46, $n = 31$), and their convexity indices were significantly larger than zero (median = 0.25, $n = 31$, Wilcoxon sign-rank test, $p < 10^{-3}$). Thus, DS cells showed nonlinear spatial integration under both natural and artificial stimuli.

Finally, using drifting gratings with higher speed, we identified OS ganglion cells (Fig. 9*C*). We found both ON- and OFF-type OS cells, possibly corresponding to the recently described different classes in the mouse retina (Nath and Schwartz, 2016, 2017). OS cells displayed characteristics of spatial integration that lay in between IRS and DS cells (Fig. 9*D*). Compared with DS cells, for example, OS cells showed better LN model performance (Wilcoxon rank-sum test, $p < 10^{-3}$), yet many cells still revealed poor model predictions (median = 0.85, $n = 63$). Likewise, SC sensitivity was lower than for DS cells, yet still significantly larger than zero (median = 0.15, $n = 63$, Wilcoxon sign-rank test, $p < 10^{-3}$). This was also reflected in the OS cells' responses to the checkerboard flashes, with lower full-field rectification (median = 0.46, $n = 50$) and convexity indices (median = 0.13, $n = 50$) compared with DS cells, yet with both indices significantly larger than zero (Wilcoxon sign-rank test, $p < 10^{-3}$ for rectification and $p = 0.005$ for convexity indices).

Examining the distributions of these measures for OS cells more closely, we observed that they appeared to be bimodal (Fig. 9*E*), which may indicate that different types of OS cells differ in how nonlinear their spatial integration is. Indeed, we found that ON-type OS cells showed fairly linear spatial integration characteristics (Fig. 9*E*), whereas OFF-type OS cells could be clustered into two separate groups: one with linear spatial integration and

good LN model performance and another with nonlinear spatial integration and poorer LN model performance (Fig. 9*F*). Interestingly, the linear and nonlinear OFF OS cells also differed systematically in their preferred orientations (Wilcoxon rank-sum test, $p = 0.006$; Fig. 9*G*); cells from different clusters (i.e., one linear and one nonlinear cell) often had orthogonal preferred orientations, whereas cells from the same cluster more often also had similar preferred orientations. Although we do not know the preferred orientations of these cells in absolute retinal coordinates, our data suggest that linear and nonlinear OFF-type OS cells have different preferred orientations and might correspond to the previously described classes of OS cells with preference for horizontal or vertical orientations (Nath and Schwartz, 2016, 2017).

Discussion

In this work, we directly addressed the question whether mouse RGC responses to natural images are consistent with a linear RF. This was the case only for a subset of cells (Fig. 1), as SC inside the RF influenced responses for many cells beyond the mean stimulus intensity (Fig. 3). Interestingly, classical identification of sensitivity to contrast-reversing high-frequency gratings provided only a moderate prediction of which cells are affected by RF nonlinearities under natural images (Fig. 4). We therefore devised a new stimulus to characterize subunit nonlinearities in detail for many simultaneously recorded cells, revealing considerable variability in the characteristics of nonlinear stimulus integration (Figs. 5, 6) and providing a better prediction of the relevance of RF nonlinearities for natural images (Fig. 7). Experiments with blurred natural images corroborated that nonlinear RFs affect responses under natural images and that specific ganglion cells are inversely sensitive to SC (Fig. 8). Finally, the relevance of nonlinear RFs appears to be cell type-specific and may help differentiate subtypes within broader functional cell classes (Fig. 9).

Diversity in natural stimulus encoding among the retina's output channels

Using a simple linear RF model, we observed multiple facets of natural image encoding in the mouse retina. We found ganglion cells that were consistent and others that were inconsistent, to different extents, with linear RFs. Of the few previous studies with natural stimuli in the mouse retina, one supports generally linear RFs in mouse ganglion cells (Nirenberg and Pandarinath, 2012). However, both spatially linear and nonlinear ganglion cell types had been identified in the mouse retina with artificial stimuli. For example, the PixON (Johnson et al., 2018) or the sustained OFF-α cells (Krieger et al., 2017) appear to have linear RFs, whereas nonlinear RF properties can be detected for ON-delayed or ON-OFF DS cells (Mani and Schwartz, 2017). Here, we showed that mouse DS cells, like several other ganglion cells, are spatially nonlinear also for natural images.

Related investigations with natural images in other species have shown, for example, that the macaque retina also contains cells with linear as well as nonlinear RFs (Turner and Rieke, 2016). Our work demonstrates that this also holds for mouse retina where, moreover, we identify a surprising diversity of spatial integration, including linear and nonlinear cells within broader cell classes (e.g., within ON-OFF or OS cells) as well as cells that are inversely sensitive to SC. In rabbit retina, ganglion cell responses to natural images had also been found to deviate from linear RFs in different ways (Cao et al., 2011). Similar to our approach, this study used the dependence of image-evoked responses on the local variability of pixel intensities beyond the mean intensity as a signature of nonlinear spatial integration under natural images. Cao et al. (2011) then went on to demonstrate that the skewed distribution of intensity values in natural scenes affects ganglion cell responses via this texture sensitivity, whereas our work here focuses on characterizing the features and variability of spatial nonlinearities across ganglion cells.

The degree and type of the spatial nonlinearity appear to differ between RGC types, yet functional cell-type classification schemes rely mostly on linear model components, such as RF size and temporal filter shapes (Chichilnisky and Kalmar, 2002; Baden et al., 2016; Franke et al., 2017; Jouty et al., 2018; Ravi et al., 2018; Rhoades et al., 2019). Including characteristics of nonlinear spatial integration, such as LN model performance, subunit rectification, or spatial scale of nonlinear integration, may help to better distinguish cell types. For example, the mouse retina contains at least four subtypes of OS cells, two of which (ON- and OFF-type) are tuned to horizontal and the other two (again ON and OFF) to vertical orientations (Nath and Schwartz, 2016, 2017). Here, we found distinct groups of OS cells with linear and nonlinear spatial integration, suggesting that OS subtypes might differ not only in contrast preference or preferred orientation, but also in how they integrate SC, which provides additional information for separating and identifying subtypes of OS cells.

When considering that synaptic transmission is often inherently nonlinear, the occurrence of linear RFs may actually be surprising (Shapley, 2009). In the salamander retina, for example, nonlinear ganglion cell RF centers and surrounds seem to be the norm (Bölinger and Gollisch, 2012; Takeshita and Gollisch, 2014). Linear RFs may be a property of mammalian retinas, as they have been described also in cat, rabbit, and macaque retinas (Enroth-Cugell and Robson, 1966; White et al., 2002; Petrusca et al., 2007; Molnar et al., 2009), and may have specifically evolved to provide raw information about illumination patterns to the cortex for further processing (Roska and Meister, 2014).

A cell class with particular sensitivity to spatial homogeneity of natural images

We identified cells in the mouse retina with particular sensitivity to spatially homogeneous regions in the images. Specifically, these cells were inversely sensitive to SC: although well described by an LN model, they respond more strongly to homogeneous stimuli than to structured stimuli of equal mean light level. This feature is not to be confused with the characteristics of suppressed-by-contrast cells (Levick, 1967; Tien et al., 2015; Jacoby and Schwartz, 2018), which are also known as uniformity detectors, and which are suppressed below baseline activity by (temporal) contrast. The homogeneity-preferring cells identified here, on the other hand, are generally activated by a new image and particularly strongly so if a spatially homogeneous region of preferred contrast falls onto the RF. This is reminiscent of the homogeneity detectors that have been described in the salamander retina (Bölinger and Gollisch, 2012), although the latter showed rectification of nonpreferred contrasts, unlike the homogeneity-sensitive cells described here. These cells, through their particular sensitivity to homogeneous stimuli, could provide information about image focus; blurring through defocusing will increase activity for this cell type and simultaneously decrease activity for spatial-contrast-sensitive cells, such as ON-delayed cells (Mani and Schwartz, 2017), which have been implicated in focus-sensing functions. A readout based on activity differences between cells of opposite tuning under image blur could provide a code

for image focus that is particularly robust, for example, to variations in contrast and spatial structure (Kühn and Gollisch, 2019).

IRS cells appear to be part of the homogeneity-sensitive cells. The IRS cells correspond to transient OFF-α cells (Krishnamoorthy et al., 2017) and should therefore also match the PV5 ganglion cells, which have been shown to be approach-sensitive (Münch et al., 2009). It seems that approach sensitivity, image-recurrence sensitivity, and sensitivity to homogeneous natural images may rely on the same circuit component: strong, local (glycinergic) ON-type inhibition, which transient OFF-α cells are known to receive (van Wyk et al., 2009), and which needs to be suppressed, perhaps below baseline level, by OFF-type stimuli for maximal activity.

Assessing nonlinear spatial integration with artificial and natural stimuli

The classical test for nonlinear spatial integration in the retina is to check for frequency-doubled responses under contrast-reversing spatial gratings whose spatial frequency is below the resolution of the linear RF (Hochstein and Shapley, 1976; Krieger et al., 2017). Yet, we found that sensitivity to fine gratings is generally not a good predictor for relevant spatial nonlinearities under natural stimuli, as measured by a failure of the LN model (Fig. 4). Several aspects likely contribute to this discrepancy. Perhaps most importantly, fine gratings isolate responses to high spatial frequencies and therefore sensitively detect nonlinear response components. Natural stimuli, on the other hand, have a broad frequency spectrum, and linear responses to the prevalent low frequencies may dominate the responses even when reversing gratings reveal nonlinear spatial integration. In addition, analyses under reversing gratings can be confounded by sensitivity to both light increments and decrements in ON-OFF cells. Finally, the high contrast typically used with reversing gratings may emphasize nonlinear effects, since higher contrast makes nonlinearities more pronounced (Turner and Rieke, 2016). The difference in temporal structure between reversing gratings and flashed images, on the other hand, did not seem to play a major role (Fig. 4*I*,*L*).

Mechanisms of linear and nonlinear spatial integration

Nonlinear spatial integration as measured with gratings is attributed to the rectified excitation that bipolar cells provide to the ganglion cell (Demb et al., 2001). The same mechanism likely also dominates the nonlinear response characteristics under natural images, as underscored by the relation between signal rectification and LN model failure (Fig. 7). Biophysically, rectification of bipolar cell signals seems to originate presynaptically from a nonlinear dependence of vesicle release on calcium concentration in the synaptic terminal (Singer and Diamond, 2003; Jarsky et al., 2011). The nonlinear integration of preferred contrast signals, which we quantified in the CI, may have a similar origin, as vesicle exocytosis and postsynaptic currents increase supralinearly with increasing calcium concentration, at least for moderate levels (Jarsky et al., 2011).

Yet, the degree of nonlinear spatial integration varied widely across cells, suggesting different levels of partial rectification in the signal transmission from bipolar to ganglion cells. What is the origin of this variability in the degree of rectification among ganglion cell types? Presynaptically, baseline activity of bipolar cell synapses may vary, allowing some synapses to modulate transmitter release in both directions and precluding others from decreasing activity much below baseline, thus causing rectified transmitter release. For example, regarding inputs to Y-type ganglion cells in guinea pig and mouse, the basal glutamate release is higher at the more linear ON (compared with OFF) bipolar cell terminals (Zaghloul et al., 2003; Borghuis et al., 2013). Further mechanisms, such as postsynaptic receptor dynamics and inhibition, may also contribute in shaping signal transmission between bipolar and ganglion cells. Crossover inhibition from glycinergic amacrine cells (Werblin, 2010), for example, can provide response suppression below baseline for nonpreferred contrast and thereby (partially) linearize the rectified pure bipolar cell signals. The crossover inhibition can act on bipolar cell terminals (Molnar et al., 2009) or directly on the ganglion cell, and its gain, relative to the gain of the preferred-contrast excitation, would determine the degree of nonlinearity in spatial integration.

The spatial scale of nonlinear spatial integration that we identified through the presentation of blurred images (100-120 μm; Fig. 8) is somewhat larger than typical bipolar cell RFs in mouse retina of ~40-60 μm (Berntson and Taylor, 2000; Schwartz et al., 2012; Franke et al., 2017). This might be an effect of the spatial correlations in natural images, which reduce the impact of blurring. Furthermore, electrical coupling between bipolar cells may increase the spatial scale, especially for stimuli with considerable spatiotemporal correlations (Kuo et al., 2016).

Limitations of this study

The lack of anatomic or genetic information often complicates the clear identification of individual ganglion cell types in extracellular multielectrode array recordings. On the other hand, these high-throughput recordings can provide an overview that highlights the diversity of response properties in a way not easily possible with targeted single-cell recordings. Another issue with multielectrode array recordings is the distribution of recorded ganglion cell RFs over the broad range of the recording sites. Under natural images, different cells are stimulated by different image regions, which contributes variability among cells of the same type, as some cells may experience more spatial structure within the presented images than others. The lack of RF information may also present a problem for artificial stimuli that should target, for example, the RF center. However, our application of locally sparse stimulus presentations shows that this can be overcome, allowing high-throughput investigations of center-surround effects (Figs. 5–7), as previously used in single-cell patch-clamp recordings (Turner et al., 2018).

To analyze the cells' sensitivity to spatial structure beyond mean light intensity, we analyzed the effect of SC within the RF center, defined via the variance of pixel intensities. This simple and straightforward quantification of spatial structure fails to capture which aspects of natural images provide the most relevant SC, which may result, for example, from object boundaries, textures, or gradients in light intensity. Once sensitivity to SC is established, follow-up investigations may ask which of these natural image features might be most relevant for mediating the SC effects.

We used flashed image presentations because our study was focused on spatial integration. Thus, while the applied stimuli had natural spatial structure, they lacked, for example, motion components that are induced by eye movements. This simplification of actual natural stimuli allowed us to specifically target spatial nonlinearities without having to explicitly consider the influence of temporal filtering and adaptation on the responses. It seems likely that nonlinear spatial integration observed under our flashed natural images will also shape responses to natural movies. On the other hand, our approach is insensitive to

nonlinearities triggered through specific temporal stimulus features. For example, IRS cells, which we here reported as being rather linear for the encoding of images flashed in isolation, can reveal nonlinearities when rapid image transitions are considered, for which disinhibitory interactions mediate a sensitivity to recurring spatial patterns (Krishnamoorthy et al., 2017). Additionally, we focused on a single light level, but spatial nonlinearities may change with light level: sustained ON-α ganglion cells in the mouse retina, for example, become more linear with decreasing light intensity (Grimes et al., 2014).

The presentation of full-field natural images stimulates both the RF center and surround. Our modeling approach aimed at capturing effects of surround suppression by using a difference of Gaussians as a spatial filter, which could contain positive and negative values. Yet, this might not reflect the actual surround strength under natural images because the surround might be underestimated with spatiotemporal white noise (Wienbar and Schwartz, 2018) and because the extraction of the spatial filter from the spatiotemporal STA, which often lacks space-time separability (Cowan et al., 2016), may further diminish the surround component. We thus cannot exclude that disregarded surround effects contribute to shortcomings of the LN model. However, the following two arguments indicate that, regardless of surround effects, nonlinear integration in the RF center is a main factor in LN model performance. First, our SC sensitivity analysis showed that across-cell differences in LN model performance could be explained to a large degree by considering SC only in the RF center. And second, rectification indices obtained from the full-field and from the locally restricted checkerboard flashes worked about equally well to explain the performance differences of the LN model.

More generally, the good correspondence between measures of spatial nonlinearity and LN model performance, in particular the fact that cells with little SC sensitivity (Fig. 3) or rectification (Fig. 7) displayed model performance near unity, supports the suitability of our approach to apply a parameterized LN model with fitted spatial filters and nonlinearities for assessing spatial nonlinearities in the encoding of natural images. It also underscores the reliability of the recorded data, reaffirming that observed variability in a cell's response to different images results from the cell's differential activation by the images and not from drift or rundown over the course of the long *in vitro* recordings.

Implications for neuronal modeling

Proposed improvements to LN-type models go in many directions (Latimer et al., 2019; Shi et al., 2019). Here we demonstrated that the incorporation of sensitivity to fine spatial structure into models (e.g., with spatial subunits) should be significant for natural stimuli. We found that cells with low LN model performance mostly showed nonlinear spatial integration and that rectification of nonpreferred contrast in the RF center was particularly important. This observation of the importance of rectification agrees with results from nonlinear subunit modeling of ganglion cells: in the macaque retina, the rectification of subunits determines the degree of nonlinear integration under natural images (Turner and Rieke, 2016); and in a model of salamander RGCs under white-noise stimulation, threshold-linear rectification of subunit signals worked nearly as well as more elaborate, fitted shapes (Real et al., 2017). However, we here also found that there is considerable variability in the type of subunit nonlinearities with different degrees of rectification and convexity as well as cells with inverse sensitivity to SC. This suggests that not all cells will be well captured by a standard subunit model with summation over half-wave rectified local signals. Our approach of analyzing SC and iso-response stimuli rather than assessing the performance of an explicit subunit model allowed us to capture this diversity. Furthermore, the checkerboard flash stimulation introduced here can be used to efficiently estimate the characteristics of subunit nonlinearities for many RGCs simultaneously. Given the recently developed techniques for estimating subunit locations (Liu et al., 2017; Maheswaranathan et al., 2018; Shah et al., 2020), this paves the way for building more detailed models for different ganglion cell types. Our results also indicate that such cell type-specific approaches may be needed as there might not be a satisfactory single "standard model" (Carandini et al., 2005).

References

Arbeláez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. IEEE Trans Pattern Anal Mach Intell 33:898–916.

Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. Neural Comput 2:308–320.

Baccus SA, Ölveczky BP, Manu M, Meister M (2008) A retinal circuit that computes object motion. J Neurosci 28:6807–6817.

Baden T, Berens P, Franke K, Román Rosón M, Bethge M, Euler T (2016) The functional diversity of retinal ganglion cells in the mouse. Nature 529:345–350.

Berntson A, Taylor WR (2000) Response characteristics and receptive field widths of on-bipolar cells in the mouse retina. J Physiol 524:879–889.

Bölinger D, Gollisch T (2012) Closed-loop measurements of iso-response stimuli reveal dynamic nonlinear stimulus integration in the retina. Neuron 73:333–346.

Bomash I, Roudi Y, Nirenberg S (2013) A virtual retina for studying population coding. PLoS One 8:e53363.

Borghuis BG, Marvin JS, Looger LL, Demb JB (2013) Two-photon imaging of nonlinear glutamate release dynamics at bipolar cell synapses in the mouse retina. J Neurosci 33:10972–10985.

Burkhardt DA, Fahey PK, Sikora M (1998) Responses of ganglion cells to contrast steps in the light-adapted retina of the tiger salamander. Vis Neurosci 15:219–229.

Burton GJ, Moorhead IR (1987) Color and spatial structure in natural scenes. Appl Opt 26:157–170.

Cao X, Merwine DK, Grzywacz NM (2011) Dependence of the retinal ganglion cell's responses on local textures of natural scenes. J Vis 11:11.

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? J Neurosci 25:10577–10597.

Carcieri SM, Jacobs AL, Nirenberg S (2003) Classification of retinal ganglion cells: a statistical approach. J Neurophysiol 90:1704–1713.

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network 12:199–213.

Chichilnisky EJ, Kalmar RS (2002) Functional asymmetries in ON and OFF ganglion cells of primate retina. J Neurosci 22:2737–2747.

Cowan CS, Sabharwal J, Wu SM (2016) Space-time codependence of retinal ganglion cells can be explained by novel and separable components of their receptive fields. Physiol Rep 4:e12952.

de Vries SE, Lecoq JA, Buice MA, Groblewski PA, Ocker GK, Oliver M, Feng D, Cain N, Ledochowitsch P, Millman D, Roll K, Garrett M, Keenan T, Kuan L, Mihalas S, Olsen S, Thompson C, Wakeman W, Waters J, Williams D, et al. (2020) A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. Nat Neurosci 23:138–151.

Demb JB, Haarsma L, Freed MA, Sterling P (1999) Functional circuitry of the retinal ganglion cell's nonlinear receptive field. J Neurosci 19:9756–9767.

Demb JB, Zaghloul K, Haarsma L, Sterling P (2001) Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (Y) ganglion cell in mammalian retina. J Neurosci 21:7447–7454.

Enroth-Cugell C, Robson JG (1966) The contrast sensitivity of retinal ganglion cells of the cat. J Physiol 187:517–552.

Felsen G, Dan Y (2005) A natural approach to studying vision. Nat Neurosci 8:1643–1646.

Fischer BJ, Anderson CH, Peña JL (2009) Multiplicative auditory spatial receptive fields created by a hierarchy of population codes. PLoS One 4: e8015.

Franke K, Berens P, Schubert T, Bethge M, Euler T, Baden T (2017) Inhibition decorrelates visual feature representations in the inner retina. Nature 542:439–444.

Freeman J, Field GD, Li PH, Greschner M, Gunning DE, Mathieson K, Sher A, Litke AM, Paninski L, Simoncelli EP, Chichilnisky EJ (2015) Mapping nonlinear receptive field structure in primate retina at single cone resolution. Elife 4:e05241.

Gollisch T, Herz AV (2012) The iso-response method: measuring neuronal stimulus integration with closed-loop experiments. Front Neural Circuits 6:104.

Gollisch T, Meister M (2010) Eye smarter than scientists believed: neural computations in circuits of the retina. Neuron 65:150–164.

Grimes WN, Schwartz GW, Rieke F (2014) The synaptic and circuit mechanisms underlying a change in spatial encoding in the retina. Neuron 82:460–473.

Hawrylycz M, Anastassiou C, Arkhipov A, Berg J, Buice M, Cain N, Gouwens NW, Gratiy S, Iyer R, Lee JH, Mihalas S, Mitelut C, Olsen S, Reid RC, Teeter C, de Vries S, Waters J, Zeng H, Koch C, MindScope (2016) Inferring cortical function in the mouse visual system through large-scale systems neuroscience. Proc Natl Acad Sci USA 113:7337–7344.

Heitman A, Brackbill N, Greschner M, Sher A, Litke AM, Chichilnisky EJ (2016) Testing pseudo-linear models of responses to natural scenes in primate retina. bioRxiv. doi: 10.1101/045336.

Hochstein S, Shapley RM (1976) Quantitative analysis of retinal ganglion cell classifications. J Physiol 262:237–264.

Jacoby J, Schwartz GW (2017) Three small-receptive-field ganglion cells in the mouse retina are distinctly tuned to size, speed, and object motion. J Neurosci 37:610–625.

Jacoby J, Schwartz GW (2018) Typology and circuitry of suppressed-by-contrast retinal ganglion cells. Front Cell Neurosci 12:269.

Jacoby J, Zhu Y, DeVries SH, Schwartz GW (2015) An amacrine cell circuit for signaling steady illumination in the retina. Cell Rep 13:2663–2670.

Jarsky T, Cembrowski M, Logan SM, Kath WL, Riecke H, Demb JB, Singer JH (2011) A synaptic mechanism for retinal adaptation to luminance and contrast. J Neurosci 31:11003–11015.

Johnson KP, Zhao L, Kerschensteiner D (2018) A pixel-encoder retinal ganglion cell with spatially offset excitatory and inhibitory receptive fields. Cell Rep 22:1462–1472.

Jouty J, Hilgen G, Sernagor E, Hennig MH (2018) Non-parametric physiological classification of retinal ganglion cells in the mouse retina. Front Cell Neurosci 12:481.

Khani MH, Gollisch T (2017) Diversity in spatial scope of contrast adaptation among mouse retinal ganglion cells. J Neurophysiol 118:3024–3043.

Krieger B, Qiao M, Rousso DL, Sanes JR, Meister M (2017) Four alpha ganglion cell types in mouse retina: function, structure, and molecular signatures. PLoS One 12:e0180091.

Krishnamoorthy V, Weick M, Gollisch T (2017) Sensitivity to image recurrence across eye-movement-like image transitions through local serial inhibition in the retina. Elife 6:e22431.

Kühn N, Gollisch T (2016) Joint encoding of object motion and motion direction in the salamander retina. J Neurosci 36:12203–12216.

Kühn NK, Gollisch T (2019) Activity correlations between direction-selective retinal ganglion cells synergistically enhance motion decoding from complex visual scenes. Neuron 101:963–976.e7.

Kuo SP, Schwartz GW, Rieke F (2016) Nonlinear spatiotemporal integration by electrical and chemical synapses in the retina. Neuron 90:320–332.

Latimer KW, Rieke F, Pillow JW (2019) Inferring synaptic inputs from spikes with a conductance-based neural encoding model. Elife 8:e47012.

Levick WR (1967) Receptive fields and trigger features of ganglion cells in the visual streak of the rabbit's retina. J Physiol 188:285–307.

Litke AM, Bezayiff N, Chichilnisky EJ, Cunningham W, Dabrowski W, Grillo AA, Grivich M, Grybos P, Hottowy P, Kachiguine S, Kalmar RS, Mathieson K, Petrusca D, Rahman M, Sher A (2004) What does the eye tell the brain? Development of a system for the large-scale recording of retinal output activity. IEEE Trans Nucl Sci 51:1434–1440.

Liu JK, Schreyer HM, Onken A, Rozenblit F, Khani MH, Krishnamoorthy V, Panzeri S, Gollisch T (2017) Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. Nat Commun 8:149.

Maheswaranathan N, Kastner DB, Baccus SA, Ganguli S (2018) Inferring hidden structure in multilayered neural circuits. PLoS Comput Biol 14: e1006291.

Mani A, Schwartz GW (2017) Circuit mechanisms of a retinal ganglion cell with stimulus-dependent response latency and activation beyond its dendrites. Curr Biol 27:471–482.

Mazurek M, Kager M, Van Hooser SD (2014) Robust quantification of orientation selectivity and direction selectivity. Front Neural Circuits 8:92.

McIntosh LT, Maheswaranathan N, Nayebi A, Ganguli S, Baccus SA (2017) Deep learning models of the retinal response to natural scenes. Adv Neural Inf Process Syst 29:1369–1377.

Mel GC, Ramachandra CA, Mel BW (2018) Detecting object boundaries in natural images requires 'incitatory' cell-cell interactions. bioRxiv. doi: 10.1101/436949.

Meyer AF, O'Keefe J, Poort J (2020) Two distinct types of eye-head coupling in freely moving mice. Curr Biol 30:2116–2130.e6.

Michaiel AM, Abe ET, Niell CM (2020) Dynamics of gaze control during prey capture in freely moving mice. Elife 9:e57458.

Molnar A, Hsueh HA, Roska B, Werblin FS (2009) Crossover inhibition in the retina: circuitry that compensates for nonlinear rectifying synaptic transmission. J Comput Neurosci 27:569–590.

Münch TA, Da Silveira RA, Siegert S, Viney TJ, Awatramani GB, Roska B (2009) Approach sensitivity in the retina processed by a multifunctional neural circuit. Nat Neurosci 12:1308–1316.

Murgas KA, Wilson AM, Michael V, Glickfeld LL (2020) Unique spatial integration in mouse primary visual cortex and higher visual areas. J Neurosci 40:1862–1873.

Nath A, Schwartz GW (2016) Cardinal orientation selectivity is represented by two distinct ganglion cell types in mouse retina. J Neurosci 36:3208–3221.

Nath A, Schwartz GW (2017) Electrical synapses convey orientation selectivity in the mouse retina. Nat Commun 8:2025.

Nirenberg S, Pandarinath C (2012) Retinal prosthetic strategy with the capacity to restore normal vision. Proc Natl Acad Sci USA 109:15012–15017.

Olmos A, Kingdom FA (2004) A biologically inspired algorithm for the recovery of shading and reflectance images. Perception 33:1463–1473.

Ölveczky BP, Baccus SA, Meister M (2003) Segregation of object and background motion in the retina. Nature 423:401–408.

Petrusca D, Grivich MI, Sher A, Field GD, Gauthier JL, Greschner M, Shlens J, Chichilnisky EJ, Litke AM (2007) Identification and characterization of a Y-like primate retinal ganglion cell type. J Neurosci 27:11019–11027.

Pouzat C, Mazor O, Laurent G (2002) Using noise signature to optimize spike-sorting and to assess neuronal classification quality. J Neurosci Methods 122:43–57.

Ravi S, Ahn D, Greschner M, Chichilnisky EJ, Field GD (2018) Pathway-specific asymmetries between ON and OFF visual signals. J Neurosci 38:9728–9740.

Real E, Asari H, Gollisch T, Meister M (2017) Neural circuit inference from function to structure. Curr Biol 27:189–198.

Rhoades CE, Shah NP, Manookin MB, Brackbill N, Kling A, Goetz G, Sher A, Litke AM, Chichilnisky EJ (2019) Unusual physiological properties of smooth monostratified ganglion cell types in primate retina. Neuron 103:658–672.e6.

Roska B, Meister M (2014) The retina dissects the visual scene into distinct features In: The new visual neurosciences (Werner JS, Chalupa LM, eds), pp 163–182. Cambridge, MA: Massachusetts Institute of Technology.

Sabbah S, Gemmer JA, Bhatia-Lin A, Manoff G, Castro G, Siegel JK, Jeffery N, Berson DM (2017) A retinal code for motion along the gravitational and body axes. Nature 546:492–497.

Schoppe O, Harper ON, Willmore BD, King AJ, Schnupp JW (2016) Measuring the performance of neural models. Front Comput Neurosci 10:10.

Schwartz GW, Okawa H, Dunn FA, Morgan JL, Kerschensteiner D, Wong RO, Rieke F (2012) The spatial structure of a nonlinear receptive field. Nat Neurosci 15:1572–1580.

Shah NP, Brackbill N, Rhoades C, Kling A, Goetz G, Litke AM, Sher A, Simoncelli EP, Chichilnisky EJ (2020) Inference of nonlinear receptive field subunits with spike-triggered clustering. Elife 9:e45743.

Shapley R (2009) Linear and nonlinear systems analysis of the visual system: why does it seem so linear? Vision Res 49:907–921.
Shi Q, Gupta P, Boukhvalova AK, Singer JH, Butts DA (2019) Functional characterization of retinal ganglion cells using tailored nonlinear modeling. Sci Rep 9:8713.
Singer JH, Diamond JS (2003) Sustained Ca^{2+} entry elicits transient postsynaptic currents at a retinal ribbon synapse. J Neurosci 23:10923–10933.
Takeshita D, Gollisch T (2014) Nonlinear spatial integration in the receptive field surround of retinal ganglion cells. J Neurosci 34:7548–7561.
Tien NW, Pearson JT, Heller CR, Demas J, Kerschensteiner D (2015) Genetically identified suppressed-by-contrast retinal ganglion cells reliably signal self-generated visual stimuli. J Neurosci 35:10815–10820.
Turiel A, Parga N (2000) The multifractal structure of contrast changes in natural images: from sharp edges to textures. Neural Comput 12:763–793.
Turner MH, Rieke F (2016) Synaptic rectification controls nonlinear spatial integration of natural visual inputs. Neuron 90:1257–1271.
Turner MH, Schwartz GW, Rieke F (2018) Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. Elife 7:252148.
van Hateren JH, van der Schaaf A (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. Proc Biol Sci 265:359–366.
van Wyk M, Wässle H, Taylor WR (2009) Receptive field properties of ON- and OFF-ganglion cells in the mouse retina. Vis Neurosci 26:297–308.
Werblin FS (2010) Six different roles for crossover inhibition in the retina: correcting the nonlinearities of synaptic transmission. Vis Neurosci 27:1–8.
White AJ, Sun H, Swanson WH, Lee BB (2002) An examination of physiological mechanisms underlying the frequency-doubling illusion. Invest Ophthalmol Vis Sci 43:3590–3599.
Wienbar S, Schwartz GW (2018) The dynamic receptive fields of retinal ganglion cells. Prog Retin Eye Res 67:102–117.
Zaghloul KA, Boahen K, Demb JB (2003) Different circuits for ON and OFF retinal ganglion cells cause different contrast sensitivities. J Neurosci 23:2645–2654.
Zhang Y, Kim IJ, Sanes JR, Meister M (2012) The most numerous ganglion cell type of the mouse retina is a selective feature detector. Proc Natl Acad Sci USA 109:E2391–E2398.

4 THE NONLINEAR STRUCTURE OF RECEPTIVE FIELDS IN THE MAMMALIAN RETINA

Attached manuscript. Research article, currently in preparation. Reprinted below is the manuscript at its current version.

Authors

Dimokratis Karamanlis[1,2,3], Tim Gollisch[1,2]

1. University Medical Center Göttingen
2. Bernstein Center Göttingen
3. International Max Planck Research School for Neurosciences, Göttingen

Own contributions

- Conceptualization
- Data curation
- Formal analysis
- Investigation
- Methodology
- Project administration (with TG)
- Software
- Validation (with TG)
- Visualization
- Writing – original draft
- Writing – reviewing and editing (with TG)

The nonlinear structure of receptive fields in the mammalian retina

Dimokratis Karamanlis[1,2,3], Tim Gollisch[1,2]

Affiliations

[1]University Medical Center Gottingen; Göttingen, Germany.

[2]Bernstein Center for Computational Neuroscience; Göttingen, Germany.

[3]International Max Planck Research School for Neurosciences; Göttingen, Germany.

Abstract

The first step in vision involves complex neural processing in the retinal circuit. However, typically used computational models of retinal ganglion cells, the output channels of the circuit, only capture part of this processing and may provide inaccurate predictions of retinal responses to natural stimuli. We developed a class of models that captures nonlinear processing upstream of ganglion cells and fitted these models to data from multielectrode-array recordings of isolated mouse retinas. Compared to existing linear models, nonlinear models improved predictions to natural images and movies with imprinted mouse gaze shifts. We used the models to compare the presynaptic circuits of different ganglion cells, revealing new asymmetries between major cell types in the mouse retina. Finally, we suggest that nonlinear processing may require surround inhibition for capturing pairwise correlations between ganglion cells under naturalistic stimulation. Together, we show how different facets of nonlinear processing appear in retinal output under naturalistic stimulation.

Main Text

Visual processing of natural scenes begins in the retina, and the results of this processing are communicated to the brain (Martersteck et al., 2017) through output neurons, the retinal ganglion cells, which are anatomically (Bae et al., 2018) and functionally distinct (Baden et al., 2016). Each ganglion cell receives excitatory input from multiple retinal bipolar cells, which together form the ganglion cell's receptive field center. Ganglion cells also receive inhibitory inputs representing the receptive field surround, which is implemented either directly through amacrine cells (Johnson et al., 2018), or indirectly through the bipolar cell inputs (McMahon et al., 2004). Although this center-surround structure can be derived by the theory of efficient coding (Barlow, 1961) and the spatial statistics of natural scenes (Atick and Redlich, 1990, 1992), it often fails when used as a linear filter in computational models that aim to predict ganglion cell responses to actual natural stimuli (Heitman et al., 2016; Karamanlis and Gollisch, 2021; Turner and Rieke, 2016).

It turns out that bipolar cell excitation is often nonlinear (Borghuis et al., 2013; Demb et al., 2001; Enroth-Cugell and Robson, 1966), and recent computational models implemented nonlinear excitation showing promising results with naturalistic stimuli (Liu et al., 2017; Shah et al., 2020). These nonlinear models split the receptive field center into functional subunits whose signals are nonlinearly summed. Existing methods for fitting nonlinear subunit models focus on the extraction of subunit filters in space (Liu et al., 2017; Maheswaranathan et al., 2018; Shah et al., 2020), but these filters may not capture the bipolar cell scale on which nonlinear summation acts upon. Furthermore, these methods may overlook that the functional form of the nonlinear summation can differ among ganglion cell types (Bölinger and Gollisch, 2012), affecting each type's sensitivity to the spatial structure of natural scenes (Karamanlis and Gollisch, 2021). Here, we develop a new class of subunit models to holistically describe nonlinear receptive fields in the mouse retina, where ganglion cell types are thoroughly catalogued based on anatomy and physiology (Baden et al., 2016; Bae et al., 2018; Goetz et al., 2021). We test our models with natural stimuli, compare model parameters between common cell types, and ask whether nonlinear receptive fields assist in the efficient coding of natural scenes.

Subunit grid models capture nonlinear receptive field properties

Retinal encoding models can provide predictions of the retinal output to arbitrary visual stimuli. The parameters of such models need to be fit to experimental data, and this process is relatively established in the case of linear receptive fields (Turner et al., 2018). An example is the difference-of-Gaussians (DoG) model, where spatially linear center and surround signals are summed (under opposite signs) and mapped to a spiking response through an output nonlinearity (Fig. 1A). In contrast, methods to fit nonlinear receptive fields to data are still being developed (Liu et al., 2017; Maheswaranathan et al., 2018; Shah et al., 2020). All proposed nonlinear methods are based on retinal stimulation with spatio-temporal white noise, which may only drive the cells weakly and thus require length visual stimulation (Wienbar and Schwartz, 2018). Here, we introduce the subunit grid model (Fig. 1B), a new computational framework for describing nonlinear receptive fields, based on the simplifying assumption that ganglion cells receive inputs from identical bipolar cells (Bleckert et al., 2014; Schwartz et al., 2012) that are spaced semi-regularly, thus forming a "grid". We fit subunit grid models with a stimulus that can strongly drive cell responses: sinusoidal gratings of varying orientation and spatial frequency (Fig. 1C), typically used to characterize nonlinear spatial processing (Crook et al., 2008; Krieger et al., 2017; Turner and Rieke, 2016).

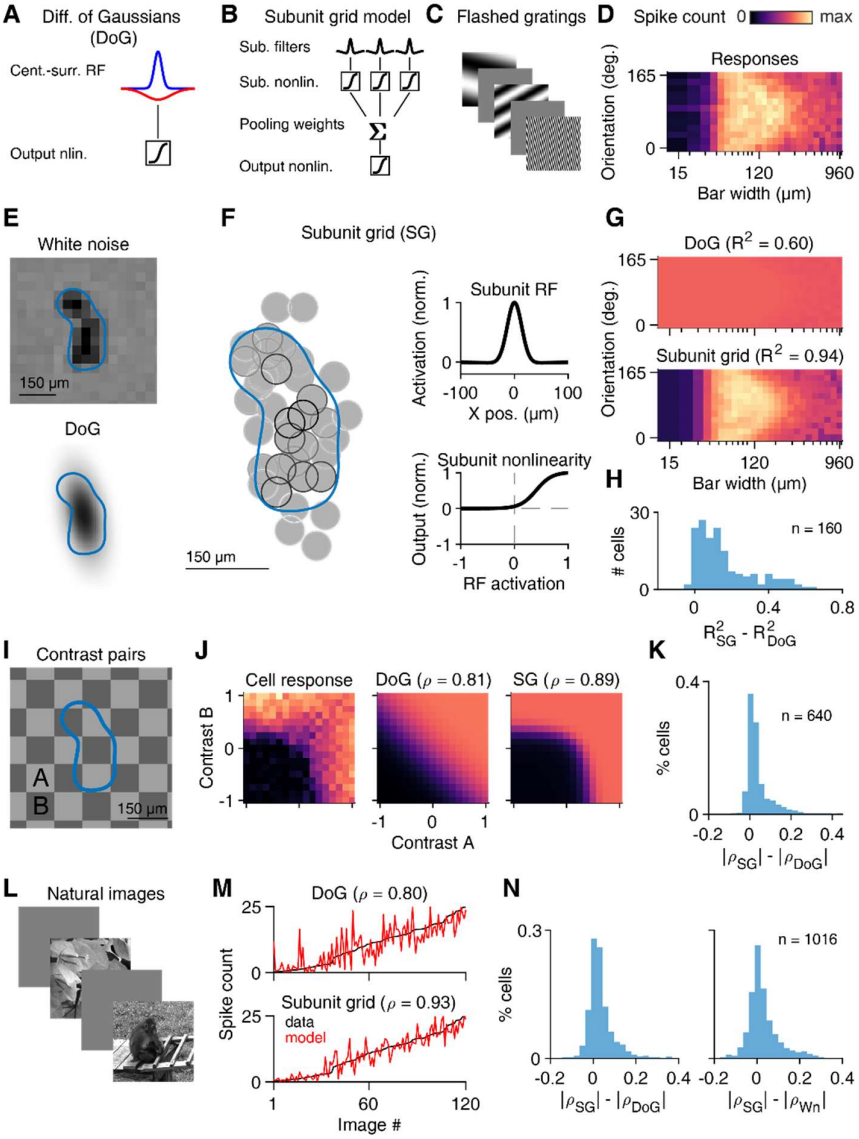

Fig. 1. The subunit grid model captures the nonlinear receptive field of retinal ganglion cells. (**A**) The difference-of-Gaussians model pools visual space linearly by combining a Gaussian center profile (blue) with a (negative) Gaussian surround profile (red). The resulting signal is passed through an output nonlinearity to generate a spike count. (**B**) In contrast, the subunit grid (SG) model pools visual space nonlinearly before the output nonlinearity, by combining the outputs of multiple identical subunit filters after they pass through a nonlinear operation. (**C**) We fitted both DoG and SG models to ganglion cell responses obtained under a sequence of 200-ms grating flashes. The gratings were sinusoidal and varied in spatial frequency, orientation and phase. (**D**) An orientation-spatial frequency tuning surface of a single mouse retinal ganglion cell, generated by averaging spike-count responses to gratings over trials and phases. The cell responds already at bar widths of ~30 um. (**E**) Spatial filter (top) of the cell in (D), measured with white noise stimulation. Darker regions correspond to higher weight. On top (blue) is a receptive-field contour, roughly outlining the receptive field center. The DoG spatial filter (bottom), determined by fitting a DoG model to the grating responses, roughly matched the location and size of the white-noise filter. (**F**) Internal subunit map (left), with multiple subunits over the receptive field surface, each with a different weight (darker outline corresponds to higher weight). Overlaid is the contour from (E) for comparison. The subunits were modelled with a DoG profile and show a diameter around 60 μm with a relatively weak surround (right, top), and a strongly rectified nonlinearity (right, bottom). (**G**) Compared to the prediction from the DoG model

(top), the SG model precisely matches the tuning surface. (**H**) Compared to DoG models, SG models could capture more of the variance in the grating responses (0.57 ± 0.30 vs. 0.73 ± 0.22, mean ± SD, $p < 10^{-26}$, Wilcoxon sign-rank test). (**I**) We stimulated the mouse retina with flashed full-field checkerboard patterns (200-ms flashes), formed by varying the contrast of A- and B-tiles. The contour from the single cell is overlaid. (**J**) We averaged spike count responses to each flashed pair of A and B contrasts (left) and tested the DoG (middle) and SG model predictions (right). We evaluated model performance by calculating the Spearman rank correlation between model predictions (omitting the output nonlinearity) and the actual cell responses. (**K**) The histogram of the difference in model performance reveals that SG models outperformed DoG models for most of the tested cells (0.81 ± 0.15 vs. 0.77 ± 0.19, mean ± SD, $p < 10^{-68}$, Wilcoxon sign-rank test). (**L**) We stimulated the retina with a random sequence of 120 natural images (200-ms flashes). (**M**) Ranking natural image responses based on the model signals (red) matches spike counts (dark) better for subunit grid model compared to the DoG model. (**N**) Model performance (Spearman's ρ) was higher for the SG model (0.75 ± 0.21, mean ± SD) both compared to the DoG model fit to gratings (left, 0.71 ± 0.24, mean ± SD, $p < 10^{-73}$, Wilcoxon sign-rank test), but also compared to models using spatial filters determined with white noise (right, 0.72 ± 0.24, mean ± SD, $p < 10^{-22}$, Wilcoxon sign-rank test).

The *a priori* definition of a grid allows for many potential subunit locations, out of which only a subset is selected through model fitting. The subunits are all identical, and arranged in a hexagonal grid with 16-μm spacing, corresponding to the highest cone bipolar cell density in the mouse retina (Wässle et al., 2009). We assumed a circular DoG profile for each subunit, which allowed us to write down cell responses as a smooth function over their defining grating parameters (Soodak, 1986), notably the subunit diameter and surround strength. Following the subunit profile definition, we applied a novel regularization method to the pooling weights (Figure S1) to constrain the selection of actual subunits. This density-based regularization helped to avoid subunits that are spaced too close to one another.

Using multielectrode array recordings from the mouse retina, we flashed sinusoidal gratings following a gray background (Fig. 1C) and recorded spike count responses. Many cells responded vigorously to gratings with a bar width of 30 μm, well below the typical diameter of a receptive field center (Fig. 1D). To capture this sensitivity to fine spatial structure, we fitted subunit grid models to the responses of all recorded cells. The weight map of the fitted subunit layout matched the weight map obtained by the classically used method of white noise (Fig. 1E), and the subunit nonlinearity often showed rectification (Fig. 1F), mirroring the contrast-response functions of ganglion cell excitatory inputs, which can be measured with patch-clamp recordings (Johnson et al., 2018; Schwartz et al., 2012; Turner and Rieke, 2016). The subunit grid model could reliably reproduce sensitivity to gratings of high spatial frequency (Fig. 1G), while a spatially linear receptive field (DoG) failed to capture this sensitivity, and this asymmetry was the rule for most recorded cells (Fig. 1H),

We proceeded to test how well subunit grid models generalize to other visual stimulus classes. Firstly, to test whether subunit grids correctly capture the overall shape of subunit nonlinearities, we used a spatially structured stimulus with two components (Fig. 1I) that we previously introduced (Karamanlis and Gollisch, 2021). This stimulus is a checkerboard of a large spatial scale that splits mouse receptive fields approximately in half, thus mimicking edges in natural scenes, and can isolate the effects of nonlinearities in spatial integration. We flashed on the retina multiple checkerboards in which we systematically varied the brightness values of the two sets of tiles, effectively examining contrast interactions within the receptive field. For many cells, spiking responses to all contrast pairs revealed a nonlinear sensitivity (Fig. 1J), and this could be better captured by the subunit grid model compared to the DoG

model (Fig. 1K). We also flashed natural images to the retina, and recorded spike count responses (Fig. 1L). The subunit grid model again predicted the retinal output better than the DoG model (Fig. 1M), which we tested by examining match accuracy between sorted image responses and sorted input activations that the models predicted. Subunit grid models were not only superior to DoG models fitted on the grating stimulus, but also to models utilizing spatial linear filters extracted from white noise stimulation (Fig. 1N), as is typically done (Heitman et al., 2016; Pillow et al., 2008). Together, these results suggest that subunit grids can accurately capture the spatial aspects of the nonlinear receptive field.

Nonlinear receptive field properties consistently vary between cell types

Ganglion cells are typically categorized in different types, based not only on their morphology, but also their functional properties (Baden et al., 2016; Goetz et al., 2021). Such functional differences are typically reported for linear receptive field properties, such as spatial and temporal filters of the receptive field center, and have led to descriptions of ganglion cell types in multiple species (Chichilnisky and Kalmar, 2002; Ravi et al., 2018), including humans (Kling et al., 2020; Soto et al., 2020). However, many of these types are spatially nonlinear and may even differ in their nonlinear properties (Bölinger and Gollisch, 2012; Karamanlis and Gollisch, 2021; Turner and Rieke, 2016). Based on responses to multiple artificial stimuli, we reliably identified four of the major retinal ganglion cell types in the mouse, ON and OFF brisk transient and sustained types (Fig. 2A-C). We matched the responses of these types to functional properties from other mouse databases (Fig. S2), concluding that our types correspond to the well-characterized alpha cells (Krieger et al., 2017). In the same retinal recording, we found that the fitted parameters of subunit grid models differed between types. (Fig. 2E-G). To an extent, these differences were evident already by examining the tuning of spiking responses with respect to spatial frequency and orientation, which the subunit grid model could quite well capture for all four types (Fig. 2H).

To identify potential asymmetries between the four cell types, we systematically compared the parameters of the fitted models. Firstly, subunit diameters typically laid between 40-80 µm for all types (Fig. 2I), matching previous results obtained with reversing gratings (Krieger et al., 2017). Interestingly, the subunits of ON brisk transient cells were larger than their sustained counterparts, while the opposite was the case for the OFF cells. Cells from all types effectively collected inputs from 15-25 subunits, an estimate in line to what is expected for the large alpha cells (Freed and Sterling, 1988), but higher than what previous methods provided for ganglion cells of other species, including the mouse (Liu et al., 2017; Shah et al., 2020). How are these subunits distributed over the receptive field? We calculated subunit coverages, which capture the amount of overlap between subunit receptive fields. Coverage values typically laid between one and two for the best fitted models (Fig. 2K), which matches the exact same range measured for the dendritic fields of bipolar cell mosaics (Behrens et al., 2016). The subunits of OFF brisk transient cells had lower coverage compared to OFF alpha sustained cells, matching the coverage asymmetry of type 3A and type 2 bipolar cells (Behrens et al., 2016), which primarily provide input to the two corresponding alpha types (Yu et al., 2018).

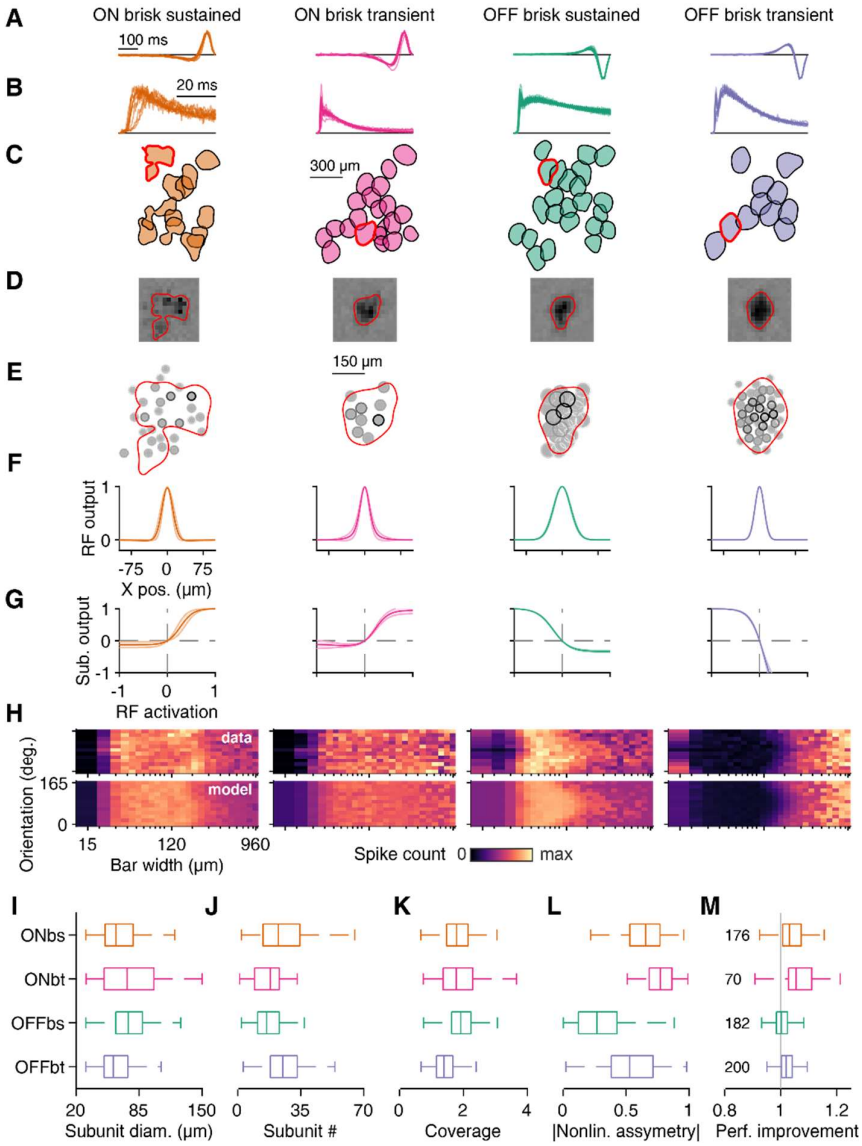

Fig. 2. Nonlinear receptive fields consistently differ between mouse ganglion cell types. We extracted four major cell types from the mouse retina, and the first panels (A-H) depict an example recording with a good representation of all four. (**A**) Temporal filters (normalized to unit norm) determined from white noise. Note that transient cells have more biphasic filters. (**B**) Spike-train autocorrelograms (normalized to unit sum) for the four cell types. (**C**) Receptive field mosaics for the four types show tiling. Contours were shrunk by 75% for clarity. (**D**) White-noise spatial filter for a highlighted cell from the mosaic. (**E**) The subunit grid map for the example cells highlighted in (D). (**F**) Subunit profiles for the cells in the mosaic. Shaded error bars show 95% confidence intervals. (**G**) Nonlinearities of the fitted models. (**H**) All subunit grid models could fit responses reasonably well, as demonstrated by the actual tuning surface (top) vs. its prediction from the SG model (bottom), for all cells. (**I**) Average subunit diameters for all four types. ON brisk sustained cells had smaller subunit diameters relative to ON transient (64.3 ± 25.4 vs. 76.1 ± 30.8 µm, mean ± SD, $p = 0.003$, Wilcoxon rank-sum test), and the inverse relationship was true between OFF brisk sustained and transient cells (76.1 ± 22.7 vs. 61.4 ± 20.0 µm, mean ± SD, $p < 10^{-12}$, Wilcoxon rank-sum test). (**J**) Subunit numbers for all types. (**K**) Coverage of the subunit mosaic. Overall, OFF brisk transient cells had the lowest coverage, significantly larger than OFF brisk sustained cells (1.50 ± 0.87 vs. 1.94 ± 0.62, mean ± SD, $p < 10^{-20}$, Wilcoxon rank-sum test). (**L**) Compared to ON brisk sustained cells (mean ± SD, Wilcoxon rank-sum test), the absolute valued of the nonlinearity asymmetry was higher in ON brisk transient

cells (0.64 ± 0.17 vs. 0.76 ± 0.15, $p < 10^{-6}$). We found the same relationship between OFF brisk sustained and transient cells (0.31 ± 0.22 vs. 0.54 ± 0.22, $p < 10^{-20}$). (**M**) Performance improvement for natural images was higher than unity (mean ± SD, Wilcoxon sign-rank test) for all types (ON brisk sustained: 1.07 ± 0.22, $p < 10^{-17}$, ON brisk transient: 1.10 ± 0.28, $p < 10^{-8}$, OFF brisk transient: 1.03 ± 0.04, $p < 10^{-17}$), except for OFF brisk sustained cells (1.03 ± 0.45, $p = 0.13$). Performances were calculated as the absolute value of the Spearman's ρ of the subunit grid model over the DoG model. Numbers denote the number of cells in each group.

Although subunit nonlinearities were fit with sigmoid functions, as is generally observed for excitation contrast-response functions (Schwartz et al., 2012; Turner and Rieke, 2016), they differed in their relative threshold and saturation relative to the subunit input. Nonlinearities of OFF brisk sustained cells were quite symmetric around zero (Fig. 2L). A subunit with a symmetric nonlinearity provides equal amounts of activation and deactivation to the ganglion cell when activated with preferred and non-preferred contrasts, respectively. Thus, preferred and non-preferred contrast contributions can cancel themselves out, leading to relatively linear OFF brisk sustained cells. ON brisk transient cells had the most asymmetric nonlinearities, also compared to their sustained counterpart, a relationship previously highlighted (Kuo et al., 2016). The asymmetry of nonlinearities was a good indicator of whether subunit grid models improved response predictions (relative to DoG models) for natural images: indeed, the highest improvement came for ON brisk transient cells, while there was almost no improvement for OFF brisk sustained cells (Fig. 2M).

An interesting observation occurred when we examined the population of OFF brisk transient cells: nonlinearities looked strikingly different between recordings from the dorsal and ventral retina (Fig. S3). Dorsal OFF transient cells showed strong saturation, and had a negative asymmetry index, while ventral cells showed a more typical rectification profile. This difference was also evident in their encoding on spatial contrast in natural images, which we previously measured (Karamanlis and Gollisch, 2021). Dorsal transient OFF alpha ganglion cells have more sustained light responses than their ventral counterparts (Warwick et al., 2018), and we verified this dichotomy in our recordings by examining differences in response transiency between the dorsal and ventral retina.

Together, our results suggest that subunit grid models offer compact descriptions of nonlinear processing within the receptive fields of ganglion cells, and they can be used to discover within-type differences across the retinal surface. Furthermore, we establish that cell-type specific nonlinear models may be needed to improve response predictions for naturalistic stimuli.

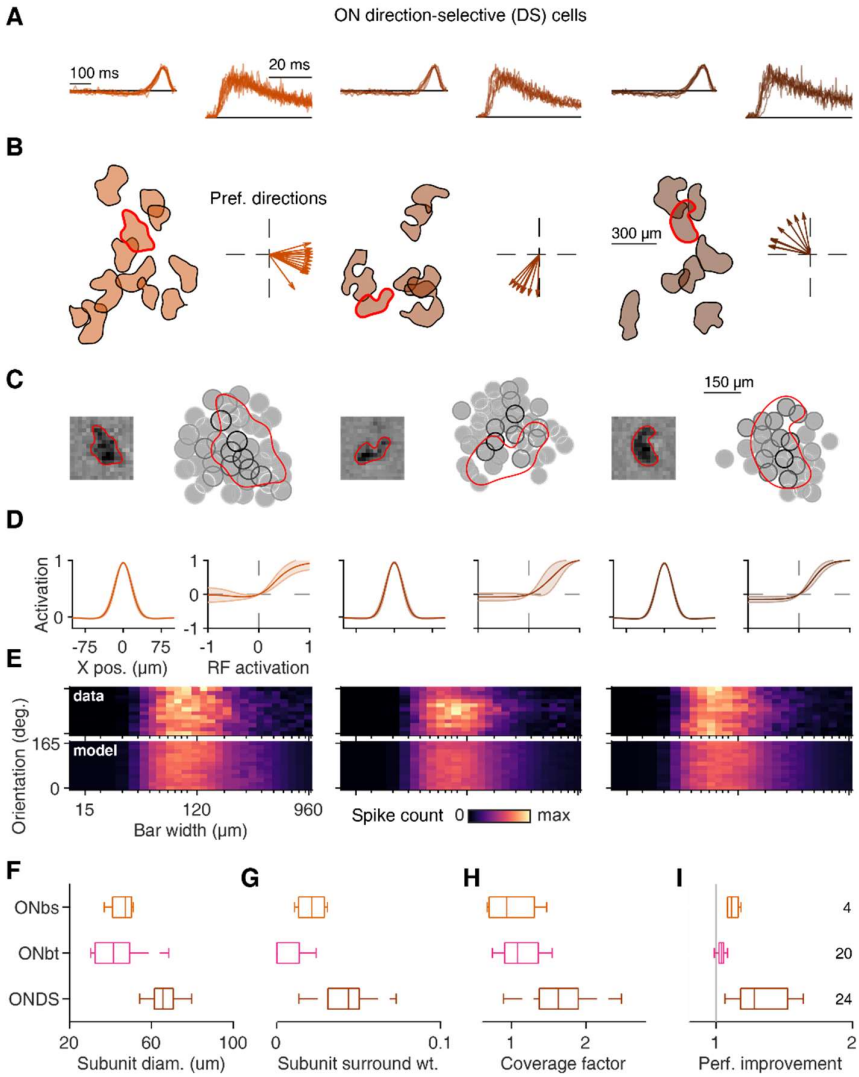

Fig. 3. Subunit grid models of ON direction-selective ganglion cells. (A) ON direction-selective (DS) cells have monophasic filters, with autocorrelograms suggesting sustained spiking responses. (**B**) ON DS cells were split into three groups, based on their preferred directions under a drifting grating stimulus. (**C**) Note that the subunit map only roughly matches the white-noise contour. (**D**) Subunit receptive fields of ON-OFF DS cells have stronger surrounds compared to alpha cells. Shaded error bars depict 95% confidence intervals. (**E**) The tuning surfaces of ON DS cells reveal strong suppression for higher spatial scales. Compared to ON brisk sustained cells (mean ± SD, Wilcoxon rank-sum test), ON DS cell from the same recording have (**F**) larger subunit diameters (45.6 ± 6.3 µm vs. 65.8 ± 7.4 µm, p = 0.003), (**G**) larger weights of subunit surround (0.021 ± 0.009 vs. 0.043 ± 0.015, p = 0.017), (**H**) larger coverage factors (1.00 ± 0.38 vs. 1.63 ± 0.39, p = 0.014), and (**I**) larger performance improvement for natural images (1.12 ± 0.05 vs. 1.50 ± 0.73, p = 0.028). Numbers denote the number of cells in each group.

The subunit grid model captures the existence of a strong subunit surround

ON direction selective (DS) ganglion cells gave us a way to examine the nonlinear receptive field of another type besides the four alpha types, because DS cells could be readily identified in some of our recordings and were sufficiently many. DS cells are comprised of three

subtypes, each with a different preferred direction, which we could extract from responses to drifting gratings (Figure 3A-B). Despite differing in preferred direction, all subtypes had very similar nonlinear receptive fields, with strongly rectifying nonlinearities (Fig. 3C-D). Closely examining the tuning of responses with respect to spatial frequency and orientation (Fig. 3E), we found that ON DS cells were mostly silent for wide-bar gratings. Such a strong surround suppression has been previously measured in ON DS cells, both with flashed spots of different sizes and drifting gratings (Dhande et al., 2013; Mani and Schwartz, 2017). The subunit grid model could reliably capture this dependence with subunits that had a pronounced surround (Fig. 3D), indicating that the strongest component of surround suppression is presynaptic. The subunits of ON DS cells were larger than the subunits of both ON brisk sustained and transient cells (Fig. 3I). This size asymmetry may reflect previous measurements: anatomical data suggest that ON DS cells receive input mainly from type 5 bipolar cells (Matsumoto et al., 2019), while ON alpha cells from type 6 (Schwartz et al., 2012), and type 5 bipolar cell have larger receptive fields (Franke et al., 2017). As also suggested by the spatial frequency and orientation tuning, the surround suppression of ON DS subunits was stronger than ON brisk sustained and transient cells (Fig. 3G), and the subunits densely covered the receptive field (Fig. 3H). Given their strong nonlinearities, the improvement for natural images was the highest (Fig. 3I) amongst the ON types. All these observations point to the fact that it is possible to distinguish many presynaptic parameters just by examining nonlinear receptive field properties.

Naturalistic movies with imprinted mouse gaze shifts drive receptive field nonlinearities

An ultimate goal of computational encoding models is to predict responses to naturalistic stimuli, not only with natural spatial structure, but also with natural temporal dynamics. To this end, we generated movies (Fig. 4A) by shifting natural images according to movements from the horizontal component of mouse gaze (Arne Meyer, personal communication), previously collected from freely-moving mice (Meyer et al., 2020). We presented these natural movies to the retina and collected ganglion cell spiking responses over multiple trials (Fig. 4B). We first examined how well linear-nonlinear (LN) models, with filters fitted from white noise, predicted responses to a repeated natural movie (Fig. 4C). We quantified model performance with a correlation coefficient between cell responses and model predictions, corrected for the retinal response variability (Schoppe et al., 2016) and this correction allowed us to compare prediction performance between natural images and movies. Overall, the performance to movies was lower compared to images (Fig. 4D), given that temporal dynamics present in the movies may drive multiple other processes like light intensity or contrast adaptation (Rieke and Rudd, 2009). Cells with more asymmetric subunit nonlinearities, estimated from flashed gratings, showed lower LN model performance for the natural movie (Fig. 4D). Thus, a model with subunit nonlinearities could potentially improve response predictions for spatially nonlinear types.

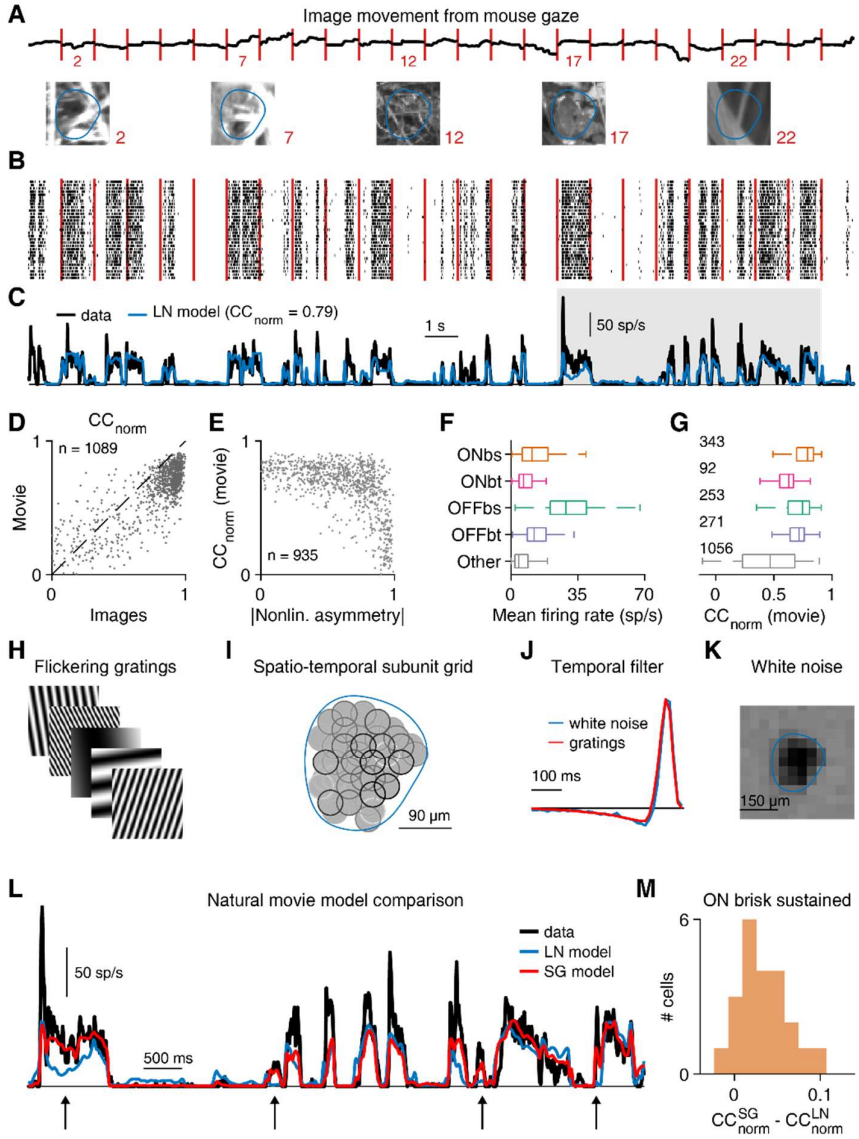

Fig. 4. Subunit grid models improve predictions of responses to movies with naturalistic temporal dynamics. (A) Natural movies were presented to the retina. The movies were made by shifting natural images. (**B**) Raster plot of an ON brisk sustained cell response to 30 repeats of the same movie. Image transitions are noted with the red lines. (**C**) Average firing rate (binned at the screen refresh rate of 75 Hz) for a single cell. Prediction for an LN model (blue), where the linear part came from white noise but the nonlinearity from held-out natural movies. (**D**) LN model predictions were worse for natural movies compared to natural images (0.62 ± 0.22 vs. 0.75 ± 0.28, mean ± SD, $n = 2000$, $p < 10^{-100}$, Wilcoxon sign-rank test). **(E)** Model performance negatively correlates with the absolute value of the nonlinearity asymmetry, which was measured from models fitted with flashed gratings (Spearman's rho = -0.57, $n = 935$, $p < 10^{-100}$). (**F**) Average firing rate during the presentation of the movie for the four alpha cell types (colored) and the rest of the recorded cells (gray). The firing rate of ON brisk transient was higher than the rest of the cells**.** (**G**) Performance for movies depends on cell type, with both ON (0.59 ± 0.13 vs. 0.76 ± 0.11, mean ± SD, $p < 10^{-27}$, Wilcoxon rank-sum test) and OFF transient (0.70 ± 0.10 vs. 0.70 ± 0.15, mean ± SD, $p = 0.012$, Wilcoxon rank-sum test) cells showing significantly lower performance than their sustained counterparts. Numbers denote the number of cells in each group. (**H**) Flickering gratings were presented to the retina at 75 Hz refresh rate. (**I**) A subunit map obtained by fitting responses to flickering gratings. (**J**) The temporal component of the fitted model matches the temporal component estimated from white noise

stimulation. (**K**) The spatial component of the cell determined from white-noise stimulation. (**L**) Predictions of the subunit grid model (CCnorm = 0.88) were better compared to the Gaussian model (CCnorm = 0.79). The highlighted part of (C) is shown, and arrows point to response parts captured by the subunit grid but not the Gaussian model. (**M**) For a population of ON brisk sustained cells (n = 22), response predictions of the subunit grid model were better compared to the white noise LN model (0.86 ± 0.03 vs. 0.83 ± 0.05, mean ± SD, $p < 10^{-3}$, Wilcoxon sign-rank test).

Most of the recorded cells responded vigorously to the movie, but the four types of brisk cells responded more strongly than the rest (Fig. 4F), as expected from their high evoked firing rates and relatively weak receptive field surrounds (Krieger et al., 2017). Standard LN model predictions differed in their accuracy depending on the cell type (Fig. 4G), highlighting that LN models may not be sufficient for describing responses of all cell types (Karamanlis and Gollisch, 2021). For example, ON brisk sustained cells had the highest prediction performance, while ON brisk transient the lowest. To close the gap between model predictions and cell responses, we fitted spatiotemporal subunit grid models to ON brisk sustained cell responses under sinusoidal gratings flickering in a rapid succession (Fig. 4H). Besides a weight map of subunits (Fig. 4I), we also obtained a temporal component associated with the subunits. The obtained temporal components typically matched temporal filters obtained from white noise stimulation (Fig. 4J), probably because flickering gratings were also uncorrelated in time. Similar to flashes, the receptive field was decomposed into small functional subunits of a diameter around 40-50 μm, matching the subunit inputs to ON alpha cells (Schwartz et al., 2012).

We then used spatiotemporal subunit grid models to predict responses to natural movies. Compared to the LN model, subunit grid models could better predict both response transients, but also periods of silence (Fig. 4L). Overall, prediction performance was higher compared to LN models obtained by white noise (Fig. 4M), but also LN DoG models fit with gratings (data not shown). These results suggest that the nonlinear receptive field structure is not only relevant under spatial naturalistic stimulation, but also under temporal naturalistic stimulation.

Subunit structure alone is insufficient for explaining decorrelation in the retinal output

What are the functional consequences of a nonlinear receptive field for the retinal encoding of natural scenes? The theory of efficient coding (Barlow, 1961) predicts that the outputs of retinal neurons will be more decorrelated compared to the highly correlated (in space and time) naturalistic stimuli. However, ganglion cell responses to simpler stimuli may retain a large part of these correlations (Pitkow and Meister, 2012; Simmons et al., 2013). We aimed to test the contribution of retinal nonlinearities to the efficient encoding of natural scenes. We first confirmed that the retinal output retains pairwise correlations within single cell types (Fig. 1A-B), using our natural movie with both natural spatial structure and animal-specific temporal dynamics. These correlations are evident for all four brisk types (Fig. 1C), and they have a strong dependence on distance. We then compared pairwise correlations between cell responses and predictions of LN models fit to the same cells: LN models were always more correlated than the actual responses, indicating that more elaborate mechanisms are needed to lower the redundancy in the responses of adjacent cells.

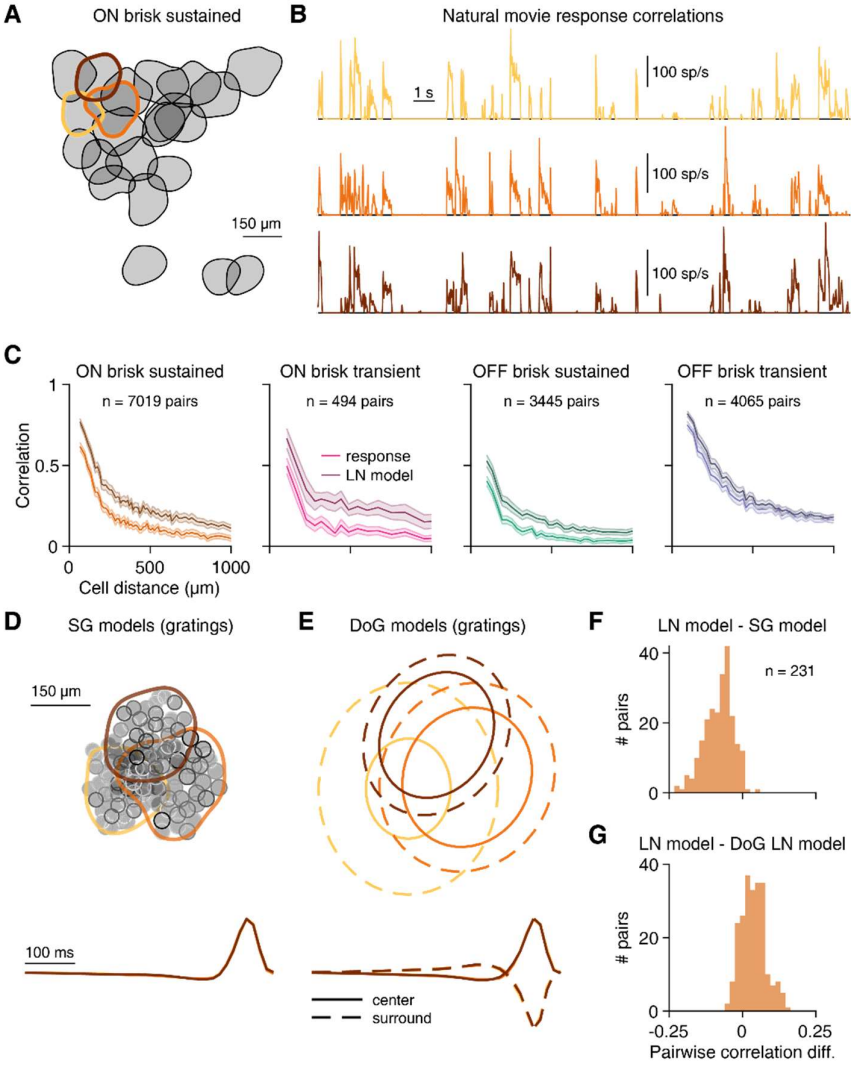

Fig. 5. Response decorrelation to a natural movie requires a surround more than a nonlinear receptive field. (**A**) A mosaic of ON brisk sustained cells. (**B**) The responses of three neighboring cells are correlated. (**C**) Response correlations over receptive field distance for the three cells. For all types and over all distances, LN model correlations were higher than the actual response correlations (mean ± SD, Wilcoxon sign-rank test): ON brisk sustained (0.24 ± 0.19 vs 0.14 ± 0.16, $p < 10^{-100}$), ON brisk transient (0.28 ± 0.19 vs 0.14 ± 0.14, $p < 10^{-100}$), OFF brisk sustained (0.16 ± 0.15 vs 0.08 ± 0.12, $p < 10^{-100}$), OFF brisk transient (0.31 ± 0.20 vs 0.28 ± 0.19, $p < 10^{-100}$). (**D**) Maps of subunit grid (SG) models (top) for the three cells outlined in (A). Overlaid are white-noise contours for comparison. Each subunit has a timecourse which matches for all ON brisk sustained cells (bottom). (**E**) Difference-of-Gaussians models for the three cells, built from responses to flickering gratings. Each cell has a Gaussian center (solid line) and surround (dotted line), here shown at 2sigma. Center and surround have different temporal filters (bottom), that match between the three neighboring cells. (**F**) The SG models are more correlated than the LN models (0.44 ± 0.20 vs 0.52 ± 0.19, mean ± SD, $p < 10^{-38}$, Wilcoxon sign-rank test).). (**G**) The opposite relationship exists between the LN model from white noise and an LN model with a spatiotemporal DoG architecture (0.44 ± 0.20 vs 0.40 ± 0.19, mean ± SD, $p < 10^{-25}$, Wilcoxon sign-rank test).

Because nonlinearities in the retinal output are responsible for a lot of the decorrelation (Pitkow and Meister, 2012), we also hypothesized that also nonlinearities upstream of

ganglion cells might decorrelate responses, an idea also suggested from nonlinear receptive field simulations (Maheswaranathan et al., 2018). Here, we used fitted subunit grid models to test whether nonlinearities in spatial integration could lower pairwise correlations relative to the LN model. However, we observed that the outputs of subunit models were more correlated that the LN model (Fig. 5F), despite providing more accurate response predictions. Because our spatiotemporal subunit models were missing a surround component, we hypothesized that a surround might be the source of the missing decorrelation, as originally predicted in the seminal works of Atick and Redlich (1990, 1992). We used the flickering gratings to fit difference-of-Gaussians LN models that had a separate temporal filter for the surround, which matched between ON brisk sustained cells (Fig. 5E). DoG LN models could indeed decrease pairwise correlations (Fig. 5G) relative to the baseline LN model from white noise. Given these results, nonlinear subunit models may require either subunit or global surrounds to correctly capture pairwise correlations between the output channels of the retina, and potentially further increase prediction performance. Thus, nonlinear processing in the retina likely fulfills computational roles extending beyond decorrelation.

Discussion

Our data show that nonlinear receptive field properties form an axis of variation in the mammalian retina that is parallel to classical linear receptive field characterizations. The nonlinear receptive fields of different cell types may differ not only in how they integrate subunits, but also in the properties of the subunits themselves. Furthermore, we reveal new asymmetries between mammalian ON and OFF cells (Ravi et al., 2018), with spatially nonlinear ON and spatially linear OFF alpha sustained cells, and small-subunit OFF and large-subunit ON alpha transient cells. We also show that the particular sensitivity of OFF alpha transient cells to homogenous stimuli is a property of the dorsal retina, further supporting the idea of regional variation within cell types (Heukamp et al., 2020).

While nonlinear subunit models are typically fit with ganglion cell responses to white noise (Liu et al., 2017; Shah et al., 2020), most experimentalists use gratings to evaluate spatial nonlinearities (Borghuis et al., 2013; Crook et al., 2008; Turner and Rieke, 2016). This structured stimulation offers the ability to efficiently sample the underlying system, and has been used to probe linear receptive fields in other parts of the visual system (Mineault et al., 2016; Ringach et al., 1997). Here, sinusoidal gratings allowed us to sample receptive field structure with an increased resolution and to decompose said structure into smaller nonlinear subunits. A similar principle is also employed by structured illumination microscopy (Gustafsson, 2000), a technique that uses grating-shaped laser light and computational reconstruction to generate super-resolved fluorescence images.

Besides in the retina, our convolutional approach of fitting nonlinear receptive field models can also be used in illuminating the receptive field substructure of neurons in early visual cortices, which may also be subunit-like (Liu et al., 2016; Vintch et al., 2015). These subunits have receptive fields more complex than difference-of-Gaussians and are typically Gabor-like. However, analytic calculations of a Gabor subunit's activation in response to sinusoidal gratings already exist (Soodak, 1986) and can be used in conjunction with our subunit grid method to build subunit models and infer integration nonlinearities.

The subunit grid method is a general-purpose nonlinear modeling framework that can capture ganglion cell responses to natural images even when the cells have quite different properties. Our method improves on previous nonlinear subunit extraction methods that have only been applied to small populations of ganglion cells, typically consisting of a few cell types (Liu et

al., 2017; Shah et al., 2020). Subunit grid models applied to natural movies could also improve response predictions, but we found that upstream nonlinearities alone are insufficient to explain response decorrelation under a natural movie. Instead, the existence of a strong inhibitory surround better captures decorrelation. Previous work already hinted at decorrelation happening at the level of bipolar cells through inhibition (Franke et al., 2017). We conclude that enhancing spatiotemporal subunit models with surround inhibition is expected to further decrease response correlations, to match the ones existent intrinsically existent in the retinal output. Since surround inhibition may also affect spatial integration (Turner et al., 2018), its implementation in subunit grid models may also improve response predictions to naturalistic stimuli with temporal dynamics.

Taken together, our results provide a basis for a new class of computational models of the retina. These models have a high predictive power for natural stimuli, show generality over cell types and are also able to capture functional variation across the retinal surface. Because our models can be easily implemented with presynaptic surround inhibition, they are expected to further improve predictions to naturalistic movies and explain the decorrelation happening in the retinal output.

Acknowledgments: We thank Arne Meyer for providing mouse horizontal gaze traces, and Alexander Ecker for discussion on fitting the subunit grid models.

Funding:

This work was supported by the Boehringer Ingelheim Fonds (DK), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 724822) and by Deutsche Forschungsgemeinschaft (DFG; German Research Foundation), project IDs 154113120 (SFB 889, project C01) and 432680300 (SFB 1456, project B05).

Author contributions:

DK: Conceptualization, Data curation, Methodology, Investigation, Formal Analysis, Project administration, Software, Validation, Visualization, Writing – original draft

TG: Methodology, Funding, Project administration, Resources, Supervision, Writing – review & editing

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data and analysis code will be made available upon the publication of this manuscript.

Materials and Methods

Animals

We used 12 retina pieces from 8 adult wild-type mice of female sex (C57BL/6J), mostly between 7-15 weeks old (except for one 23-week-old). All mice were housed in a 12-hour light/dark cycle. Experimental procedures were in accordance with national and institutional guidelines and approved by the institutional animal care committee of the University Medical Center Göttingen, Germany. No statistical methods were used to predetermine sample size.

Tissue preparation, electrophysiology, and spike-sorting

We cut the globes along the ora serrata, removing the cornea, lens, and vitreous humor. The resulting eyecups were cut in half to allow two separate recordings. Based on anatomical landmarks, we performed the cut along the midline, and marked dorsal and ventral eyecups. Before the start of each recording, we isolated retina pieces from the eyecups. We placed the pieces ganglion cell-side-down on planar multielectrode arrays (Multichannel Systems; 252 electrodes; 30 mm diameter, either 60- or 100-mm minimal electrode distance) with the help of a semipermeable dialysis membrane (Spectra Por), stretched across a circular plastic holder (removed before the recording). The arrays were coated with poly-D-lysine (Merck Millipore). Dissection and mounting were performed under infrared light on a stereo microscope equipped with night-vision goggles.

Throughout the recording, retinal pieces were continuously superfused with the oxygenated Ames solution flowing at 5-6 ml/min. The bath solution was heated to a constant temperature of 34°C-35°C via an inline heater in the perfusion line and a heating element below the array. Extracellular voltage signals were amplified, bandpass filtered between 300 Hz and 5 kHz, and digitized at 25 kHz sampling rate. We used Kilosort (Pachitariu et al., 2016) for spike-sorting. To ease manual curation, we implemented a channel selection step from Kilosort2 in the pipeline, by discarding channels that did not contain any threshold crossings. We curated Kilosort's output through phy, a graphical user interface for visualization, and selected only well-separated units with clear refractory periods in the autocorrelograms. In a few cases, we had to merge units with temporally misaligned templates; we aligned the spike times by finding the optimal shift through the cross-correlation of the misaligned templates.

Visual stimulation

Visual stimuli were generated and controlled through custom-made software, based on Visual C++ and OpenGL. Different stimuli were presented sequentially to the retina through a gamma-corrected monochromatic white OLED monitor (eMagin) with 800 x 600 square pixels and 75 Hz refresh rate. The monitor image was projected through a telecentric lens (Edmund Optics) onto the photoreceptor layer of the retina, and each pixel's side measured 7.5 mm on the retina. All stimuli were presented on a background of low photopic light levels (~4000 R*/rod/s), and their mean intensity was always equal to the background. We fine-tuned the focus of stimuli on the photoreceptor layer before the start of each experiment by visual monitoring through a light microscope and by inspection of spiking responses to contrast-reversing gratings with a bar width of 30 μm.

Receptive field characterization

To characterize spatial and temporal response properties of the recorded ganglion cells, we used a spatiotemporal binary white-noise stimulus (100% contrast) consisting of a checkerboard layout with flickering squares, 37.5 μm on the side. The stimulus update rate was 75 Hz. We then calculated spike-triggered averages (STAs) over a 500 ms time window (Chichilnisky, 2001), and extracted spatial and temporal filters for each cell as previously described (Freeman et al., 2015; Rhoades et al., 2019). Briefly, the temporal filter was calculated from the average of STA elements whose absolute peak intensity exceeded 4.5 robust standard deviations of all elements. The robust standard deviation of a sample was defined as 1.4826 times the median absolute deviation of all elements. The spatial filter was obtained by projecting the spatiotemporal STA on the temporal filter. We also calculated spike-train autocorrelation functions under white noise, using a discretization of 0.5 ms. For plotting and subsequent analyses, all autocorrelations were normalized to unit sum.

For each cell, a contour was used to summarize the spatial receptive field (RF). We upsampled the spatial RF to single-pixel resolution, and then blurred it with a circular Gaussian of $\sigma = 4$ pixels. We extracted RF contours using MATLAB's "contour" function at 25% of the maximum value in the blurred filter. In some cases, noisy STAs would cause the contour to contain points that laid further away from the actual spatial RF. Thus, we triaged the contour points, and removed points that exceeded 20 robust standard deviations of all consecutive distances. The center of each RF was defined as the median of all contour points, and its area as the area enclosed by the contour.

Ganglion cell type identification

We used responses to a barcode stimulus (Drinnenberg et al., 2018) to cluster cells in functional types within each single recording. The barcode is a one-dimensional variation of light intensity that moves across the screen. In particular, the barcode pattern had a length of 12,495 mm and was generated by superimposing sinusoids of different spatial frequencies (f) with an 1/f weighting. The constituent sinusoids had spatial frequencies between 1/12495 and 1/120 μm^{-1} (separated by 1/12495 μm^{-1} steps) and had pseudorandom phases. The final barcode pattern was normalized so that the brightest (and dimmest) values corresponded to 100% (and -100%) Weber contrast from the background. The pattern moved horizontally across the screen at a constant speed of 1125 μm/s, and the stimulus was repeated 10 to 20 times. Obtained spike trains were converted into firing rates using 20-ms bins, and Gaussian smoothing with a $\sigma = 20$ ms. We quantified cell reliability with a symmetrized coefficient of determination (R^2), as described previously (Karamanlis and Gollisch, 2021). We only included cells with a symmetrized R^2 value of at least 0.1, that were not putative direction-selective cells (see below).

We used barcode average responses to generate a pairwise similarity matrix, as described previously (Drinnenberg et al., 2018). We defined the similarity between each pair of cells as the peak of the normalized cross-correlation function between the spike rate profiles of the two cells. To obtain a final similarity matrix, we multiplied the barcode similarity matrix with three more similarity matrices, obtained from RF response properties. The first two were generated by computing pairwise correlations between both the temporal filters and the autocorrelation functions of each cell. The third one used RF areas and was defined as the ratio of the minimum of the two areas over their maximum.

We converted the combined similarity matrix to a distance matrix by subtracting it from one. We then computed a hierarchical cluster tree with MATLAB's "linkage" function and the largest distance between cells (complete linkage) as a measure for cluster distance. The tree was used to generate 45 clusters. This procedure yielded clusters with uniform temporal components and autocorrelations, and RF overlaps expected from tiling, but typically resulted in oversplitting functional ganglion cell types. Thus, we manually merged clusters with at least two cells, based on the similarity of properties used for clustering and RF tiling. To incorporate cells that were left out of the clustering because of the barcode quality criterion, we expanded the clusters obtained after merging. For each unclustered cell, we calculated Mahalanobis distances to all the obtained clusters. A cell was assigned to a cluster if its distance from the cluster was at most 5σ, but at least 10σ for all other clusters. Our method could consistently identify types with tiling RFs, such as the four brisk types we refer to throughout this work.

Matching cell types to previously identified functional ganglion cell types

We also presented the chirp stimulus, with parameters matching the original description (Baden et al., 2016). From complete darkness, to twice the mean brightness of our OLED screen. The stimulus was presented 10-20 times. To compare to the calcium traces in the database, we convolved our spiking data with the calcium kernel reported in the original paper (Baden et al., 2016). We then calculated "match-indices" (Román Rosón et al., 2019), by computing correlations to the average traces of each cluster in the database.

For some experiments, we used the responses under spot stimuli (Goetz et al., 2021). Briefly, we flashed one-second-long spots over the retina at different locations and five different diameters (100, 240, 480, 960 and 1200 μm). Between spot presentations, illumination was set to complete darkness, and the spots had an intensity of 100-200 R*/rod/s. For each cell, we estimated a response center, by finding which location drove stronger responses over all five different spots. We only used cells whose estimated response center for the spots lay no further than 75 μm from the RF center. To calculate similarities to the database, we concatenated vectors of firing rate responses to the five different spots for all types in the database. We then used correlation to match our ganglion cells to the database templates.

In some experiments, we also used saccade gratings to detect image-recurrence-sensitive cells, that correspond OFF transient alpha cells in the mouse retina (Karamanlis and Gollisch, 2021; Khani and Gollisch, 2017; Krishnamoorthy et al., 2017).

Extraction of direction-selective (DS) ganglion cells

To identify DS ganglion cells, we used drifting sinusoidal gratings of 100% contrast, 240 mm spatial period, and a temporal frequency of 0.6 Hz (Sabbah et al., 2017), and analyzed responses as previously described (Karamanlis and Gollisch, 2021). Cells with a mean firing rate of at least 1 Hz and a direction-selectivity-index (DSI) of at least 0.2 (significant at 1% level) were considered putative DS cells. The DSI was defined as the magnitude of the normalized complex sum $\sum_\theta r_\theta e^{i\theta} / \sum_\theta r_\theta$. The preferred direction was obtained as the argument of the same sum. The statistical significance of the DSI was determined through a Monte Carlo permutation approach (Karamanlis and Gollisch, 2021; Liu et al., 2017).

To separate ON from ON-OFF DS cells, we used a moving bar stimulus. The bars had 100% contrast and were moved in eight different directions with a speed of 1125 μm/s. We extracted an response profile to all bars through singular value decomposition, as previously described (Baden et al., 2016), and calculated an ON-OFF index, to determine whether cells responded only to the bar onset (ON), or to both onset and offset (ON-OFF). Cells with an ON-OFF index above 0.4 were assigned as ON DS cells and were grouped into three clusters based on their preferred directions.

Flashed gratings

We generated 1200 different gratings with 25 or 30 different spatial frequencies (f), with half-periods between 15 and 1200 μm, approximately logarithmically spaced. For each grating, we generated 12 or 10 orientations (θ) and 4 spatial phases (φ_o). For a given grating, the contrast value for each pixel with (x, y) coordinates was generated based on the following equation:

$$C(x, y) = \sin(2\pi f(x\cos\theta + y\sin\theta + \varphi_o))$$

Gratings were presented as 200-ms flashes on the retina, separated by a 600-ms or 800-ms gray screen in between. The order of presentation was pseudorandom. We collected spike-count responses to the flashes by counting spikes 20 ms after stimulus onset up to 20 ms after stimulus offset, and typically collected three to five trials per grating. We used tuning surfaces to summarize responses (Fig. 1D), which we generated by averaging responses over trials and spatial phases for each frequency-orientation pair. We then calculated symmetrized R^2 values for the spike counts, and only cells with an R^2 of at least 0.2 were used for further analyses.

Difference-of-Gaussians model

We analytically estimated the activation of a difference-of-Gaussians (DoG) receptive field, by considering the activations of both center and surround elliptical Gaussians to the grating, based on previous calculations (Soodak, 1986). Concretely, the response of a DoG receptive field (r_{DoG}) centered at (x_o, y_o), to a parametric sinusoidal grating (f, θ, φ_o) is

$$r_{DoG}(f, \theta, \varphi_o; x_o, y_o, \sigma_x, \sigma_y, \theta_{DoG}, k_s, w_s) = A_{DoG}(f; \sigma_x, \sigma_y, \theta_{DoG}, k_s, w_s) * \cos\Theta_{DoG}(f, \theta, \varphi_o; x_o, y_o)$$

with the amplitude A_{DoG} given by

$$A_{DoG}(f; \sigma, k_s, w_s) = e^{-2\pi\sigma_{DoG}^2 f^2} - w_s e^{-2\pi(k_s\sigma_{DoG})^2 f^2}$$

with

$$\sigma_{DoG} = \sqrt{\sigma_y^2 \sin^2(\theta + \theta_{DoG}) + \sigma_x^2 \cos^2(\theta + \theta_{DoG})}$$

where standard deviations σ_x and σ_y at x- and y-axes, θ_{DoG} the orientation of one of the principal axes, k_s the scaling for the subunit surround, and w_s a factor determining the relative strength of the surround. The subunit phase Θ_s is given by

$$\Theta_{DoG}(f, \theta, \varphi_o; x_o, y_o) = 2\pi f\sqrt{x_o^2 + y_o^2}\cos(\theta - \tan^{-1}\frac{y_o}{x_o}) + \varphi_o - \pi/2$$

The full DoG response model was

$$R = aN(\beta_{DoG} r_{DoG} + \gamma_{DoG})$$

where $N(x) = (1 + e^{-x})^{-1}$ is a logistic function, β_{DoG} and γ_{DoG} are parameters determining the steepness and threshold of the output nonlinearity, and a is a response scaling factor.

All model parameters $(x_o, y_o, \sigma_x, \sigma_y, \theta_{DoG}, k_s, w_s, \beta_{DoG}, \gamma_{DoG}, a)$ were optimized simultaneously using constrained gradient descent in MATLAB with the following constraints: $\sigma_x, \sigma_y > 7.5$ μm, $-\pi/4 < \theta_{DoG} < \pi/4$, $1 < k_s < 6$, $a > 0$. The cost function we used was negative log-likelihood.

Subunit grid model

We fit all subunit grid models with 1200 subunits, placed in a hexagonal grid around a given RF center location. The center was taken as the fitted center of the DoG model. The subunits were spaced 16 μm apart. Each subunit had a circular DoG profile, with standard deviation of σ and centered at (x_{os}, y_{os}), and its activation was given by

$$r_s(f, \theta, \varphi_o; x_{os}, y_{os}, \sigma, k_s, w_s) = A_s(f; \sigma, k_s, w_s) * \cos \Theta_s(f, \theta, \varphi_o; x_{os}, y_{os})$$

where both amplitude and phase are given by the DoG receptive field formulas for $\sigma_x = \sigma_y = \sigma$.

The full response model was

$$R_{SG} = aN(\sum_{s=1}^{N_{sub}} w_s N(\beta r_s + \gamma) + b)$$

where $N(x) = (1 + e^{-x})^{-1}$ is a logistic function, β and γ are parameters determining the steepness and threshold of the subunit nonlinearity, b determines the baseline activation, w_s are non-negative subunit weights, and a is a response scaling factor.

Fitting and model selection

We optimized subunit grid models using ADAM (Kingma and Ba, 2014), with the following parameters: batch size = 64, $\eta = 0.005$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-6}$. The cost function we minimized was

$$-\frac{1}{N_{sp}} \ln L(\boldsymbol{s_G}, \boldsymbol{r_G}; \boldsymbol{\theta_p}, \boldsymbol{w}) + \lambda \sum_{s=1}^{N_{sub}} w_s \sum_{i \neq s} \frac{w_i}{d_{si}^2}$$

with N_{sp} being the total number of spikes, L the Poisson likelihood, $\boldsymbol{s_G}$ the vector of all grating parameters used, $\boldsymbol{r_G}$ the corresponding spike-count response vector, $\boldsymbol{\theta_p} = (\sigma, k_s, w_s, \beta, \gamma, b, a)$ all the shared model parameters, and $\boldsymbol{w} = (w_1, \ldots, w_{N_{sub}})$ the vector containing all subunit weights. λ controls the regularization strength, which depends on the pairwise subunit distances d_{si}.

After the end of the optimization, we pruned subunit weights with small contributions or weights that ended up outside the receptive field. At first, we set to zero every weight smaller than 5% of the maximum subunit weight. We then fitted a two-dimensional Gaussian to an

estimate of the receptive field, obtained by summing subunit receptive fields weighted by the subunit weights. The weight corresponding to any subunit center lying more than 2 sigma outside that Gaussian was set to zero. To ensure proper scaling of the output nonlinearity, we refitted a scaling factor to the weights after weight pruning.

We used the Bayesian Information Criterion (BIC) to select for the right amount of regularization. In particular, we calculated BIC for a subunit grid model as

$$N_{sub} \ln(N_{data}) - 2\ln(L)$$

where N_{sub} is the number of non-zero subunits, N_{data} is the number of grating-response pairs used to fit the model, and L the likelihood of the fitted model. The model with the lowest BIC balanced good prediction performance and realistic receptive field substructure.

Parameter characterization of the subunit grid model

A coverage value was calculated if there were at least 3 subunits with non-zero weights in the model. It was calculated as the ratio of A over B, where A was the subunit diameter (4σ of the Gaussian), and B was the average subunit distance. For a particular cell, the average subunit distance was calculated over all subunit pairs, weighted by each pair's average subunit weight.

To plot and characterize nonlinearities, we first added an offset so that at an input of zero they show zero output. We then scaled them with the maximum value over the [-1,1] range. Following offsetting and scaling, we calculated nonlinearity asymmetries to quantify the response linearity of subunits:

$$\frac{1 - |\min(N(x))|}{1 + |\min(N(x))|}$$

where $x \in [-1,1]$ and $N(x)$ is the nonlinearity defined over the same range. In most cases we used the absolute value of the nonlinearity asymmetry, to also cover cases where negative values dominated, such as in OFF brisk transient cells.

Checkerboard flashes and responses predictions

The checkerboard flash stimulus was used to test response predictions of the different receptive field models, and we described it before (Karamanlis and Gollisch, 2021). For our experiments, we used a square sampling space instead of a polar one. The checkerboards were flashed for 200 ms, interleaved with either 600 or 800 ms of grey background. Average spike counts were collected as in the case of flashed gratings and only cells with symmetrized R^2 of at least 0.2 were used for further analyses. Response predictions were evaluated by calculating the rank correlation between a cell's responses and model predictions without considering the output nonlinearity. For generating the response maps of Fig. 1J, we fit a logistic nonlinearity between model predictions and cell responses.

Natural images and response predictions

We flashes a series of 120 natural images to the retina, as described previously (Karamanlis and Gollisch, 2021). We used images from the Van Hateren database, which were cropped to their central 512x512 square, and flashed over the multielectrode array. All images were multiplicatively scaled to have the same mean intensity (same as the background).

Interspersed with the natural images, we also presented 80 artificial images. The images were generated as black-and-white random patterns at a single-pixel level, and then blurred with Gaussians of eight different spatial scales (Schwartz et al., 2012). The images were flashes for 200 ms, interleaved with either 600 or 800 ms of grey background. Average spike counts were collected as in the case of flashed gratings and only cells with symmetrized R^2 of at least 0.2 were used for further analyses.

To calculate response predictions for white noise, we used the output of spatial filters that were involved with the natural images. The filters were upsampled to match the resolution of the presented images and normalized with the sum of their absolute values. For models obtained by flashed gratings, DoG receptive fields and subunit receptive fields were instantiated at a single pixel resolution, and then convolved with the natural images. For the subunit model, the convolution outputs were passed through the fitted nonlinearity, and then summed under the subunit weights. The performance for each model was calculated as the rank correlation between the model output (without an explicit output nonlinearity) and cell responses to the natural images.

Flickering gratings and baseline spatiotemporal LN models

We generated 4800 different gratings with 30 different spatial frequencies, between 15 and 1200 μm bar widths, approximately logarithmically spaced. For each grating, we generated 20 orientations and 8 spatial phases. The gratings were presented in a pseudorandom sequence, updated at a 75 Hz refresh rate. Every 1200 frames, we interleaved a unique sequence of 600 frames that was repeated throughout the recording to evaluate response quality.

We fit a spatio-temporal DoG LN model to the grating responses. The temporal filters spanned a duration of 500 ms and were modeled as a linear combination of ten basis functions. The response delay was accounted for with two square basis functions spanning the period of two frames before a spike. The remaining eight of were chosen from a raised cosine basis (Latimer et al., 2019).

Concretely, the spatio-temporal DoG model had the form

$$R = aN(\boldsymbol{r_C^T k_{Ct}} + \boldsymbol{r_S^T k_{St}} + b)$$

where $N(x) = (1 + e^{-x})^{-1}$ is a logistic function, $\boldsymbol{k_{Ct}}$ and $\boldsymbol{k_{St}}$ are separate temporal filters for the center and the surround, b determines the baseline activation, and a is a response scaling factor. The vectors $\boldsymbol{r_C}$ and $\boldsymbol{r_S}$ contain DoG receptive field activations for 500 ms before a particular frame and were calculated based on the same calculations we used for the flashed gratings. The model was fit with nonlinear constrained optimization, with DoG constraints similar as in the case of flashed gratings, and $a > 0$.

Spatio-temporal subunit grid model

We also fit a spatio-temporal subunit grid model to the grating responses. Our strategy was very similar to the grating flash case. We fit all subunit grid models with 1000 subunits, placed in a hexagonal grid around a given RF center location. The center was taken as the fitted center of the DoG model. The subunits were spaced 16 μm apart.

$$R = aN(\sum_{s=1}^{N_{sub}} w_s N(\boldsymbol{r_s^T k_t} + \gamma) + b)$$

where $N(x)$ is a logistic function, $\boldsymbol{k_t}$ is a temporal filter shared for all subunits, γ determines the nonlinearity threshold, b determines the model's baseline activation, a is a response scaling factor, and w_s are non-negative subunit weights. The vectors $\boldsymbol{r_s}$ contain Gaussian subunit activations for 500 ms before a particular frame and for each subunit. The only parameter that is required to fit Gaussian subunits is the standard deviation σ.

We used block-cyclic stochastic gradient descent to fit spatio-temporal models. For each batch of 512 samples, we only updated the parameters belonging to one block. The blocks we used updated:

- the basis coefficients for the temporal filter, the Gaussian standard deviation σ, and the threshold parameter γ.
- the subunit weights w_s.
- The output nonlinearity parameters a and b.

Each block was updated using ADAM gradients. We used the same regularization as in the case of flashed gratings to control for subunit density, and we performed model selection through the BIC.

Natural movies, response predictions, and response correlations

We constructed natural movies based on rationale previously applied for the primate retina (Heitman et al., 2016; Shah et al., 2020). Briefly, the movies consisted of 325 images from the Van Hateren database (van Hateren and van der Schaaf, 1998), shown for one second each, and jittered according to the horizontal gaze component (Meyer et al., 2020) of freely moving mice (Arne Meyer, personal communication). Because the original gaze traces were sampled at 60 Hz, we resampled traces to produce a movie with a refresh rate of 75 Hz. Horizontal gaze is a one-dimensional movement on the nasotemporal axis of the retina: since the relative position of image movement and the retina was unknown during the experiment, we randomly assigned each one-second movement to one of four orientations (0, 45, 90 or 135 degrees). The amplitude of the original movement was given in visual degrees and we transformed it to μm on the retina, using a retinal magnification factor of 31 μm/deg for the mouse. All images were multiplicatively scaled to have the same mean intensity (same as the background). The presented natural movie consisted of multiple cycles of training and test stimuli. The training stimuli consisted of 35 images out of the 300 (sampled with replacement), each paired with a unique movement trajectory. Testing stimuli consisted of 25 distinct natural images, again paired with unique movement trajectories. To extract firing rates for the testing stimulus, spike trains were binned at a single frame resolution, and only cells with a symmetrized R^2 of at least 0.2 were used for following analyses. Model performance was calculated on the testing stimulus, using the normalized correlation coefficient (Schoppe et al., 2016).

All model predictions for natural movies used the testing part for evaluation, and the training part for estimating the output nonlinearity. The output nonlinearity was built a through a histogram method (40 bins) applied on model generator signals (Liu et al., 2015) and was

applied to testing generator signals through linear interpolation. For an LN model built from white noise, we projected movie frames on the upsampled spatial filter (to single-pixel resolution) and convolved the result with the temporal filter. For difference-of-Gaussians LN models built from flickering gratings, we instantiated center and surround filters to single pixel resolution. Again, we projected movie frames on both filters separately, convolved each result with the corresponding temporal filter, and summed the two outputs for obtaining the final generator signal. For subunit grid models, Gaussian filters for each subunit were instantiated at a single pixel resolution and then applied to movie frames, followed by temporal convolution. The subunit nonlinearity fitted from the gratings was then applied to linear subunit outputs, which were then summed with the non-negative pooling weights to obtain the final generator signal.

We calculated movie response correlations between all cell pairs of the same type, using the trial-averaged firing rates of the testing stimulus. We also performed the same analyses for model predictions. To generate correlation-distance curves (Fig. 5C), we sorted pairs by ascending distance, and averaged pair correlations over groups of 90 pairs (for ON brisk transient cells we used 30, because of the limited number of available pairs).

Supplementary Figures

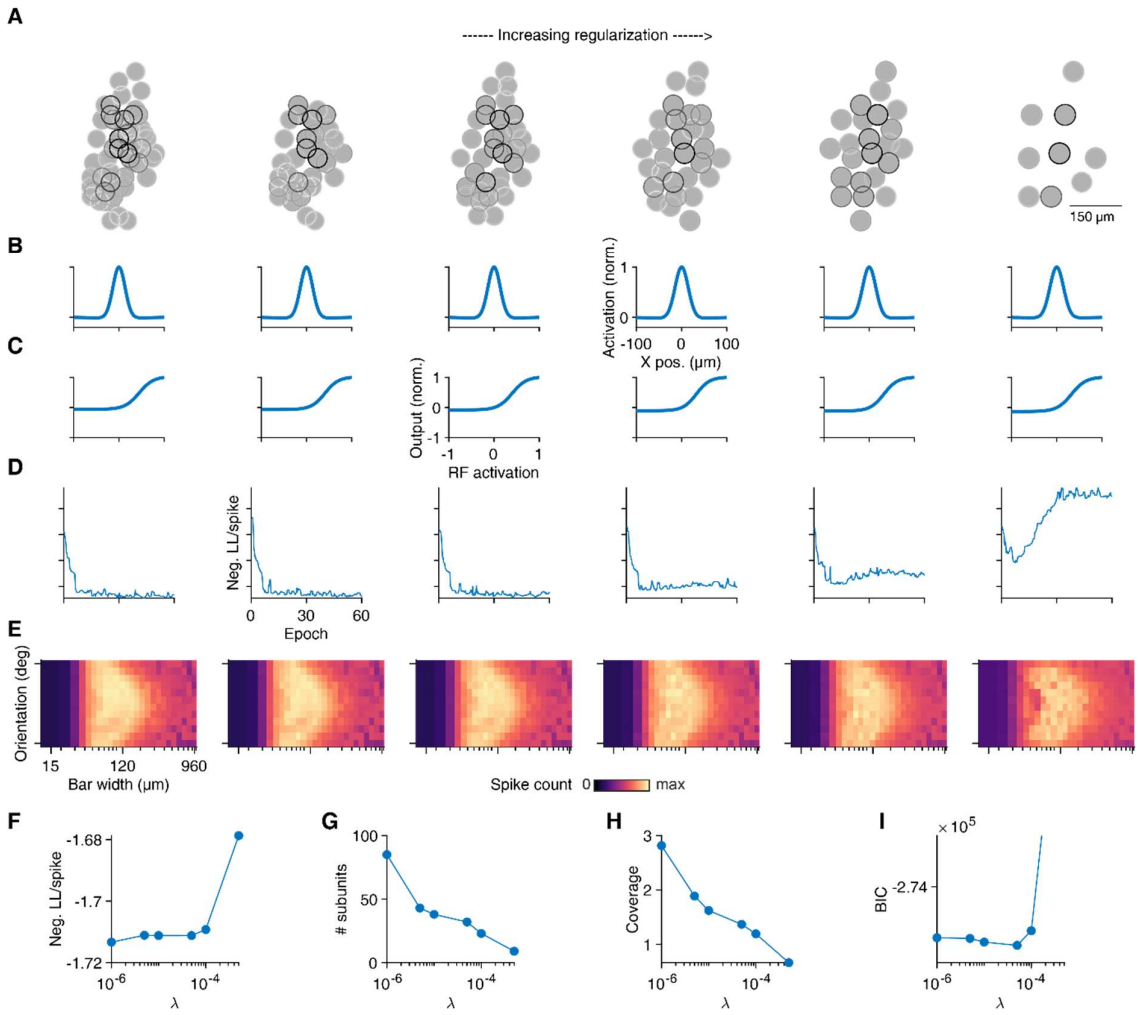

Fig. S1. Weight density regularization decreases subunit coverage with increasing regularization strength. (**A**) Subunit grids fitted to spiking responses of an example retinal ganglion cell (Fig. 1) for six different regularization values. (**B**) Subunit receptive field profiles. (**C**) Subunit nonlinearities. (**D**) Model training curve. (**E**) Tuning curve prediction of the fitted model (**F**) The cost function at the end of the optimization versus the regularization strength. (**G**) The number of subunits decreases with increasing regularization. (**H**) Receptive field coverage decreases with the strength of the applied regularization. (**I**) We used the Bayesian Information Criterion (BIC) to select the best model. We used the number of non-negative subunits as the number of parameters for specifying the BIC.

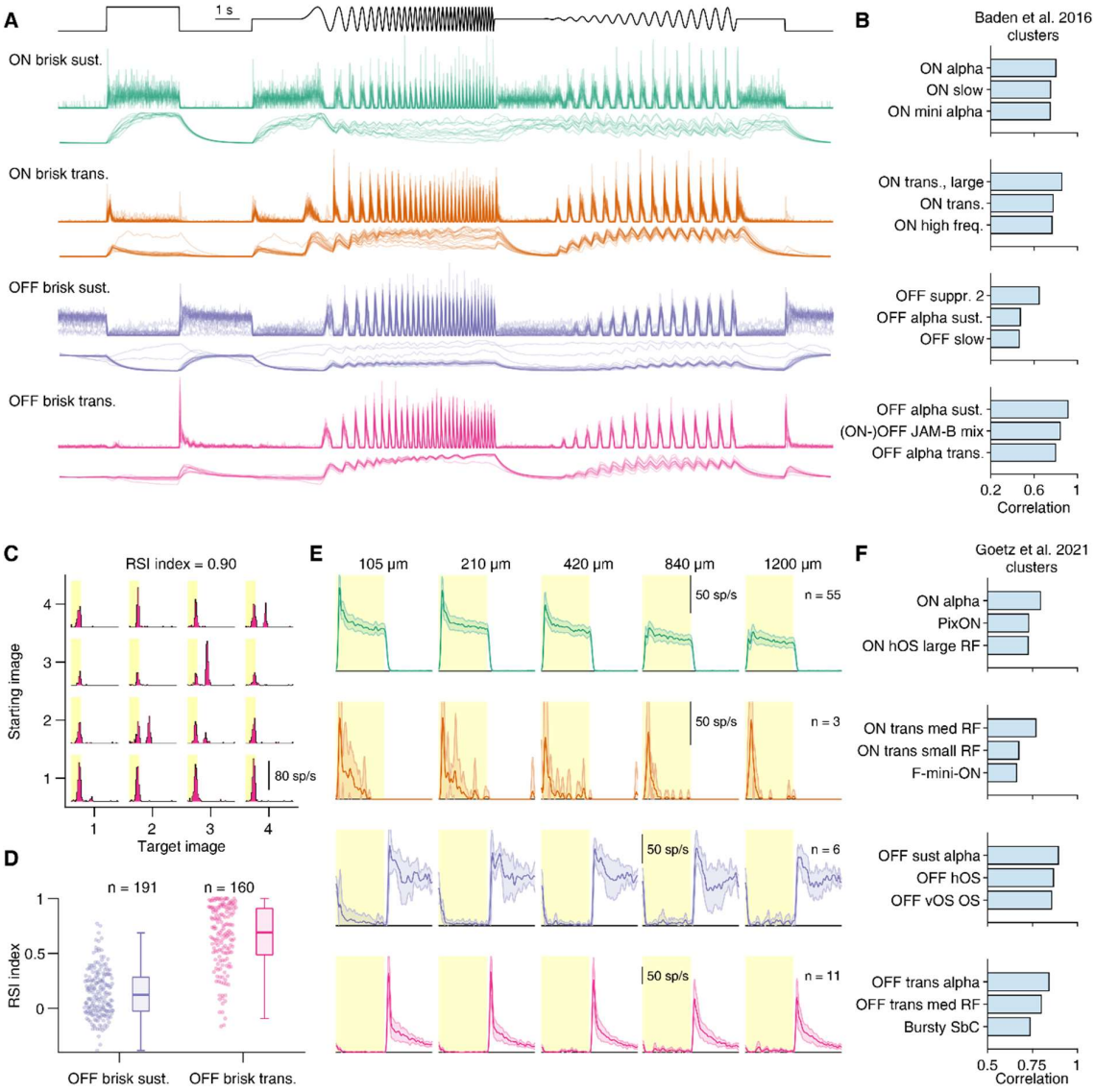

Fig. S2. Mapping identified ganglion cells mosaics to previously characterized cell types in the mouse retina. (**A**) Responses of the four brisk types to the chirp stimulus (top), previously used to classify mouse retinal ganglion cells (Baden et al., 2016). The spiking responses were converted to a calcium-equivalent signal and compared (with Pearson correlation) to the reported means of different clusters (right). (**B**) Average correlation to all clusters. Shown are the top three hits. (**C**) An OFF brisk transient cell showing image recurrence sensitivity, measured with saccade gratings (Krishnamoorthy et al., 2017). This sensitivity was quantified with the recurrence sensitivity index (RSI). (**D**) OFF brisk transient cells had significantly higher RSI indices than OFF brisk sustained cells (mean ± SD vs. mean ± SD, mean ± SD, $p<10^{-50}$, Wilcoxon rank-sum test), and the indices were significantly higher than 0.5 ($p = 0.038$, Wilcoxon sign-rank test), the threshold used for the original characterization. (**E**) Average responses of the four main types to flashed spots of five different sizes. The spots were flashed either within, or very close to the receptive field centers of the selected cells. Shaded error bars are 95% confidence intervals. (**F**) Spot responses were compared to a functional database (Goetz et al., 2021) with correlation. Top three hits are shown, and all match the alpha types. For ON brisk transient cells, the match is medium RF, hypothesized to match the original description of the ON alpha transient cell (Krieger et al., 2017).

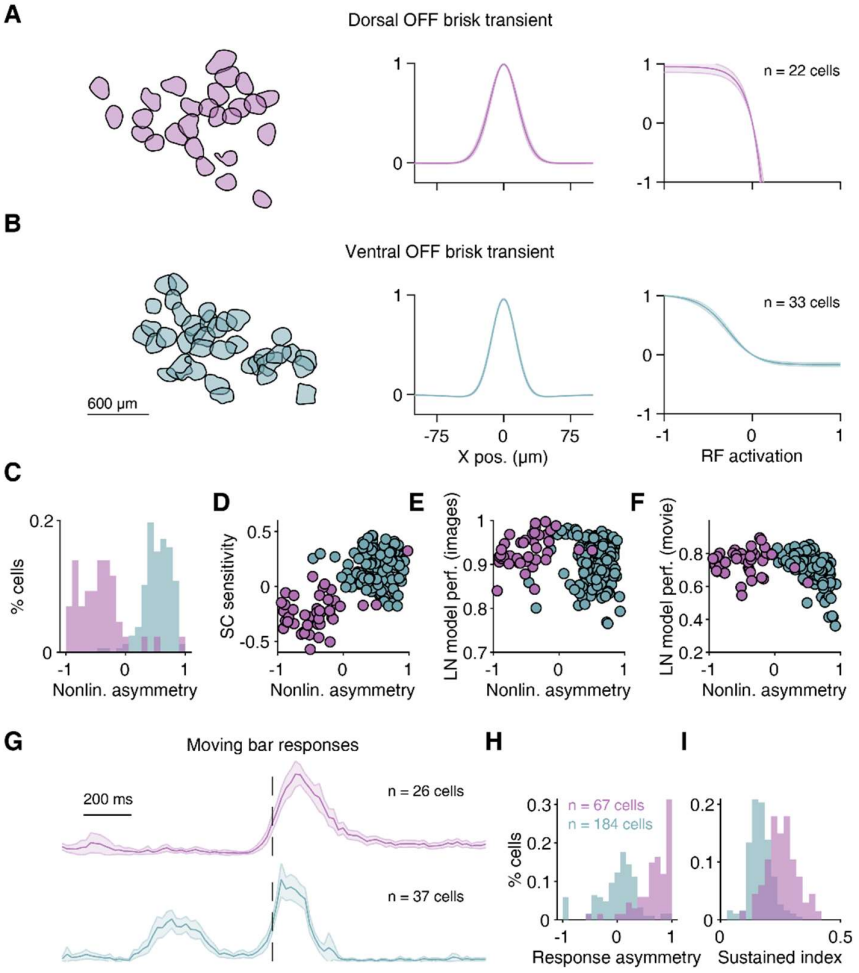

Fig. S3. Subunit nonlinearities differ between dorsal and ventral OFF alpha transient cells. (**A**) Subunit model parameters for a transient OFF alpha cell mosaic in the dorsal retina. (**B**) Same as (A), but for a recording from the same eye coming from the ventral retina. (**C**) Asymmetry in the nonlinearities is evident for all recorded OFF alpha cells. (**D**) The asymmetry affects sensitivity to spatial contrast (SC) in natural scenes. SC sensitivity was defined in our previous work (Karamanlis and Gollisch, 2021). Nonlinearity asymmetry was related to both model performance calculated for natural images (**E**) but also natural movies (**F**). Both dorsal and ventral cells were better predicted with LN models if their nonlinearity asymmetries were close to zero. (**G**) Responses of dorsal (top) and ventral (bottom) OFF alpha transient cells to a moving bar stimulus. The responses correspond to the average bar response over eight different directions. The bars had an ON contrast and approximately entered the receptive field of the cells at the plots start and left the receptive field approximately at the timepoint marked by the dashed lines. (**H**) The responses in the ventral retina showed a peak following the onset of the bar, which we quantified with a response asymmetry index. The index was (Roff - Ron)/(Roff + Ron) was significantly higher for the dorsal retina (0.67 ± 0.31 vs 0.00 ± 0.38, mean ± SD, $p < 10^{-23}$, Wilcoxon rank-sum test). (**I**) Moving bar offset responses in the dorsal retina were more sustained compared to the ventral retina (0.26 ± 0.06 vs 0.23 ± 0.22, mean ± SD, $p < 10^{-12}$, Wilcoxon rank-sum test). The sustained index was defined as the ratio of the average response over the maximum response in the time window following the bar leaving the receptive field center.

References and Notes

Atick, J.J., and Redlich, A.N. (1990). Towards a Theory of Early Visual Processing. Neural Comput. *2*, 308–320.

Atick, J.J., and Redlich, A.N. (1992). What Does the Retina Know about Natural Scenes? Neural Comput. *4*, 196–210.

Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., and Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. Nature *529*, 345–350.

Bae, J.A., Mu, S., Kim, J.S., Turner, N.L., Tartavull, I., Kemnitz, N., Jordan, C.S., Norton, A.D., Silversmith, W.M., Prentki, R., et al. (2018). Digital Museum of Retinal Ganglion Cells with Dense Anatomy and Physiology. Cell *173*, 1293-1306.e19.

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. Sens. Commun. *6*, 57–58.

Behrens, C., Schubert, T., Haverkamp, S., Euler, T., and Berens, P. (2016). Connectivity map of bipolar cells and photoreceptors in the mouse retina. ELife *5*, e20041.

Bleckert, A., Schwartz, G.W., Turner, M.H., Rieke, F., and Wong, R.O.L. (2014). Visual space is represented by nonmatching topographies of distinct mouse retinal ganglion cell types. Curr. Biol. *24*, 310–315.

Bölinger, D., and Gollisch, T. (2012). Closed-Loop Measurements of Iso-Response Stimuli Reveal Dynamic Nonlinear Stimulus Integration in the Retina. Neuron *73*, 333–346.

Borghuis, B.G., Marvin, J.S., Looger, L.L., and Demb, J.B. (2013). Two-Photon Imaging of Nonlinear Glutamate Release Dynamics at Bipolar Cell Synapses in the Mouse Retina. J. Neurosci. *33*, 10972–10985.

Chichilnisky, E.J. (2001). A simple white noise analysis of neuronal light responses. Netw. Comput. Neural Syst. *12*, 199–213.

Chichilnisky, E.J., and Kalmar, R.S. (2002). Functional Asymmetries in ON and OFF Ganglion Cells of Primate Retina. J. Neurosci. *22*, 2737–2747.

Crook, J.D., Peterson, B.B., Packer, O.S., Robinson, F.R., Troy, J.B., and Dacey, D.M. (2008). Y-cell receptive field and collicular projection of parasol ganglion cells in macaque monkey retina. J. Neurosci. *28*, 11277–11291.

Demb, J.B., Zaghloul, K., Haarsma, L., and Sterling, P. (2001). Bipolar Cells Contribute to Nonlinear Spatial Summation in the Brisk-Transient (Y) Ganglion Cell in Mammalian Retina. J. Neurosci. *21*, 7447–7454.

Dhande, O.S., Estevez, M.E., Quattrochi, L.E., El-Danaf, R.N., Nguyen, P.L., Berson, D.M., and Huberman, A.D. (2013). Genetic dissection of retinal inputs to brainstem nuclei controlling image stabilization. J. Neurosci. *33*, 17797–17813.

Drinnenberg, A., Franke, F., Morikawa, R.K., Jüttner, J., Hillier, D., Hantz, P., Hierlemann, A., Azeredo da Silveira, R., and Roska, B. (2018). How Diverse Retinal Functions Arise from Feedback at the First Visual Synapse. Neuron *99*, 117-134.e11.

Enroth-Cugell, C., and Robson, J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. J. Physiol. *187*, 517–552.

Franke, K., Berens, P., Schubert, T., Bethge, M., Euler, T., and Baden, T. (2017). Inhibition decorrelates visual feature representations in the inner retina. Nature *542*, 439–444.

Freed, M. a, and Sterling, P. (1988). The ON-alpha ganglion cell of the cat retina and its presynaptic cell types. J. Neurosci. *8*, 2303–2320.

Freeman, J., Field, G.D., Li, P.H., Greschner, M., Gunning, D.E., Mathieson, K., Sher, A., Litke, A.M., Paninski, L., Simoncelli, E.P., et al. (2015). Mapping nonlinear receptive field structure in primate retina at single cone resolution. ELife *4*, e05241.

Goetz, J., Jessen, Z.F., Jacobi, A., Mani, A., Cooler, S., Greer, D., Kadri, S., Segal, J., Shekhar, K., Sanes, J.R., et al. (2021). Unified classification of mouse retinal ganglion cells using function, morphology, and gene expression. BioRxiv 447922v1.

Gustafsson, M.G.L. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J. Microsc. *198*, 82–87.

van Hateren, J.H., and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. B Biol. Sci. *265*, 359–366.

Heitman, A., Brackbill, N., Greschner, M., Sher, A., Litke, A.M., and Chichilnisky, E.J. (2016). Testing pseudo-linear models of responses to natural scenes in primate retina. BioRxiv 045336v2.

Heukamp, A.S., Warwick, R.A., and Rivlin-Etzion, M. (2020). Topographic Variations in Retinal Encoding of Visual Space. Annu. Rev. Vis. Sci. *6*, 237–259.

Johnson, K.P., Zhao, L., and Kerschensteiner, D. (2018). A Pixel-Encoder Retinal Ganglion Cell with Spatially Offset Excitatory and Inhibitory Receptive Fields. Cell Rep. *22*, 1462–1472.

Karamanlis, D., and Gollisch, T. (2021). Nonlinear spatial integration underlies the diversity of retinal ganglion cell responses to natural images. J. Neurosci. *41*, 3479–3498.

Khani, M.H., and Gollisch, T. (2017). Diversity in spatial scope of contrast adaptation among mouse retinal ganglion cells. J. Neurophysiol. *118*, 3024–3043.

Kingma, D.P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–15.

Kling, A., Gogliettino, A.R., Shah, N.P., Wu, E.G., Brackbill, N., Sher, A., Litke, A.M., Silva, R.A., and Chichilnisky, E.J. (2020). Functional Organization of Midget and Parasol Ganglion Cells in the Human Retina. BioRxiv 240762v1.

Krieger, B., Qiao, M., Rousso, D.L., Sanes, J.R., and Meister, M. (2017). Four alpha ganglion cell types in mouse retina: Function, structure, and molecular signatures. PLoS One *12*, e0180091.

Krishnamoorthy, V., Weick, M., and Gollisch, T. (2017). Sensitivity to image recurrence across eye-movement-like image transitions through local serial inhibition in the retina. ELife *6*, e22431.

Kuo, S.P., Schwartz, G.W., and Rieke, F. (2016). Nonlinear Spatiotemporal Integration by Electrical and Chemical Synapses in the Retina. Neuron *90*, 320–332.

Latimer, K.W., Rieke, F., and Pillow, J.W. (2019). Inferring synaptic inputs from spikes with a conductance-based neural encoding model. ELife *8*, e47012.

Liu, J.K., Gollisch, T., and Gollisch, Tim; Liu Jian, K.. (2015). Spike-Triggered Covariance Analysis Reveals Phenomenological Diversity of Contrast Adaptation in the Retina. PLoS Comput. Biol. *11*, 1–30.

Liu, J.K., Schreyer, H.M., Onken, A., Rozenblit, F., Khani, M.H., Krishnamoorthy, V., Panzeri, S., and Gollisch, T. (2017). Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. Nat. Commun. *8*, 149.

Liu, L., She, L., Chen, M., Liu, T., Lu, H.D., Dan, Y., and Poo, M. (2016). Spatial structure of neuronal receptive field in awake monkey secondary visual cortex (V2). Proc. Natl. Acad. Sci. *113*, 1913–1918.

Maheswaranathan, N., Kastner, D.B., Baccus, S.A., and Ganguli, S. (2018). Inferring hidden structure in multilayered neural circuits. PLOS Comput. Biol. *14*, e1006291.

Mani, A., and Schwartz, G.W. (2017). Circuit Mechanisms of a Retinal Ganglion Cell with Stimulus-Dependent Response Latency and Activation Beyond Its Dendrites. Curr. Biol. *27*, 471–482.

Martersteck, E.M., Hirokawa, K.E., Evarts, M., Bernard, A., Duan, X., Li, Y., Ng, L., Oh, S.W., Ouellette, B., Royall, J.J., et al. (2017). Diverse Central Projection Patterns of Retinal Ganglion Cells. Cell Rep. *18*, 2058–2072.

Matsumoto, A., Briggman, K.L., and Yonehara, K. (2019). Spatiotemporally Asymmetric Excitation Supports Mammalian Retinal Motion Sensitivity. Curr. Biol. *29*, 3277-3288.e5.

McMahon, M.J., Packer, O.S., and Dacey, D.M. (2004). The classical receptive field surround of primate parasol ganglion cells is mediated primarily by a non-GABAergic pathway. J. Neurosci. *24*, 3736–3745.

Meyer, A.F., O'Keefe, J., and Poort, J. (2020). Two Distinct Types of Eye-Head Coupling in Freely Moving Mice. Curr. Biol. *30*, 2116-2130.e6.

Mineault, P.J., Tring, E., Trachtenberg, J.T., and Ringach, D.L. (2016). Enhanced Spatial Resolution During Locomotion and Heightened Attention in Mouse Primary Visual Cortex. J. Neurosci. *36*, 6382–6392.

Pachitariu, M., Steinmetz, N.A., Kadir, S.N., Carandini, M., D., H.K., and Harris, K.D. (2016). Fast and accurate spike sorting of high-channel count probes with KiloSort. In Advances in Neural Information Processing Systems 29, D.D. Lee, M. Sugiyama, U. V Luxburg, I. Guyon, and R. Garnett, eds. (Curran Associates, Inc.), pp. 4448–4456.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature *454*, 995–999.

Pitkow, X., and Meister, M. (2012). Decorrelation and efficient coding by retinal ganglion cells. Nat. Neurosci. *15*, 628–635.

Ravi, S., Ahn, D., Greschner, M., Chichilnisky, E.J., and Field, G.D. (2018). Pathway-Specific Asymmetries between ON and OFF Visual Signals. J. Neurosci. *38*, 9728–9740.

Rhoades, C.E., Shah, N.P., Manookin, M.B., Brackbill, N., Kling, A., Goetz, G., Sher, A., Litke, A.M., and Chichilnisky, E.J. (2019). Unusual Physiological Properties of Smooth Monostratified Ganglion Cell Types in Primate Retina. Neuron *103*, 658-672.e6.

Rieke, F., and Rudd, M.E. (2009). The Challenges Natural Images Pose for Visual Adaptation. Neuron *64*, 605–616.

Ringach, D.L., Sapiro, G., and R.Shapley (1997). A subspace reverse correlation method for the study of visual neurons. Vision Res. *37*, 2455–2464.

Román Rosón, M., Bauer, Y., Kotkat, A.H., Berens, P., Euler, T., and Busse, L. (2019).

Mouse dLGN Receives Functional Input from a Diverse Population of Retinal Ganglion Cells with Limited Convergence. Neuron *102*, 462-476.e8.

Sabbah, S., Gemmer, J.A., Bhatia-Lin, A., Manoff, G., Castro, G., Siegel, J.K., Jeffery, N., and Berson, D.M. (2017). A retinal code for motion along the gravitational and body axes. Nature *546*, 492–497.

Schoppe, O., Harper, O.N.S., Willmore, B.D.B., King, A.J., and Schnupp, J.W.H. (2016). Measuring the Performance of Neural Models. Front. Comput. Neurosci. *10*, 1–11.

Schwartz, G.W., Okawa, H., Dunn, F.A., Morgan, J.L., Kerschensteiner, D., Wong, R.O., and Rieke, F. (2012). The spatial structure of a nonlinear receptive field. Nat. Neurosci. *15*, 1572–1580.

Shah, N.P., Brackbill, N., Rhoades, C., Kling, A., Goetz, G., Litke, A.M., Sher, A., Simoncelli, E.P., and Chichilnisky, E. (2020). Inference of nonlinear receptive field subunits with spike-triggered clustering. ELife *9*, e45743.

Simmons, K.D., Prentice, J.S., Tkačik, G., Homann, J., Yee, H.K., Palmer, S.E., Nelson, P.C., and Balasubramanian, V. (2013). Transformation of Stimulus Correlations by the Retina. PLoS Comput. Biol. *9*, e1003344.

Soodak, R.E. (1986). Two-dimensional modeling of visual receptive fields using Gaussian subunits. Proc. Natl. Acad. Sci. *83*, 9259–9263.

Soto, F., Hsiang, J., Rajagopal, R., Piggott, K., Harocopos, G.J., Couch, S.M., Custer, P., Morgan, J.L., and Kerschensteiner, D. (2020). Efficient Coding by Midget and Parasol Ganglion Cells in the Human Retina. Neuron *107*, 656-666.e5.

Turner, M.H., and Rieke, F. (2016). Synaptic Rectification Controls Nonlinear Spatial Integration of Natural Visual Inputs. Neuron *90*, 1257–1271.

Turner, M.H., Schwartz, G.W., and Rieke, F. (2018). Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. ELife *7*, 252148.

Vintch, B., Movshon, J.A., and Simoncelli, E.P. (2015). A Convolutional Subunit Model for Neuronal Responses in Macaque V1. J. Neurosci. *35*, 14829–14841.

Warwick, R.A., Kaushansky, N., Sarid, N., Golan, A., and Rivlin-Etzion, M. (2018). Inhomogeneous Encoding of the Visual Field in the Mouse Retina. Curr. Biol. *28*, 655-665.e3.

Wässle, H., Puller, C., Müller, F., and Haverkamp, S. (2009). Cone contacts, mosaics, and territories of bipolar cells in the mouse retina. J. Neurosci. *29*, 106–117.

Wienbar, S., and Schwartz, G.W. (2018). The dynamic receptive fields of retinal ganglion cells. Prog. Retin. Eye Res. *67*, 102–117.

Yu, W.Q., El-Danaf, R.N., Okawa, H., Pacholec, J.M., Matti, U., Schwarz, K., Odermatt, B., Dunn, F.A., Lagnado, L., Schmitz, F., et al. (2018). Synaptic Convergence Patterns onto Retinal Ganglion Cells Are Preserved despite Topographic Variation in Pre- and Postsynaptic Territories. Cell Rep. *25*, 2017-2026.e3.

5 GENERAL DISCUSSION

Using a combination of multielectrode-array recordings from the isolated mouse retina and computational modeling, we showed how nonlinear spatial processing sculpts the retinal response to natural stimuli. In particular, we argued that experiments with naturalistic visual stimuli can aid us identify mechanisms relevant for natural vision (Chapter 2), we showed that nonlinear spatial integration endows retinal ganglion cells with sensitivity to the spatial structure of natural images (Chapter 3), and we built general-purpose nonlinear subunit models, based on subunit grids, that can capture this sensitivity to spatial structure (Chapter 4). Here, we compare our results to other established findings about nonlinear spatial integration, and discuss implications of our modeling approach for retinal neuroscience in particular, and vision overall.

5.1 NATURAL STIMULI DRIVE NONLINEAR SPATIAL INTEGRATION

The *circuit* perspective we introduced in Chapter 2 supports that studying retinal activity under natural visual stimulation can help us understand the relevant circuit mechanisms. What do high-dimensional natural stimuli have to offer relative to parametric artificial stimuli (Rust and Movshon 2005)? Simple artificial designs allow the control of a few stimulus parameters that isolate the circuit components under scrutiny; however the same components may be driven in completely different regimes (or not at all) when multiple stimulus parameters change in unison. Such discrepancies are becoming evident downstream of the retina, where heterogeneous circuit components interact, as in the case of V4 color coding (Benjamin et al. 2020), or corticothalamic feedback at the lateral geniculate nucleus (Benjamin et al. 2020). However, even at the simple retinal circuit, this natural-artificial dichotomy has been observed in the case of nonlinear spatial processing. ON parasol cells in the macaque retina show nonlinear spatial integration with contrast-reversing gratings, but are linear under naturalistic movies (Turner and Rieke 2016). Careful measurements reveal that the dichotomy appears due to the saccade-and-fixate temporal dynamics of natural movies, which drive the circuit upstream of ON parasol cells in a linear regime (Yu et al. 2021). Thus, experiments with natural stimuli can reveal how the neural code operates in functional regimes that are ethologically relevant. Naturalistic stimulus approaches have assisted in understanding the neural code of other sensory systems (Theunissen and Elie 2014), and their utility extends to cognitive neuroscience (Allen et al. 2021; Sonkusare et al. 2019).

We showed that spatial nonlinearities can shape the output of the mouse retina in response to natural images. Similar to primates, mice shift their horizontal gaze in fixation-like steps, with a median fixation lasting around 200 ms (Meyer et al. 2020; Michaiel et al. 2020). Thus, even simple image flashes can serve as a prototypical stimulus that controls for history effects and can dissect nonlinear spatial integration (Turner and Rieke 2016). We

found that spatial integration under particular nonlinearities is both necessary and sufficient to explain ganglion cell responses to natural images. These nonlinearities were typically rectifying, as originally described (Victor and Shapley 1979), and promoted sensitivity to fine spatial structure within the receptive field center, as previously noted in the rabbit (Cao et al. 2011). In a small subset of ganglion cells, we described saturating spatial nonlinearities that grant cells with a particular sensitivity to spatially homogeneous stimuli. Experiments with blurred natural images helped us estimate the spatial scale of both sensitivity types at approximately half of the receptive field center diameter. This scale property may be the reason why we could explain across-cell differences in spatial structure sensitivity just by examining the responses to an artificial stimulus with two spatial components that split the receptive field in half, but not as well with responses to contrast-reversing gratings. When incorporated into subunit grid models, subunit nonlinearities improved predictions relative to linear receptive fields for both the stimulus with the two components and natural images. Furthermore, cells with stronger nonlinearities also showed greater increase in prediction performance, such in the case of ON alpha transient or ON direction-selective cells. Overall, our results suggest that predictive models of the retinal output can benefit from subunits with cell-type-specific nonlinearities.

We extended our investigations to stimuli also natural in their temporal dynamics. Following the example from the primate retina (Heitman et al. 2016; Turner and Rieke 2016), and due to the recent availability of mouse gaze data, we combined natural images (Hateren and Schaaf 1998) with global motion traces from mouse gaze shifts. Similar to our findings with natural images (Karamanlis and Gollisch 2021), we observed cell-type-specific LN model performance, and, relative to retinal reliability, the movie performances were much lower compared to images. We hypothesize that this is due to the diverse temporal mechanisms that movies may drive, such as mean and variance adaptation (Rieke and Rudd 2009). ON alpha sustained cells were the type with the best LN model predictions, but subunit grid models could further improve response predictions, indicating that the nonlinear receptive field is also driven under stimuli with naturalistic temporal dynamics. While the movies we described here could drive strong and reliable responses to all alpha types, they drove all of the other types less. A potential reason is that our movies lacked object motion, which: may be of behavioral relevance to the mouse when hunting prey (Holmgren et al. 2021; Michaiel et al. 2020), is a specific driver for several mouse cell types (Jacoby and Schwartz 2017; Zhang et al. 2012), and can be utilized to learn object motion sensitivity in convolutional neural networks of the retina (Maheswaranathan et al. 2018b).

To fully capture the retinal output to natural movies, computational models may need to incorporate mechanisms beyond nonlinear spatial integration. Here we focused on subunit grid model predictions for a single cell type with an established nonlinear receptive field (Schwartz et al. 2012). We avoided transient ganglion cell types for which response predictions may suffer without the implementation of post-spike filters that can better capture transient firing events (Heitman et al. 2016; Pillow 2005). Mechanisms that dynamically adjust nonlinear spatial integration may also underlie some of

those firing events. For example, signal flow through electrical synapses between bipolar cell terminals enlarges the scale of spatial integration (Kuo et al. 2016), and selectively enhances responses to stimuli with spatio-temporal correlations (Manookin et al. 2018), such as natural scenes. Photoreceptor adaptation is also strongly driven under stimuli with natural gaze dynamics (Angueyra et al. 2021) and was shown to modulate nonlinear spatial integration in downstream ON parasol cells (Yu et al. 2021). Finally, the receptive field surround of bipolar cells gates the spatial nonlinearity in the receptive field center only when center and surround light intensities are matched, a feature common in natural scenes (Turner et al. 2018). Our subunit grid formalism allows for the explicit modeling of such mechanisms, exemplified by our inclusion of the subunit surround in the case of natural images. To facilitate understanding about the missing model components, we also need to bridge the gap between natural images and movies: a way to proceed is by studying image-to-image transitions that can reveal non-trivial history effects, such as image-recurrence sensitivity in the mouse retina (Krishnamoorthy et al. 2017).

5.2 DIVERSE NONLINEAR RECEPTIVE FIELDS IN THE RETINA

The systematic treatment of spatial integration allowed us to summarize many observations related to the nonlinear receptive field in a compact model. We could then readily compare fitted model parameters between cell types, as is typically done with white-noise-based LN models of the retina (Chichilnisky and Kalmar 2002; Jouty et al. 2018; Ravi et al. 2018; Segev et al. 2006). In our recordings, we commonly encountered a handful of cell types, namely the brisk transient and sustained cells. Using previously reported functional response properties (Baden et al. 2016; Goetz et al. 2021), we matched our cell types to the four alpha types of the mouse retina (Krieger et al. 2017). This procedure allowed us to compare nonlinear receptive field properties of alpha cells to previous anatomical and physiological measurements.

Sustained ON alpha cells are perhaps the best-characterized ganglion cell type in the mouse retina (Cui et al. 2016; Grimes et al. 2014; Kuo et al. 2016; Schmidt et al. 2014; Schwartz et al. 2012), due to their large soma size which makes them perfect for targeted single-cell recordings. We confirmed that their receptive field comprised of small subunits that were summed under a mostly rectifying nonlinearity (Schwartz et al. 2012); their nonlinear receptive field structure was also driven under natural movies. ON brisk transient cells in our recordings matched the ON transient with medium receptive field type of Goetz et al. (2021) and the large ON transient type of Baden et al. (2016). We thus concluded that our ON brisk transient cells correspond to the recently described transient ON alpha cells (Krieger et al. 2017). We recorded from transient ON alpha cells less often compared to the other alpha types, but these transient cells formed consistent mosaics with relatively large receptive fields. Transient ON alpha cells had larger subunits and more rectified nonlinearities compared to their sustained counterparts, and thus lower LN model performance for both natural images and movies. These differences in presynaptic response components probably mirror

anatomical differences in synaptic inputs, as both alpha types stratify in different parts of the inner plexiform layer (Krieger et al. 2017). Interestingly, neurotransmitter release from bipolar cell terminals is more nonlinear in the middle part of the inner plexiform layer (Borghuis et al. 2013), where ON transient cells collect their inputs from. Because of their large cell bodies (Krieger et al. 2017), ON transient alpha cells probably comprise a large portion of the "ON transient" physiological types reported in single-cell studies. Based on this assumption, our data match the fact that ON transient cells show larger subunits compared to sustained ON alpha cells (Mani and Schwartz 2017), as they display a relatively enlarged scale of their spatial integration (Kuo et al. 2016). In total, subunit grid model parameters may reflect some aspects of functional connectivity between bipolar and ganglion cells.

We also observed differences in the nonlinear receptive field organization of OFF alpha ganglion cells. Sustained OFF alpha cells show relatively linear spatial integration, a finding supported by both our findings and previous measurements (Krieger et al. 2017; Wienbar and Schwartz 2021). Here we found that this linearity is coordinated through multiple mechanisms: sustained OFF alpha cells have few large subunits that are overlapping and their outputs are relatively linear. Subunit overlap in sustained OFF alpha cells was larger compared to their transient counterpart. This asymmetry mirrors the difference in the dendritic field coverage of their major bipolar cell inputs (Behrens et al. 2016), type 2 for sustained and type 3a for transient OFF alpha cells (Yu et al. 2018). We extensively investigated the receptive field substructure of transient OFF alpha cells, for which regional differences in response transiency have been reported between dorsal and ventral retina (Warwick et al. 2018; Werginz et al. 2020). We confirmed those differences, and found that while ventral cells showed rectifying nonlinearities, the dorsal ones showed saturating nonlinearities and concomitant sensitivity to spatially homogeneous stimulation. Such a saturated contrast-response function was also observed in the total synaptic currents of dorsal transient OFF cells in the guinea pig retina (Manookin et al. 2008). The mechanism behind this shape of input nonlinearity is the activation of AII amacrine cells that provide strong ON-type inhibition to the OFF alpha transient cell. Among all alpha types, transient OFF alpha cells had the lowest subunit overlap, an observation explaining their spatially independent dendritic receptive fields (Ran et al. 2020).

The nonlinear subunits of ON direction-selective (DS) ganglion cells were distinct from the subunits providing input to ON alpha cells. Besides having more rectifying nonlinearities, ON DS subunits had strong surround suppression, which we predict comes from presynaptic inhibition. When measured both with expanding spots or drifting gratings, ON DS cells have relatively strong surrounds (Dhande et al. 2013), which matches our tuning curve measurements. All of the above properties were different from both ON alpha transient and sustained cells. These differences may reflect the differential connectivity of those types, as ON DS cells receive input from type 5 bipolar cells (Matsumoto et al. 2019), whereas ON alpha sustained cells receive input from type 6 (Schwartz et al. 2012). Thus our model could capture the presynaptic surround, and could allow for potential extensions.

5.3 A SYSTEMATIC EXTENSION OF NONLINEAR SUBUNIT MODELS

Building models that improve over the LN model and contain interpretable components is a challenge that extends beyond retinal neuroscience. Systems identification for the visual system has been traditionally based on white noise characterization (Marmarelis and Marmarelis 1978), because the statistical properties of white noise simplify LN inference (Chichilnisky 2001). The hope was that with a correct statistical treatment of natural stimuli, LN models could be fit to and explain responses elicited by natural stimuli; yet, such LN models provide mediocre response predictions even at the level of the retinal output (Heitman et al. 2016). It is now becoming clear that nonlinear model structure is necessary for the retina (Gollisch and Meister 2010), but fitting nonlinear models of any kind to neural responses becomes increasingly challenging due to the larger number of parameters that need to be specified. Historically, spike-triggered covariance (STC) methods (Schwartz et al. 2006), that involve multiple linear filters whose outputs are nonlinearly combined, have been used to model responses to natural stimuli, with reported performance improvement over LN models in V1 (Touryan et al. 2005). However, STC filters are not directly interpretable: in a simulation of V1 neurons with multiple localized filter inputs, obtained STC filters show up as a mixture of the underlying filters (McFarland et al. 2013). Instead, subunit models are the main drivers of nonlinear receptive field modeling, not only because of their directly explainable components in terms of presynaptic neurons (Liu et al. 2017), but also because they tend to perform better than STC-based or other types of nonlinear models (Vintch et al. 2015). Thus, nonlinear subunit models emerge in the sweet spot of prediction performance and biological interpretability.

Extracting nonlinear subunits from retinal ganglion cell responses is now possible, but existing methods may be challenged with providing interpretable components in the mammalian retina. All proposed methods are based on stimulation with white noise (Liu et al. 2017; Maheswaranathan et al. 2018a; Shah et al. 2020), which may weakly drive the receptive field center when the size of stimulus pixels is small. Thus subunit estimation may require lengthy visual stimulation and can be inaccurate for small subunits and insufficient data. This problem has been demonstrated by simulations with the spike-triggered clustering method which picks up combinations of bipolar cells as nonlinear subunits (Shah et al. 2020) . Furthermore, spike-triggered clustering applied to OFF parasol cells recovers fewer subunits than what is expected from anatomical estimates. Another method, spike-triggered non-negative matrix factorization (STNMF) correctly recovers subunits in simple simulations (Liu et al. 2017) and provides subunits that match bipolar cell receptive fields in the salamander retina, where spatial nonlinearities are very prominent (Bölinger and Gollisch 2012). When applied to mouse ganglion cells, STNMF returns subunits of diameters >100 µm (Liu et al. 2017), clearly above what we observed here and the diameters of 40–70 µm expected by typical bipolar cell receptive fields (Franke et al. 2017). Such clumping of subunits can be avoided by explicitly constraining the underlying scale, which has been achieved in the special case of the peripheral primate retina, where single-cone receptive fields can be distinguished

Non-negative matrix factorization algorithms have an inherent clustering property, which clusters spike-triggered stimuli

(Freeman et al. 2015).

The subunit grid method we introduced in this thesis overcomes the problems stated above by explicitly modeling subunit size and introducing a stimulus that strongly drives the nonlinear receptive field. We took advantage of retinal anatomy to drastically lower the number of parameters required to fit a nonlinear subunit model. Instead of estimating subunit filters separately, we shifted the problem focus to inferring subunit weights, and robustly learned the correct subunit scale through the stimulation with gratings of variable spatial frequencies. Indeed, grating stimuli are commonly used to determine the scale of nonlinear spatial integration (Crook et al. 2008a; Crook et al. 2008b; Krieger et al. 2017; Turner and Rieke 2016). Our approach resulted in models whose subunit sizes matched anatomical measurements and also generalized in other stimulus classes including natural stimuli. The analytical framework we introduced for defining and fitting subunit grid models allows for multiple extensions that model circuit mechanisms, namely subunit surrounds with temporal dynamics (Turner et al. 2018), and electrical synapses between subunits (Kuo et al. 2016; Manookin et al. 2018). Furthermore, our approach can be used to construct models with inputs from multiple subunit grids, mirroring ganglion cell types that receive inputs from multiple bipolar cell types, such as ON-OFF ganglion cells.

Subunit grids fitted with flashed gratings include an explicit subunit nonlinearity that can be directly compared to experimental data. We focused on how different shapes of this nonlinearity vary across cell types and how these shapes may affect natural stimulus encoding. Because the nonlinearity in our model is shared across all subunits, it will dictate how the synaptic inputs respond to light when visual stimulation of the receptive field center is homogeneous. We propose that the subunit nonlinearity can be measured by determining the contrast-response function of the synaptic current entering or leaving a cell under spots of different contrasts. The current measurements should be done close to the resting potential, to reflect the combined contribution of excitatory and inhibitory currents. In the case of dorsal transient OFF alpha cells, our predicted nonlinearities may match the contrast-response function of previously measured input currents (Manookin et al. 2008). If the inhibitory contributions onto the ganglion cell are small, the nonlinearity reduces to the contrast-response function of the excitatory inputs (Schwartz et al. 2012; Turner and Rieke 2016), and the subunit profile is the size-response function of the excitation (Jacoby and Schwartz 2017; Mani and Schwartz 2017), both measurements typically conducted in whole-cell patch clamp recordings. It thus becomes clear that our model can be validated in single cell recordings, first by sampling ganglion cell spikes to fit the subunit grid model, and then measuring the contrast response function of the total synaptic current. Another model parameter, namely the distribution of subunit weights, could be validated anatomically through the quantification of bipolar cell synaptic contacts with ganglion cell dendrites (Schwartz et al. 2012).

The subunit grid approach can be readily extended to visual neurons with more complex receptive fields, such as those encountered in the visual cortices. Compared to white noise, grating stimuli provide a stronger drive to V1 neurons, thus facilitating the estimation of their linear receptive fields

(Ringach et al. 1997). We also used gratings to fit subunit grid models, whose responses were analytically derived based on the original calculations of Soodak (1986). The same work also contains calculations predicting the responses of Gabor-type receptive fields to gratings. Such Gabor-type receptive fields emerge from other nonlinear subunit methods based on white noise (McFarland et al. 2013; Shah et al. 2020; Vintch et al. 2015). Thus, our method could be readily extended to determine presynaptic parameters in other parts of the visual system, even up to V2 (Liu et al. 2016).

A Gabor receptive field extracts specific spatial frequency and orientation content in an image region

5.4 RETINAL CODING OF NATURAL SPATIAL STRUCTURE

In Chapter 3, we showed that sensitivity to the spatial structure of natural images is ubiquitous in the retinal output. Besides spatial correlations, natural scenes contain additional structure, such as sharp edges, gradients and textures (Turiel and Parga 2000). Nonlinear spatial integration can distinguish between those structures and homogeneous illumination within the receptive field center, but such a separation disappears when contrast differences between center and surround dominate (Turner et al. 2018). These facts may point to an encoding strategy that prioritizes edge detection in a wide range of spatial scales, but also raises the question of whether information about spatial structure is used by downstream visual circuits. For example, it may be used to guide gaze, as humans make saccadic eye movements towards natural image regions that have high spatial contrast or low spatial correlation (Baddeley and Tatler 2006; Reinagel and Zador 1999).

How to understand what kind of structure ganglion cells might be extracting? The *coding* perspective we introduced in Chapter 2 may guide us in answering this question, by trying to find how populations of ganglion cells encode spatial structure. It is clear that center-surround structure is involved in decoding raw natural images from the retina (Brackbill et al. 2020), but the reconstructions are missing high spatial frequency components. Spike-train information beyond the spike count can help decode some of these finer features of natural scenes (Kim et al. 2021). However, multiple experiments including ours (Cao et al. 2011; Karamanlis and Gollisch 2021; Turner et al. 2018; Turner and Rieke 2016), suggest that there is a significant amount of spatial structure information in the spike count itself. What features of natural images beyond raw light intensity can we decode using the spike count? A way to approach this question is to decode filtered versions of images, using filters that facilitate, for example, edge detection, such as the Sobel operator or the local standard deviation, which we also used in our definition of spatial contrast (Karamanlis and Gollisch 2021; Liu and Gollisch 2021). To isolate the spiking responses relevant to spatial structure, a prediction from a linear receptive field could be subtracted from the response. If a particular feature emerges as a likely candidate, further questions can be posed. Do nonlinear subunit models account for these decoded features related to spatial structure? And what is the spatial scale relevant for the decoded features?

The Sobel operator approximates the gradient of an image by convolving the image with filters in the horizontal and vertical directions

Using natural movies, we measured the redundancy in the neural code of the retina. The efficient coding theory (Barlow 1961) postulates that the goal of the retina is to reduce such redundancy, which is usually measured with

pairwise response correlations of cells belonging to the same type. Using artificial stimuli that match the correlation profile of natural scenes, previous measurements revealed that pairwise correlations in the salamander retina were much lower than those of the underlying stimulus, but remained substantial in close distances (Pitkow and Meister 2012). We confirmed this finding in the mouse retina using actual natural movies, and in a precise cell-type-specific manner. Additionally, we found that pairwise correlations calculated from white-noise-fitted LN models overestimated the actual pairwise correlations, confirming a similar finding from the guinea pig retina (Simmons et al. 2013). LN models with components fitted under white noise underestimate the receptive field surround (Wienbar and Schwartz 2018). For this reason, we estimated LN models from flickering gratings, with a separate filter for the surround. These models could more accurately capture pairwise correlations, highlighting the importance of retinal inhibition in decorrelating neural responses.

Do subunit nonlinearities also play any role in retinal decorrelation? Simulations of nonlinear subunit models suggested that subunit nonlinearities are responsible for most of the response decorrelation (Maheswaranathan et al. 2018a). We also tested this hypothesis by calculating pairwise correlations between subunit grid models fitted to ON alpha sustained cells under flickering gratings. Surprisingly, subunit grid models were more correlated on average than simpler LN models, even though they predicted spiking responses better. A potential reason is that these correlations are driven by common bipolar cell input, which is well captured by subunit grid models. The addition of a subunit surround with its own temporal dynamics is expected not only to lower these pairwise correlations, but also to further improve ganglion cell responses (Turner et al. 2018).

5.5 APPLICATIONS OF RETINAL ENCODING MODELS

Computational models of the retinal circuit can summarize our current knowledge about the functional organization of retinal output under naturalistic vision. The potential uses of models extend beyond providing such a compact description, and here we provide two major examples. The first is related to basic research: subunit grid models of the receptive field can provide a window to the functional connectivity between bipolar and ganglion cell types, i.e. how specific connections between them are coordinated to generate computations. The second is translational: models with ethologically-relevant components can guide the design of neural prosthetic devices that replace parts of a diseased retina.

5.5.1 *Functional connectivity between bipolar and ganglion cells*

Understanding connectivity on a functional level may refine our definitions of cell types and cell-type-specific computations (Seung and Sümbül 2014; Vlasits et al. 2019). The subunit grid model provides a template for understanding bipolar-to-ganglion cell connectivity: presynaptic bipolar cells of a particular type tile visual space (the grid), and their outputs are pooled by a

ganglion cell. Properties of the subunit grid, such as the coverage factor or subunit surrounds and nonlinearities, reflect both functional properties of presynaptic bipolar cells, but also the inhibition acting on them. We found that these properties are largely consistent within ganglion cell types, with the exception of regional variation in OFF alpha transient cells. Thus, just by examining connectivity properties, or simply spatial integration, we can infer functional differences between ganglion cell types. For example, we discovered spatially linear and nonlinear subtypes of OFF orientation-selective cells (Karamanlis and Gollisch 2021), which also had different preferred orientations, possibly matching the previously identified subtypes (Nath and Schwartz 2017).

The subunit grid framework we introduced can also be used to guide modeling when the wiring diagram between bipolar and ganglion cells is known. If a ganglion cell receives excitatory inputs from two types of bipolar cells, it can be modeled as receiving input from two independent subunit grids whose properties would be then fit to data. An example is the case of ON-OFF cells which receive both ON- and OFF-type excitation (Jacoby and Schwartz 2017) and could be represented by two grids constrained to have opposite response polarities. Another example is ON direction-selective cells in the mouse retina, which receive excitation asymmetric in its temporal dynamics and aligned with the preferred-null direction axis (Matsumoto et al. 2019). This direction-selective circuit could be modeled with two types of ON-type subunits (fast and slow), and direction-selectivity could be partially captured just by the model itself. The subunit grid framework may also find direct application on the primate retina, where only one or two bipolar cell types converge onto a ganglion cell type and the connectivity diagram is mostly known (Grünert and Martin 2021). Besides offering functional bipolar cell characterizations, one can infer the involvement of different inhibitory components in ganglion cell processing by examining fitted subunit surrounds or nonlinearities. Thus, subunit grids can be used as templates for generating connectomically-exact models of the retina, which will probably further improve response predictions to natural stimuli.

5.5.2 *Vision restoration*

Neural prostheses could achieve vision restoration in some forms of retinal degeneration, such as *retinitis pigmentosa*, where the photoreceptors progressively degenerate, but the ganglion cell layer remains mostly intact (Sahel et al. 2021). Currently, there are two major approaches for stimulating the remaining ganglion cells. In the first one, electrical current is applied to ganglion cells through a prosthetic device to mimic bipolar cell excitation (Goetz and Palanker 2016). This strategy mirrors the successful cochlear implant, where electrical stimulation bypasses the diseased cochlea and directly drives the spiral ganglion neurons that form the auditory nerve. The second approach is optogenetic, where light-sensitive ion channels are expressed in the ganglion cells (Ferrari et al. 2020; Sahel et al. 2021; Sengupta et al. 2016), and driven by laser stimulation through the cornea. Both approaches use an external camera and convert the visual scene to either electrical or laser signals that in turn activate the ganglion cells. While both approaches

can achieve reasonable spatial resolution, they end up stimulating all ganglion cell types in the same way, completely disregarding cell-type-specific differences that arise from the cells' synaptic inputs. This cell-type indifference may be the reason that patients with such treatments only "see" by correlating the qualities of phosphene-like percepts to known objects in their environment (Erickson-Davis and Korzybska 2021), an experience strikingly different from natural vision. A way to resolve this issue is through cell-type-specific activation by selective electrical stimulation (Shah and Chichilnisky 2020). Such stimulation protocols are being actively developed for parasol and midget cells (Jepson et al. 2013; Madugula et al. 2020; Sekirnjak et al. 2006) which are the types responsible for the majority of the retinal output in primates.

Computational models are key to designing the activation pattern of ganglion cells for any type of prosthesis. Indeed, the existence of an encoder between the excitation device and the camera input, based on the LN model, could achieve responses that better match the natural patterns of ganglion cell activity (Nirenberg and Pandarinath 2012). If prostheses ever achieve cell-type-specific stimulation, fully restoring the retinal output will require encoders that can capture subtle differences in the nonlinear response properties of cell types, such as those captured by our subunit grid models. These encoders may be especially relevant for highly nonlinear cell types, such as the ON DS cells involved in non-image-forming functions. Our modeling efforts, expanding upon very rich previous work (Chichilnisky 2001; Liu et al. 2017; Pillow 2005; Schwartz et al. 2012; Shah et al. 2020; Turner and Rieke 2016), may thus find application in creating type-specific excitation patterns that match the ones existing in healthy retinas driven with naturalistic stimulation.

5.6 CONCLUSION

In this thesis, I showed that the results of nonlinear spatial processing appear in the retinal output under naturalistic stimulation. Nonlinear processing may differ between ganglion cell types and thus affects how they process natural visual scenes. To capture this nonlinear processing, I developed a novel framework for fitting nonlinear subunit models to ganglion cell responses. Using the fitted models, I thoroughly described the types of nonlinearities present in the retinal output, and could further improve response predictions under naturalistic stimulation. The modeling framework I introduced incorporates many reasonable assumptions about signal processing in the retina, can be readily extended with various biophysical mechanisms, and can potentially be applied to responses in downstream visual areas. Together, my investigations offer new avenues for understanding nonlinear processing in different cell types of the retina, and provide new questions about how these signals might be used downstream.

BIBLIOGRAPHY

Allen, E.J., St-Yves, G., Wu, Y., Breedlove, J.L., Prince, J.S., Dowdle, L.T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J.B., Naselaris, T., and Kay, K. (2021). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. Nat. Neurosci., doi: 10.1038/s41593–021–00962–x.

Angueyra, J.M., Baudin, J., Schwartz, G.W., and Rieke, F. (2021). Multiple time scales of adaptation allow cones to encode the inputs created by visual exploration of natural scenes. bioRxiv, 431101v1.

Applebury, M.L., Antoch, M.P., Baxter, L.C., Chun, L.L., Falk, J.D., Farhangfar, F., Kage, K., Krzystolik, M.G., Lyass, L.A., and Robbins, J.T. (2000). The murine cone photoreceptor: A single cone type expresses both S and M opsins with retinal spatial patterning. Neuron *27*.3, 513–523.

Atick, J.J. and Redlich, A.N. (1992). What Does the Retina Know about Natural Scenes? Neural Comput. *4*.2, 196–210.

Baccus, S.A., Ölveczky, B.P., Manu, M., and Meister, M. (2008). A retinal circuit that computes object motion. J. Neurosci. *28*.27, 6807–17.

Baddeley, R.J. and Tatler, B.W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. Vision Res. *46*.18, 2824–2833.

Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., and Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. Nature *529*.7586, 345–350.

Baden, T., Euler, T., and Berens, P. (2019). Understanding the retinal basis of vision across species. Nat. Rev. Neurosci. *21*.1, 5–20.

Bae, J.A., Mu, S., Kim, J.S., Turner, N.L., Tartavull, I., Kemnitz, N., Jordan, C.S., Norton, A.D., Silversmith, W.M., Prentki, R., Sorek, M., David, C., Jones, D.L., Bland, D., Sterling, A.L., Park, J., Briggman, K.L., and Seung, H.S. (2018). Digital Museum of Retinal Ganglion Cells with Dense Anatomy and Physiology. Cell *173*.5, 1293–1306.e19.

Barlow, H.B. (1953). Summation and inhibition in the frog's retina. J. Physiol. *119*.1, 69–88.

— (1961). Possible principles underlying the transformations of sensory messages. Sens. Commun. *6*.2, 57–58.

Behrens, C., Schubert, T., Haverkamp, S., Euler, T., and Berens, P. (2016). Connectivity map of bipolar cells and photoreceptors in the mouse retina. eLife *5*, e20041.

Behrens, C., Yadav, S.C., Korympidou, M.M., Zhang, Y., Haverkamp, S., Irsen, S., Schaedler, A., Lu, X., Liu, Z., Lause, J., St-Pierre, F., Franke, K., Vlasits, A., Dedek, K., Smith, R.G., Euler, T., Berens, P., and Schubert, T. (2021). Retinal horizontal cells use different synaptic sites for global feedforward and local feedback signaling. Curr. Biol., doi: 10.1016/j.cub.2021.11.055.

Benjamin, A.S., Ramkumar, P., Fernandes, H., Smith, M., and Kording, K.P. (2020). Hue tuning curves in V4 change with visual context. bioRxiv, 780478v2.

Berson, D.M., Dunn, F.A., and Takao, M. (2002). Phototransduction by retinal ganglion cells that set the circadian clock. Science *295*.5557, 1070–1073.

Bicknell, B.A. and Häusser, M. (2021). A synaptic learning rule for exploiting nonlinear dendritic computation. Neuron *109*.24, 4001–4017.e10.

Bimbard, C., Sit, T.P., Lebedeva, A., Harris, K.D., and Carandini, M. (2021). Behavioral origin of sound-evoked activity in visual cortex. bioRxiv, 450721v1.

Bleckert, A., Schwartz, G.W., Turner, M.H., Rieke, F., and Wong, R.O.L. (2014). Visual space is represented by nonmatching topographies of distinct mouse retinal ganglion cell types. Curr. Biol. *24*.3, 310–315.

Bölinger, D. and Gollisch, T. (2012). Closed-Loop Measurements of Iso-Response Stimuli Reveal Dynamic Nonlinear Stimulus Integration in the Retina. Neuron *73*.2, 333–346.

Boone, H.C., Samonds, J.M., Crouse, E.C., Barr, C., Priebe, N.J., and McGee, A.W. (2021). Natural binocular depth discrimination behavior in mice explained by visual cortical activity. Curr. Biol. *31*.10, 2191–2198.e3.

Borghuis, B.G., Marvin, J.S., Looger, L.L., and Demb, J.B. (2013). Two-Photon Imaging of Nonlinear Glutamate Release Dynamics at Bipolar Cell Synapses in the Mouse Retina. J. Neurosci. *33*.27, 10972–10985.

Brackbill, N., Rhoades, C., Kling, A., Shah, N.P., Sher, A., Litke, A.M., and Chichilnisky, E. (2020). Reconstruction of natural images from responses of primate retinal ganglion cells. eLife *9*, e58516.

Cajal, S.R. (1893). La rétine des vertébrés. La Cellule *9*, 119–257.

Cao, X., Merwine, D.K., and Grzywacz, N.M. (2011). Dependence of the retinal Ganglion cell's responses on local textures of natural scenes. J. Vis. *11*.6, 11.

Carcieri, S.M., Jacobs, A.L., and Nirenberg, S. (2003). Classification of Retinal Ganglion Cells: A Statistical Approach. J. Neurophysiol. *90*.3, 1704–1713.

Cembrowski, M.S. and Menon, V. (2018). Continuous Variation within Cell Types of the Nervous System. Trends Neurosci. *41*.6, 337–348.

Cembrowski, M.S. and Spruston, N. (2019). Heterogeneity within classical cell types is the rule: lessons from hippocampal pyramidal neurons. Nat. Rev. Neurosci. *20*.4, 193–204.

Chichilnisky, E.J. (2001). A simple white noise analysis of neuronal light responses. Netw. Comput. Neural Syst. *12*.2, 199–213.

Chichilnisky, E.J. and Kalmar, R.S. (2002). Functional Asymmetries in ON and OFF Ganglion Cells of Primate Retina. J. Neurosci. *22*.7, 2737–2747.

Collewijn, H. (1970). The normal range of horizontal eye movements in the rabbit. Exp. Neurol. *28*.1, 132–143.

Cook, P.B. and McReynolds, J.S. (1998). Lateral inhibition in the inner retina is important for spatial tuning of ganglion cells. Nat. Neurosci. *1*.8, 714–719.

Crook, J.D., Peterson, B.B., Packer, O.S., Robinson, F.R., Gamlin, P.D., Troy, J.B., and Dacey, D.M. (2008a). The Smooth Monostratified Ganglion Cell: Evidence for Spatial Diversity in the Y-Cell Pathway to the Lateral Geniculate Nucleus and Superior Colliculus in the Macaque Monkey. J. Neurosci. *28*.48, 12654–12671.

Crook, J.D., Peterson, B.B., Packer, O.S., Robinson, F.R., Troy, J.B., and Dacey, D.M. (2008b). Y-cell receptive field and collicular projection of parasol ganglion cells in macaque monkey retina. J. Neurosci. *28*.44, 11277–11291.

Cui, Y., Wang, Y.V., Park, S.J.H., Demb, J.B., and Butts, D.A. (2016). Divisive suppression explains high-precision firing and contrast adaptation in retinal ganglion cells. eLife *5*, e19460.

De Franceschi, G., Vivattanasarn, T., Saleem, A.B., and Solomon, S.G. (2016). Vision Guides Selection of Freeze or Flight Defense Strategies in Mice. Curr. Biol. *26*.16, 2150–2154.

Demb, J.B., Zaghloul, K., and Sterling, P. (2001). Cellular basis for the response to second-order motion cues in Y retinal ganglion cells. Neuron *32*.4, 711–721.

Dhande, O.S., Estevez, M.E., Quattrochi, L.E., El-Danaf, R.N., Nguyen, P.L., Berson, D.M., and Huberman, A.D. (2013). Genetic dissection of retinal inputs to brainstem nuclei controlling image stabilization. J. Neurosci. *33*.45, 17797–813.

Dong, D. and Atick, J. (1995). Statistics of natural time-varying images. Netw. Comput. Neural Syst. *6*.3, 345–358.

Drinnenberg, A., Franke, F., Morikawa, R.K., Jüttner, J., Hillier, D., Hantz, P., Hierlemann, A., Azeredo da Silveira, R., and Roska, B. (2018). How Diverse Retinal Functions Arise from Feedback at the First Visual Synapse. Neuron *99*.1, 117–134.e11.

Easter, S.S. (1975). The time course of saccadic eye movements in goldfish. Vision Res. *15*.3, 405–409.

Enroth-Cugell, C. and Freeman, A.W. (1987). The receptive-field spatial structure of cat retinal Y cells. J. Physiol. *384*.1, 49–79.

Enroth-Cugell, C. and Robson, J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. J. Physiol. *187*.3, 517–552.

Erickson-Davis, C. and Korzybska, H. (2021). What do blind people "see" with retinal prostheses? Observations and qualitative reports of epiretinal implant users. PLoS One *16*.2, e0229189.

Felsen, G. and Dan, Y. (2005). A natural approach to studying vision. Nat. Neurosci. *8*.12, 1643–1646.

Fernandez, D.C., Fogerson, P.M., Lazzerini Ospri, L., Thomsen, M.B., Layne, R.M., Severin, D., Zhan, J., Singer, J.H., Kirkwood, A., Zhao, H., Berson, D.M., and Hattar, S. (2018). Light Affects Mood and Learning through Distinct Retina-Brain Pathways. Cell *175*.1, 71–84.e18.

Ferrari, U., Deny, S., Sengupta, A., Caplette, R., Trapani, F., Sahel, J.A., Dalkara, D., Picaud, S., Duebel, J., and Marre, O. (2020). Towards optogenetic vision restoration with high resolution. PLoS Comput. Biol. *16*.7, 1–17.

Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A. *4*.12, 2379–94.

Field, G.D. and Rieke, F. (2002). Nonlinear Signal Transfer from Mouse Rods to Bipolar Cells and Implications for Visual Sensitivity. Neuron *34*.5, 773–785.

Field, G.D., Sher, A., Gauthier, J.L., Greschner, M., Shlens, J., Litke, A.M., and Chichilnisky, E.J. (2007). Spatial properties and functional organization of small bistratified ganglion cells in primate retina. J. Neurosci. *27*.48, 13261–13272.

Fox, M. (1965). The visual cliff test for the study of visual depth perception in the mouse. Anim. Behav. *13*.2-3, 232–233.

Franke, K., Berens, P., Schubert, T., Bethge, M., Euler, T., and Baden, T. (2017). Inhibition decorrelates visual feature representations in the inner retina. Nature *542*.7642, 439–444.

Franke, K., Willeke, K.F., Ponder, K., Galdamez, M., Muhammad, T., Patel, S., Froudarakis, E., Reimer, J., Sinz, F., and Tolias, A. (2021). Behavioral state tunes mouse vision to ethological features through pupil dilation. bioRxiv, 458870v1.

Frazor, R.A. and Geisler, W.S. (2006). Local luminance and contrast in natural images. Vision Res. *46*.10, 1585–1598.

Freeman, J., Field, G.D., Li, P.H., Greschner, M., Gunning, D.E., Mathieson, K., Sher, A., Litke, A.M., Paninski, L., Simoncelli, E.P., and Chichilnisky, E.J. (2015). Mapping nonlinear receptive field structure in primate retina at single cone resolution. eLife *4*, e05241.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. *36*.4, 193–202.

Gastinger, M., Tian, N., Horvath, T., and Marshak, D. (2006). Retinopetal axons in mammals: Emphasis on histamine and serotonin. Curr. Eye Res. *31*.7-8, 655–667.

Ghazanfar, A.A. and Schroeder, C.E. (2006). Is neocortex essentially multisensory? Trends Cogn. Sci. *10*.6, 278–85.

Gilchrist, I.D., Brown, V., and Findlay, J.M. (1997). Saccades without eye movements. Nature *390*.6656, 130–131.

Goetz, G.A. and Palanker, D.V. (2016). Electronic approaches to restoration of sight. Rep. Prog. Phys. *79*.9, 096701.

Goetz, J., Jessen, Z.F., Jacobi, A., Mani, A., Cooler, S., Greer, D., Kadri, S., Segal, J., Shekhar, K., Sanes, J.R., and Schwartz, G.W. (2021). Unified classification of mouse retinal ganglion cells using function, morphology, and gene expression. bioRxiv, 447922v1.

Goldin, M., Lefebvre, B., Virgili, S., Ecker, A., Mora, T., Ferrari, U., and Marre, O. (2021). Context-dependent selectivity to natural scenes in the retina. bioRxiv, 462157v1.

Gollisch, T. (2013). Features and functions of nonlinear spatial integration by retinal ganglion cells. J. Physiol. *107*.5, 338–348.

Gollisch, T. and Meister, M. (2010). Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina. Neuron *65*.2, 150–164.

Greschner, M., Field, G.D., Li, P.H., Schiff, M.L., Gauthier, J.L., Ahn, D., Sher, A., Litke, A.M., and Chichilnisky, E.J. (2014). A Polyaxonal Amacrine Cell Population in the Primate Retina. J. Neurosci. *34*.10, 3597–3606.

Grimaldi, A., Kane, D., and Bertalmío, M. (2019). Statistics of natural images as a function of dynamic range. J. Vis. *19*.2, 1–19.

Grimes, W.N., Schwartz, G.W., and Rieke, F. (2014). The synaptic and circuit mechanisms underlying a change in spatial encoding in the retina. Neuron *82*.2, 460–473.

Grünert, U. and Martin, P.R. (2021). Morphology, Molecular Characterization, and Connections of Ganglion Cells in Primate Retina. Annu. Rev. Vis. Sci. *7*, 73–103.

Hartline, H.K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. Am. J. Physiol. Content *121*.2, 400–415.

Hateren, J.H. van and Schaaf, A. van der (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. B Biol. Sci. *265*.1394, 359–366.

Heitman, A., Brackbill, N., Greschner, M., Sher, A., Litke, A.M., and Chichilnisky, E.J. (2016). Testing pseudo-linear models of responses to natural scenes in primate retina. bioRxiv, 045336v2.

Heukamp, A.S., Warwick, R.A., and Rivlin-Etzion, M. (2020). Topographic Variations in Retinal Encoding of Visual Space. Annu. Rev. Vis. Sci. *6*, 237–259.

Holmgren, C.D., Stahr, P., Wallace, D.J., Voit, K.-m., Matheson, E.J., Sawinski, J., Bassetto, G., and Kerr, J.N. (2021). Visual pursuit behavior in mice maintains the pursued prey on the retinal region with least optic flow. eLife *10*, e70838.

Hong, G., Fu, T.-M., Qiao, M., Viveros, R.D., Yang, X., Zhou, T., Lee, J.M., Park, H.-G., Sanes, J.R., and Lieber, C.M. (2018). A method for single-neuron chronic recording from the retina in awake mice. Science *360*.6396, 1447–1451.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks *2*.5, 359–366.

Hoy, J.L., Yavorska, I., Wehr, M., and Niell, C.M. (2016). Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice. Curr. Biol. *26*.22, 3046–3052.

Hubel, D.H. and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. *148*.3, 574–591.

— (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*.1, 106–154.

Huberman, A.D. and Niell, C.M. (2011). What can mice tell us about how vision works? Trends Neurosci. *34*.9, 464–473.

Hutmacher, F. (2019). Why Is There So Much More Research on Vision Than on Any Other Sensory Modality? Front. Psychol. *10*, 2246.

Jacoby, J. and Schwartz, G.W. (2017). Three Small-Receptive-Field Ganglion Cells in the Mouse Retina Are Distinctly Tuned to Size, Speed, and Object Motion. J. Neurosci. *37*.3, 610–625.

Jepson, L.H., Hottowy, P., Mathieson, K., Gunning, D.E., Dabrowski, W., Litke, A.M., and Chichilnisky, E.J. (2013). Focal electrical stimulation of major ganglion cell types in the primate retina for the design of visual prostheses. J. Neurosci. *33*.17, 7194–7205.

Johnson, K.P., Fitzpatrick, M.J., Zhao, L., Wang, B., McCracken, S., Williams, P.R., and Kerschensteiner, D. (2021). Cell-type-specific binocular vision guides predation in mice. Neuron *109*.9, 1527–1539.e4.

Johnson, K.P., Zhao, L., and Kerschensteiner, D. (2018). A Pixel-Encoder Retinal Ganglion Cell with Spatially Offset Excitatory and Inhibitory Receptive Fields. Cell Rep. *22*.6, 1462–1472.

Jouty, J., Hilgen, G., Sernagor, E., and Hennig, M.H. (2018). Non-parametric Physiological Classification of Retinal Ganglion Cells in the Mouse Retina. Front. Cell. Neurosci. *12*, 481.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydin, Ç., Barbic, M., Blanche, T.J., Bonin, V., Couto, J., Dutta, B., Gratiy, S.L., Gutnisky, D.A., Häusser, M., Karsh, B., Ledochowitsch, P., Lopez, C.M., Mitelut, C., Musa, S.,

Okun, M., Pachitariu, M., Putzeys, J., Rich, P.D., Rossant, C., Sun, W.L., Svoboda, K., Carandini, M., Harris, K.D., Koch, C., O'Keefe, J., and Harris, T.D. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*.7679, 232–236.

Kaplan, E. and Shapley, R.M. (1982). X and Y cells in the lateral geniculate nucleus of macaque monkeys. J. Physiol. *330*.1, 125–143.

Karamanlis, D. and Gollisch, T. (2021). Nonlinear spatial integration underlies the diversity of retinal ganglion cell responses to natural images. J. Neurosci. *41*.15, 3479–3498.

Keat, J., Reinagel, P., Reid, R.C., and Meister, M. (2001). Predicting every spike: a model for the responses of visual neurons. Neuron *30*.3, 803–17.

Kim, T., Shen, N., Hsiang, J.-C., Johnson, K., and Kerschensteiner, D. (2020). Dendritic and parallel processing of visual threats in the retina control defensive responses. Sci. Adv. *6*.47, eabc9920.

Kim, T., Soto, F., and Kerschensteiner, D. (2015). An excitatory amacrine cell detects object motion and provides feature-selective input to ganglion cells in the mouse retina. eLife *4*, e08025.

Kim, Y.J., Brackbill, N., Batty, E., Lee, J., Mitelut, C., Tong, W., Chichilnisky, E.J., and Paninski, L. (2021). Nonlinear decoding of natural images from large-scale primate retinal ganglion recordings. Neural Comput. *33*.7, 1719–1750.

Kling, A., Gogliettino, A.R., Shah, N.P., Wu, E.G., Brackbill, N., Sher, A., Litke, A.M., Silva, R.A., and Chichilnisky, E. (2020). Functional Organization of Midget and Parasol Ganglion Cells in the Human Retina. bioRxiv, 240762v1.

Krause, M., Distler, C., and Hoffmann, K.P. (2014). Retinal ganglion cells projecting to the accessory optic system in optokinetic blind albinotic rats are direction-selective. Eur. J. Neurosci. *40*.1, 2274–2282.

Krieger, B., Qiao, M., Rousso, D.L., Sanes, J.R., and Meister, M. (2017). Four alpha ganglion cell types in mouse retina: Function, structure, and molecular signatures. PLoS One *12*.7, e0180091.

Krishnamoorthy, V., Weick, M., and Gollisch, T. (2017). Sensitivity to image recurrence across eye-movement-like image transitions through local serial inhibition in the retina. eLife *6*, e22431.

Kuffler, S.W. (1953). Discharge patterns and functional organization of mammalian retina. J. Neurophysiol. *16*.1, 37–68.

Kühn, N.K. and Gollisch, T. (2019). Activity Correlations between Direction-Selective Retinal Ganglion Cells Synergistically Enhance Motion Decoding from Complex Visual Scenes. Neuron *101*.5, 963–976.e7.

Kuo, S.P., Schwartz, G.W., and Rieke, F. (2016). Nonlinear Spatiotemporal Integration by Electrical and Chemical Synapses in the Retina. Neuron *90*.2, 320–332.

Land, M. (2019). Eye movements in man and other animals. Vision Res. *162*, 1–7.

Land, M.F. (2015). Eye movements of vertebrates and their relation to eye form and function. J. Comp. Physiol. A *201*.2, 195–214.

LeCun, Y., Kavukcuoglu, K., and Farabet, C. (2010). Convolutional networks and applications in vision. In Proc. 2010 IEEE Int. Symp. Circuits Syst. (IEEE), pp. 253–256.

Lee, J.H., Mitelut, C., Shokri, H., Kinsella, I., Dethe, N., Wu, S., Li, K., Reyes, E.B., Turcu, D., Batty, E., Kim, Y.J., Brackbill, N., Kling, A., Goetz, G., Chichilnisky, E.J., Carlson, D., and Paninski, L. (2020). YASS: Yet another spike sorter applied to large-scale multi-electrode array recordings in primate retina. bioRxiv, 997924v1.

Leonardo, A. and Meister, M. (2013). Nonlinear dynamics support a linear population code in a retinal target-tracking circuit. J. Neurosci. *33*.43, 16971–16982.

Lettvin, J., Maturana, H., McCulloch, W., and Pitts, W. (1959). What the Frog's Eye Tells the Frog's Brain. Proc. IRE *47*.11, 1940–1951.

Liang, L., Fratzl, A., Goldey, G., Ramesh, R.N., Sugden, A.U., Morgan, J.L., Chen, C., and Andermann, M.L. (2018). A Fine-Scale Functional Logic to Convergence from Retina to Thalamus. Cell *173*.6, 1343–1355.e24.

Liang, L., Fratzl, A., Reggiani, J.D., El Mansour, O., Chen, C., and Andermann, M.L. (2020). Retinal Inputs to the Thalamus Are Selectively Gated by Arousal. Curr. Biol. *30*.20, 3923–3934.e9.

Lindsey, J., Ocko, S.A., Ganguli, S., and Deny, S. (2019). A Unified Theory of Early Visual Representations from Retina to Cortex through Anatomically Constrained Deep CNNs. arXiv, 1901.00945v1.

Liu, J.K., Schreyer, H.M., Onken, A., Rozenblit, F., Khani, M.H., Krishnamoorthy, V., Panzeri, S., and Gollisch, T. (2017). Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. Nat. Commun. *8*.1, 149.

Liu, J.K. and Gollisch, T. (2021). Simple Model for Encoding Natural Images by Retinal Ganglion Cells with Nonlinear Spatial Integration. bioRxiv, 458067v1.

Liu, L., She, L., Chen, M., Liu, T., Lu, H.D., Dan, Y., and Poo, M.-m. (2016). Spatial structure of neuronal receptive field in awake monkey secondary visual cortex (V2). Proc. Natl. Acad. Sci. *113*.7, 1913–1918.

Madugula, S., Gogliettino, A.R., Zaidi, M., Aggarwal, G., Shah, N.P., Vilkhu, R., Hays, M., Nguyen, H., Silva, R.A., Chichilnisky, E.J., Kling, A., Shah, N.P., Vilkhu, R., Hays, M., Nguyen, H., Fan, V., Wu, E.G., Hottowy, P., Sher, A., Litke, A.M., Silva, R.A., and Chichilnisky, E.J. (2020). Focal Electrical Stimulation of Human Retinal Ganglion Cells. bioRxiv, 263608v1.

Maheswaranathan, N., Kastner, D.B., Baccus, S.A., and Ganguli, S. (2018a). Inferring hidden structure in multilayered neural circuits. PLOS Comput. Biol. *14*.8, e1006291.

Maheswaranathan, N., McIntosh, L.T., Tanaka, H., Grant, S., Kastner, D.B., Melander, J.B., Nayebi, A., Brezovec, L., Wang, J., Ganguli, S., and Baccus, S.A. (2018b). The dynamic neural code of the retina for natural scenes. bioRxiv, 340943v5.

Mangel, S.C. (1991). Analysis of the horizontal cell contribution to the receptive field surround of ganglion cells in the rabbit retina. J. Physiol. *442*.1, 211–234.

Mani, A. and Schwartz, G.W. (2017). Circuit Mechanisms of a Retinal Ganglion Cell with Stimulus-Dependent Response Latency and Activation Beyond Its Dendrites. Curr. Biol. *27*.4, 471–482.

Manookin, M.B., Beaudoin, D.L., Ernst, Z.R., Flagel, L.J., and Demb, J.B. (2008). Disinhibition combines with excitation to extend the operating range of the off visual pathway in daylight. J. Neurosci. *28*.16, 4136–4150.

Manookin, M.B., Patterson, S.S., and Linehan, C.M. (2018). Neural Mechanisms Mediating Motion Sensitivity in Parasol Ganglion Cells of the Primate Retina. Neuron *97*.6, 1327–1340.e4.

Marmarelis, P.Z. and Marmarelis, V.Z. (1978). Analysis of Physiological Systems (Boston, MA: Springer US).

Martersteck, E.M., Hirokawa, K.E., Evarts, M., Bernard, A., Duan, X., Li, Y., Ng, L., Oh, S.W., Ouellette, B., Royall, J.J., Stoecklin, M., Wang, Q., Zeng, H., Sanes, J.R., and Harris, J.A. (2017). Diverse Central Projection Patterns of Retinal Ganglion Cells. Cell Rep. *18*.8, 2058–2072.

Martinez-Conde, S., Macknik, S.L., and Hubel, D.H. (2004). The role of fixational eye movements in visual perception. Nat. Rev. Neurosci. *5*.3, 229–240.

Masland, R.H. (2012a). The Neuronal Organization of the Retina. Neuron *76*.2, 266–280.

— (2012b). The tasks of amacrine cells. Vis. Neurosci. *29*.01, 3–9.

Masri, R.A., Percival, K.A., Koizumi, A., Martin, P.R., and Grünert, U. (2017). Survey of retinal ganglion cell morphology in marmoset. J. Comp. Neurol. *527*.1, 236–258.

Matsumoto, A., Briggman, K.L., and Yonehara, K. (2019). Spatiotemporally Asymmetric Excitation Supports Mammalian Retinal Motion Sensitivity. Curr. Biol. *29*.19, 3277–3288.e5.

McFarland, J.M., Cui, Y., and Butts, D.A. (2013). Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. PLoS Comput. Biol. *9*.7, e1003143.

McIntosh, L.T., Maheswaranathan, N., Nayebi, A., Ganguli, S., and Baccus, S.A. (2016). Deep Learning Models of the Retinal Response to Natural Scenes. Adv. Neural Inf. Process. Syst. *29*, 1369–1377.

McMahon, M.J., Packer, O.S., and Dacey, D.M. (2004). The classical receptive field surround of primate parasol ganglion cells is mediated primarily by a non-GABAergic pathway. J. Neurosci. *24*.15, 3736–45.

Meister, M., Pine, J., and Baylor, D.A. (1994). Multi-neuronal signals from the retina: acquisition and analysis. J. Neurosci. Methods *51*.1, 95–106.

Meyer, A.F., O'Keefe, J., and Poort, J. (2020). Two Distinct Types of Eye-Head Coupling in Freely Moving Mice. Curr. Biol. *30*.11, 2116–2130.e6.

Michaiel, A.M., Abe, E.T., and Niell, C.M. (2020). Dynamics of gaze control during prey capture in freely moving mice. eLife *9*, e57458.

Morin, L.P. and Studholme, K.M. (2014). Retinofugal projections in the mouse. J. Comp. Neurol. *522*.16, 3733–3753.

Mostofi, N., Zhao, Z., Intoy, J., Boi, M., Victor, J.D., and Rucci, M. (2020). Spatiotemporal Content of Saccade Transients. Curr. Biol. *30*.20, 3999–4008.e2.

Münch, T.A., Da Silveira, R.A., Siegert, S., Viney, T.J., Awatramani, G.B., and Roska, B. (2009). Approach sensitivity in the retina processed by a multifunctional neural circuit. Nat. Neurosci. *12*.10, 1308–1316.

Murphy-Baum, B.L. and Taylor, W.R. (2015). The Synaptic and Morphological Basis of Orientation Selectivity in a Polyaxonal Amacrine Cell of the Rabbit Retina. J. Neurosci. *35*.39, 13336–13350.

Nakajima, Y., Iwakabe, H., Akazawa, C., Nawa, H., Shigemoto, R., Mizuno, N., and Nakanishi, S. (1993). Molecular characterization of a novel retinal

metabotropic glutamate receptor mGluR6 with a high agonist selectivity for L-2-amino-4- phosphonobutyrate. J. Biol. Chem. *268*.16, 11868–11873.

Nath, A. and Schwartz, G.W. (2017). Electrical synapses convey orientation selectivity in the mouse retina. Nat. Commun. *8*.1, 2025.

Nirenberg, S. and Pandarinath, C. (2012). Retinal prosthetic strategy with the capacity to restore normal vision. Proc. Natl. Acad. Sci. *109*.37, 15012–15017.

Ölveczky, B.P., Baccus, S.A., and Meister, M. (2003). Segregation of object and background motion in the retina. Nature *423*.6938, 401–408.

Pachitariu, M., Steinmetz, N.A., Kadir, S.N., Carandini, M., D., H.K., and Harris, K.D. (2016). "Fast and accurate spike sorting of high-channel count probes with KiloSort". Adv. Neural Inf. Process. Syst. 29, D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, and R. Garnett, ed. (Curran Associates, Inc.), pp. 4448–4456.

Peng, Y.-R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G.S., Zyl, T. van, Do, M.T.H., Regev, A., and Sanes, J.R. (2019). Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina. Cell *176*.5, 1222–1237.e22.

Pillow, J.W. (2005). Prediction and Decoding of Retinal Ganglion Cell Responses with a Probabilistic Spiking Model. J. Neurosci. *25*.47, 11003–11013.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature *454*.7207, 995–9.

Pitkow, X. and Meister, M. (2012). Decorrelation and efficient coding by retinal ganglion cells. Nat. Neurosci. *15*.4, 628–635.

Qiu, Y., Zhao, Z., Klindt, D., Kautzky, M., Szatko, K.P., Schaeffel, F., Rifai, K., Franke, K., Busse, L., and Euler, T. (2021). Natural environment statistics in the upper and lower visual field are reflected in mouse retinal specializations. Curr. Biol. *31*.15, 3233–3247.e6.

Ran, Y., Huang, Z., Baden, T., Schubert, T., Baayen, H., Berens, P., Franke, K., and Euler, T. (2020). Type-specific dendritic integration in mouse retinal ganglion cells. Nat. Commun. *11*.1, 2101.

Ravi, S., Ahn, D., Greschner, M., Chichilnisky, E.J., and Field, G.D. (2018). Pathway-Specific Asymmetries between ON and OFF Visual Signals. J. Neurosci. *38*.45, 9728–9740.

Reinagel, P. and Zador, A.M. (1999). Natural scene statistics at the centre of gaze. Netw. Comput. Neural Syst. *10*.4, 341–350.

Reinhard, K., Li, C., Do, Q., Burke, E.G., Heynderickx, S., and Farrow, K. (2019). A projection specific logic to sampling visual inputs in mouse superior colliculus. eLife *8*, e50697.

Reinhard, K. and Münch, T.A. (2021). Visual properties of human retinal ganglion cells. PLoS One *16*.2, e0246952.

Rheaume, B.A., Jereen, A., Bolisetty, M., Sajid, M.S., Yang, Y., Renna, K., Sun, L., Robson, P., and Trakhtenberg, E.F. (2018). Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes. Nat. Commun. *9*.1, 2759.

Rhoades, C.E., Shah, N.P., Manookin, M.B., Brackbill, N., Kling, A., Goetz, G., Sher, A., Litke, A.M., and Chichilnisky, E.J. (2019). Unusual Physiological

Properties of Smooth Monostratified Ganglion Cell Types in Primate Retina. Neuron *103*.4, 658–672.e6.

Rieke, F. and Rudd, M.E. (2009). The Challenges Natural Images Pose for Visual Adaptation. Neuron *64*.5, 605–616.

Ringach, D.L., Sapiro, G., and R.Shapley (1997). A subspace reverse correlation method for the study of visual neurons. Vision Res. *37*.17, 2455–2464.

Román Rosón, M., Bauer, Y., Kotkat, A.H., Berens, P., Euler, T., and Busse, L. (2019). Mouse dLGN Receives Functional Input from a Diverse Population of Retinal Ganglion Cells with Limited Convergence. Neuron *102*.2, 462–476.e8.

Rompani, S.B., Müllner, F.E., Wanner, A., Zhang, C., Roth, C.N., Yonehara, K., and Roska, B. (2017). Different Modes of Visual Integration in the Lateral Geniculate Nucleus Revealed by Single-Cell-Initiated Transsynaptic Tracing. Neuron *93*.4, 767–776.e6.

Roska, B. and Meister, M. (2014). "The Retina Dissects the Visual Scene into Distinct Features". New Vis. Neurosci. J.S. Werner and L.M. Chalupa, ed. (Cambridge, MA: The MIT Press), pp. 163–182.

Roska, B. and Werblin, F. (2003). Rapid global shifts in natural scenes block spiking in specific ganglion cell types. Nat. Neurosci. *6*.6, 600–608.

Roy, S., Jun, N.Y., Davis, E.L., Pearson, J., and Field, G.D. (2021). Inter-mosaic coordination of retinal receptive fields. Nature *592*.7854, 409–413.

Rucci, M. and Poletti, M. (2015). Control and Functions of Fixational Eye Movements. Annu. Rev. Vis. Sci. *1*, 499–518.

Rust, N.C. and Movshon, J.A. (2005). In praise of artifice. Nat. Neurosci. *8*.12, 1647–1650.

Sabbah, S., Gemmer, J.A., Bhatia-Lin, A., Manoff, G., Castro, G., Siegel, J.K., Jeffery, N., and Berson, D.M. (2017). A retinal code for motion along the gravitational and body axes. Nature *546*.7659, 492–497.

Sahel, J.A., Boulanger-Scemama, E., Pagot, C., Arleo, A., Galluppi, F., Martel, J.N., Esposti, S.D., Delaux, A., de Saint Aubert, J.B., Montleau, C. de, Gutman, E., Audo, I., Duebel, J., Picaud, S., Dalkara, D., Blouin, L., Taiel, M., and Roska, B. (2021). Partial recovery of visual function in a blind patient after optogenetic therapy. Nat. Med. *27*.7, 1223–1229.

Salisbury, J.M. and Palmer, S.E. (2016). Optimal Prediction in the Retina and Natural Motion Statistics. J. Stat. Phys. *162*.5, 1309–1323.

Schmidt, T.M., Alam, N.M., Chen, S., Kofuji, P., Li, W., Prusky, G.T., and Hattar, S. (2014). A Role for Melanopsin in Alpha Retinal Ganglion Cells and Contrast Detection. Neuron *82*.4, 781–788.

Schreyer, H.M. and Gollisch, T. (2021). Nonlinear spatial integration in retinal bipolar cells shapes the encoding of artificial and natural stimuli. Neuron *109*.10, 1692–1706.e8.

Schröder, S., Steinmetz, N.A., Krumin, M., Pachitariu, M., Rizzi, M., Lagnado, L., Harris, K.D., and Carandini, M. (2020). Arousal Modulates Retinal Output. Neuron *107*.3, 487–495.e9.

Schwartz, G. and Rieke, F. (2011). Nonlinear spatial encoding by retinal ganglion cells: when 1 + 1 # 2. J. Gen. Physiol. *138*.3, 283–290.

Schwartz, G.W., Okawa, H., Dunn, F.A., Morgan, J.L., Kerschensteiner, D., Wong, R.O., and Rieke, F. (2012). The spatial structure of a nonlinear receptive field. Nat. Neurosci. *15*.11, 1572–1580.

Schwartz, O., Pillow, J.W., Rust, N.C., and Simoncelli, E.P. (2006). Spike-triggered neural characterization. J. Vis. *6*.4, 13.

Seabrook, T.A., Burbridge, T.J., Crair, M.C., and Huberman, A.D. (2017). Architecture, Function, and Assembly of the Mouse Visual System. Annu. Rev. Neurosci. *40*.1, 499–538.

Segal, I.Y., Giladi, C., Gedalin, M., Rucci, M., Ben-Tov, M., Kushinsky, Y., Mokeichev, A., and Segev, R. (2015). Decorrelation of retinal response to natural scenes by fixational eye movements. Proc. Natl. Acad. Sci. U. S. A. *112*.10, 3110–3115.

Segev, R., Puchalla, J., and Berry, M.J. (2006). Functional organization of ganglion cells in the salamander retina. J. Neurophysiol. *95*.4, 2277–92.

Sekirnjak, C., Hottowy, P., Sher, A., Dabrowski, W., Litke, A.M., and Chichilnisky, E.J. (2006). Electrical stimulation of mammalian retinal ganglion cells with multielectrode arrays. J. Neurophysiol. *95*.6, 3311–3327.

Sengupta, A., Chaffiol, A., Macé, E., Caplette, R., Desrosiers, M., Lampič, M., Forster, V., Marre, O., Lin, J.Y., Sahel, J.-A., Picaud, S., Dalkara, D., and Duebel, J. (2016). Red-shifted channelrhodopsin stimulation restores light responses in blind mice, macaque retina, and human retina. EMBO Mol. Med. *8*.11, 1248–1264.

Seung, H.S. and Sümbül, U. (2014). Neuronal cell types and connectivity: Lessons from the retina. Neuron *83*.6, 1262–1272.

Shah, N.P., Brackbill, N., Rhoades, C., Kling, A., Goetz, G., Litke, A.M., Sher, A., Simoncelli, E.P., and Chichilnisky, E. (2020). Inference of nonlinear receptive field subunits with spike-triggered clustering. eLife *9*, e45743.

Shah, N.P. and Chichilnisky, E.J. (2020). Computational challenges and opportunities for a bi-directional artificial retina. J. Neural Eng. *17*.5, 055002.

Sherrington, C.S. (1906). Observations on the scratch-reflex in the spinal dog. J. Physiol. *34*.1-2, 1–50.

Simmons, K.D., Prentice, J.S., Tkačik, G., Homann, J., Yee, H.K., Palmer, S.E., Nelson, P.C., and Balasubramanian, V. (2013). Transformation of Stimulus Correlations by the Retina. PLoS Comput. Biol. *9*.12, e1003344.

Sinha, R., Hoon, M., Baudin, J., Okawa, H., Wong, R.O.L., and Rieke, F. (2017). Cellular and Circuit Mechanisms Shaping the Perceptual Properties of the Primate Fovea. Cell *168*.3, 413–426.e12.

Smith, I.T., Townsend, L.B., Huh, R., Zhu, H., and Smith, S.L. (2017). Stream-dependent development of higher visual cortical areas. Nat. Neurosci. *20*.2, 200–208.

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. Trends Cogn. Sci. *23*.8, 699–714.

Soodak, R.E. (1986). Two-dimensional modeling of visual receptive fields using Gaussian subunits. Proc. Natl. Acad. Sci. *83*.23, 9259–9263.

Sterling, P. and Laughlin, S. (2015). Principles of Neural Design (The MIT Press).

Szatko, K.P., Korympidou, M.M., Ran, Y., Berens, P., Dalkara, D., Schubert, T., Euler, T., and Franke, K. (2020). Neural circuits in the mouse retina support color vision in the upper visual field. Nat. Commun. *11*.1, 3481.

Szél, Röhlich, P., Gaffé, A.R., Juliusson, B., Aguirre, G., and Van Veen, T. (1992). Unique topographic separation of two spectral classes of cones in the mouse retina. J. Comp. Neurol. *325*.3, 327–342.

Theunissen, F.E. and Elie, J.E. (2014). Neural processing of natural sounds. Nat. Rev. Neurosci. *15*.6, 355–366.

Touryan, J., Felsen, G., and Dan, Y. (2005). Spatial structure of complex cell receptive fields measured with natural images. Neuron *45*.5, 781–791.

Turiel, A. and Parga, N. (2000). The Multifractal Structure of Contrast Changes in Natural Images: From Sharp Edges to Textures. Neural Comput. *12*.4, 763–793.

Turner, M.H., Schwartz, G.W., and Rieke, F. (2018). Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. eLife *7*, 252148.

Turner, M.H. and Rieke, F. (2016). Synaptic Rectification Controls Nonlinear Spatial Integration of Natural Visual Inputs. Neuron *90*.6, 1257–1271.

Ujfalussy, B.B., Makara, J.K., Branco, T., and Lengyel, M. (2015). Dendritic nonlinearities are tuned for efficient spike-based computations in cortical circuits. eLife *4*, e10056.

Ujfalussy, B.B., Makara, J.K., Lengyel, M., and Branco, T. (2018). Global and Multiplexed Dendritic Computations under In Vivo-like Conditions. Neuron *100*.3, 579–592.e5.

Victor, J.D. and Shapley, R.M. (1979). The nonlinear pathway of Y ganglion cells in the cat retina. J. Gen. Physiol. *74*.6, 671–89.

Vintch, B., Movshon, J.A., and Simoncelli, E.P. (2015). A Convolutional Subunit Model for Neuronal Responses in Macaque V1. J. Neurosci. *35*.44, 14829–14841.

Vlasits, A.L., Euler, T., and Franke, K. (2019). Function first: classifying cell types and circuits of the retina. Curr. Opin. Neurobiol. *56*, 8–15.

Wang, F., Li, E., De, L., Wu, Q., and Zhang, Y. (2021). OFF-transient alpha RGCs mediate looming triggered innate defensive response. Curr. Biol. *31*.11, 2263–2273.e3.

Wang, Q., Gao, E., and Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. J. Neurosci. *31*.5, 1905–1918.

Warwick, R.A., Kaushansky, N., Sarid, N., Golan, A., and Rivlin-Etzion, M. (2018). Inhomogeneous Encoding of the Visual Field in the Mouse Retina. Curr. Biol. *28*.5, 655–665.e3.

Werblin, F.S. (2010). Six different roles for crossover inhibition in the retina: correcting the nonlinearities of synaptic transmission. Vis. Neurosci. *27*.1-2, 1–8.

Werginz, P., Raghuram, V., and Fried, S.I. (2020). Tailoring of the axon initial segment shapes the conversion of synaptic inputs into spiking output in OFF-α T retinal ganglion cells. Sci. Adv. *6*.37, abb6642.

Wienbar, S. and Schwartz, G.W. (2018). The dynamic receptive fields of retinal ganglion cells. Prog. Retin. Eye Res. *67*, 102–117.

— (2021). Differences in spike generation instead of synaptic inputs determine the feature selectivity of two retinal cell types. bioRxiv, 464988v1.

Yan, W., Laboulaye, M.A., Tran, N.M., Whitney, I.E., Benhar, I., and Sanes, J.R. (2020). Mouse Retinal Cell Atlas: Molecular Identification of over Sixty Amacrine Cell Types. J. Neurosci. *40*.27, 5177–5195.

Yedutenko, M., Howlett, M.H.C., and Kamermans, M. (2021). Enhancing the dark side: asymmetric gain of cone photoreceptors underpins their discrimination of visual scenes based on skewness. J. Physiol., doi: 10.1113/JP282152.

Yeh, C.I., Stoelzel, C.R., and Alonso, J.M. (2003). Two different types of Y cells in the cat lateral geniculate nucleus. J. Neurophysiol. *90*.3, 1852–1864.

Yger, P., Spampinato, G.L., Esposito, E., Lefebvre, B., Deny, S., Gardella, C., Stimberg, M., Jetter, F., Zeck, G., Picaud, S., Duebel, J., and Marre, O. (2018). A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. eLife *7*, e34518.

Yilmaz, M. and Meister, M. (2013). Rapid innate defensive responses of mice to looming visual stimuli. Curr. Biol. *23*.20, 2011–2015.

Yu, W.Q., El-Danaf, R.N., Okawa, H., Pacholec, J.M., Matti, U., Schwarz, K., Odermatt, B., Dunn, F.A., Lagnado, L., Schmitz, F., Huberman, A.D., and Wong, R.O. (2018). Synaptic Convergence Patterns onto Retinal Ganglion Cells Are Preserved despite Topographic Variation in Pre- and Postsynaptic Territories. Cell Rep. *25*.8, 2017–2026.e3.

Yu, Z., Turner, M.H., Baudin, J., and Rieke, F. (2021). Adaptation in cone photoreceptors contributes to an unexpected insensitivity of On parasol retinal ganglion cells to spatial structure in natural images. bioRxiv, 450295v1.

Zhang, Y., Kim, I.-J., Sanes, J.R., and Meister, M. (2012). The most numerous ganglion cell type of the mouse retina is a selective feature detector. Proc. Natl. Acad. Sci. *109*.36, E2391–E2398.

ACKNOWLEDGMENTS

Although only my name is written on it, this thesis is by no means the result of a single person's effort. I only made it this far because I had a group of awesome people that supported me.

First and foremost, I would like to thank Tim Gollisch for all the opportunities he offered me. I managed to grow as a scientist, because he gave me space to try out new ideas and was always there to provide critical and useful feedback. He sets the bar high for being a good supervisor, and this is evident in the culture of the lab.

This thesis is also built on the contributions of my current and past colleagues at the Gollisch lab. Vidhya Krishnamoorthy taught me his magical preparation techniques and introduced me to the retinal literature. Fernando Rozenblit also taught me how to perform experiments, often lent me his sharp debugging skills, assisted me with proofreading this thesis, and our discussions sparked my interest in ganglion cell types. Mohammad Khani shared his passion for science with me, initiated our exciting collaboration on the primate retina, and always took good care of the lab. Helene Schreyer often found nice papers that precisely answered my questions, and was a great source of information on bipolar cells and nonlinear spatial integration. Sören Zapp and Yunus Erol wrote the `phy` plugins that turned manual curation of spike sorting results into a fun game. Michael Weick, Norma Kühn, JD Prieto, Steffen Nitsche, Varsha Ramakrishna, Shashwat Sridhar also aided me with lab work and participated in many useful discussions. I was fortunate to be in the skillful hands of Christiane Westermann and Daniela Proto, who took care of all the little details that made my scientific work possible. Thank you all for the great times we spent inside and outside of the lab and for always taking the time (and your headphones off) to answer my questions.

I am also grateful to the vision scientists I had the chance to (even shortly) collaborate with. Dario Protti allowed me to get spikes from the marmoset retina for the first time in the lab, even if I messed up the recording. Arne Meyer shared mouse gaze data which made the mouse natural movies possible. Same goes for Jude Mitchell and Jacob Yates, who eagerly shared marmoset eye movement data.

Alexander Gail and Marion Silies supported me as thesis committee members, provided me with useful feedback and encouraged me after a failed project. I would also like to thank Hansjörg Scherberger, Tobias Moser and Jan Clemens for accepting to be part of my examination board and taking time to read this thesis.

I would not have ended up in Göttingen without the IMPRS Neuroscience program, supported by Sandra Drube, Michael Hörner and Jonas Barth. I also acknowledge the support of the Boehringer Ingelheim Fonds that not only financed a large part of my project, but also brought me in touch with a group of amazing people, from which I have learned a lot.

I am also indebted to the people in my life that I consider as my family. To Nikos for being my ever-present companion in Göttingen and maybe the

person I trust the most. Thank you for proofreading this thesis and for all our scientific discussions which I really enjoy. To Theo, for all the online cooking evenings, for working together in Zürich, and for all of the great C++ advice. To the rest of the Kamystes, Stefanos, Kostas, Spyros, Panos, Thanasis, for all the great discussions, our trips, the loud Among Us evenings, and believing in me. To the friends I made in Göttingen, Myrto, thank you for always listening to me, and Haris, thank you for all of the cozy evenings. To Eleonora, your infinite positivity (and golden stars) kept me standing countless days. Thank you for proofreading the thesis, and I look forward to growing even further together with you.

Last but not least, this thesis would not have been possible without the life-long lessons and support of my parents, Lina and Aris. Thank you for everything, especially during the last weeks of thesis writing. I dedicate this thesis to you.

DECLARATION

Herewith I declare, that I prepared the Doctoral thesis "How nonlinear processing shapes natural stimulus encoding in the retina" on my own and with no other sources and aids than quoted.

Göttingen, May 23, 2022

Dimokratis Karamanlis