THE UNIVERSITY *of* EDINBURGH

Coding of multivariate stimuli and contextual interactions in the visual cortex

Sander Wessel Keemink

Doctor of Philosophy

Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh

2017

Abstract

The primary visual cortex (V1) has long been considered the main low level visual analysis area of the brain. The classical view is of a feedfoward system functioning as an edge detector, in which each cell has a receptive field (RF) and a preferred orientation. Whilst intuitive, this view is not the whole story. Although stimuli outside a neuron's RF do not result in an increased response by themselves, they do modulate a neuron's response to what's inside its RF. We will refer to such extra-RF effects as contextual modulation. Contextual modulation is thought to underlie several perceptual phenomena, such as various orientation illusions and saliency of specific features (such as a contour or differing element). This gives a view of V1 as more than a collection of edge detectors, with neurons collectively extracting information beyond their RFs. However, many of the accounts linking psychophysics and physiology explain only a small subset of the illusions and saliency effects: we would like to find a common principle. So first, we assume the contextual modulations experienced by V1 neurons is determined by the elastica model, which describes the shape of the smoothest curve between two points. This single assumption gives rise to a wide range of known contextual modulation and psychophysical effects. Next, we consider the more general problem of encoding and decoding multi-variate stimuli (such as center surround gratings) in neurons, and how well the stimuli can be decoded under substantial noise levels with a maximum likelihood decoder. Although the maximum likelihood decoder is widely considered optimal and unbiased in the limit of no noise, under higher noise levels it is poorly understood. We show how higher noise levels lead to highly complex decoding distributions even for simple encoding models, which provides several psychophysical predictions. We next incorporate more updated experimental knowledge of contextual modulations. Perhaps the most common form of contextual modulations is center surround modulation. Here, the response to a center grating in the RF is modulated by the presence of a surrounding grating (the surround). Classically this modulation is considered strongest when the surround is aligned with the preferred orientation, but several studies have shown how many neurons instead experience strongest modulation whenever center and surround are aligned. We show how the latter type of modulation gives rise to stronger saliency effects and unbiased encoding of the center. Finally, we take an experimental perspective. Recently, both the presence and the underlying mechanisms of contextual modulations has been increasingly studied in mice using calcium imaging. However, cell signals extracted with calcium imaging are often highly contaminated by other sources. As contextual

effects beyond center surround modulation can be subtle, a method is needed to remove the contamination. We present an analysis toolbox to de-contaminate calcium signals with blind source separation. This thesis thus expands our understanding of contextual modulation, predicts several new experimental results, and presents a toolbox to extract signals from calcium imaging data which should allow for more in depth studies of contextual modulation.

Acknowledgements

My first thanks go to my supervisors Mark van Rossum, Clemens Boucsein, and Ad Aertsen.

Mark for thinking much alike, but on a level beyond what I was capable of.

Clemens for giving me some much needed biological basis and realism, despite having only been able to supervise me for a year.

Ad for taking over Clemens' supervision, and for being there for an endless supply of signatures for any Freiburg paperwork.

All of the Rochefort lab members for further biological basis, and getting an unintentionally large (but very welcome!) experience in coding software for real-world experimental scenarios.

My fellow DTC and Eurospin program students, in particular:

Miha Pelko for preceding me in everything through the years, since he was the only other Edinburgh-Freiburg student, and putting up with my endless supply of bureaucracy related questions.

Scott Lowe (mostly unintentionally) forcing me to finally take some time to learn proper coding practices.

Matt Graham for having read most of this thesis at some point or another and providing valuable feedback.

Susan Lechelt and Martin Asenov for graciously spending their Honours research projects investigating some further questions I did not have time for myself.

Many thanks also to my previous mentors-in-research: Toshihide Nabatame and Chikyo Toyohiro for providing my first real research experience, Richard van Wezel and Chris Klink for my first neuroscience research, Ken Britten and Mark Goldman and Seth Egger for a solid basis in both experimental data and theory, and finally Eric Shea-Brown and Nicholas Cain for pure theory.

Udo Ernst and Peggy Seriès for being examiners at my Edinburgh thesis examination. They managed to give me more confidence in this work, despite as well as thanks

to criticisms of many of the choices herein. Without them this final version would not have been possible.

And finally, and most importantly, thanks to my family, who not only supported me through my PhD years, but also the many years before, without whom none of the above would even have been possible.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text. Apart from being examined at the partner university, Albert-Ludwigs-Universität Freiburg im Breisgau, as part of the requirements for the EuroSPIN programme, this work was not submitted or examined for any other degree or professional qualification.

(*Sander Wessel Keemink*)

Lay Summary

You are reading this thesis. You see and understand the words written on this page, which appear sharp and clear. Through our eyes, and presumably brains, we are able to perceive the world in a way which seemingly reflects reality. Yet a few readers might have seen some of the many funky visual illusions thought up by artists and scientists (see Fig. 1.1 on page 2 for some examples). These illusions seem to trick the mind into perceiving something which is not actually the same in reality. While such illusions can be very entertaining, the very specific ways in which our perception differs from reality also give us more insight into the inner workings of our brain.

How does our vision work anyway? Our brain consists of brain cells, or neurons, which can send out electrical signals. Many neurons are specialized for seeing, meaning they send out many electrical signals (or 'respond') based on things in our vision. For example, neurons in the eye respond to light hitting them. We will be more concerned with neurons in the primary visual area of the brain however (the primary visual cortex). These neurons are each responsible for a specific area of the visual field (all that you see), at 50 cm distance this area would be roughly 1cm wide. They will only respond if there is an edge of something in their area, for example the outer edge of a candle (they also care about color, but that is outside the scope of this thesis). Additionally, each neuron has a favorite orientation, and the closer the edge covering their area is oriented to the favorite orientation, the more a neurons will respond. As such a theory about the function of the primary visual cortex is that of an edge-detector. The area analyzes the visual scene, and its neurons tell the rest of the brain what edges, and at what orientations, there are in the scene. As with many things in science however, this turned out to be far from the whole story of the primary visual cortex.

A neuron in the primary visual cortex will not respond to anything presented outside its area. However, things outside its area it will change its response to what is *inside* its area. For example, let's consider a neuron that really likes vertical edges. Normally it would send a lot of electrical signals for an vertical edge in its area, but it

turns out that if there are additional vertical edges nearby (outside its area), it will send out less signals. This is called contextual effect, as this neuron changes its response based on the surrounding context of what it normally responds to.

Contextual effects are thought to underlie many perceptual illusions. This is because they can change what neurons are representing. Let's imagine a group of neurons responsible for the same area, but with different favorite orientations. There is nothing of interest in their area of the visual field. Suddenly, a vertical edge appears! The neurons that like vertical edges all start sending out electrical signals, and other areas of the brain now 'know' that there is an upright edge in their area. But soon other edges appear outside their area, and all the neurons start responding differently due to the contextual effects, some less, some more. Neurons that like horizontal things are now responding more than the vertical-loving neurons. The rest of the brain now thinks there's an horizontal edge, even though there actually is an horizontal edge! An illusion is born. Many types of contextual effects exist, and they have profound impact on how visual neurons respond. This thesis studies these effects.

How much, and when, should neurons change their response based on edges outside their own area? Chapter 2 tries to define a mathematical model for what the contextual effects should be. The theory of 'elastica' describes the smoothest shape possible between a start and end point. This chapter assumes that the elastica theory governs how much neurons change their response based on contextual edges, and this turns out to explain a wide range of known visual illusions.

If you know how neurons respond to specific images, it is possible to try to estimate the original image from their signals. However, neurons are very 'noisy', meaning that they don't always respond the same. If a neuron is very noisy, it will almost always send out a different set of electrical signals, each time the same image is shown. If a neuron is not noisy at all, it will always respond the same to the same image. When trying to estimate an image from a group of neurons, this has to be taken into account. Consider a movie of two groups of overlapping dots, each moving in a different direction. You are asked to estimate how different their directions are. Chapter 3 finds that if neurons in the primary visual cortex are very noisy and are under the influence of contextual effects to represent such a movie, they will sometimes represent no motion direction difference at all, and sometimes overestimate it slightly.

Chapter 4 compares two types of contextual effects. In the first type, neurons will respond less if there are orientations that they like outside of their own area of the visual field. In the second type, they respond less if the outside edges are the same as

the edge inside of their own area (whether they like its orientation or not). The first type has been well studied and results in an illusion much like the vertical bar turned horizontal described above. However, the first chapter shows that although the second type still changes how neurons respond to their area, it does not lead to a change in which neurons are most active, and thus leads to no illusion.

As a neuron sends out electrical signals, it releases a lot of calcium. One common way to measure how a neuron responds to an image is therefore to measure the amount of calcium in and around a neuron. However, when this is done in living animals, there will be many other neurons nearby, each also releasing electrical signals and calcium. When measuring the signal of one neuron, you effectively measure a mix of several neurons. Chapter 5 presents a software package which demixes these signals.

Contents

Chapter 1

Introduction

The words on this page appear sharp and clear. Indeed, our perception of the world seems accurate and precise in general. Yet when we systematically study perception we find that it does not always fully reflect reality, often resulting interesting perceptual illusions. One of the major challenges of neuroscience has been, and still is, understanding how the brain's perception functions, or in more specific terms: how is the perceptual information encoded and decoded? Many different mechanisms might be possible explaining how perception could work well, yet it is when the system seemingly breaks (as with the illusions) when we get the best clues for the inner workings and possible underlying strategies.

For example, the Herman grid illusion, where illusory dark spots appear at the crossings between black squares (Fig. 1.1A), has been explained by on-center ganglion cells. At the crossings, these get more inhibition from bright surroundings than in the lanes, leading to a relatively smaller response, giving the illusion of black dots where there are none. Indeed, this model was confirmed by retinal recordings with responses being lower at the crossings than in lanes (Baumgartner, 1960). Interestingly, a sinusoidal grid given this model would predict the same perception, but does not produce the same perception, Fig. 1.1A right, indicating we do not understand the full story yet (Geier et al., 2004).

In this thesis we are mostly concerned with orientation coding in the cortex. One of the most studied and best understood orientation illusions are the many forms of the tilt illusion where the orientation of some element of the stimulus is misjudged, usually due to the presence of some context. The clearest example of this type of illusion can be seen with a center grating surrounded by a second grating, Fig. 1.1B. The central gratings in the example should have the same orientation, but to our perception are

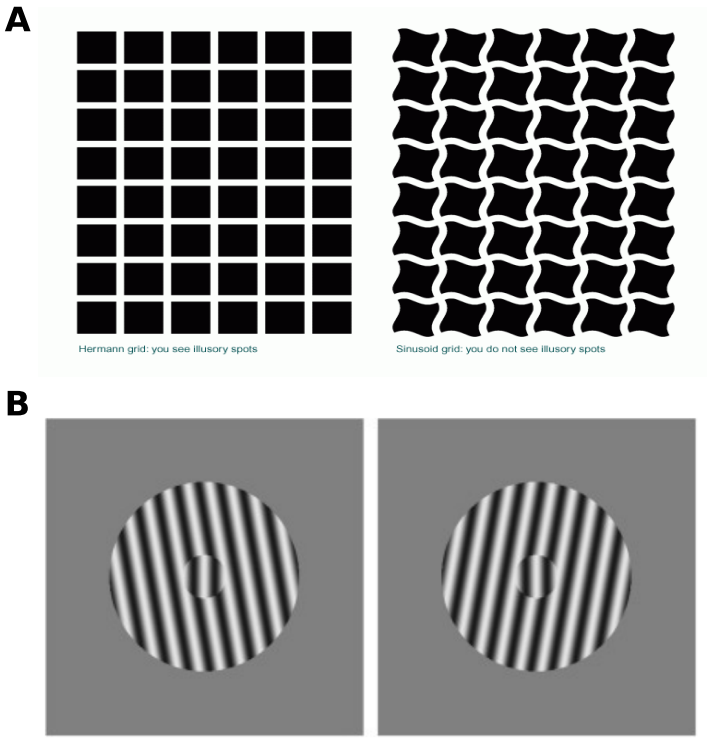

Figure 1.1: Example visual illusions. (A) In the Herman grid illusion illusory dark spots appear at the crossings between black squares. This is believed to happen on the level of the retina due to lower responses in the crossings, but see the right figure for a counter argument (Geier et al., 2004) (B) The tilt illusion. The two center gratings are both fully vertical (90 degrees), but appear tilted away from their surrounding gratings (this effect can be subtle, but is usually clearest when you try to see if the two central gratings have to same orientation or not). Herman grids were copied from `http://www.michaelbach.de/ot/`.

tilted away from their surround (for a recent review of this phenomenon see Clifford, 2014). The prevailing model for this is contextual modulation, where neurons responding to the center grating are modulated by the surrounding grating in such a way that their response becomes biased, leading to an illusion. This thesis deals mainly with such contextual modulations, their impact on how visual information is encoded and decoded, and how that links our perception to contextual modulations.

I will now first detail the physiological background and psychophysics, then consider the currently known links between them, and will finally pose the central hypothesis of this thesis as well as summarize the main results.

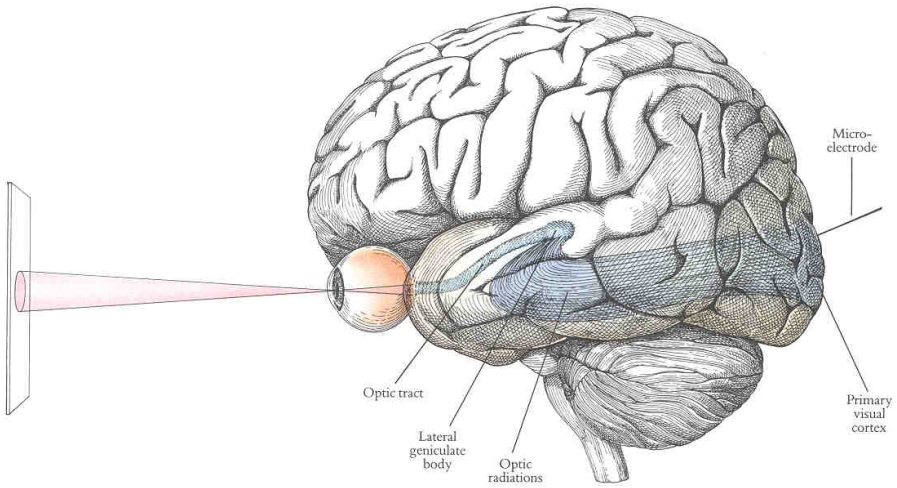

Figure 1.2: Information flow from outside scene, to eyes, to the primary visual cortex. Image source is Hubel (1995).

1.1 Physiology background

1.1.1 The primary visual cortex

Viewing the letters and figures of this thesis, this information first passes through the retina. Retinal ganglion cells respond to differences in local contrasts, within a restricted region of the visual field, termed the receptive field (RF) (Sherrington, 1906; Hartline, 1938). This visual activity is passed on towards the back of the brain, passing through the lateral geniculate nucleus (LGN), and arriving at the primary visual cortex (V1). While retina RFs cover a very small region of the visual field and are quite simple, they get progressively larger and complex as one progresses downstream to V1 and beyond (Alonso and Chen, 2009).

Hubel and Wiesel were the first to electrophysiologically characterize V1 neuron RFs. Their initial idea was that V1 neurons respond to the presence or absence of light a specific region of the visual field, similarly to retinal ganglion cells. They presented cats and monkeys with dots in different positions using glass microscope slides with small metal circles glued on, but weren't having much success. At some point they accidentally moved the edge of the slide over a neuron's RF, and could hear a sudden increase in response (Hubel and Wiesel, 1962). In their own words:

> 'Suddenly, just as we inserted one of our glass slides into the ophthalmoscope, the cell seemed to come to life and began to fire impulses like a machine gun.' - Hubel and Wiesel (2005)

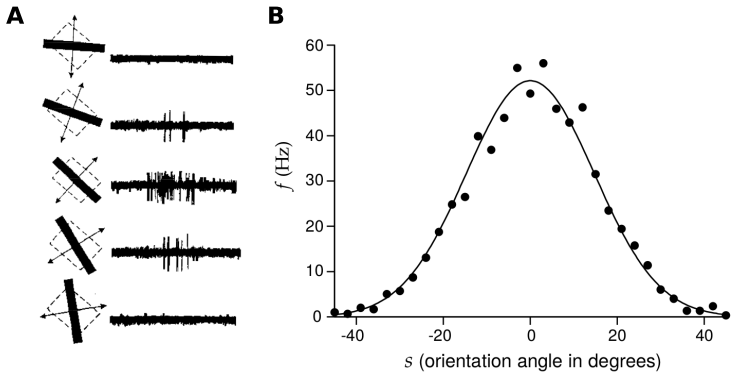

Figure 1.3: Orientation tuning in the primary visual cortex. (A) Recordings of a macaque V1 neuron. Note the increased response for a particular stimulus orientation. (B) The average firing rate of a cat V1 neuron as a function of stimulus orientation. The original image source for A and B is Dayan and Abbott (2001).

It turned out that it was the edge of the slide that was eliciting the response! Working from this assumption and moving slides at different orientations they generally found V1 neurons to respond most strongly to a given (the preferred) orientation in their receptive field, and less so to other orientations, as is nicely illustrated by a so-called tuning curve, Fig. 1.3.

These and subsequent experiments lead to a view of V1 neurons as edge detectors, Fig. 1.4. In this view V1 neurons are single units representing to what degree their preferred orientation is present in their RF. Across many neurons this should result in an orientation representation of the visual scene, which V1 passes on to downstream areas where further computations take place. However, this was far from the whole story of orientation coding in V1.

1.1.2 Contextual modulation

Anatomical studies demonstrated the existence of connections within, from and to V1 (e.g. Bosking et al., 1997; Nassi et al., 2013), and there are well documented changes to neural responses when stimuli outside the RF are used. V1 neurons specifically respond to orientations within their RF, and have no response to stimuli outside their RF (by its very definition). However, such extra-RF stimuli will often modulate a neuron's RF-only response. Hubel and Wiesel already described this effect in the phenomenon of end-stopping, where a bar extended outside some neurons' RF will reduce its response (Hubel and Wiesel, 1965, 1968). The same will happen when a

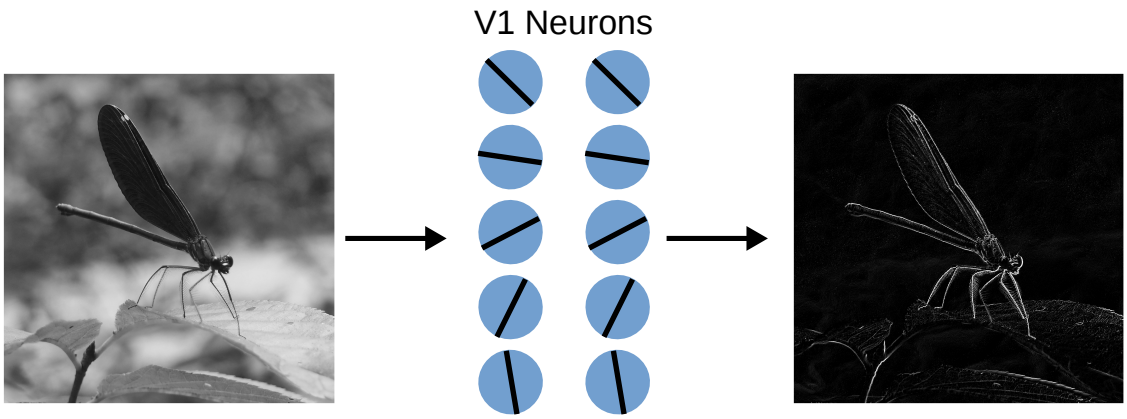

Figure 1.4: Edge detection model of V1. Neurons with different receptive fields and preferred orientations process a scene and pass on an orientation representation.

grating is extended beyond a neuron's RF (Sceniak et al., 1999; Girman et al., 1999; Freeman et al., 2001; Cavanaugh et al., 2002b). These extra-RF effects will be referred to as contextual modulation.

How contextual modulations are mediated biophysically is still under much debate. Contributions to contextual modulations have been show to come from feed-forward connections (surround modulation exists to a lesser degree in LGN (Alitto and Usrey, 2008)), horizontal connections within the same region (Bosking et al., 1997), and feedback connections from higher areas (Nassi et al., 2013). None of these studies were able to fully account for V1's contextual modulations, and in my view it is likely to be a combination of all. When this thesis deals with contextual effects we will usually not consider whether these come from feed-forward, horizontal, or feedback connections, and concern ourselves only with the functional impact.

Perhaps the most extensively studied example of contextual modulation is surround modulation (Blakemore and Tobin, 1972; Sillito et al., 1995; Freeman et al., 2001; Bair et al., 2003; Seriès et al., 2003). Here a grating is shown to a neuron's RF (the center), and a surrounding grating is shown outside the RF (the surround). If the center and surround orientation are kept the same, this is the same as the extended gratings described above, but now the orientation of the surrounding grating is also varied, as in Fig. 1.5A, left. The surrounding grating modulates (usually suppresses, but for low center contrast sometimes facilitates (e.g. Sillito and Jones, 1996)) the response to the center. The modulation is orientation dependent; when the center grating is at the preferred orientation, the modulation will be strongest when the preferred and surround gratings are aligned, Fig. 1.5A, right. As I will show soon, this property is often used

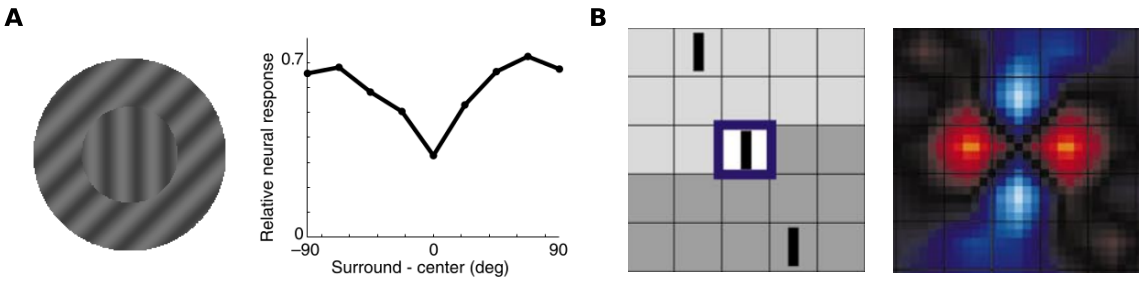

Figure 1.5: Contextual modulation examples. (A) In center-surround modulation, a grating is covering a neuron's RF (center), and surrounded by a second grating (the surround). The surround modulates the response to the center. The modulation is orientation dependent, with the strongest effect when surround and center are aligned (although this is usually tested with the center at the preferred orientation). The modulation curve figure source is Schwartz et al. (2009). (B) Besides orientation dependent, surround modulation is also position dependent. For example, in Kapadia et al. (2000b) a central bar was presented to a neuron's RF (illustrated by the blue square), with two mirrored flankers shown in surrounding locations (gray areas). Relative to the response with just the single central bar, this resulted in increased responses when the bars were aligned (blue color), and decreased responses when they were on the sides (red color). Image source for B is Kapadia et al. (2000b).

to explain orientation illusions.

Investigation of subfields of the surround revealed further complexities. First, analogously to center-surround grating, several studies use a single bar as the center, and then surround this center with sets of bars (Knierim and van Essen D.C., 1992; Kastner et al., 1999; Nothdurft et al., 1999). Such stimuli give similar results as with gratings. But perhaps more interestingly, beyond orientation dependence, such stimuli also reveal a strong positional dependence. For example, when presenting a central bar with two mirrored flankers, Fig. 1.5B, aligned bars increase the response, whereas bars on the sides cause a decrease (Kapadia et al., 1995, 2000b; Bauer and Heinze, 2002; Li et al., 2006), giving a butterfly-like effect. Small gratings also show facilitation at low contrasts, with suppression for aligned gratings at higher contrasts (Polat et al., 1998).

1.1.3 Contextual modulations in mice

The vast majority of contextual modulations results have been obtained from cats, monkeys, or sometimes ferrets or rats. Classically, mice have been considered to have

no or very small contextual effects, due to already large receptive fields, and only weak effects from anything beyond (Van den Bergh et al., 2010). However, this was usually studied in anesthetized mice, and with the ascend of more awake imaging more contextual interactions have been found, mainly using full center-surround gratings (Haider et al., 2013; Vaiceliunaite et al., 2013b; Self et al., 2014).

Calcium imaging in mouse V1 allows for measuring large field of views and many neurons at once, which could potentially let us measure neurons responding to both some central stimulus and the contextual stimulus at once. The mouse model now also has a wide array of genetic tools that allow for direct manipulation of specific neural subtypes, which could give us further insights into the mechanisms and specifics of contextual modulation, as in Adesnik et al. (2012), where they identified somatostatin-expressing inhibitory neurons as a specific neural subgroup involved in surround suppression.

However, the more subtle effects from smaller stimuli such as sets of bars have yet to be studied using 2-photon imaging, partly because such subtle effects are still quite hard to measure. Better signal extraction tools will be needed before such an attempt can be made.

1.2 Psychophysics background

1.2.1 Orientation illusions

An illusion is defined as a discrepancy between physical reality and the perceptual representation (Wenderoth and Johnston, 1987). The initial intuition this often gives rise to is that a particular illusion 'breaks' perception, but this is not necessarily true. After all, the brain's task is not to to represent the world as *accurately* as possible, but rather as *effectively* as possible. Although it is of course imaginable that the brain misinterprets something due to sheer non-optimality, it is considered likely that illusions are not a result of something breaking, but of some other processing going on (Barlow and Földiák, 1988; Clifford et al., 2000). In either case illusions are still hugely informative of both the inner physiological workings of the brain, and its strategies.

Illusions where some visual element's orientation is misjudged go back a long way. It was initially mainly studied in the tilt after effect, where a line's orientation is misjudged after looking at a differently oriented line (the adapter stimulus) for some time before (Gibson and Radner, 1937). Whereas an adapter stimulus is a temporal con-

text (Clifford et al., 2000), spatial contexts turned out to give much the same effect (O'Toole and Wenderoth, 1977). From there, the study of tilt went in roughly two directions. Based on gratings (Wenderoth and Johnston, 1988; Clifford et al., 2000; Qiu et al., 2013; Clifford, 2014), where a center grating is misjudged in the presence of a surround grating, Fig. 1.6A. Or based on bars (Westheimer, 1990; Kapadia et al., 1995, 2000b), where a bar's orientation is misjudged in the presence of some combination of other bars, Fig. 1.6B+C.

For center surround gratings, the center appears repulsed by the surround for most surround orientations, and attracted for larger orientations, Fig. 1.6A third column (Qiu et al., 2013). When a central bar is used with a hexagon of flanking bars which rotate in place, a very similar illusion arises (Westheimer, 1990), Fig. 1.6B. Interestingly, the data from Westheimer (1990) is often used as an example of the grating-based illusion, as for example in Schwartz et al. (2009) and Clifford (2014). This is understandable, as the two illusions are quite similar, despite the quite different stimuli used, but do note that the grating illusion usually persists over time (Wenderoth and Johnston, 1988), whereas the Westheimer et al. experiment was done by only briefly showing the bars. These two illusions are jointly referred to as the tilt illusion.

As with contextual modulation, more complexity arises as we look at the influence of spatial location of the surrounding bars (Kapadia et al., 2000b). When two bars are shown on the sides of the central bar, and rotated, they consistently give a repulsive illusion, Fig.1.6 C third column, solid curve. When they are placed at the ends of the central bar, they give an attractive illusion for small orientations, and again repulsion for larger orientations, Fig.1.6 C third column, dashed curve.

1.2.2 Salient features

In visual search experiments the task it to find a deviant element among other identical elements. An easily findable element is said to have a high salience (Itti, 2007), which is usually measured by search time (how long does it take to find some differing element? Treisman and Gormican, 1988) or the proportion of correct indications of the differing element location (Field et al., 1993). In the context of the bar-stimuli described above, one might have a single bar at a different orientation from the rest, or a set of bars might be grouped in some way like a contour, as in Fig. 1.6D.

As with contextual modulation, such pop-out effects are orientation dependent. Straight contours are easy to extract from noisy background, whereas more windy

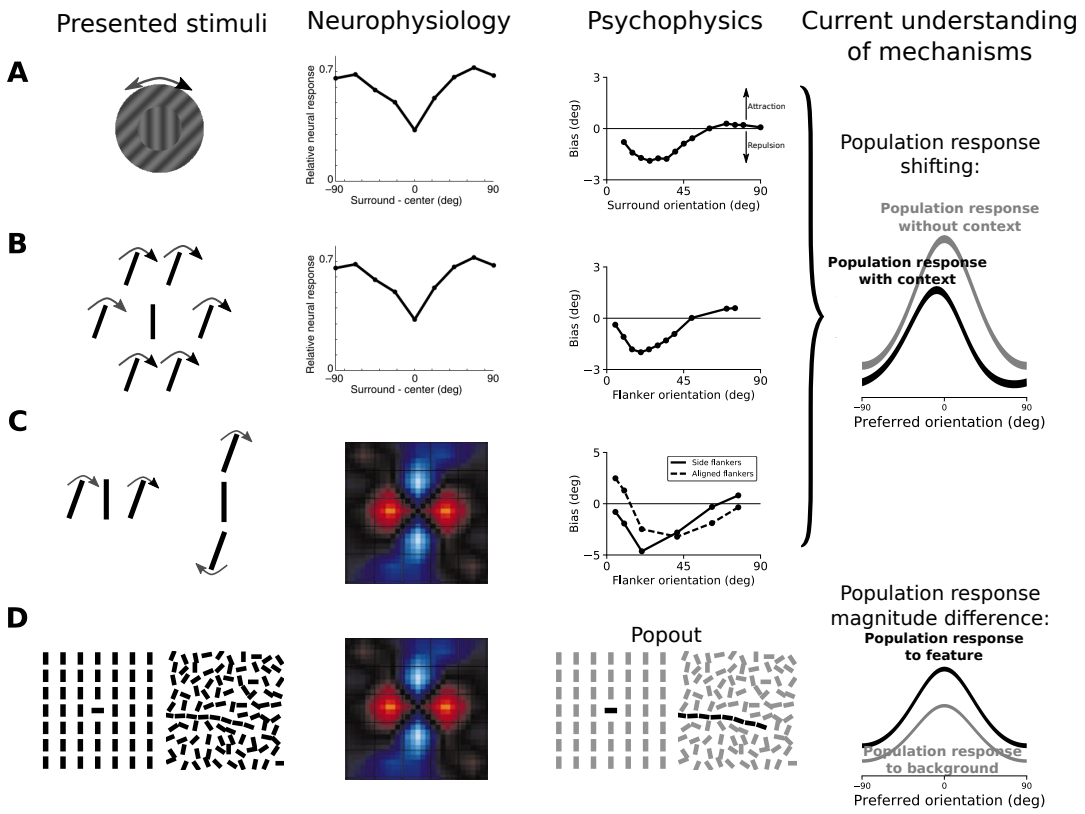

Figure 1.6: Overview of the current understanding of basic neurophysiology and psychophysics for the main stimuli studied in this thesis. The first column shows the types of stimuli studied in this thesis. The second column shows the commonly assumed underlying contextual interactions for these stimuli. The third column shows the main psychophysical effect, as considered in this thesis. And finally, the last column shows the current understanding of how the physiological results possibly explain the psychophysical results. (A) Center-surround stimuli. Here the perceived orientation of the center is biased as a function of the surround grating. A common model for this is a shift of V1 population responses due to the surround modulation (see also Fig. 1.5). The bias values were extracted from Qiu et al. (2013), Fig. 2. (B) A central bar, with a hexagon of surrounding flankers. The perceived orientation of the central bar has a similar illusion as with the gratings in A. The bias values were extracted from Clifford et al. (2000), Fig 2B (see also Westheimer, 1990). (C) A central bar accompanied by two mirrored flankers, either on the sides or aligned. Bias values were extracted from Kapadia et al. (2000b), Fig. 4C. (D) More extensive bar layouts, with some feature consisting of a set of bars, and a background of identically or randomly oriented bars. Here the an observer will quickly pick out the feature if its a clearly differing feature (as in the left example), or a because it is a clear contour (as in the right example). Note that the darker color in the psychophysics column only indicates the feature's saliency, not the perceived color. This effect is generally modeled by a higher V1 response to the feature than the background, due to the contextual modulations.

contours (until the angles between successive bars in the contour are 90 degrees) get progressively harder to extract (Field et al., 1993).

In terms of center-surround gratings salience does not make as much sense, as one would never ask to find the central grating. A similar effect does however occur in terms of the perceived contrast of the central grating, which changes as a function of the surround grating: the perceived contrast will be highest when the center and surround gratings are perpendicular, and will progressively decline as the center and surround are more closely aligned (Schwabe et al., 2010).

1.3 Linking physiology and psychophysics

The physiology section described how neurons respond to various types of stimulation, while the psychophysics section described how humans perceive the various stimuli. Presumably, some link should exist between these two. This is quite difficult to test experimentally directly, as we can't easily adjust the physiology and observe the effects on the psychophysics described above. Most attempt to link physiology and psychophysics have therefore relied on at least some degree of either purely hypothetical or explicit mathematical modeling.

1.3.1 Orientation illusions

A wide range of functional explanations exist for the orientation illusions described above (Schwartz et al., 2007; Clifford, 2014). Early theories of tilt perception hypothesized that the perception of tilt was normalized to the one or both of the major axes (i.e. vertical and horizontal) (e.g. Gibson and Radner, 1937; Beh et al., 1971). An influential more recent model claims that the illusions are more a result of self-calibration of the neural response, i.e. mostly in the form of a mechanistic explanation (Clifford et al., 2000). Using image statistic to group image elements based on the likelihood that they belong to the same object also turns out to give a good account (Schwartz et al., 2009).

The orientation illusion of a central bar with two flankers, as in Fig. 1.6C, was explained in the original paper as a consequence of the butterfly-shaped neural interactions (Kapadia et al., 2000b), which shifted the population response attractively for aligned flankers, or repulsively for the side-flankers. A more functional explanation in Schwartz et al. (2009) hypothesizes that the perception of the central bar is biased

towards the best continuation between the two flankers, which gives a similar result, with an attractive bias for two aligned flankers, and a repulsive bias for side-flankers.

While some of the above mechanisms can seem rather high level, they usually have a possible low V1-level implementation. These implementations all have in common that they rely on contextual modulations which 'shift' the population response from what it would normally be without the context, as in 1.6, top right. For example, for the grating based and bar-hexagon based tilt illusion, the surround modulation would be such that the population response is shifted away from the surround orientation, as neurons with preferred orientations close to the surround orientation are modulated more, resulting in a repulsive illusion.

1.3.2 Saliency

The main psychophysical principle used to explain contour extraction comes from the Gestalt-law of good continuation (Wertheimer, 1923; Koffka, 1935). Here, oriented elements that form 'good' continuations are more strongly associated together than oriented elements that form 'bad' continuations. These interactions are commonly referred to as the association field (Field et al., 1993), which is already quite similar to the butterfly-style modulations reported in Kapadia et al. (2000b). Association fields thus emphasize smooth contours, which is strikingly similar to the Schwartz et al. (2006a) model for the repulsive and attractive illusions for bars with a flanker pair.

As with the tilt illusion models this might at first sight seem like rather high-level processing, but again can be implemented at a V1-level through butterfly-like neural connectivity, giving rise to the 'saliency map' hypothesis of V1 (Zhaoping, 1998, 1999). Usually, when trying to explain saliency from just V1 responses, studies usually use some measure of the difference in neural response. For example, in the Li model, contextual interactions reduce the response of a background element more than a feature, or increase the response in a contour through butterfly-like interactions, and thus create a difference in response to a feature or a background element, 1.6, bottom right. Similar pop-out effects can also be obtained from general smooth contours theory (Ernst et al., 2012) or with the predictive coding hypothesis (Spratling, 2012).

1.3.3 Modulation tuning

Importantly, the current explanations for both the illusion and saliency effects all rely on the assumption that the contextual modulation depends on the preferred orientation

of a neuron, and the contextual orientation. This assumption is understandable for two reasons. First, contextual effects are often quite subtle, and hard to measure if a neuron is firing at a low rate. Therefore, most studies only studied contextual effects with optimal stimulation of a neuron's receptive field. Secondly, neurons with similar preference preferentially connect to each other both through long range (Bosking et al., 1997) and short range (Ko et al., 2011) connections. However, a long series of studies has shown that many neurons instead experience strongest modulation when the center and surround are aligned, irrespective of the neuron's preferred orientation (center dependent modulation)(Sillito et al., 1995; Cavanaugh et al., 2002a,b; Shushruth et al., 2012). It is as of yet unclear what affect, if any, this change will have on our understanding of the link between neurophysiology and the illusion and saliency effects.

1.3.4 Coding/decoding framework

Models for the perception of tilt often rely on the coding/decoding framework, Fig. 1.7. In this framework, a stimulus is assumed, a set of neurons that encode this stimulus in some way (tuning curves and some noise model), and finally some decoder then estimates the original stimulus using *only* the observed responses through some decoder process. In this thesis two recurring decoders will be the population vector (Georgopoulos et al., 1986a; Schwartz et al., 2009) and the maximum likelihood (ML) decoder (Kay, 1993). The population vector simply measures the average orientation encoded by a group of neurons' activity. As it basically measures location of the population response in terms of the neurons' preferred orientations, it is a natural choice for explaining the tilt illusion (as for example in Schwartz et al., 2009). The ML decoder meanwhile does something a bit more sophisticated, and gives the most probable encoded stimulus, given full knowledge of the neurons' encoding properties.

There are roughly two ways the coding/decoding framework is used; First, given an encoding model constrained by known neural properties, can we explain known psychophysical properties? For example, if one implements surround modulation, the surround grating moves the population response resulting in a biased representation of the center stimulus. This is the main use in the models for the tilt illusion described above. Secondly, generative models; given some constraints, what should the encoding model be? For example Bayesian theory (Deneve et al. (1999)), predictive coding (Spratling, 2012), or image statistics (Schwartz et al., 2009; Coen-Cagli et al., 2012). Generally studies using coding/decoding framework have identified various properties

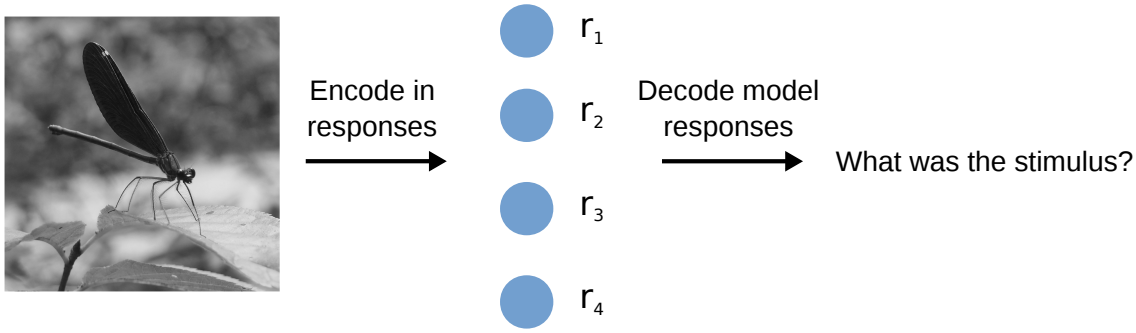

Figure 1.7: The coding/decoding framework. Some stimulus is encoded in a group of neurons using a know encoding model (such as each neuron responding to a specific orientation). Then, some decoding model is used (such as center of mass, or a probabilistic model based on the encoding model) to estimate the encoded stimulus.

such as optimal tuning width (Zhang and Sejnowski, 1999), that ML decoding is optimal (meaning unbiased and with decoding variance reaching the Cramér–Rao bound) (Kay, 1993), that population vector and simple center-of-mass decoders perform close to an ML decoder (Snippe, 1996), and that the ML decoder is (in principle) implementable in a neural network (Deneve et al., 1999; Jazayeri and Movshon, 2006b).

Although the encoding/decoding framework is widely used in tilt illusion studies, these tend to view a given group of neurons as still a one-dimensional encoder, while in reality they are a multiple-dimension encoder, as the contextual stimuli change the encoder behavior, and the contextual orientation could be seen as just another variable. Indeed, in most cases, the illusion might be a result of using a decoder that does not take this into account (Seriès et al., 2009). Additionally, many of the principled coding/decoding studies in particular have focused on this simple feed-forward system, with every neuron only representing single orientations, with no contextual effects and under asymptotic (infinite neurons/low noise) assumptions.

1.4 Central hypothesis and questions

I have so far outlined the main physiological and psychophysical for contextual effects, as well as the current state of the link between them based on modeling. A central conclusion we can draw so far is that most psychophysical results presented can be explained from known low-level V1 responses and contextual interactions. However, especially in regards to the illusions and saliency results, the effects are mostly explained individually.

In this thesis we develop this further and state the following hypothesis:

> The psychophysical observations can be explained from known V1 responses and contextual interactions through an encoding/decoding approach, and those contextual interactions from a single unifying principle.

As this is a rather broad statement, we must approach it in several steps:

First, in Chapter 2, we attempt to find a single principle to explain most of the contextual interactions and psychophysics discussed above. For contour saliency, some principle of smooth contours is clearly a strong contender through the association field, which also seems to have some success to explain some orientation illusions. I therefore hypothesize that the principle of smooth contours is the unifying principle we seek.

Secondly, in Chapter 3, we take a deeper look at the encoding/decoding framework on which many of the tilt illusion explanations (and also Chapter 2) rely, by using a population vector decoder which does not take into account the effects of contextual modulation. We know that the ML decoder is considered optimal and unbiased, although this has not been as extensively tested for multiple-variable encoding models (which we are essentially dealing with here). Chapter 3 extends the encoding/decoding framework to study this. Will decoding biases, and therefore orientation illusions, always fully disappear?

Thirdly, in Chapter 4, we take a look at the latest neurophysiological results. Most tilt illusion and saliency models, including what I will present in Chapter 2 and 3, rely on the dependence of contextual modulation on a neuron's preferred orientation. However, significant evidence now exists that a majority of V1 neurons actually experience center dependent contextual modulation (Shushruth et al., 2012). Do the tilt illusion and saliency results still hold under this assumption?

Lastly, in Chapter 5, we lay the basis for contextual modulation investigations in mouse V1 using 2-photon calcium imaging. Although the mouse model theoretically allows for both direct manipulation of specific cell types, as well as measuring many neurons at once, a significant problem is still extracting a clean calcium signal. Before we can have any hope of measuring the more subtle contextual effects, we will need to solve this problem first. Chapter 5 presents a toolbox to tackle this problem.

I will now briefly summarize the results for each chapter. Note that each chapter is either a published paper, or a paper in preparation, and will therefore mostly stand on its own.

1.5 Chapter 2: a unifying principle

In chapter 2 we tackle the problem of unifying the the psychophysical effects of saliency and tilt illusion under a single principle. We do this using elastica, a mathematical theory describing how curves should behave based on 'smoothness', which is widely used for contour extraction (e.g. Sharon et al., 1997) and has been linked to human psychophysics (e.g. Schwartz et al., 2006a; Ernst et al., 2012). However, these phenomena were mostly explained using separate models. In this chapter we present a single neural model based on elastica, which explains both known neurophysiology, and a large range of psychophysical effects, given simple assumptions of how bars surrounding a central bar modulate the neural responses, Fig. 1.8. This chapter was recently published as a paper in Vision Research (Keemink and van Rossum, 2016).

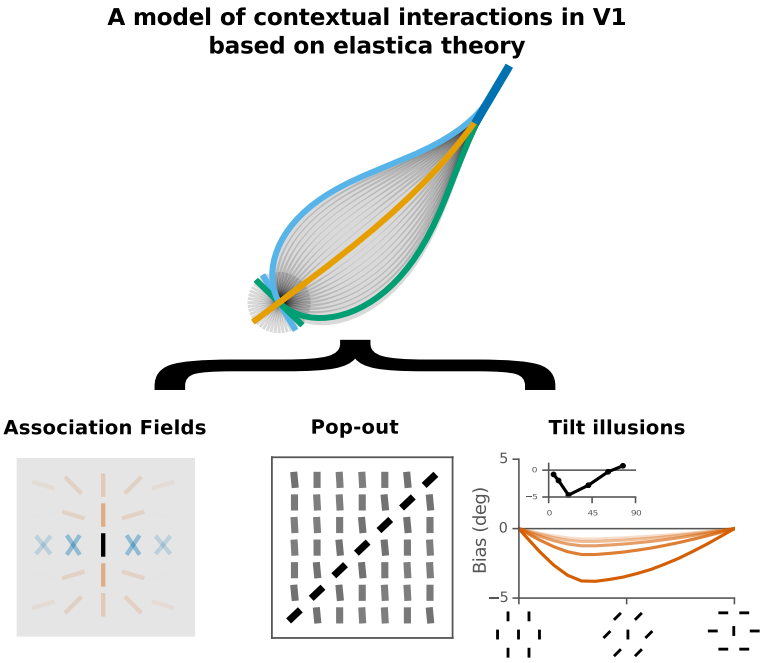

Figure 1.8: Chapter 2 visual abstract. Assuming that contextual modulations is informed by a theory of curve smoothness (elastica), a wide range of phenomena can be explained.

1.6 Chapter 3: Coding/decoding framework for contextual modulations

This chapter investigates the use of a maximum likelihood decoder for multivariate coding models (such as a group of V1 neurons responding to a center grating and

modulated by a surround grating). Although is known that decoding variance diverges from the Cramér–Rao bound for higher noise levels, it is still considered unbiased (Xie, 2002). This has furthermore mostly been studied for one dimensional stimuli. In Chapter 4 we seek to remedy this. We use two types of multivariate stimuli; two superimposed stimuli and the now familiar center-surround gratings. The resulting decoding distributions are highly non-uniform, leading to both biases and high variance, Fig. 1.9. We find that this holds true for a simple single-variate encoding model as well. The non-uniformity of the decoding distributions turns out to be due to ambiguities of the encoding models.

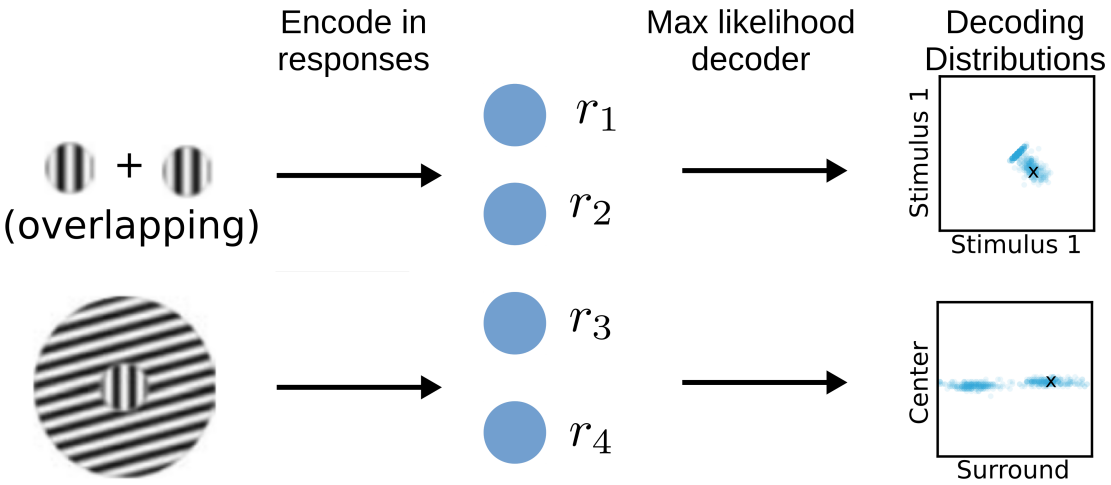

Figure 1.9: Chapter 3 visual abstract. While the maximum likelihood decoder is often considered the best possible decoder, this is only true in the limit of no noise. Chapter 3 explores several types of encoded stimuli in the presence of substantial noise, and finds that the decoding distributions become strongly biased even given simple encoding models.

1.7 Chapter 4: Implementing center dependent modulation

The many explanations for the tilt illusion all have relied on the fixed modulation model, Fig. 1.10 top row, which has often been linked to both psychophysical saliency effects (Shushruth et al., 2013) and the tilt illusion (for a review see Clifford (2014)). However, as described in the physiology section, most cells are instead center dependent, Fig. 1.10 bottom row. The functional consequences of center dependent modulation remain unknown.

The two types of modulation reflect two different coding models. In Chapter 4 we will show how the two models have two main functional differences. First, whilst we confirm that under fixed modulation biases arise, no such bias exists under center dependent modulation. Secondly, pop out effects due to surround modulation exist with both types of modulation, but are stronger under center dependent modulation, Fig. 1.10, suggesting a clear function for center dependent modulation.

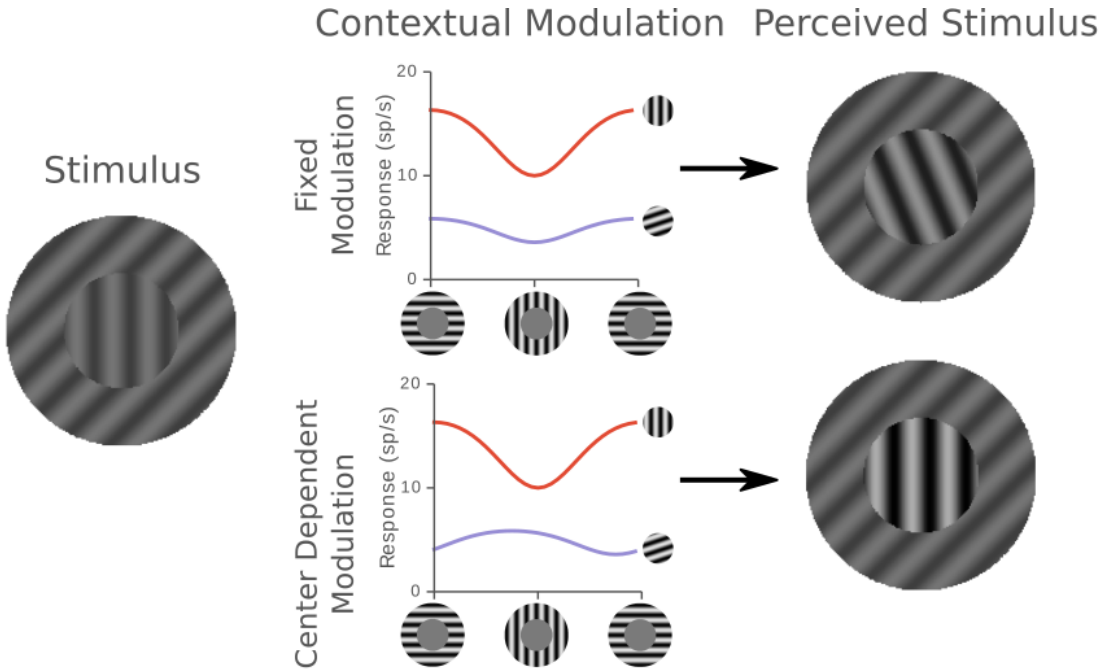

Figure 1.10: Chapter 4 visual abstract. When presented with a center surround stimulus, roughly two types of modulation exist in the primary visual cortex (middle). In fixed modulation, the modulation is strongest when the surround is aligned with the preferred orientation. For center dependent modulation, strongest modulation occurs when the surround is aligned with the center grating. In Chapter 4 we reproduce the well known effect that fixed modulation leads to pop-out of a center different from its surround, and of a biased representation of the center orientation (top-right). When the contextual modulation is made center dependent however, the pop-out effect is strengthened, while no bias of the center orientation occurs (bottom-right).

1.8 Chapter 5: Calcium-imaging signal separation

We would like to test several of the above results in the mouse brain using 2-photon calcium imaging. However, the contextual modulations effects discussed in this thesis

can be subtle, and one of the main problems with calcium imaging is that a cell's measured signal is often heavily contaminated by other nearby signals. While extraction methods exist that can deal with this contamination, they are slow and rely on automatic cell detection which is not always good enough yet; a better decontamination method is needed. Chapter 5 presents a toolbox which uses blind source separation on the measured cell's and surrounding signals to extract the true signal, Fig. 1.11.

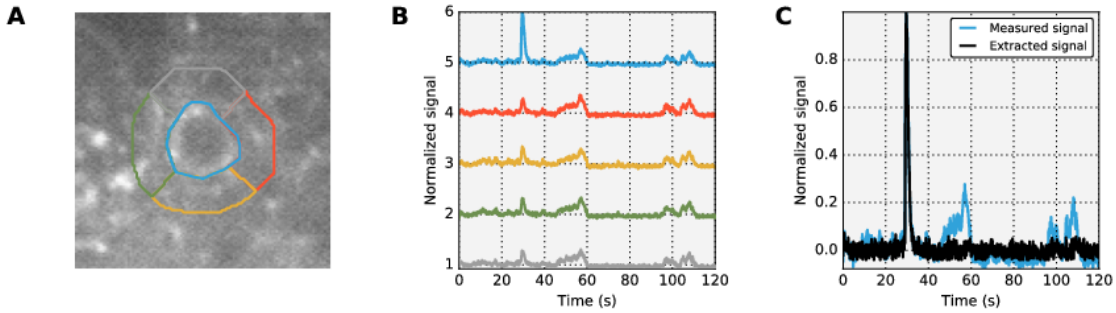

Figure 1.11: Chapter 5 visual abstract. A major problem in calcium imaging is contamination of the measured signal. Chapter 5 proposes a toolbox that is able to efficiently remove such contamination using blind source separation on the surrounding signals. (A) The region to measure (blue outline), and set of surrounding regions. (B) Extracted signals from each region. These are the signal blind source separation is applied to. (C) Measured signal (blue) and extracted signal (black).

1.9 Animations

Although this thesis is wholly self-containing and explanatory, some concepts are easiest to see in an animated fashion. Therefore this thesis is accompanied by animations at `https://swkeemink.github.io/thesis`.

Chapter 2

A unified account of tilt illusions, association fields, and contour detection based on elastica

*This chapter was published in Vision Research as **A unified account of tilt illusions, association fields, and contour detection based on elastica** (Keemink and van Rossum, 2016). All the work herein was done by me.*

As expressed in the Gestalt law of good continuation, human perception tends to associate stimuli that form smooth continuations. Contextual modulation in primary visual cortex, in the form of association fields, is believed to play an important role in this process. Yet a unified and principled account of the good continuation law on the neural level is lacking. In this study we introduce a population model of primary visual cortex. Its contextual interactions depend on the elastica curvature energy of the smoothest contour connecting oriented bars. As expected, this model leads to association fields consistent with data. However, in addition the model displays tilt-illusions for stimulus configurations with grating and single bars that closely match psychophysics. Furthermore, the model explains not only pop-out of contours amid a variety of backgrounds, but also pop-out of single targets amid a uniform background. We thus propose that elastica is a unifying principle of the visual cortical network.

2.1 Introduction

The Gestalt psychologists emphasized that human perception should be understood as a whole, rather than as the sum of individual elements. In the context of contour recog-

nition, they proposed the law of good continuation, in which collinear or curvilinear line elements are associated together (Wertheimer, 1923; Koffka, 1935). This principle presumably underlies the ease with which humans extract smooth contours in natural as well as artificial images. The detection of contours does not necessarily require high level vision or receptive fields that encompass the complete contour. Instead, Field et al. (1993) argued that association fields in early vision boost the response to collinear line elements, Fig. 2.1A, and that these local interactions are sufficient for contour detection. Computational models have used association fields to explain contour extraction and completion (e.g. Zhaoping, 1998, 1999; Williams and Thornber, 2001; Tang et al., 2007).

Evidence for the association field has been found in psychophysics (Field et al., 1993; Kapadia et al., 1995, 2000b; Ernst et al., 2012), neural recordings (Kapadia et al., 1995, 2000b; Bauer and Heinze, 2002; Li et al., 2006), and in neural connectivity (Bosking et al., 1997), and supports the view that contour detection happens as early as primary visual cortex (V1). Furthermore, association fields have been linked to image statistics (e.g. Sigman et al., 2001; Geisler et al., 2001) and predictive coding (Spratling, 2012).

From a functional point of view, association fields emphasize smooth contours. Indeed, in Fig. 2.1B the red bars are smooth continuations from the black bar, while the blue bars are more tortuous continuations. Smoothness can be quantified using the elastica principle, which measures the bending energy needed to connect two bars with a curve. In computer vision the curve with the lowest energy is better known as a spline (historically, a thin wooden rod used by draftsmen to create smooth curves). Line elements that can be connected with a low energy spline likely belong to the same object contour and are therefore particularly relevant to higher level vision. The elastica curve can also be seen as the maximum likelihood path of a stochastic contour completion process based on drifting particles, with the two bars as source and sink elements (Mumford, 1994; Williams and Jacobs, 1997; Williams and Thornber, 2001).

In this paper we propose that smoothness, as formalized by elastica, is the underlying principle for contextual modulations in V1. We first derive an efficient calculation of the elastica energy between two oriented line elements. We next assume that the contextual modulation in a neural population is determined by this elastica energy. We find that the resulting model has 1) realistic association fields and neural responses, 2) produces various tilt illusions for both small sets of bars and grating-like stimuli, and, 3) leads to robust contour extraction, as well as single target pop-out from uniform

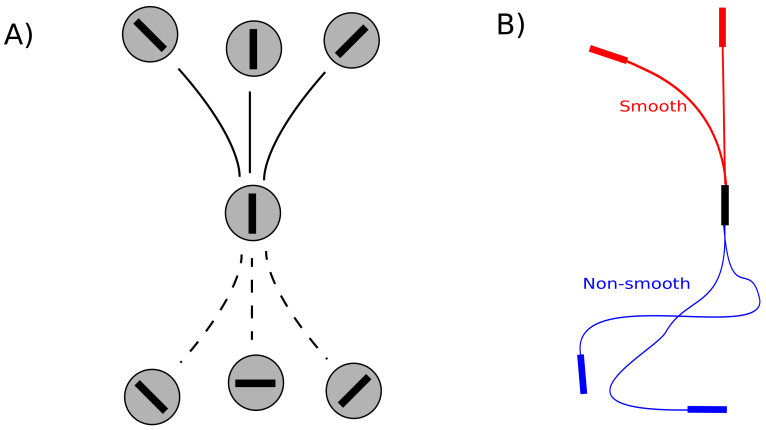

Figure 2.1: Visual association fields. (A) The basic idea of an association field, as theorized based on psychophysical experiments in Field et al. (1993). Solid lines connect collinear elements and correspond to strong associations, while dotted lines correspond to weak associations. (B) Association fields and the relation to smooth contours. When trying to connect the center bar to any of the 4 presented flankers using smooth lines, the connections to the red bars are smoother continuations and the connecting curve will have lower curvature energy than the blue bars.

backgrounds.

While links between these phenomena and elastica have been shown before individually, from contour extraction (Sharon et al., 1997; Ernst et al., 2012) to the tilt illusion (Schwartz et al., 2006a), there has been considerably variation in the precise implementation and biological realism of these studies. Here we present a unified account that, to our knowledge, is the first to collect such a wide range of phenomena in a single model with elastica at its core. Yet, the model is straightforward and biophysically realistic, relying only on independent contextual modulation terms. These results suggest that elastica is a core principle underlying the contextual interactions in V1.

2.2 Methods

2.2.1 Elastica

In this section we quantify smoothness according to the elastica principle. Extending earlier results we derive an accurate approximation for the smoothness of a curve connecting two line elements. Consider a scene with two bars, Fig.2.2A: a center bar (red) and a flanker bar (blue). We define the positions and orientations as follows: The

orientation of the center bar relative to the vertical is θ_c. The flanker is a distance r_f away from the center, and placed at a position which has an angle φ_f with the vertical. The orientation of the flanker is given by θ_f. The angles of the center and flanker bar relative to the line connecting them, are

$$\beta_c = \varphi_f - \theta_c$$
$$\beta_f = \theta_f - \varphi_f,$$

where the minus signs signifies circular differences, so that they lie in the interval $[-\pi, \pi]$.

We would like to know the smoothest curve connecting the two line elements. This minimization problem, known as elastica, has a long history dating back several hundred years (Mumford, 1994; Levien, 2008). In particular in computer vision and computer graphics this problem has been studied extensively, where the smooth curves are known as splines. According to the elastica principle, the two line elements are imagined to be connected using a flexible rod. Since most human perception seems to be scale invariant, we use the scale invariant version of the elastica energy (Bruckstein and Netravali, 1990; Sharon et al., 1997)

$$E_{true} = L \int_0^L \left[\frac{d\Psi(s)}{ds} \right]^2 ds, \tag{2.1}$$

with s signifying the position along the curve, Ψ the relative angle of the curve at location s, and L the total length of the curve. The energy has a minimal value of zero when the curve is straight. As a curve becomes more tortuous, and thus has more curvature, the energy increases.

To find the smoothest curve, this energy has to be minimized w.r.t. $\Psi(s)$ and subject to the conditions $\Psi(0) = \beta_c$ and $\Psi(L) = \beta_f$. The length of the bars is assumed much shorter than r_f so that the curve between two bars goes from the middle of one bar to the other (see Discussion). The elastica energy can be found using the minimization method outlined in Sharon et al. (1997).

Typically elastica assumes a start and end direction. However, in our case, the elements have an orientation only, and we thus need a direction invariant energy, or the minimal energy across situations where the center and/or flanker angles can be flipped 180 degrees. This results in the direction invariant elastica energy

$$E_{\text{inv}}(\theta_c, \theta_f, \varphi_f) = \min_{k,l=\{-1,0,1\}} E(\theta_c + k\pi, \theta_s + l\pi). \tag{2.2}$$

Fig. 2.2B shows the family of curves that minimize this energy as a function of the orientation of the center. The corresponding energy for the curves is depicted in panel C. Notably, the energy has cusps where the solution switches from the green to the purple curve.

Finding the true curvature energy is computationally expensive. However, the energy of the smoothest curve between a given center and flanker pair is approximated by (Leung and Malik, 2001; Sharon et al., 1997)

$$E_{\mathrm{S}}(\theta_c, \theta_f, \varphi_f) \approx 4(\beta^2 + \beta_f^2 - \beta_c\beta_f), \tag{2.3}$$

While equation 2.3 was derived from the assumption of small angles, it turned out to be a very good approximation for larger angles (Sharon et al., 1997). Furthermore, the direction invariant energy (Eq.2.2) based on this approximation is an exceptionally good match with the true direction invariant energy. Fig. 2.3 compares the approximated and true energies across different center orientations for several flanker orientations. The error is small - certainly for our purposes - and free of qualitative differences.

2.2.2 Neural model

Here we implement how the neural responses in V1 are modulated by the surround using the elastica principle. For each bar, we assume that there is a population of $N = 32$ neurons, with preferred orientations ϕ_i in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Initially we shall be mainly interested in the population that encodes the center bar. The response of a neuron in the absence of flankers is modeled by a von Mises function

$$g(\phi_i, \theta_c) = A_c \exp(K_c \cos 2[\phi_i - \theta_c]), \tag{2.4}$$

where A_c is the response amplitude, which we set to 1 Hz without loss of generality, as we only consider stimuli with identical contrast (see Discussion for a possible extension to stimuli with heterogeneous contrasts). θ_c is the orientation of the stimulus in the neuron's receptive field (the center bar). K_c sets the width of the neural tuning, with narrower curves for higher values. We use $K_c = 1$. The function $g()$ across the population is illustrated in Fig. 2.2D by the red curve.

The smoothest curve connecting a flanker to a center bar with a neuron's preferred orientation ϕ_i has a curvature energy $E(\phi_i, \theta_f, \varphi_f)$. We propose that the neural response is modulated by a flanker through a modulation term

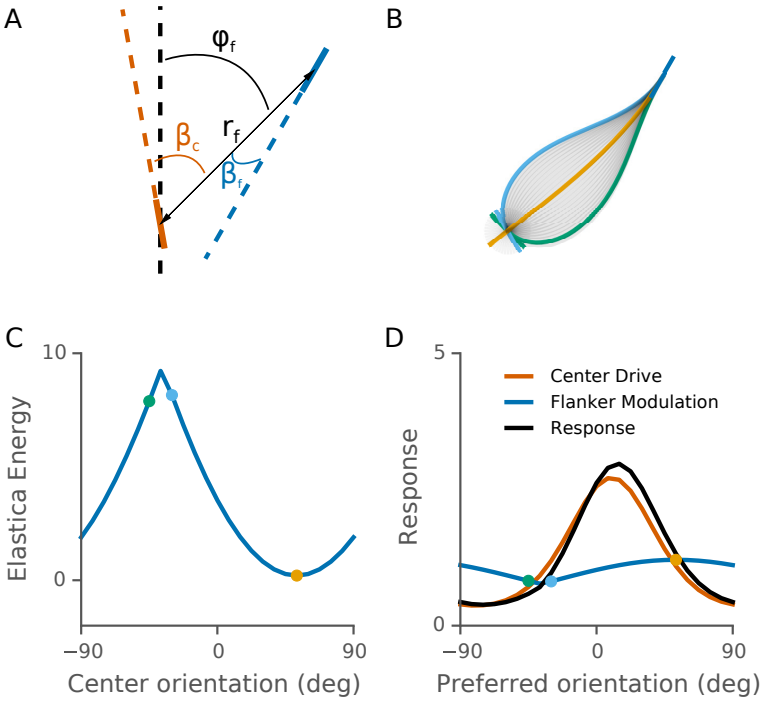

Figure 2.2: The elastica smoothness and its proposed effect on surround modulation. (A) Setup of the center and flanker bars, illustrating the various angles used in the calculation of the curvature energy. (B) The family of minimum energy curves connecting the flanker to the center bar for various center orientations. (C) The energy of the curves in panel B, as a function of center orientation θ_c. The colored dots correspond to the colored curves in panel B. The cusp occurs when the minimum energy curve switches sides (purple and green curve). (D) The corresponding population response of neurons with the center bar as their receptive field as a function of the neuron's preferred orientation. Each neuron receives a feed-forward drive from the center bar with a strength dependent on the neuron's preferred orientation (red). Each neuron also receives contextual modulation from the flanker (blue) which depends on the curvature energy of the elastica curve connecting the flanker to a center bar with the neuron's preferred orientation. The input drive is multiplied by the modulation to give the resulting population response (black), which is both deformed and shifted compared to the feed-forward drive.

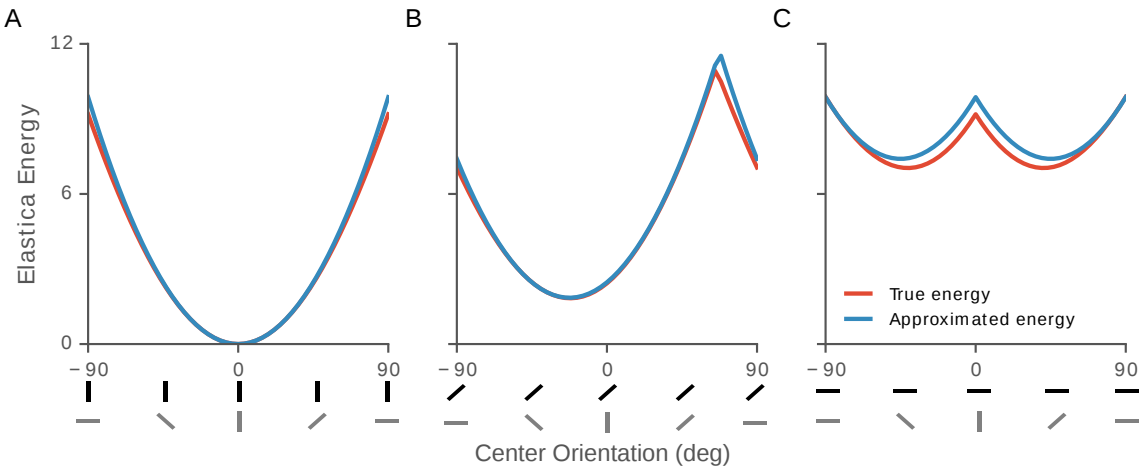

Figure 2.3: The Sharon approximation (blue curve, Eq.2.3) compares well to the true elastica energy (red curve, Eq.2.1). Both energies were made direction invariant (Eq.2.2). (A-C) The elastica energy across different center orientations, for fixed flanker orientations of 0, 45 and 90 degrees respectively. For -90, -45, 0, 45 and 90 degrees a situation sketch is shown, with the gray bar indicating the center orientation, and the black bar the flanker orientation. Note that for a perpendicular flanker, relative to the center, the energy is minimal at two center orientations resulting in two dips in the energy, panel C.

$$h(\phi_i, \theta_f, \varphi_f, r_f) = \exp\left[-\frac{a}{r_f}\left(E(\phi_i, \theta_f, \varphi_f) - E_0\right)\right].$$

This modulation is illustrated across the population (i.e. versus ϕ_i) in Fig. 2.2D by the blue curve. This formula was arrived at as follows. The elastica energy is always positive, but, in order to be consistent with physiology, we want flankers with a low curvature energy (i.e. high smoothness) to facilitate the response. Since $E \geq 0$ for all cases, we subtract an offset energy E_0 from the elastica energy, so that $h() > 1$ for smooth contours. In addition, we divide the energy by r_f, the distance to the flanker, so that far away flankers do not modulate the response, similar to previous elastica studies (Sharon et al., 1997; Bruckstein and Netravali, 1990). The gain parameter a determines the strength of the modulation. In the limit when $a = 0$, one has $h() = 1$ and the response is independent of any flankers. For all simulations we use $a = 0.1$, and $E_0 = 4$. The exact values of these parameters have little qualitative effect, as will be discussed at the end of the results.

The modulation $h()$ acts multiplicatively, consistent with physiology (Cavanaugh et al., 2002b). We assume that each flanker contributes independently to the modula-

tion, so that the final response of a neuron to a stimulus with n flankers is

$$r_i = g(\phi_i, \theta_c) \prod_{j=1}^{n} h(\phi_i, \theta_f^j, \varphi_f^j, r_f^j).$$

An example for $n = 1$ flankers is shown by the black line in Fig. 2.2D.

2.2.3 Population vector

We read out the orientation encoded by the population using the population vector method. The population vector is a 2D vector given by the sum of the preferred orientation vectors of the neurons weighted by their firing rate (Georgopoulos et al., 1986a), representing the estimated center orientation vector $\hat{\mathbf{v}}_c$

$$\hat{\mathbf{v}}_c = \sum_i r_i \mathbf{u}_i,$$

where r_i is the firing rate, and $\mathbf{u}_i = (\sin 2\phi_i, \cos 2\phi_i)$ is the unit vector pointing in neuron i's preferred orientation (multiplied by two to ensure circularity). The estimated center orientation $\hat{\theta}_c$ follows from the angle of the population vector

$$\hat{\theta}_c = \frac{1}{2} \angle \hat{\mathbf{v}}_c \tag{2.5}$$

where $\angle$ denotes a vector's angle.

2.2.4 Computations

All data analysis and models were implemented in Python 2.7.5, using the Numpy 1.7.1, SciPy 0.12.0 and Matplotlib 1.2.1 toolboxes. Model code is available at `https://github.com/swkeemink/elastica`.

2.3 Results

In the context of the Gestalt law of good continuation, we study a neural network in which the contextual interactions are based on pairwise optimal smoothness. For clarity we restrict ourselves to images composed of bars with various orientations all with identical contrast, and at a single spatial scale. Each bar is assumed to fall in the classical receptive field of a population of neurons with preferred orientations $\phi_i = -\pi/2 \ldots \pi/2$. We will refer to the bar under consideration as the center bar, and the

other bars as the flankers. As an example, consider the scene in Fig. 2.2A, where the red bar represents the center and the blue bar a single flanker. We write the neural response of neuron i with preferred orientation ϕ_i as

$$r_i = g(\phi_i, \theta_c) h(\phi_i, \theta_f, \varphi_f, r_f).$$

The function $g()$ models the classical V1 orientation response to the center stimulus using a von Mises function, where θ_c is the orientation of the center stimulus. The function $h()$ models the contextual modulation from the flanker as follows (Methods):

$$h(\phi_i, \theta_f, \varphi_f, r_f) = \exp\left\{-\frac{a}{r_f}\left[E(\phi_i, \theta_f, \varphi_f) - E_0\right]\right\},$$

where θ_f, φ_f an r_f are the flanker orientation, angular position, and distance from the center respectively. $E()$ describes the curvature energy of the smoothest curve connecting the flanker to a bar of orientation ϕ_i in the center. The parameters a and E_0 set the strength and offset of the modulation (see Methods and below). The smoother the curve from the flanker to the center, the more positive the modulation (for smooth curves for which $E() < E_0$ it become facilitatory, $h() > 1$), whereas when the curve is tortuous and thus has a high elastica energy, the modulation is inhibitory. In the Methods the functional form of the modulation is derived from the elastica principle, and an efficient approximation for the curvature energy is presented. The center drive $g()$, contextual modulation $h()$ and resulting population response $\mathbf{r}$ are illustrated in Fig. 2.2D. It can be observed that due to the modulation the population response is a deformed version of the center drive.

2.3.1 Single flanker association field

We first analyze the association field that the model predicts by considering the effect of a single flanker on the response of a neuron with preferred orientation $\phi_i = 0$, as represented by the black bar in Fig. 2.4. A flanker in a given surround location will modulate the response of this neuron by a factor $h(\phi_i = 0, \theta_f, \varphi_f, r_f)$. In Fig. 2.4A we indicate at each location which flanker orientation yields the most positive modulation, $\text{argmax}_{\theta_f} h(0, \theta_f, \varphi_f, r_f)$. The color of a bar indicates excitation (red, $h() > 1$) or inhibition (blue, $h() < 1$), and its opacity is proportional to the modulation strength. The bars that increase the response (red bars in panel A), correspond to flankers that form a smooth contour. Flankers with an inhibitory effect correspond to more tortuous

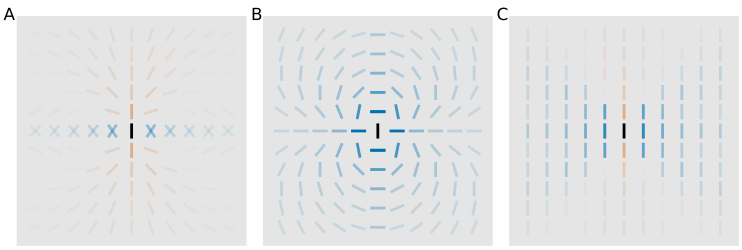

Figure 2.4: Association fields as derived from the elastica principle. (A) The association field for a neuron with preferred orientation indicated by the black bar. In each location, the flanker orientation that most positively modulates the activity is shown. Red (blue) flankers correspond to facilitatory (inhibitory) modulation; the opacity indicates the strength of the modulation. (B) Same as in A, but instead the orientations that most **decrease** the activity are shown (the "dis-association field"). (C) The effect of a flanker of the same orientation as the preferred orientation in the center.

contours. The parameter E_0 determines for which amount of curvature the modulation become inhibitory (see also Fig. 2.8).

To get further insight in the model we plot the bar which leads to the most suppressive modulation, $\text{argmin}_{\theta_f} h(0, \theta_f)$, which can be called the 'dis-association field' in panel B. Consistent with Fig. 2.3, the strongest suppressing flankers are approximately rotated 90deg from the most facilitating ones. Furthermore, inhibition exists across a wider range of orientations.

Finally, we plot the effect of flankers oriented the same way as the preferred orientation to more directly test our association field against neural measurements, in which the flanker orientation is usually kept the same as the preferred orientation of the neuron being measured, panel C.

The butterfly shapes in panel A and C corresponds to both psychophysics in the form of the association field (Field et al., 1993), as well as electrophysiological results in monkeys where, in particular at low contrast, flankers collinear with the preferred orientation in the center excite (Kinoshita et al., 2009), and flankers parallel to the center inhibit (Kapadia et al., 1995, 2000b).

2.3.2 The tilt illusion: two flankers

Next we study the case when two flankers are placed oppositely each other, around the center. To extend the model to this situation we assume that each flanker independently modulates the response, thus the response of a neuron in the center population

is $r_i = g(\phi_i, \theta_c)h(\phi_i, \theta_f^1)h(\phi_i, \theta_f^2)$. We decode the center population response using the population vector (Methods). Due to deformation of the population response caused by the modulation, the population vector is no longer aligned with the stimulus orientation. Psychophysically this presumably leads to a tilt illusion in the percept of the center bar.

We illustrate several configurations in Fig. 2.5. In each panel we show two example rotations in the top row. The elastica curves connect the flankers to neurons in the center with different preferred orientations; the curve's opacity is proportional to that neuron's response, indicative of the resulting population response.

For parallel flankers rotating around the center, panel A, the elastica energy is smallest for neurons with preferred orientations ± 45 deg from the flankers' orientations Fig. 2.3C, resulting in two "valleys" in the energy. Accordingly, the elastica curves from these flankers fan in two directions, Fig. 2.5A top. The resulting modulation shifts the population response depending on the flanker orientation. In the example of a 30 degree flanker rotation, top left, the modulation is mostly counter-clockwise. However, when the flanker is rotated to 60 degrees, top right, the same curves pass the vertical before reaching the center, resulting in the population responses being shifted clockwise. These effects result in a repulsive illusion for flanker rotations 0 to 45 degrees and an attractive tilt for 45 to 90 degrees, panel A bottom.

When the lateral flankers are rotated in place, Fig. 2.5B top, the effect is always repulsive. With the flankers either at 30 degrees (left) or 60 degrees (right), the most influential elastica curves from the left flanker always move up first before going down (and vice versa for the right flanker), ending in an orientation which is repulsed away from the flanker orientation.

For the aligned flankers rotating around the center, panel C top, elastica curves connect smoothly to the flankers, with the lowest energy curve being a straight line. As a result the modulation is excitatory for neurons with a preferred orientation close to the flanker orientation (Fig. 2.4A), and suppressive far away, and the population response follows the flankers. Across flanker orientations this generates an attractive illusion. Note that the attractive effect does not require excitation, it is sufficient that the modulation from both flankers is least suppressive for the neurons with a preferred orientation equal to the flanker orientation.

Next, we rotate the flankers in place above and below the center, panel D top. When the flankers are tilted 45 degrees clockwise, right example, the elastica curves from the top flanker start off towards the left, then bend back towards to center. The

population response is thus shifted away from the flanker orientation, resulting in a repulsive illusion.

The two bar illusions in the model are in close accordance with known psychophysics. Westheimer (1990) reported repulsion from tilted flankers on the sides (Fig.2.5B), or above and below (panel D). The orientation dependence of the repulsion from lateral flankers (panel B) closely matches Kapadia et al. (2000b) who also found repulsion maximal at around 30 deg (except for the small attractive effect they reported for larger flanker orientations), inset panel B. Kapadia et al. (2000b) also placed flankers above and below the center, and tilted them (as in panel D). But in addition the flankers were displaced so that the near end of the flankers aligned with the center bar. In this case they found an attractive effect for small flanker orientations, and a repulsive effect other orientations, inset panel E. The attractive effect went away quickly as the flanker distance was increased.

We reproduced their setup by assuming a finite bar length for the flanker rotation, and displacing the flanker accordingly, as in Fig. 2.5E top. Since the bar lengths in our model are assumed infinitesimally short, the curves are drawn from and to the centers of the bars. This turns out to be key to understanding the attractive part of this illusion. As long as the flankers are close enough, smaller orientations will seem aligned as the origin of the elastica curves is slightly displaced, panel E top left. When the flanker orientation is larger, this is no longer the case and the curves first need to pass the vertical before reaching the center, resulting in a repulsive effect. When the flankers are moved further away (i.e. displaced in the y direction), the situation becomes more analogous to panel D, as the x-displacement of the flanker becomes insignificant, and the illusion turns repulsive for all flanker orientations. Although the precise angular dependence in the model does not match the data, it is surprising that it can exhibit both tilt effects.

All illusions and modulations weaken as the flanker distance was increased, as illustrated by curves of decreasing opacity in the bias graphs. This is in accordance with most tilt illusion studies which note that the illusions decreases in strength as the contextual stimuli are placed further way (Westheimer, 1990; Kapadia et al., 2000b).

In the context of elastica, the illusions in Fig. 2.5B and D were previously explained by noting that the lowest energy curve connecting the two flankers is either oriented towards the flankers' orientation, as in A, or away from them, as in B (Schwartz et al., 2006a). Bayesian estimation then results in an orientation between this smooth orientation and the presented center orientation, resulting in the repulsive or attractive

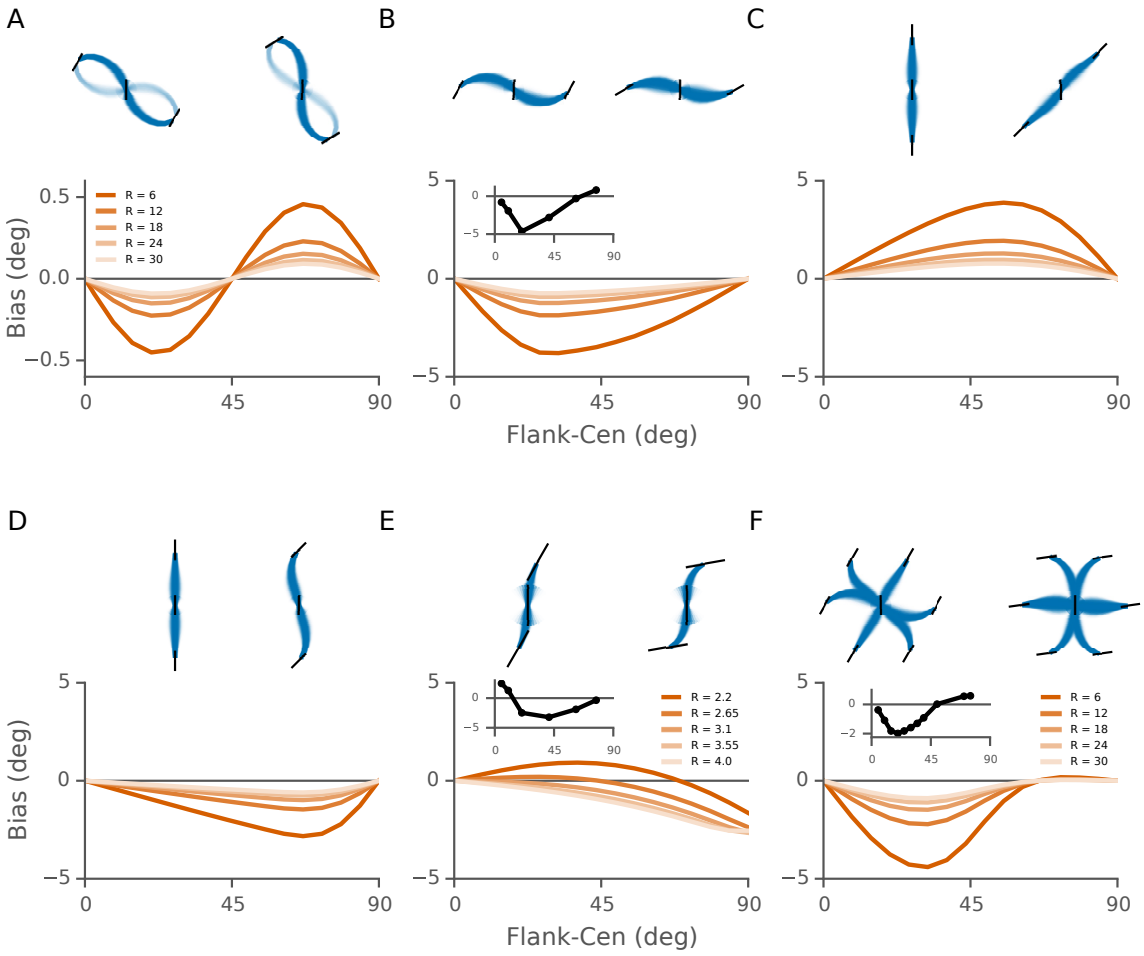

Figure 2.5: Tilt illusions from elastica-based contextual modulation by flanker pairs. Negative bias indicates repulsion, positive bias indicates attraction. Top rows: example configurations, the elastica curves connect the flankers to neurons in the center population with a specific preferred orientation. The opacity is proportional to the response of the corresponding neuron. (For clarity the modulation strength was increased to $a = 0.5$). Bottom rows: tilt bias as a function of flanker orientation. (A) Two parallel lateral flankers are rotated around the center (at 0 degrees the flankers are on the sides). The illustrated flanker tilts are 30 and 60 degrees from left to right. (B) Same as panel A, but with the flankers kept in place. The tilts are 30 and 60 degrees left and right. Inset: illusion data extracted from Kapadia et al. (2000b), Fig.4C. (C) Two aligned flankers, rotating around the center. The examples have a rotation of 0 and 45 degrees respectively. The elastica curves follow the flankers, resulting in an attractive illusion. (D) Keeping the flankers in place and tilting them in place. The elastica curves follow the flanker tilt and then bend back towards the center. (E) Reproducing the alignment in experiment in Kapadia et al. (2000b), the ends of the flankers were aligned with the center bar. Inset: illusion data extracted from Kapadia et al. (2000), Fig.4C. (F) Reproducing the setup in Westheimer (1990), with 6 flankers in a hexagon which are rotated in place. Inset: averaged illusion data extracted from Westheimer's Fig.1.

illusion. However, an additional displacement of the center bar was allowed to produce an attractive solution for small angles in panel C.

2.3.3 The tilt illusion: full surround.

The two flanker stimuli described above, lead to both repulsive and attractive effects. However when either a hexagon of surround bars Westheimer (1990), or a surround grating Clifford (2014) is used a repulsive tilt illusion occurs for most center-surround orientation differences, while a weak attractive effect occurs for larger orientation differences (inset Fig. 2.5F), which has been speculated to have a different origin Clifford (2014). Mechanistically, the repulsive tilt illusion has been explained by the fact that a surround grating results in orientation tuned suppression, with most suppression when the surround is the same as a neuron's preferred orientation (e.g. Clifford et al., 2000; Schwartz et al., 2007).

To examine these illusions in our model, we first turn to Westheimer's experiment, which we can reproduce exactly. There a hexagon of 6 flankers was placed around a center bar, evenly spaced so that there are two parallel bars on the sides, as in Fig. 2.5F top. The bars were then rotated in place and we measured the effect on the neural population response at the center location. As above we assume that each flanker independently modulates the center responses and the orientation of the center bar was again decoded from the neural activities using a population vector. For most orientations, the net effect from the elastica curves to the center is repulsive. However, when they are close to perpendicular to the center, the four top and bottom flankers win out with a small attractive effect (also see Fig. 2.8). Thus the model explains both the repulsive illusion, and the attractive effect for larger orientation differences.

We next approximate a center grating by a single oriented bar, and a surround grating as a large set of 16 identically oriented bars, Fig. 2.6A. At first glance perhaps a weak attractive tilt would be expected again. However, the net modulation from all flankers is inhibitory, panel B, and strongest when a neuron's preferred orientation is the same as the surround orientation, in close accordance with known neural responses. As a result, the decoded orientation is repulsed away from the surround orientation, corresponding to a repulsive tilt illusion, panel C. We also varied the number of flankers and found much the same effects, with still a weak attractive effect for 8 bars, but repulsion otherwise (not shown). In summary, tilt illusions in stimuli with surround gratings and with pairs of flankers can be unified under the elastica principle.

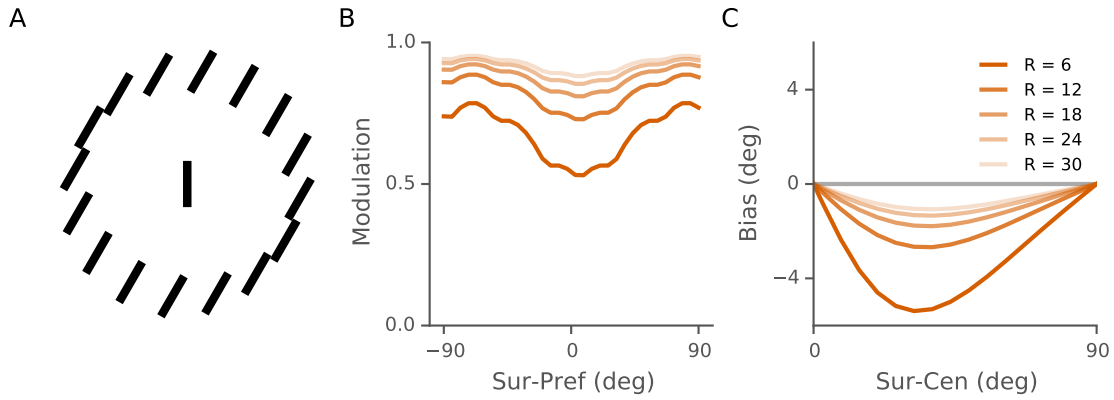

Figure 2.6: Tilt illusion from elastica-based surround modulation by a grating type stimulus. (A) The stimulus consists of the center bar surrounded by a grating consisting of equally orientated bars. (B) The net modulation from the surrounding flankers across the population, relative to the surround orientation. As the orientation of the surrounding bars is varied relative the center orientation, the net response varies in a way reminiscent of surround modulation in V1, with strong suppression when the preferred orientation of a neuron and surround are aligned. (C) The bias is repulsive across all surround orientations.

2.3.4 Contour detection

So far we have focused on the effect of flankers on the decoded orientation of a center bar. We now turn our attention to larger scenes consisting of several bars, where we find the response to each bar in succession by taking that bar as the center, and considering all other bars as the flankers. The principle of elastica and smooth contours has classically been used to extract contours from images. These implementations typically explicitly calculate the curvature energy for all element combinations, rather than incorporating the energy in a modulation term as we do here. While association fields in general have been used to facilitate contour detection through contextual modulations (e.g. Field et al., 1993; Zhaoping, 1998, 1999; Bauer and Heinze, 2002), here we examine if elastica based modulation also leads to contour extraction.

To study contour extraction in our model we measure the apparent saliency s of the contour (as in Zhaoping, 1999). The apparent saliency of a bar is defined as the maximal response in the population responding to that bar (i.e. the maximum over their preferred orientations ϕ_i), compared to the maximal responses to bars in other

locations

$$s = \frac{\max_{\phi_i}(r)}{\langle \max_{\phi_i}(r) \rangle_{\text{image}}}.$$

where the average in the denominator is taken over the whole image. The saliency of a complete contour is defined as the mean maximal response to a contour relative to the whole image

$$s = \frac{\langle \max_{\phi_i}(r) \rangle_{\text{contour}}}{\langle \max_{\phi_i}(r) \rangle_{\text{image}}}.$$

where the average in the nominator is over the bars that constitute the contour. If $s > 1$, the responses are higher in the contour and it is salient. We also implemented a mean based saliency measure, which uses mean responses instead of maximal responses. This resulted in weaker saliencies, but no qualitative differences (not shown).

In the plots that follow, we show the encoded image above its modeled percept. The opacity of the decoded bars is proportional to its saliency. Consider, first, a lone target bar amidst a homogeneous background, Fig. 2.7A. While this is not a true contour, from psychophysics and neural measurements we expect the target to be salient (Nothdurft, 1993; Shushruth et al., 2013). Indeed, the neurons in the background are inhibited more than the center flanker, similar to the effect in Fig. 2.6B. Due to the resulting higher neural response of the center bar it jumps out from the background, as signified by its darker color. As the target is rotated towards the surrounding orientations, the saliency decreases until it is no longer salient (not shown).

Next, we embed a simple contour in a homogeneous background, as in Fig. 2.7B. Here again, for the decoded image the bars of the feature of interest are darker than those in the background, indicating a salient contour. In this case, although actually all bars in the image experience suppression, those in the background are suppressed more, since they are surrounded by more bars of similar orientations. However, bars that are part of the contour enhance each others responses, resulting in the high saliency of the contour. Further note that the decoded bar orientations differ from the stimulus orientations.

We finally examine saliency of more general contours in random backgrounds. We use the method described in Field et al. (1993) to generate random images containing a random contour of length 8. Briefly, the contour is generated with a starting orientation and location, after which a set orientation change is made in a random direction (left or right), and a new bar is placed following the new orientation. An example stimulus containing a contour with orientation changes of 11.25 degrees, is shown in Fig. 2.7C,

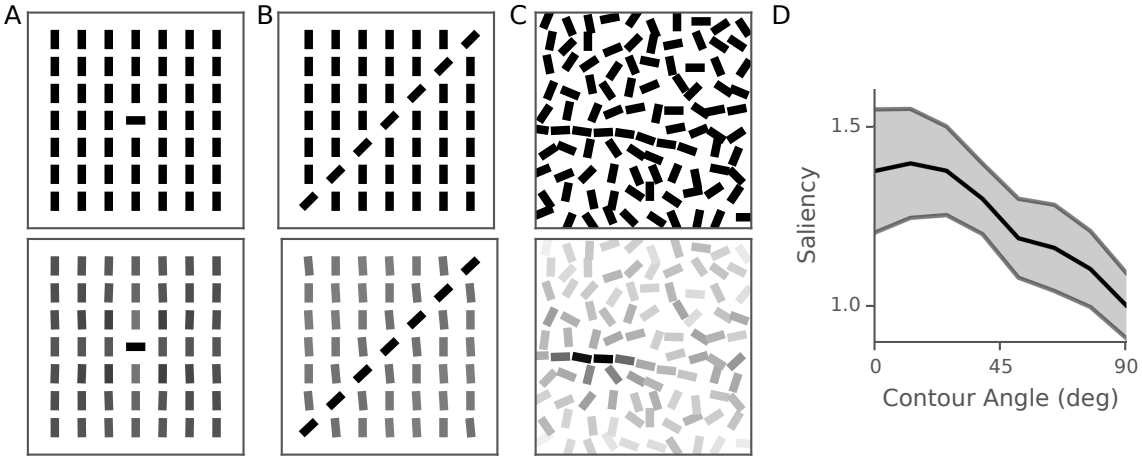

Figure 2.7: Saliency effects resulting from elastica and contextual modulations. Top: input stimulus, bottom, perceived stimulus where the opacity codes for the maximal response strength of the corresponding population/or the saliency (A) A single target stimulus pops out from the background, and does so more strongly based on how different the orientations are (now shown). The bars are placed on a grid such that the distance between two neighboring bars is 5. (B) Pop out of a simple contour in a uniform background. (C) Example of a noisy scene, when the angles between each bar in the contour are 11.25 degrees. The spacing between the centers of each box in which a bar is randomly placed is 3. (D) How well the extraction works for different angles in the contour: as the angles are increased, saliency goes down, corresponding to the detection effects in Field et al. (1993). Gray error region indicates standard deviation of the mean.

with the decoded image in B. In Fig. 2.7D we quantify the contour saliency by calculating the average contour saliency over 50 different contour and background configurations, for different contour angle changes. As this angle increases, the saliency drops quickly, mirroring psychophysical contour detection probability (Field et al., 1993).

2.3.5 Dependency on model parameters

The elastica based contextual modulation has two parameters, E_0 and a, here we show that the essential features of the model do not depend on them. First, we fix $a = 0.1$ and vary E_0 to be 0, 4, or 8. Due to the fact that both contour extractions and the tilt illusion rely on only relative changes, they are fully invariant to changes in E_0. However, the association field varies strongly with the E_0 parameter, as would be expected from a parameter which mainly varies excitation versus inhibition, Fig. 2.8, top.

Next, we fix $E_0 = 1$, and set $a = 0.02$, 0.1 and 0.5, Fig. 2.8, middle. As the gain parameter a changes the strength of the modulation, we see no change in the shape of the association field, but quantitative changes of the saliency of features and the decoding biases, displaying a trade-off between saliency strength and coding biases.

The increased bias with large a can be partially counteracted by narrowing the neural tuning width K_c (Eq2.4.), as contextual modulation results in a smaller shift of the population response when the tuning curves are sharp. Interestingly, the tuning width K_c also has an effect on the attraction effect in the tilt illusion with 6 flankers, Fig. 2.8, bottom row. As the tuning curve becomes sharper (i.e. larger K_c), the total illusion becomes weaker as expected, but the attractive reduces more. For $K_c = 1.5$ the attractive illusion completely disappears. This is because the attractive pull of the flankers is felt most strongly by neurons with a preferred orientation close to 90 degrees. When the tuning curves are too narrow, these neurons do not respond, and thus the population response is not shifted.

2.4 Discussion

We have developed a computational model of V1 that implements the Gestalt Law of good continuation on a neural level through contextual modulations that were determined by the elastica energy. More specifically, the modulation by each bar outside a neuron's receptive field is governed by the curvature energy of the smoothest curves connecting to it. This quite naturally lead to contour extraction, but more surprisingly

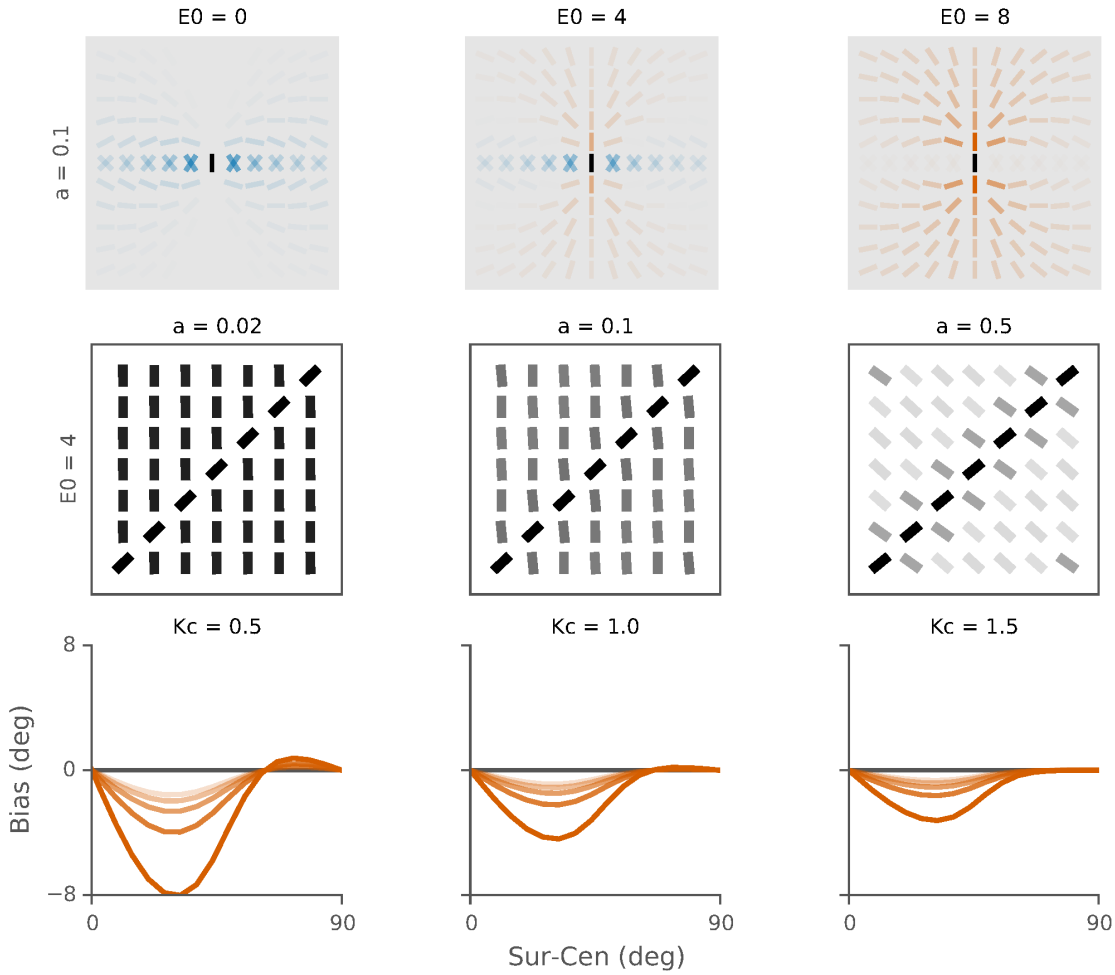

Figure 2.8: Exploring the effect of the a, E_0 and K_c parameters in model. (Top row) Modulation strength a is kept constant, while E_0 is varied. The E_0 parameter determines the relative presence of excitation and inhibition in the association field. The illusions and saliency results depend on the relative difference of modulation and are not affected (not shown). (Middle row) Modulation offset E_0 is kept constant, while a is varied. While the relative presence of excitation and inhibition is not affected by this parameter, the strength of modulation is directly changed, resulting in larger differences in response. This results in stronger bias and saliency effects. (Bottom row) Neural tuning width K_c is varied while both E_0 and a are kept constant. While this generally affects the magnitude of the tilt illusions, the attractive illusion in the Westheimer experiment (Fig. 2.5F) disappears with large K_C.

also explains saliency detection, association fields, and various forms of the tilt illusion.

Our work builds on a large body of literature linking these various aspects of visual processing. Association fields have been derived from an image statistics perspective (Geisler et al., 2001; Sigman et al., 2001), and contour detection has been linked to association fields (Zhaoping, 1998, 1999; Hansen and Neumann, 2008; Spratling, 2012). Elastica and association fields have been linked before and used for contour detection and completion (Williams and Jacobs, 1997; Sharon et al., 1997; Williams and Thornber, 2001; Ernst et al., 2012). Finally, the tilt illusion for two flankers was explained from elastica in a Bayesian framework, but using a different expression for the modulation from us (Schwartz et al., 2006a). Our model proposes a new explanation for several forms of the tilt illusions as following from individual elastica contour completions, including some counter-intuitive attractive effects. However, more importantly our model combines the phenomena described in earlier studies, both contour and illusion related, and links them together under the same basic elastica principle.

Not all aspects of the tilt illusion are captured by the model. Most prominently, the exact shape of the tilt illusion for collinear flankers as observed in Kapadia et al. (2000b) was not reproduced, Fig. 2.5F. However, we did find attraction for smaller angles, and repulsion otherwise, which has not been explained before. The attraction disappears as the flankers are placed further away, also in accordance with what was found in Kapadia's work. An important factor in producing these effects is calculating the elastica curves from the centers of the bars. Although the flanker ends were level with the center bar, this allowed for both the attractive and repulsive effects dependent on orientation and distance. This might suggest that neurons with a receptive field of smaller scale than the flanker bar drive the attractive effect in humans. However, such an explanation would require assumptions about how tilt estimates at different spatial scales are combined.

Interestingly, the illusion caused by a hexagon of 6 flankers was captured by the model, with both a repulsive and attractive effect. This effect was dependent on the tuning to the center bar. Although the attraction completely disappeared for very narrow tuning $K_c = 1.5$, most neurons have broad tuning corresponding to $K_c \approx 0.5$ (Cavanaugh et al., 2002a). Next, when in order to mimic gratings we increased the number of flankers and rotated the entire surround, the attractive effect disappeared for more than 8 flankers (for our parameters). However the repulsive illusion was always present. Existing explanations of the tilt illusion have taken either a mechanistic view,

where it is purely a result of the surround suppression (Clifford et al., 2000), or functional views such as arising from image statistics (Schwartz et al., 2009), Bayesian processing (Schwartz et al., 2006a), or as some form of image normalization (Clifford, 2014). We propose a new hypothesis: the smoothest continuations of the surround elements tilts the percept away from the surround and, in special cases, attracts it.

Despite the model's simplicity, we consider the model biologically feasible. The contextual modulations are effected as independent contributions from each flanker, as one would expect for modulation from individual surround neurons. The resulting modulation matches electrophysiology; both for individual flanker contributions in the form of the neural association field, with excitatory effects for collinearity and inhibitory effects for parallel bars (Kapadia et al., 2000b; Kinoshita et al., 2009), and the net effect for many surrounding bars which leads to suppressive surround modulation (Gilbert and Wiesel, 1990; Cavanaugh et al., 2002b; Seriès et al., 2003). Because the modulation relies on pairwise interactions only, it is plausible that some form of Hebbian learning shapes its tuning (Bednar, 2012). A caveat is that the statistics of natural images which include both textures and contours are dominated by parallel structures; an association field arises only when the statistics are restricted to contours (Geisler et al., 2001).

Currently we have assumed a single contrast level, which is clearly unrealistic for most natural images. In particular, association fields are known to change with contrast. Low contrast leads predominantly to excitation, while high contrast leads predominantly to inhibition (Kapadia et al., 2000b; Cavanaugh et al., 2002a). It is possible to extend the model to describe responses to stimuli with heterogeneous contrasts. The center contrast can be represented by the A_c parameter (e.g. Sclar et al., 1990). The contrast of the flanker can be coded in the a and E_0 parameters of each modulation term. In particular, changing E_0 as in Fig. 2.8 top row, qualitatively matches the observed contrast dependence of the association fields in Kapadia et al.

It is perhaps not surprising that elastica models the contextual interactions of V1 well, if the interactions do indeed exist for the purpose of detecting contours. Besides the elastica curves being especially pleasing to the eye, contours of natural objects are often well described by elastica curves. As an example in this paper, the shape in Fig. 2.2B would be a good candidate for a leaf. This is the very reason it is used in computer vision for contour completion of partially hidden objects (e.g. Mumford, 1994; Kimia et al., 2003; Zhou et al., 2012).

Our model makes several predictions: First, the inhibitory connections seem to be

broader tuned than excitatory connections. However, which inhibitory connections are the strongest is strongly dependent on relative position and orientation, Fig. 2.4. This could be tested experimentally. Secondly, our contextual modulations affect a neuron individually, and lead to contextual interactions both for small sets of bars and full surrounds. Experimentally, neither neurophysiologically nor psychophysically, it is known if these are linked. I.e., it is unknown whether the modulation by a surround built up with individual elements, can be explained from its individual contributions.

The neural character of the model allows for a number of straightforward extensions: 1) It will be interesting to include more realistic Gabor-type receptive fields at a variety of scales. 2) Currently the bars are assumed to have zero length. It is straightforward to find the elastica curves connecting the ends of the bars, with the only minor complication that β_f and β_c now become dependent on θ_f and θ_c. However, without a more realistic receptive field such an extension is rather ad hoc. 3) In the current implementation flankers modulate the center, but there is no recurrent feedback in which the modulated response change the activity of the flankers. In this sense the model performs a one-step approximation, which is valid as long as the shifts in the tuning curve are moderate. In a recurrent model, the dynamics of the illusions presented here would be of interest.

Chapter 3

Ambiguity and bias in population coding of single- and multivariate stimuli

This chapter is a paper in revision. A bioRxiv preprint is available as ***Biases in multivariate neural population codes*** *(Keemink and van Rossum, 2017). All the work herein was done by me.*

Across the brain information is often coded in population codes with broad tuning curves, so that multiple neurons respond to a single stimulus and a single neuron will respond to multiple stimuli. While research has focused on variability in decoding population activity, decoding biases remain understudied despite their possible role in explaining psychophysical data. Here we show that biases readily emerge when multiple stimuli are coded simultaneously in a population. The estimates follow non-trivial distributions and strong correlations between the estimates can arise. These effects can be understood from the particular shape of the likelihood function, which we model as a Gaussian process. The results carry concrete predictions for behavioral experiments, and provide a novel interpretation of the tilt illusion.

3.1 Introduction

In many cortical areas information coding is distributed across neurons using population codes. Such codes allow for accurate encoding of stimuli by pooling across neurons, despite the broad tuning of single neurons. Numerous studies have studied the role of the tuning curves (Zhang and Sejnowski, 1999), noise-correlations (Sompolin-

sky et al., 2002; Moreno-Bote et al., 2014), and heterogeneity on the coding accuracy of single variables (Shamir and Sompolinsky, 2006; Ecker et al., 2011). However, while it is of interest to know the theoretical limits on maximal decoding accuracy, the relation to psychometric data is not always clear.

In addition to decoding accuracy, decoders can also have a bias (a systematic difference between estimated and true value). Biases have received less attention in coding studies, as in many cases biases can be compensated for (Seriès et al., 2009) or are absent altogether. First, the decoding problem might have an intrinsic symmetry that abolishes bias (over- and underestimation of the stimulus are equally likely); this typically happens for uniformly coded one-dimensional stimuli. Secondly, in the limit of low noise, estimators such as the maximum likelihood decoder can be shown to be unbiased (Kay, 1993). Either condition by itself is sufficient to warrant unbiased estimation. For instance, while the maximum likelihood decoder is sub-optimal for high noise, it remains unbiased when decoding one dimensional stimuli (such as orientation) coded in a uniform population. In many practical situations however, neither can be guaranteed and biases can be expected.

As we shall see here, biases readily emerge when multiple variables are coded simultaneously, a situation that has only occasionally been studied (Treue et al., 2000; Orhan and Ma, 2015). These studies mainly looked at stimulus multiplicity (how many stimuli are there?) and decoding variance respectively. Our study provides insight into how biases in such situations arise, what properties they have, and what psychophysical predictions they yield.

3.2 Encoding and decoding model

We consider a set of $N = 64$ neurons with firing rates given by

$$r_i = f_i(\theta) + \eta_i,$$

where $i = 1 \ldots N$ is the neuron index, $f_i()$ is the tuning curve of neuron i, and η_i is IID Gaussian noise with mean zero and variance σ^2. Importantly, θ is a *vector* of stimulus parameters to be estimated. For concreteness we consider $f_i(\theta_1, \theta_2) = \sum_{k=1,2} g_i(\theta_k)$ where the tuning to a single stimulus is given by $g_i(\theta) = A \exp\left[-\frac{(\theta-\phi_i)^2}{2w^2}\right]$, where A is the response amplitude (set to 1), w is the width of the tuning curve (set to 1) and ϕ_i is the preferred stimulus of that neuron (with spacing $2\pi/N$ between neurons). An example of such a model is the coding of two overlapping random dot motion patterns

in area MT, where some studies report an averaging of the individual responses (van Wezel et al., 1996; Treue et al., 2000). In that case the two parameters to be estimated, i.e. the two motion directions, are circular variables. As the effects for circular variables are slightly more complex (see below), we first use unconstrained, real valued parameters θ_1 and θ_2. Note that the joint tuning curve $f()$ is given by the sum (or, equivalently, average) of the tuning curves to the individual stimuli. Some data suggest a stronger max-like competition when the stimuli are simultaneously presented (Gawne and Martin, 2002; Zoccolan et al., 2005); we will not explore this here in detail, apart from noting that such encoding models can reduce, but do not abolish the biases.

ML decoding

The setup is shown in Fig. 3.1A. We draw stochastic responses from the above model and then decode from the noisy population response using the maximum likelihood (ML) decoder (see Discussion for other decoders). The ML decoder estimates $\hat{\theta}$ by maximizing the log-likelihood,

$$\hat{\theta} = \text{argmax}_\theta \log p(\mathbf{r}|\breve{\theta}),$$

where $\mathbf{r} = [r_1, r_2, ..., r_N]$, for some true stimulus $\breve{\theta}$. For independent Gaussian noise this is equivalent to minimizing the mean squared error (MSE) E:

$$\begin{aligned} \hat{\theta} &= \text{argmin}_\theta E(\theta) \\ &= \text{argmin}_\theta \sum_{i=1}^{N} [r_i - f_i(\theta)]^2. \end{aligned}$$

In practice this minimization can be carried out using steepest descent. The stimuli to be estimated $\theta_{1,2}$ are real valued, but because the encoder loses the identity of the stimuli, we constrain the decoder to always return $\theta_2 \geq \theta_1$.

The resulting distribution of estimates $P(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2)$ is shown in Fig. 3.1B for a number of stimulus and noise settings. When the stimuli are far apart, $\theta_\Delta = \theta_2 - \theta_1 \gg w$, and in particular at low noise, the joint probability $P(\hat{\theta})$ resembles a 2D Gaussian, bottom-right panel. Here the ML decoder is an unbiased estimator, as the true value (cross) and mean estimate (star) coincide. The minimal variance follows the Cramér-Rao bound (CRB) of the individual stimuli. However, when the stimulus distance is small an interesting pattern emerges, left panels. Here the estimates fall essentially in two categories. First, either the estimates are strongly positively correlated, and cluster

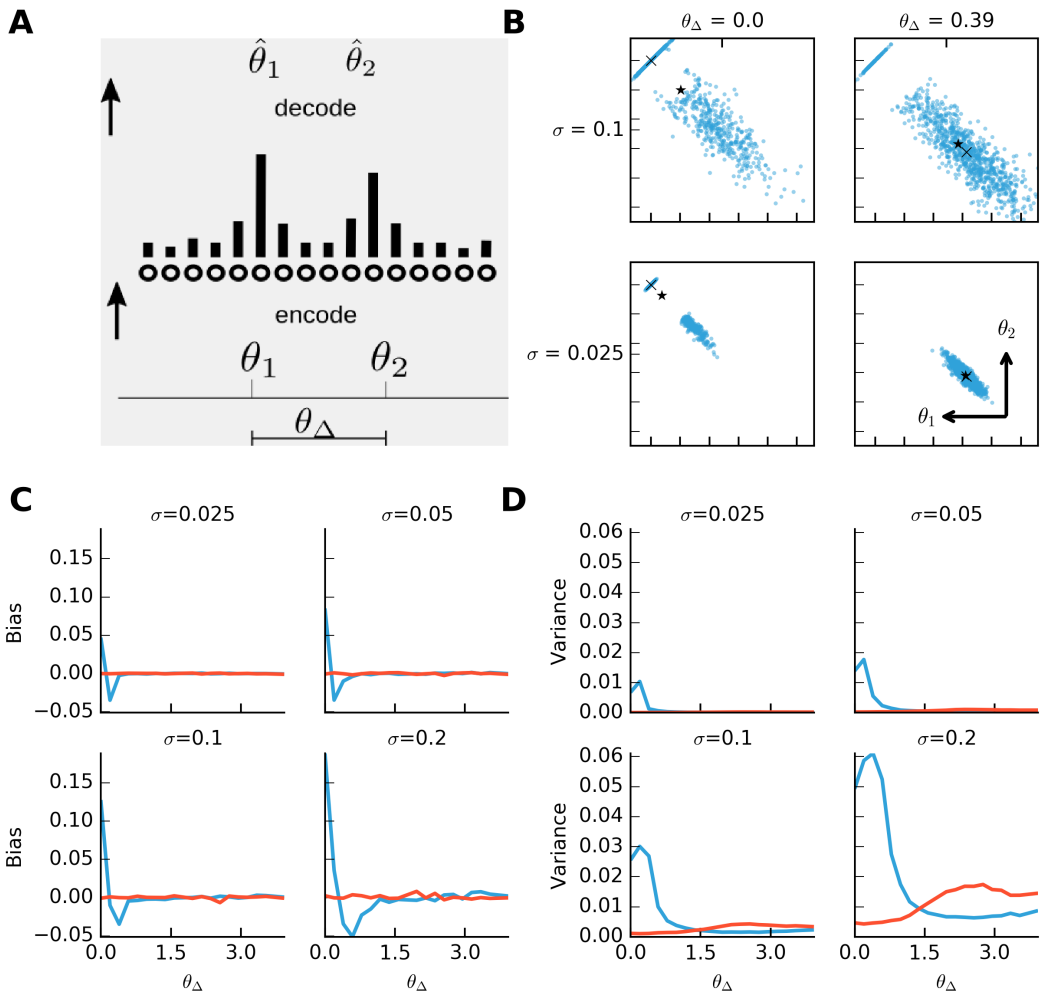

Figure 3.1: A) Basic encoding-decoding setup. A population of neurons codes for two simultaneous stimuli, parameterized by θ_1 and θ_2. The population response is the sum of the tuning curves of the individual responses. The task is to decode the stimulus parameters, their estimates denoted $\hat{\theta}_{1,2}$, from the noisy population response. B) The distribution of ML estimates $P(\hat{\theta}_1, \hat{\theta}_2)$ for different stimuli ($\theta_\Delta = 0, 0.39$) and noise levels ($\sigma = 0.025, 0.1$). The blue dots show the decoded values across 1000 realizations. The black crosses show the true values of the parameters; the stars indicate the mean of the estimates. A significant number of estimates cluster along the unity line (where $\hat{\theta}_1 = \hat{\theta}_2$). C) The resulting decoding biases as a function of the distance between the stimuli for various noise levels. While the sum of the stimuli is always decoded without bias (red), the estimated stimulus distance (blue) shows both positive and negative biases. D) The decoding variance as a function of the distance between the stimuli for various noise levels.

on the diagonal where $\hat{\theta}_1 = \hat{\theta}_2$. In this case the population response is interpreted as most likely caused by two fully overlapping stimuli. Or secondly, the estimates are negatively correlated and repelled from the true solution. When $\theta_2 = \theta_1$, exactly the same fraction of estimates fall in each category (as is shown below).

To more easily understand these results we transform the parameters to $\theta_\Sigma = \theta_1 + \theta_2$ and $\theta_\Delta = \theta_2 - \theta_1$. The Fisher Information matrix for additive, uncorrelated Gaussian noise is given by $I_{kl} = \frac{1}{\sigma^2}\sum_{i=1}^{N} \partial_{\theta_k} f_i(\theta)\partial_{\theta_l} f_i(\theta)$. In the limit of dense tuning curves it equals in the new coordinates

$$I = \frac{A^2\rho\sqrt{\pi}}{8w^3}\begin{pmatrix} 2w^2 + (\theta_\Delta^2 - 2w^2)\mathrm{e}^{-\theta_\Delta^2/4w^2} & 0 \\ 0 & 2w^2 - (\theta_\Delta^2 - 2w^2)\mathrm{e}^{-\theta_\Delta^2/4w^2} \end{pmatrix}, \quad (3.1)$$

where ρ is the coding density. I is diagonal, which confirms the intuition from Fig. 3.1B that for optimal decoders there is no correlation between the θ_Σ and θ_Δ estimates. We calculate the bias and variance of the estimator for these new variables, where the bias is defined as the expected deviation of the estimator from the true value

$$\mathbf{b}(\breve{\theta}) = \langle\hat{\theta}\rangle - \breve{\theta},$$

where $\langle x\rangle$ is the expected value of x. The estimate for θ_Σ is bias-free, as dictated by the translational invariance of the problem, Fig. 3.1C (red curves). The variance in the estimate is typically low, and weakly dependent on θ_Δ, Fig. 3.1D. For low enough noise we confirmed numerically that $var(\theta_\Sigma) = 1/I_{11}$. Thus the ML decoder for θ_Σ is efficient and unbiased.

In contrast, the estimator of θ_Δ is biased. When the stimuli overlap ($\theta_\Delta = 0$) the bias is repulsive (that is, $\hat{\theta}_\Delta > \breve{\theta}_\Delta$). But as the stimulus distance increases, the bias becomes attractive, before reducing to zero when the stimuli are distant, Fig. 3.1C (blue curves). One could argue the bias is simply caused by imposing $\theta_\Delta \geq 0$. However, the result is non-trivial as the distribution of estimates is bi-modal, with a gap between $\theta_\Delta = 0$ and a second peak. Furthermore, the bias changes sign and is attractive for larger θ_Δ, which is unexpected from such an interpretation. Whether the estimator θ_Δ is efficient, is discussed below.

The setup is translation invariant w.r.t. θ_Σ. In Fig. 3.1A it should not matter where along the axis θ_1 and θ_2 are encoded as long as there distance is the same. In other words, the coding θ_Δ should not depend of the value of θ_Σ (assuming a dense, infinite population). Therefore we set θ_Σ to 0 without loss of generality, such that

$$f_i(\theta_\Delta) = g_i(\frac{\theta_\Delta}{2}) + g_i(-\frac{\theta_\Delta}{2}). \quad (3.2)$$

We thus create a 1D system, and consider from here on only the coding of $\hat{\theta}_\Delta$.

Analytical description of the bias and variance

To better understand the decoding behavior, we must analyze the distribution of the estimates in detail. The distributions plotted in Fig. 3.1B can be estimated by performing steepest descent on each trial to find $\hat{\theta}$, and these estimates are then combined to yield $p(\hat{\theta}|\breve{\theta})$. However, there is no guarantee to find the global minimum on a given trial and the underlying distribution is only accurately reproduced for a large number of noise realizations.

Here we present an analytical numerically exact method to calculate the behavior of an ML decoder for a group of neurons with arbitrary tuning curves and Gaussian additive noise. The MSE $E(\theta)$ equals

$$\begin{aligned} E(\theta) &= \sum_{i=1}^{N} [r_i - f_i(\theta)]^2 \\ &= E_{\text{mean}}(\theta) + E_{\text{noise}}(\theta) + c_0, \end{aligned}$$

where the MSE is split into a mean $E_{\text{mean}}(\theta) = \sum_{i=1}^{N} [f_i(\breve{\theta}) - f_i(\theta)]^2$ and a noise term $E_{\text{noise}}(\theta) = -2\sum_{i=1}^{N} \eta_i f_i(\theta)$, and c_0 is a stimulus independent term. Note that $E_{\texttt{noise}}$ fully describes the stimulus dependent changes across realizations, as $E_{\texttt{mean}}$ remains constant. In other words, $E(\theta)$ is a smooth Gaussian process,

$$p(E|\breve{\theta}) = \mathcal{N}(\mu, \Sigma),$$

where $\mu = E_{\text{mean}}(\theta)$, and Σ is the covariance matrix with $\Sigma_{ab} = 4\sigma^2 \sum_{i=1}^{N} f_i(\theta_a) f_i(\theta_b)$.

The interplay between the mean and noise errors is integral to understanding the emergence of bias shown in Fig. 3.1C. $E_{\text{mean}}$and $E_{\texttt{noise}}$ are compared in Fig. 3.2 for the case that $\breve{\theta}_\Delta = 0$. While the mean error is lowest at $\theta_\Delta = 0$, as expected, it is flat near the true solution, Fig. 3.2A, black curve. Because of symmetry in the combined tuning curves, Eq. 3.2, not only the first, but also the second derivative of $E_{\texttt{mean}}$ is zero at θ_Δ (as well as all other odd derivatives). Thus $E_{\texttt{mean}}(\theta_\Delta) \sim O(\theta_\Delta^4)$, and for small θ_Δ the noise term dominates.

It is straightforward to see that the noise term is symmetric and smooth in θ_Δ. Therefore it is in leading order either an upward or downward curved parabola at the origin. If the noise term has a minimum at $\theta_\Delta = 0$, the total MSE also likely has a global minimum there, Fig. 3.2A+B, yellow curves. On the other hand, if the noise term has a maximum at $\theta_\Delta = 0$, due to the correlations, any minimum in the noise term will be further away. This results in the global minimum being repulsed away from the true solution, as in the green and red curves in Fig. 3.2A+B. The smoothness of the

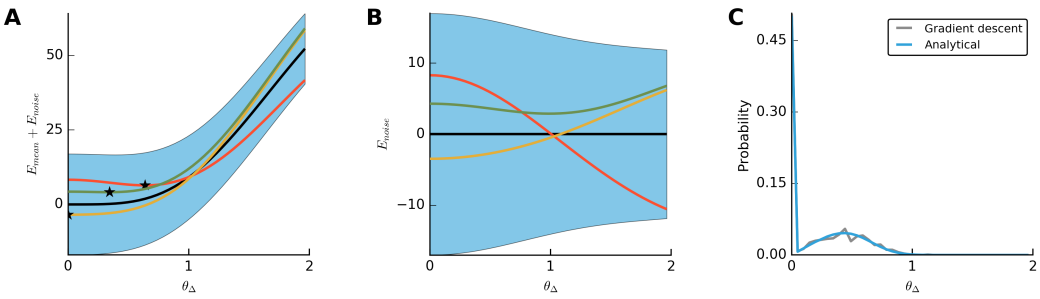

Figure 3.2: MSE behavior when the encoded stimulus distance $\breve{\theta}_\Delta = 0$, and $\sigma = 1$. A) The total stimulus dependent MSE ($E_{\text{mean}} + E_{\text{noise}}$), as a function of the stimulus distance. The black curve illustrates the mean error, E_{mean}. The blue region indicates the standard deviation. The colored curves are MSE's for different noise realizations. The stars indicate the locations of the minima. Note how the mean error is particularly flat near the true solution. B) The E_{noise} term of the MSE. The colors correspond to those in A. C) The decoding distribution by applying gradient descent over many noise realizations (blue) and from our theory (gray).

Gaussian process is given by the tuning width of the neurons; the wider the tuning, the longer the correlation length in $E(\theta)$ will be.

As a result the decoding distribution shows a peak at 0, and a smooth peak further away, Fig. 3.2C gray curve. This analysis also implies that for $\breve{\theta}_\Delta = 0$ exactly half of the estimates will be at $\theta_\Delta = 0$ (i.e. fall on the diagonal in 2D, as in Fig. 3.1B) and the other half are not, as can be seen in Fig. 3.2C at $\theta_\Delta = 0$. As the stimulus distance increases, the probability to find $\theta_\Delta = 0$ will decrease and the second distribution will gain more mass. As a result this will first decrease the repulsion, then turn into an attractive effect, and finally disappear.

Analyzing the MSE directly does not only yield qualitative insight in the biases, it can also be used to calculate the distribution of decoding estimates directly. Given a noisy response $\mathbf{r}$, we want to find the probability that a given stimulus value θ minimizes the error, i.e. that $E(\theta)$ is smaller than all other $E(\theta')$

$$\begin{aligned} p(E(\theta) \text{ is smallest}) &= p\left(E(\theta) < E(\theta')\right) \forall \theta' \neq \theta \\ &= p\left(E(\theta) - E(\theta') < 0\right) \forall \theta' \neq \theta. \end{aligned}$$

Because $E(\theta)$ is a smooth Gaussian process, and nearby E's are correlated, we can find this probability by finely discretizing θ. We define a set of M candidate stimuli $\Theta = [\theta_0, ..., \theta_M]$ for the multivariate normal distribution. For a given stimulus θ_j

from this set, the $M-1$ other stimuli the error is to be compared with is described by $\Phi_j = \{\theta_0, ..., \theta_M\} \backslash \theta_j$, and the set of error differences by $\triangle \mathbf{E}_j = E(\theta_j) - E(\Phi_j)$. The distribution of $\triangle \mathbf{E}_j$ is a multivariate normal

$$p(\triangle \mathbf{E}_j | \breve{\theta}) = \mathcal{N}(\triangle \mu^j, \triangle \Sigma^j),$$

where $\triangle \mu^j = E_{\text{mean}}(\theta_j) - E_{\text{mean}}(\Phi_j)$ and $\triangle \Sigma^j$ is an $M-1$ by $M-1$ covariance matrix with entries $\triangle \Sigma^j_{ab} = 4\sigma^2 \sum_{i=1}^{N} [f_i(\theta_j) - f_i(\theta_a)][f_i(\theta_j) - f_i(\theta_b)]$. For each candidate stimulus we thus have a multivariate distribution $\mathcal{N}(\triangle \mu^j, \triangle \Sigma^j)$, describing the difference in MSE. The probability that θ_j has the lowest MSE across all sampled stimuli is

$$\begin{aligned} p(E(\theta_j) \text{ is smallest}) &= p(\triangle \mathbf{E}_j < 0 | \breve{\theta}) \\ &= \int_{-\infty}^{0} ... \int_{-\infty}^{0} p(\triangle \mathbf{E}_j | \breve{\theta}) d\triangle E_{j,0} ... d\triangle E_{j,M-1}, \end{aligned} \tag{3.3}$$

which is a multi-variate cumulative normal distribution. While this is not itself analytically tractable, several efficient algorithms exist that calculate it to a high precision for values of M up to in the hundreds (We used the mvnun function from Scipy, based on Genz 1992, 1998).

We applied our theory using $M = 100$, and $\Theta = [\theta_{\Delta 0} = 0, ..., \theta_{\Delta M} = \pi]$ (using a larger M had negligible effects). We first apply Eq.3.3 to the example in Fig. 3.2C, and accurately reproduce the decoding distribution, blue curve. Next, we use this approach to calculate the decoding biases and variances directly. Given $p(\triangle \mathbf{E}_j < 0 | \breve{\theta})$ the average decoded stimulus follows as

$$\left\langle \hat{\theta}_\Delta \right\rangle = \sum_{j=1}^{M} \theta_{\Delta j} \, p(\triangle \mathbf{E}_j < 0 | \breve{\theta}_\Delta),$$

while the bias is

$$b(\breve{\theta}_\Delta) = \left\langle \hat{\theta}_\Delta \right\rangle (\breve{\theta}) - \breve{\theta}_\Delta$$

and the variance

$$\sum_{j=1}^{M} \left(\theta_{\Delta j} - \left\langle \hat{\theta}_\Delta \right\rangle\right)^2 p(\triangle \mathbf{E}_j < 0 | \breve{\theta}^{\vee}_\Delta).$$

As the encoded stimulus changes, the decoding distribution shifts, Fig. 3.3A. The balance between probability mass at $\theta_\Delta = 0$ and the second distribution further away results in first a repulsive and then an attractive bias, until finally settling at 0 as before. Both this decoding bias and the decoding variance are accurately calculated by the above theory, Fig. 3.3B+C.

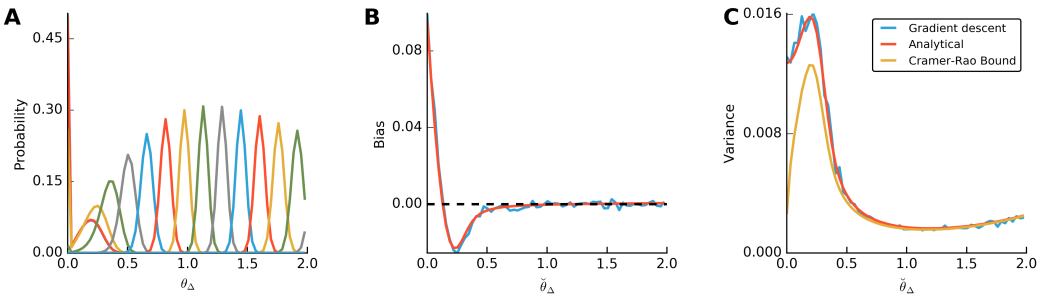

Figure 3.3: A) The distribution of ML estimates for different encoded stimuli, and different noise levels. As the encoded stimulus distance increases, the distribution moves with it. At first however, all that happens is that the probability mass at $\theta_\Delta = 0$ moves into the secondary distribution. B) Decoding biases as a function of the encoded stimulus. C) Decoding variances as a function of the encoded stimulus. In this figure the noise $\sigma = 0.2$.

We next calculate the effective CRB for this encoder. For a biased estimator the CRB is

$$CRB(\breve{\theta}_\Delta) = \frac{[1 + b'(\breve{\theta}_\Delta)]^2}{I_{22}},$$

where b' is the derivative of the bias with respect to $\breve{\theta}_\Delta$, and I_{22} is given by Eq.3.1. Away from $\theta_\Delta = 0$, it can be easily calculated. However near $\theta_\Delta = 0$ the situation is more subtle. Here I_{22} goes to zero, but because the estimator is a smooth, symmetric function in θ_Δ, one also has that $b'(0) = -1$. Thus the leading terms cancel. Furthermore it is hard to calculate $CRB(\breve{\theta}_\Delta)$ for small $\breve{\theta}_\Delta$ accurately as the bias estimate is noisy, even for a large number of realizations. With the analytical description of the bias however, the derivative is smooth, and we can estimate b' accurately. The resulting CRB is plotted in Fig. 3.3C, yellow curve. While the decoder reaches the CRB for large stimulus differences, when $\breve{\theta}_\Delta$ approaches zero the decoder variance diverges from the CRB, i.e. the estimator is no longer efficient. This is likely due to the bimodal probability distribution in this range.

3.3 Center-surround modulation in V1

So far we have considered neurons responding to combinations of the same variable (such as overlapping gratings). Another case where multiple variables are encoded occurs when neurons are modulated by other variables. One of the most prominent ex-

amples is contextual modulation, which occurs for example in the primary visual cortex when a center grating covering a neuron's receptive field is surrounded by another grating. The surrounding grating induces mostly repulsive biases in the perception of the center grating (known as the 'tilt illusion'), and reduces the response of neurons to the center grating (Clifford et al., 2000; Schwartz et al., 2007), most strongly when the surround is aligned with the preferred orientation (Seriès et al., 2003). We model this with a tuning curve of the following form

$$f_i(\theta_c, \theta_s) = g_c(\phi_i - \theta_c) g_s(\phi_i - \theta_s),$$

where θ_c and θ_s are the center and surround orientations, $g_c(x) = A_c \exp(K_c \cos x)$ is the tuning to the center orientation, and $g_s(x) = 1 - A_s \exp(K_s \cos x)$ describes the modulation from the surround. We set $A_c = 1$, $K_c = 1$, $A_s = 0.5$, and $K_s = 0.5$. When decoding the center orientation as the population vector of the population response, this type of model explains the repulsive part of the tilt illusion quite nicely (Clifford et al., 2000). However, at in the limit of zero noise or infinite neurons a maximum likelihood decoder applied to this model can recover both θ_c and θ_s without bias, suggesting that this encoding model is not inherently biased.

To examine this we plot as above the decoding distribution for different stimuli and noise levels, Fig. 3.4A. The center and surround biases are shown in respectively in B and C. Surprisingly, as the noise increases, clear biases appear. The bias in the center orientation (blue curves in B) is especially reminiscent of the repulsive part of the tilt illusion, where for most surround angles the center orientation is perceived repulsed away from the surround (e.g. Clifford et al., 2000 Fig.2b). However, rather than originating from a shifted population response, the bias results from a binary decoding distributions, as is clear from A for higher noise levels. In this case it is due to $E_{\texttt{mean}}$ having two local minima, one for the true solution, and one for a false solution, both of which give approximately the same population response. While the false solution normally has a lower likelihood (higher MSE) than the true solution, $E_{\texttt{noise}}$ can change this balance, resulting in a binary decoding distribution. Thus, this predicts that the tilt illusion not merely the consequence of a shifted population response, but from an inherent near-ambiguity in the encoding model.

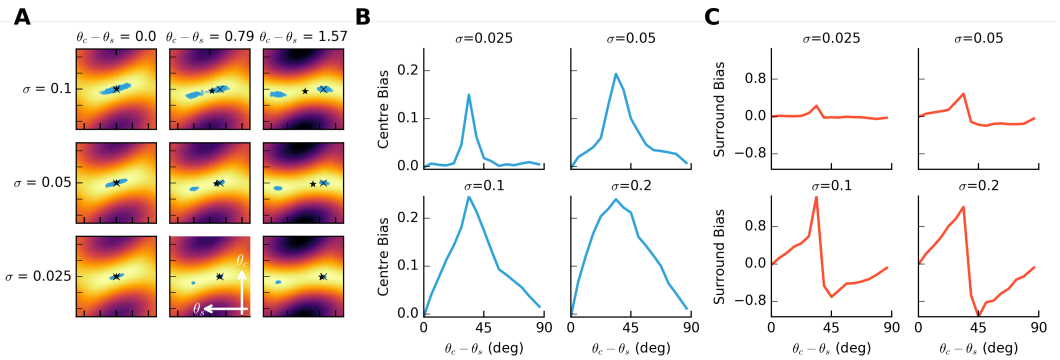

Figure 3.4: Decoding behavior for center surround modulation, where the response to a grating in the receptive field of a neuron (center), is modulated by a surrounding grating. A) Decoding distributions for different stimuli and noise levels. In the background the mean error in the absence of noise is shown ($E_{\mathtt{mean}}$). The blue dots show the decoded values over many noise realizations (1000). The crosses show the true stimuli, the stars the average decoded stimuli. θ_c is kept at zero, as θ_s is changed. While the mean error at the true $(\breve{\theta}_c, \breve{\theta}_s)$ always has mean error zero, and a part of the decoded stimuli are indeed centered around the cross, there is a also a local minimum in the error for a different (θ_c, θ_s) set, resulting in two distributions in decoded stimuli. B) Here we show how the distribution changes in A translate to a repulsive decoding bias in the center orientation. The bias for higher noise is similar to the repulsive part of the tilt illusion, Clifford et al. (2000) Fig.2b. C) Same as in B but for the surround orientation.

3.3.1 Attractive tilt illusion

The full tilt illusion as experienced psychophysically also has an attractive effect (Clifford, 2014). This can be added to our current model by adding an excitatory term which depends on the surround such that

$$f_i(\theta_c, \theta_s) = g_c(\phi_i - \theta_c) g_s(\phi_i - \theta_s) g_s^e(\phi_i - \theta_s),$$

where $g_s^e(x) = 1 + A_s^e \exp(K_s^e \cos x)$, with $K_s^e = 2$ and $K_s^e = 1$. The attractive illusion would normally arise due to the population vector being repulsed from the surround for smaller orientation differences, and then being attracted for larger differences. Surprisingly, the attractive bias persists under maximum likelihood estimation for high enough noise levels, Fig. 3.5. As for the previous model with only a repulsive illusion, this is not due to a simple population response movement, but by non-uniformity of the decoding distribution. These are not merely bimodal, but sometimes seem to have

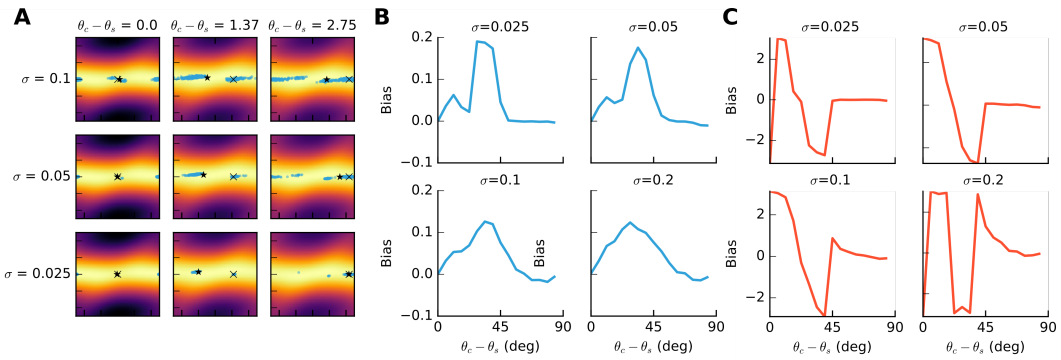

Figure 3.5: Same as Fig. 3.4, but with the attractive illusion. A) Decoding distributions for different stimuli and noise levels. In the background the mean error in the absence of noise is shown ($E_{\mathtt{mean}}$). The blue dots show the decoded values over a 1000 noise realizations. The crosses show the true stimuli, the stars the average decoded stimuli. θ_c is kept at zero, as θ_s is changed. While the mean error at the true $(\breve{\theta}_c, \breve{\theta}_s)$ always has mean error zero, and a part of the decoded stimuli are indeed centered around the cross, there is also a local minimum in the mean error for a different (θ_c, θ_s) sets, resulting in different distributions in decoded stimuli. B) Here we show how the distribution changes in A translate to a repulsive decoding center bias for smaller orientation differences, and attractive for larger differences. C) Same as in B but for the surround orientation. The sudden changes in the last subpanel are due to circularity.

up to 4 local minima (Fig.3.5.A, top left panel), the net effect being both repulsive and attractive biases at different surround orientations.

3.4 Discussion

We have shown that when multiple stimuli are coded simultaneously in population codes, biases in decoding arise. Furthermore, we have developed an analytical theory of ML decoders which allows to understand and calculate these biases. That biases occur is by itself not surprising as the absence of biases can only be proven in some conditions. For instance the ML decoder is unbiased only in the limit of many observations (or low noise), but in general the ML decoder is not unbiased (Kay, 1993; Pilarski and Pokora, 2015), nor efficient (Xie, 2002). Yet, the rich structure of the biases even in the simple models we considered is surprising.

A natural question is how general these results are. Although not shown here, similar results hold when we use a Bayesian inference decoder $\hat{\theta} = \int P(\theta|\breve{\theta})\theta\, d\theta$ (Dayan

and Abbott, 2001). While the decoding distributions are not binary in this case, the biases and variances behave the same. Furthermore, the question arises what encoding models would lead to biases. It is easy to imagine biases arising in multivariate systems such as center-surround modulation. However, but even when we kept θ_Σ fixed, effectively creating an inhomogeneous 1D encoding model, biases occurred.

While we only considered additive Gaussian noise, simulations show that our results extend to Poisson and Multiplicative Gaussian noise, as well as correlated noise. The analytical theory can easily be extended for Gaussian multiplicative noise. Early explorations of these systems indicate very similar bias and variance behaviors.

We applied our analytical theory only to a 1D ML decoder. In principle, the method can be applied to any dimension of decoder. As one needs to discretize the higher dimensional stimulus space, the only limitation is the efficient calculation of the integrals in Eq.3.3. Algorithms that calculate Cumulative Multivariate Normals for even higher dimensions do exist, e.g. Azzimonti and Ginsbourger (2016).

For the overlapping stimuli, we only considered two indistinguishable stimuli, such as dots moving in two different directions. As such, the mixing of the two tuning curves was equal: $f_i(\theta_1,\theta_2) = \sum_{k=1,2} g_i(\theta_k)$. However, if the stimuli are distinguishable (i.e. if the dots have different contrasts or colors) we could change this to $f_i(\theta_1,\theta_2) = \sum_{k=1,2} c_i g_i(\theta_k)$, where c_i is some constant which is different for each stimulus. This removes, first of all, the need to enforce $\theta_2 \geq \theta_1$, as an ML decoder can now distinguish between the two stimuli. Secondly, we predict that the bigger the difference in c_i is, the smaller the resulting bias will be, as this makes it easier to distinguish what each stimulus is contributing to the population response.

Although our bias has not been directly reported, a similar bias can be seen for stimulus distance (Oleksiak et al., 2010). The results also bear upon psycho-physical experiments where two overlapping random dot motion patterns with different directions are presented and subjects are asked to guess the angle between the two directions. In such experiments repulsive biases have commonly been observed (Marshak and Sekuler, 1979), although attractive effects have also been observed (Braddick et al., 2002). Several effects have been hypothesized to underlie these biases, including adaptation (Rauber and Treue, 1999), cortical interactions (Carandini and Ringach, 1997) and repulsion from the cardinal directions (Rauber and Treue, 1998). The bias described here, is not at odds with those explanations, but presents a novel contribution to the total bias. It should be most prominent at small angles and for short presentation times, when the signal-to-noise ratio is presumably small.

The estimated decoding distribution can be seen to reflect an ambiguity between the presence of one or two stimuli. After all, when the stimuli are equal ($\theta_1 = \theta_2$), this is the same as just encoding one stimulus. Apart from predicting a bias, the theory predicts a bi-modal distribution of direction difference estimates and for small angles about half the time the two motions should be perceived as one. In experiments the number of stimuli that can simultaneously be perceived using overlapping motions is limited (e.g. Edwards and Greenwood, 2005) and when three or five overlapping motions are presented, they can sometimes be perceived as two (so called metamers, Treue et al., 2000); an effect which previously has been explained using the probabilistic population code framework (Zemel et al., 1998; Zemel and Dayan, 1999). The results here suggest that differences in the numerosity between presented and perceived stimuli already emerge with maximum likelihood decoders. Quantitative verification of this prediction of our study should be possible but might be challenging as attention, participants' expectations, and natural priors for perceiving a single motion direction instead of two directions can influence results.

Regarding priors, this chapter has assumed that there is a flat prior. However, it is imaginable that there is a prior which is taken into account in the decoding process (for example, that overlapping stimuli will tend to have the same direction). In our model this would bias the results further towards that prior. In terms of a prior in the center-surround decoding process, this would bias the overall decoding likelihood landscape towards the prior as well. However, depending on the strength of the prior, this would not remove the ambiguity nor the bimodality of the decoding distributions fully.

In all the encoding models used in this chapter, the stimuli were also various stimuli of the same dimension: the neurons has only a single preferred value for this dimension. A second type of multivariate stimuli would be two different dimensions; for example contrast and orientation or color and direction. If that would also lead to some kind of bias depends on the exact nature of the neural response, and the spread of the neurons over the two dimensions. If the neurons have independent tuning curves to both stimuli, and cover the full range of possible stimuli, we would predict that decoding is perfectly possible, as there are no possible ambiguities.

Our findings lead to predictions that can be tested psychophysically. An experiment would be to present two overlapping random dot motion patterns with slightly different directions and measure in particular the distribution of estimates of the angle between the two directions. The findings here would predict a bi-modal distribution of direction difference estimates. In our model the bias in the estimates depends strongly on the

noise level and they disappear at zero noise. Psycho-physically the noise level might be altered by modifying the presentation time; for long presentations, there should be no bias. Already the center bias in the center-surround model has a striking resemblance to well-known tilt illusion (Clifford et al., 2000). Here we would predict that the perception of the surround grating has a significant bias, with a binary representation, and that the decoding distribution of the center orientation is not a simple distribution around the mean. By better understanding biases in population codes and linking those to psychophysics we hope that deeper insights in neural coding can be gained.

Chapter 4

Effects of the tuning of V1 surround modulation the tilt illusion and on visual saliency

This chapter is a paper in revision. All the work herein was done by me. Discussions with Dr. S. Shushruth, Dr. J. Bednar and Dr. M. Hennig are gratefully acknowledged.

It is well established that the responses of neurons in the primary visual cortex to stimuli in the receptive field (the center) are modulated by stimuli outside the receptive field (the surround). Classically, modulation is assumed strongest if the surround orientation aligns with the neuron's preferred orientation; fixed irrespective of the actual center stimulus presented ('fixed modulation'). Under this assumption surround modulation has been shown to explain a wide range of psychophysical phenomena, such as salience and tilt perception. However, several studies have measured surround modulation with non-preferred center stimulation and found that modulation is typically strongest if the surround orientation is aligned with the center stimulus ('center dependent modulation'). It is not currently known what, if any, functional differences result from these two modulation variants. Here we use computational models to explore these differences across a range of phenomena. First, we investigate orientation decoding and find that, while fixed modulation leads to biases in tilt perception (which have been used to explain the tilt illusion), center dependent modulation induces no such bias. Next, we find that center dependent modulation enhances contrast, saliency and pop-out effects. Our work shows that center dependent modulation reduces bias, while simultaneously increasing the response to salient features, and suggests that a mixture of the two modulation types is necessary to quantitatively match observed

biases.

4.1 Introduction

Neurons in the primary visual cortex (V1) of mammals respond to stimuli inside their receptive field (the center). However, while stimuli surrounding the center by themselves do not cause a response, they can modulate the center response, as commonly investigated with center-surround grating pairs as in Fig.4.1A (Blakemore and Tobin, 1972; Maffei and Fiorentini, 1976; Nelson and Frost, 1978; Fries, Albus, and Creutzfeldt, 1977; Gilbert and Wiesel, 1990; Sillito and Jones, 1996; Girman, Sauvé, and Raymond, 1999; Jones, Grieve, Wang, and Sillito, 2001; Freeman, Ohzawa, and Walker, 2001; Seriès, Lorenceau, and Frégnac, 2003; Shushruth, Mangapathy, Ichida, Bressloff, Schwabe, and Angelucci, 2012). Surround modulation is thought to underlie many perceptual phenomena, such as contrast perception (Shushruth et al., 2013), saliency detection (Sillito et al., 1995; Zhaoping, 1999; Petrov and McKee, 2006), contour integration (Zhaoping, 1998) , as well as the tilt illusion in which the center orientation is misjudged in the presence of a surround grating. (Clifford et al., 2000; Schwartz et al., 2009; Qiu et al., 2013; Keemink and van Rossum, 2016). Crucially, all these studies assume that for a given neuron, modulation depends on the difference between its preferred orientation and the surround orientation (which we call 'fixed modulation', Fig.4.1B). However, a number of studies that varied center and surround orientation showed that for the majority of V1 neurons the modulation instead depends on the orientation of the center stimulus (which we call 'center dependent modulation', Fig.4.1C) (Sillito et al., 1995; Cavanaugh et al., 2002a,b; Shushruth et al., 2012). Furthermore, such center dependent modulation arises naturally in neural models based on image statistics (Coen-Cagli et al., 2012) and Bayesian probability models (Lochmann and Deneve, 2011; Lochmann et al., 2012). Yet despite its prevalence, the functional consequence of center dependent modulation is unknown, although Shushruth et al. (2012) predicted that it might play a role in enhancing saliency.

In this study we contrast fixed vs. center dependent modulation by developing a set of phenomenological V1 models in which the surround modulation tuning can be made fixed or center dependent, without affecting other properties of the model. Any functional difference can therefore be solely attributed to the difference in surround modulation tuning. We first consider a single population of neurons responding to a center grating, and modulated by a surround. Compared to fixed modulation, center

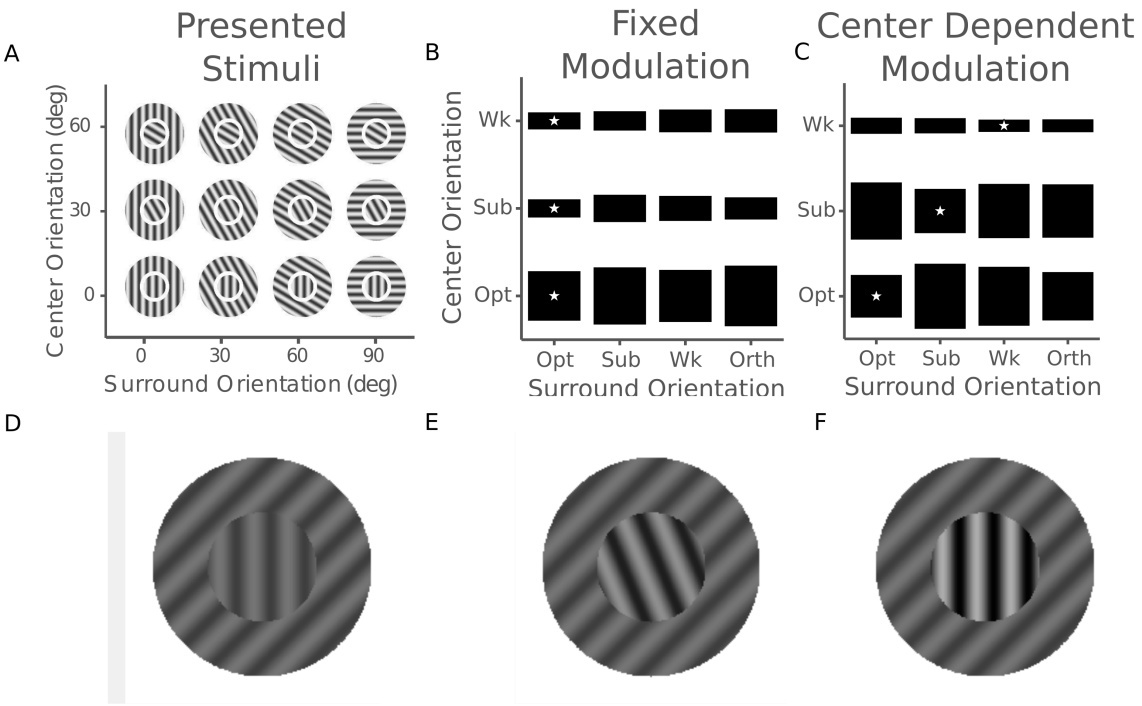

Figure 4.1: Two models of surround modulation in V1 and their consequences. (A) Sample center-surround stimuli consisting of two gratings: the center grating covers a neuron's classical receptive field, the larger grating covers the surround. We consider neural responses when both the center and surround orientations are varied. (B) Responses of an example cell with fixed modulation for varying center and surround orientation pairs. The height of a box is proportional to the response. 'Opt' corresponds to the neuron's preferred orientation, 'Orth' to the orientation orthogonal to 'Opt'. 'Sub' and 'Wk' correspond to two intermediate orientations. The stars indicate the strongest suppression for a given center orientation. For this neuron strongest suppression always occurred at the preferred center orientation, indicating fixed modulation. (C) Same as B but for a cell with center dependent modulation, which is more common. Here the stars shift with the center orientation, indicating that suppression was strongest when center and surround were aligned. Data used in panels B and C courtesy of Shushruth and Angelucci (Shushruth et al., 2012). (D-F) Cartoon illustrating the main results, by exaggerating the effects of the two modulation types on the perception of a center-surround grating pair. (E) The effect of fixed modulation on perception of the scene in D, with exaggerated effects for illustration. The center grating pops out due to its different orientation and its angle is misjudged. (F) Same as E but with center dependent modulation. There is now stronger popout, and no decoding bias.

dependent modulation (1) results in an unbiased representation of the center orientation (i.e. no tilt illusion), and (2) increases saliency effects, Fig.4.1 D-F. We next expand the model to describe the response at multiple locations to understand the effects for more complex scenes, and confirm that either type of modulation leads to saliency detection of salient elements such as boundaries, contours and single elements, with a stronger effect under center dependent modulation. Our work shows that large and crucial functional differences can emerge from seemingly small variations in contextual modulation.

4.2 Methods

4.2.1 Encoding model

4.2.1.1 Center surround model

To examine the functional differences between fixed and center dependent surround modulation, we compare two phenomenological models consisting of $N = 32$ neurons with preferred orientations equally spaced in the interval $[0, \pi]$. Presented with just a center grating, a neuron's firing rate is modeled by a von Mises function (von Mises, 1918)

$$g(\phi_i, \theta_c) = A_c \exp\left[k_c \cos 2(\phi_i - \theta_c)\right],$$

where k_c determines the neuron's tuning width, ϕ_i is the neuron's preferred orientation, and θ_c is the orientation of the center stimulus. The firing rate is maximal when the center equals the preferred orientation ($\phi_i = \theta_c$). The peak firing rate equals $A_c \exp(k_c)$ and is fixed to 20Hz, by setting $A_c = 20\text{Hz}.\exp(-k_c)$.

Presented with both a center and surround stimulus the response $f()$ is given by the center drive $g()$ as

$$f_i(\phi_i, \theta_c, \theta_s) = g(\phi_i, \theta_c) h(\theta_{ref}, \theta_s),$$

with $h()$ modeling multiplicative modulation (Cavanaugh et al., 2002b). The surround modulation is modeled using a second von Mises function

$$h(\theta_{ref}, \theta_s) = 1 - A_s \exp\left[k_s \cos 2(\theta_{ref} - \theta_s)\right], \tag{4.1}$$

where k_s determines the surround modulation tuning width, θ_s is the orientation of the surround stimulus, and θ_{ref} is the reference orientation h(see below). The maximal suppression is $1 - A_s \exp(k_s)$, to ensure that $0 \leq h \leq 1$ we restrict $0 \leq A_s \leq \exp(-k_s)$. The parameters were fitted to the average normalized tuning and modulation curves across 34 V1 cells in Fig.3D from Cavanaugh et al. (2002a) in the region $\theta = -\pi/2 \ldots \pi/2$. This fit yielded $A_s = 0.5$, $k_c = 0.6$, $k_s = 0.5$, meaning that the center and surround tuning widths were very similar. This match is not required for our findings, nor do the specific variable values impact our results qualitatively.

Crucially, θ_{ref} in the modulation function $h()$ can be set to either 1) the preferred orientation $\theta_{ref} = \phi_i$, reflecting fixed modulation, or 2) to the center orientation $\theta_{ref} = \theta_c$, reflecting center dependent modulation.

We also examined subtractive modulation, but found qualitatively similar results (not shown). In some cases noise was introduced by modeling the neural responses as Poisson processes with a rate given by the tuning functions $f()$, resulting in the distribution of possible noisy responses

$$p(r_i) = \frac{(f_i T)^{r_i T}}{(r_i T)!} \exp(-f_i T).$$

The observation time T was set to 0.5s unless indicated otherwise. Qualitatively, the results extend to Gaussian additive and multiplicative noise models.

4.2.1.2 Encoding sets of bars

The center surround model only models the response to gratings, here we extend the model to respond to sets of bars (such as in Fig. 4.8). First, to study the coding of a single bar, with orientation θ_c, amidst a uniform background, with orientation θ_s, we set the response of the neurons at the center bar to $f_i(\phi_i, \theta_c, \theta_s)$, and the background neurons to $f_i(\phi_i, \theta_s, \theta_s)$, by assuming the influence of the center bar is negligible. This allows for easy mathematical analysis, see Results.

To encode an arbitrary stimulus consisting of multiple oriented bars we extend the model to the one used in Chapter 2 as follows: 1) instead of a single surround modulation term, the surround modulation is composed of the modulation from individual bars, 2) the response is calculated for each bar separately.

Each bar has an associated population of 32 neurons tuned to its orientation. For n surrounding bars the response $f_{i,x}()$ for neuron i in location x is given by

$$f_{i,x} = g(\phi_i, \theta_x) \prod_{j=1}^{n} h(\theta_{ref}, \theta_f^j, \varphi_f^j, r_f^j),$$

where the function $h(\theta_{ref}, \theta_f, \varphi_f, r_f) = \exp\left[-\frac{a}{r_f}\left(E(\theta_{ref}, \theta_f, \varphi_f) - E_0\right)\right]$. $E()$ is the curvature energy of the optimal curve connecting the local neuron i with reference orientation θ_{ref} to the flanker bar j with orientation θ_j, a and E_0 are constants, φ_f, is the positional angle and r_f is the distance of the flanker. In the original implementation of this model (Keemink and van Rossum, 2016), $\theta_{ref} = \phi_i$, such that the modulation was effectively fixed. As for the center surround model, the surround modulation can be made center dependent by setting $\theta_{ref} = \theta_i$. For detailed interpretations of the variables and model see Keemink and van Rossum (2016) and Chapter 2.

4.2.2 Decoding models

4.2.2.1 Population vector decoding

The population vector is a 2D vector given by the sum of the preferred orientation vectors of the neurons weighted by their firing rate (Georgopoulos et al., 1986a; Schwartz et al., 2009), representing the estimated center orientation vector $\hat{\mathbf{v}}_c$ as

$$\hat{\mathbf{v}}_c = \sum_i r_i \mathbf{u}_i, \tag{4.2}$$

where r_i is the firing rate, and $\mathbf{u}_i = (\sin 2\phi_i, \cos 2\phi_i)$ is the unit vector pointing in neuron i's preferred orientation (multiplied by two to ensure circularity). The estimated center orientation $\hat{\theta}_c$ follows from the angle of the population vector with the x-axis

$$\hat{\theta}_c = \frac{1}{2} \angle \hat{\mathbf{v}}_c,$$

where $\angle$ denotes a vector's angle with the x-axis.

In the absence of surround stimulation, the estimated center orientation is bias-free. This can be shown explicitly by using that for dense coding with many identical neurons, Eq.4.2 can be written as $\sum_i g_i \mathbf{u}_i = \int_0^\pi g(\phi, \theta_c)[\sin 2\phi, \cos 2\phi]^T d\phi$, which yields that $\hat{\mathbf{v}}_c \propto \mathbf{v_c}$, where $\mathbf{v_c} = [\sin 2\theta_c, \cos 2\theta_c]^T$ describes the center orientation vector. Symmetry and circularity arguments yield that this holds for any orientation tuning curve that is symmetric around its preferred orientation, i.e. a function of $|\phi_i - \theta_c|$ only.

To see this, we write

$$\hat{\mathbf{v}}_\mathbf{c} = \int_0^\pi f(2[\phi - \theta_c])[\sin 2\phi, \cos 2\phi]^T d\phi,$$

where $f(x)$ is an arbitrary function circular in 2π and symmetric in $x = 0$. We make the substitution $x = \phi - \theta_c$ such that

$$\hat{\mathbf{v}}_\mathbf{c} = \int_0^\pi f(2x)[\sin 2(x+\theta_c), \cos 2(x+\theta_c])]dx,$$

where we can still write the integral from 0 to π, thanks to the circularity of $f(2x)$. Considering only $v_1 = \int_0^\pi f(2x)\sin 2[x+\theta_c]d\phi$,

$$\begin{aligned} v_1 &= \int_0^\pi f(2x)(\cos 2x \sin 2\theta_c + \cos 2\theta_c \sin 2x)dx \\ &= \sin 2\theta_c \int_0^\pi f(2x)\cos 2x dx + \cos 2\theta_c \int_0^\pi f(2x)\sin 2x dx \\ &\propto \sin 2\theta_c, \end{aligned}$$

using the fact that $\int_0^\pi f(2x)\cos 2x dx$ is a constant, and $\int_0^\pi f(2x)\sin 2x dx = 0$, due to the symmetry and circularity of $f(2x)$. Similarly, for $v_2 = \int_0^\pi f(2x)\cos 2[x+\theta_c]d\phi \propto \cos 2\theta_c$.Resulting finally in

$$\hat{\mathbf{v}}_\mathbf{c} \propto \mathbf{v}_\mathbf{c}.$$

Next we analytically derive the decoded vector for center dependent and fixed modulation in the presence of a surround grating. In the case of center dependent surround modulation, the estimated center orientation vector becomes

$$\begin{aligned} \hat{\mathbf{v}}_c^{dep} &= \int g(\phi,\theta_c)h(\theta_c,\theta_s)\begin{bmatrix}\sin 2\phi \\ \cos 2\phi\end{bmatrix}d\phi \\ &= h(\theta_c,\theta_s)\int g(\phi,\theta_c)\begin{bmatrix}\sin 2\phi \\ \cos 2\phi\end{bmatrix}d\phi \\ &= \alpha \mathbf{v}_c, \end{aligned}$$

where the constant $\alpha = 2\pi A_c h(\theta_c,\theta_s)I_1(k_c)$, and I_1 is the Bessel function of the first kind.

For fixed surround modulation the estimated center orientation vector is given by

$$\hat{\mathbf{v}}_c^{fix} = \int g(\phi,\theta_c)h(\phi,\theta_s)\begin{bmatrix}\sin 2\phi\\ \cos 2\phi\end{bmatrix} d\phi$$
$$= \alpha(\mathbf{v}_c - \beta\mathbf{v}_{shift}), \tag{4.3}$$

where $\beta = A_s \exp(-k_s) I_1(|k_c\mathbf{v}_c + k_s\mathbf{v}_s|)/I_1(k_c)$, α is as above, and the shift vector $\mathbf{v}_{shift} = k_c\mathbf{v}_c + k_s\mathbf{v}_s$, where $\mathbf{v_s} = \begin{bmatrix}\sin 2\theta_s\\ \cos 2\theta_s\end{bmatrix}$.

4.2.2.2 Maximum likelihood decoding

We also decode the stimulus using maximum likelihood (ML) decoding. The likelihood of finding a particular population response is maximized over all possible stimuli to find the most likely stimulus $\hat{\theta}$

$$\hat{\theta} = \text{argmax}_\theta L(\mathbf{r}|\theta),$$

where L indicates the log likelihood and $\mathbf{r}$ is the population response.

First, we assume a naive decoder which does not take the effect of surround modulation into account, resulting in an estimate of the center orientation only. Under Poisson noise the stimulus dependent part of the log likelihood L is

$$L(\mathbf{r}|\theta_c) = \sum_{i=1}^{N} r_i \log g(\phi_i,\theta_c).$$

In the limit of low noise, the estimated center orientation $\hat{\theta}_c$ can be found by setting the derivative of the likelihood with respect to θ_c to zero, resulting in

$$\hat{\theta}_c = \arctan\left(\frac{\sum_{i=1}^{N} r_i \sin\phi_i}{\sum_{i=1}^{N} r_i \cos\phi_i}\right),$$

which is the angle of the population vector. Hence the naive ML decoder and the population vector decoder give identical results.

Next, we use an ML decoder which decodes both the center and surround orientations, taking full account of the effect of surround modulation. When we include this dependence of the response on the surround, the stimulus dependent part of the log likelihood becomes

$$L(\mathbf{r}|\theta_c,\theta_s) = \sum_{i=1}^{N} r_i \log\left[g(\phi_i,\theta_c)h_i(\theta_{ref},\theta_s)\right] - \sum_{i=1}^{N} g(\phi_i,\theta_c)h_i(\theta_{ref},\theta_s).$$

We note that for the naive decoder described above, the second term simplifies to $\sum g(\phi_i, \theta_c)$ which is independent of the stimulus for dense enough tuning curves. Here, the second term is not independent of the stimulus, due to the modulation term $h(\theta_{ref}, \theta_s)$. To find the estimate of both the center and surround orientations, $(\hat{\theta}_c, \hat{\theta}_s)$, the likelihood needs to be maximized with respect to both θ_c and θ_s.

As to our knowledge there is no closed expression for the ML solution, we maximized the log likelihood numerically, starting from different initial conditions for the estimated stimulus to avoid local minima. For each trial, we generated a noisy iteration of the population response $\mathbf{r}$, and used gradient descent to find the stimulus pair which maximized the log likelihood.

4.2.3 Computations

All data analysis and models were implemented in Python 2.7.5, using the Numpy 1.7.1, SciPy 0.12.0 and Matplotlib 1.2.1 toolboxes.

4.3 Results

To examine the functional differences between fixed and center dependent modulation, we compare two phenomenological models for V1 neural responses. The response is the combination of the tuning curve for the center orientation $g()$ and a multiplicative modulation term $h()$ (see Methods). Apart from the surround orientation θ_s, the modulation term depends on either the orientation of the center stimulus θ_c or the preferred orientation of the neuron ϕ_i. This leads to two model variants for neural responses:

$$f_i^{dep}(\phi_i, \theta_c, \theta_s) = g(\phi_i, \theta_c) h(\theta_c, \theta_s) \qquad \text{center dependent} \tag{4.4}$$

$$f_i^{fix}(\phi_i, \theta_c, \theta_s) = g(\phi_i, \theta_c) h(\phi_i, \theta_s) \qquad \text{fixed} \tag{4.5}$$

The model responses $f_i(\phi_i, \theta_c)$ for the two modulation conditions are shown for a neuron with preferred orientation $\phi_i = 0$ in Fig.4.2, with the top row being fixed modulation, and the bottom row center dependent modulation. The tuning to the center alone is identical for both models (Fig.4.2A and D). The responses to arbitrary center-surround combinations with a surround grating present are plotted in Fig.4.2B and E. Sample surround tuning curves with different center orientations are plotted in Fig.4.2C and F.

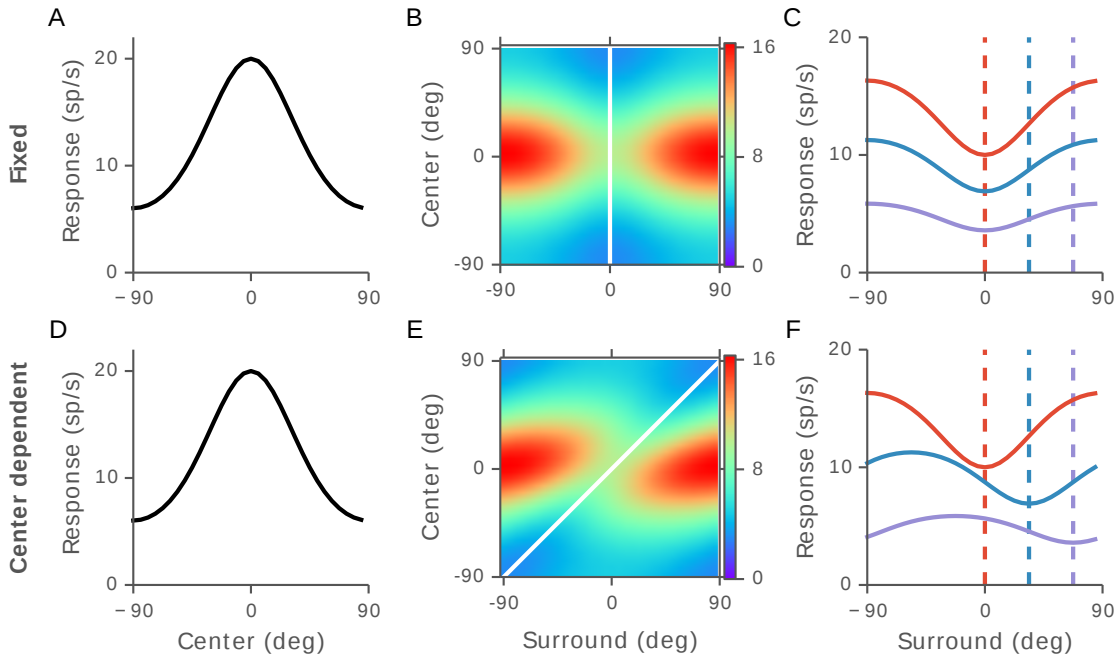

Figure 4.2: Single neuron responses for the two modulation variants: fixed surround modulation (top row) and center dependent surround modulation (bottom row). (A/D) Tuning curve of a neuron with preferred orientation 0deg to the center stimulus only (identical for both models). (B/E) The response of the neuron to all center-surround combinations. The white line indicates the lowest response for a given center orientation, or the strongest suppression. In the fixed model this results in a vertical line (since the minima are always at the same orientation as the preferred orientation), while for the center dependent model the minima are always at the same orientation as the center orientation. (C/F) Surround tuning of the neuron, for different center orientations (rows in panels B and E). For clarity the model response is shown against the surround orientation. The center orientations are indicated by the dashed lines and their respective colors. The curves do not shift for the fixed model, since modulation is fixed irrespective of the center orientation, but shift with the center orientation for the center dependent model, since modulation now depends on the center orientation.

The strongest suppression at a given center orientation is indicated by the white line in panels B and E. In the fixed model, the suppression is strongest whenever the surround is aligned to the preferred orientation of the neuron: as the surround is changed, the modulation curve shifts up and down depending on the center orientation, but does not shift left or right (Fig.4.2C). It is tuned to what the neuron prefers. In contrast, for the center dependent model, the modulation is strongest whenever the center and surround are aligned. As a result, the modulation curves shift left or right depending on the center orientation, (Fig.4.2F). This property has been succinctly expressed as: *"The surround maximally suppresses responses to what the center sees, not to what the center prefers"* (Cavanaugh et al., 2002a).

4.3.1 Population responses

To study the functional impact of the two modulation types, it is important to understand how the modulation specificity is reflected on a population level. We use a population consisting of $N = 32$ neurons with preferred orientations equally spaced in the interval $[0, \pi]$, and responses modeled by Eqs. 4.4 and 4.5. Fig.4.3 illustrates how the resulting population responses (bottom row) are formed from the center drive and the surround modulation across the population (top and middle row), using different center-surround stimulus pairs. The effect of the surround modulation on the population response is shown by the population responses in the presence (black curves) and absence (gray curves) of a surround grating.

The overall population response is suppressed by the presence of a surround grating for both models. The main difference is the way the modulation is felt across the population. In the fixed model, every neuron is modulated differently, as the reference orientation is each neuron's preferred orientation, see Eq. 4.5. In the center dependent model, however, every neuron receives identical modulation, since it is a function of the center and surround stimuli only. This difference is reflected in the population response in two ways.

First, for smaller center-surround orientation differences the mean population response is lower in the center dependent model, compared to the fixed model (Fig.4.3, left column). For higher orientation differences this is reversed and the fixed model has the overall lower response (Fig.4.3, right column). Secondly, dependent on the relative surround orientation, fixed modulation leads to a shifted population response, while center dependent modulation keeps the response stationary (Fig.4.3, middle col-

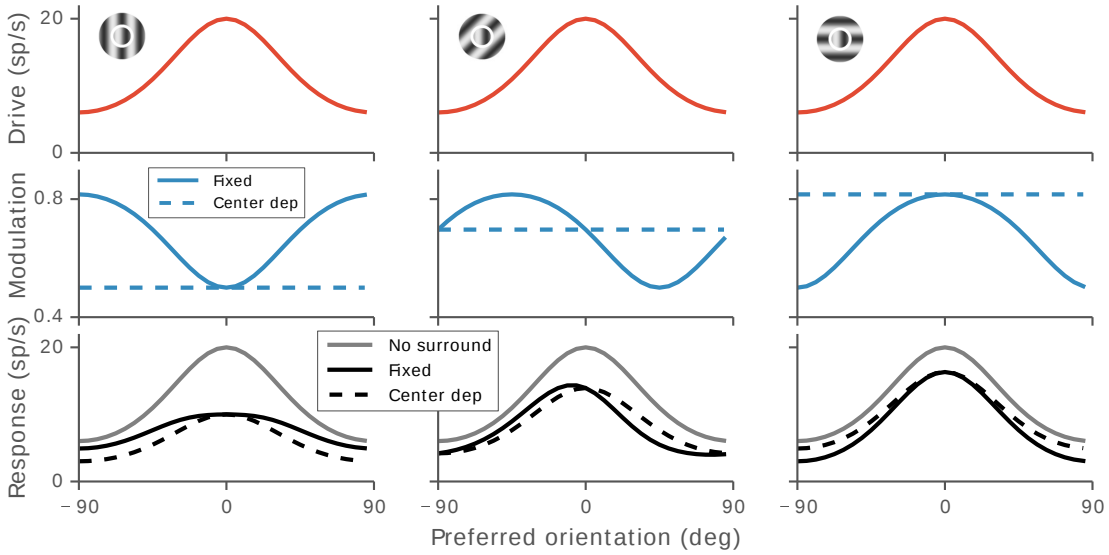

Figure 4.3: Population responses for either surround modulation model. Top row: center drive across the population. Middle row: surround modulation across the population. Fixed modulation is indicated by the solid curve, center dependent modulated by the dashed curve. Bottom row: population responses resulting from the center drive and surround modulation. For comparison, the population response in the absence of a surround grating is given by the gray curve. The center and surround orientations in each column are indicated by the grating in the top row. From left to right, the surround orientations are 0, 45, and 90 degrees. The center orientation is 0 degrees for all three examples. For both fixed and center dependent modulation the presence of the surround suppresses the population response. For small orientation differences center dependent modulation leads to a lower mean response compared to fixed modulation, and to a higher mean response for larger orientation differences. Furthermore, fixed modulation leads for certain surround orientations to a shift of the response (middle column), while center dependent modulation never does.

umn). As we shall show below, these differences in the population responses have major consequences for coding and saliency.

4.3.2 Tilt illusion and center orientation decoding

The differences in the population response have a strong effect on orientation coding. We estimate the center orientation by decoding the neural population response. Many population coding studies have looked at the decoding accuracy, i.e. the trial-to-trial variation and its relation to the neural noise model (e.g. Shamir, 2014). However, here we are particularly interested in the biases in the decoding, that is, the systematic misestimation of the stimulus (Seriès et al., 2009; Cortes et al., 2012). Psychophysically, such decoding biases are thought to be related to the tilt illusion, where the perceived orientation of a center grating is influenced by the presence of a surround grating (e.g. Westheimer, 1990; Clifford, 2014). In particular, for smaller angles the illusion is repulsive. (For larger angles the illusion becomes weakly attractive, termed the indirect tilt illusion). Several models have shown how surround modulation leads to the tilt illusion (Clifford et al., 2000; Schwartz et al., 2009; Qiu et al., 2013; Keemink and van Rossum, 2016), but these all depend on fixed surround modulation. Here we compare fixed to center dependent surround modulation in the effect on this illusion.

4.3.2.1 Population vector

First, we decode the center orientation from the population response using the population vector decoder. This decoder estimates the center orientation from the sum of the preferred orientation vectors of the neurons weighted by their firing rate (see Methods, Georgopoulos et al., 1986a; Schwartz et al., 2009). Across trials the orientation estimate will fluctuate due to noise. Bias is the systematic error in the estimate that remains when averaging over many trials. The estimation bias b_c is defined as (Cover and Thomas, 1991)

$$b_c = \langle \hat{\theta}_c \rangle - \theta_c$$

where $\hat{\theta}_c$ is the estimated center orientation, θ_c the true orientation, and $\langle\rangle$ denotes the circular average over many trials. Not surprisingly, one can show that when only a center stimulus is presented the population vector decoder is bias-free (see Methods).

Next, we present a fixed center grating and a range of surround orientations and decode the center orientation. The biases for both models are plotted against surround

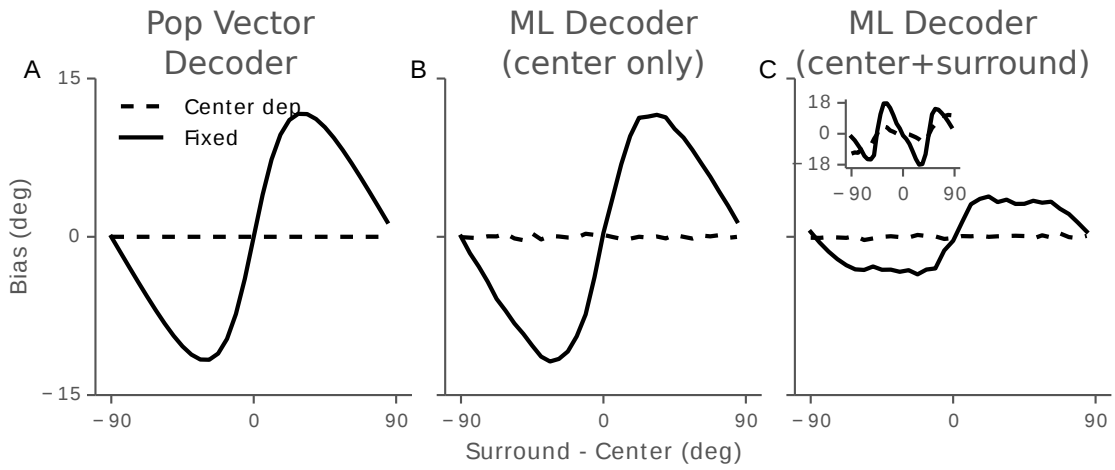

Figure 4.4: Decoding biases in the estimate of the center stimulus orientation, under fixed and center dependent modulation. (A) Center orientation bias for the population vector decoder. Fixed modulation shifts the population response, resulting in a strong repulsive bias. Center dependent modulation does not express this shift, leading to zero bias. (B) The bias for the naive maximum likelihood (ML) decoder. This decoder does not take presence of surround modulation into account, and has the same bias as a population vector for either modulation type. (C) The center bias for the full ML decoder, which takes the influence from the surround into account by decoding both the center and surround orientations. Its knowledge of the surround modulations reduces the bias compared to the population vector. Inset: bias in the estimated surround orientation. The mean bias in B and C was calculated over 5000 iterations.

orientation, Fig.4.4A. As expected and in line with previous models, we find that the fixed model has a strong repulsive bias (solid curve). However, with center dependent modulation this bias completely disappears (dashed curve).

These results can be understood from the population responses in Fig.4.3. If the modulation depends on the preferred orientations, each neuron is modulated differently, with most suppression when the surround aligns with the preferred orientation. This varying modulation shifts the population response away from where it would be for center stimulation only, resulting in a repulsive illusion (Fig.4.4A, solid curve). However, if the surround modulation instead depends on the center stimulus, irrespective of the preferred orientation, then for a given surround and center orientation each neuron is modulated identically. As a result, the population response is still modulated, but not shifted, resulting in zero bias (Fig.4.4A, dashed curve). This effect is fully analog to the results presented in Seriès et al. (2009), where bias effects were studied for

sequentially presented gratings, of which our system is the spatial analogue.

In the Methods we derived the bias analytically for both center dependent and fixed modulation (to our knowledge no analytical derivation of the tilt illusion for von Mises tuning curves was published before). This derivation shows that above results hold generally. For center dependent modulation (Methods) the estimated population vector $\hat{\mathbf{v}}_c^{dep}$ points in the same direction as the vector associated to the actual center orientation $\mathbf{v}_c$,

$$\hat{\mathbf{v}}_c^{dep} \propto \mathbf{v}_c.$$

Since the vectors point in the same direction, this estimate is unbiased, independent of model parameter values.

In contrast, for the fixed model this does not hold and instead (Methods)

$$\hat{\mathbf{v}}_c^{fix} \propto (\mathbf{v}_c - \beta \mathbf{v}_{shift}), \tag{4.6}$$

where β is a constant. The shift vector $\mathbf{v}_{shift} = k_c \mathbf{v}_c + k_s \mathbf{v}_s$ lies between $\mathbf{v}_c$ and $\mathbf{v}_s$ (except if $\theta_c = \theta_s + \frac{1}{2}\pi n$, when it points at $\mathbf{v}_c$). Due to the minus sign in Eq.4.6, the vector $\hat{\mathbf{v}}_c^{fix}$ is repelled from $\mathbf{v}_{shift}$. The amount of repulsion depends primarily on the tuning widths, as given by k_c and k_s, and the modulation strength A_s. In summary, in the presence of a differently oriented surround, a fixed surround modulation model always leads to a repulsive bias in the decoded center orientation.

The qualitative difference between the two models is independent of the precise parameters or model details. First, as the bias concerns averages, the population vector bias is independent of noise model and noise level. Furthermore, the absence of a bias extends to all models where surround modulation is felt equally across neurons, whether the surround modulation is multiplicative, subtractive or some combination of both, that is, for all models of the form $f_i(\phi_i, \theta_c, \theta_s) = g(\phi_i, \theta_c)h(\theta_c, \theta_s) + k(\theta_c, \theta_s)$, where $k()$ is a function describing subtractive modulation.

4.3.2.2 Maximum Likelihood Decoding

Next we decode the responses using a maximum likelihood (ML) decoder. The ML decoder finds the stimulus that most likely caused the observed response; it is often regarded as one of the best possible decoders, in the sense that it should be unbiased and its variance reaches the Cramér- Rao bound (Kay, 1993; Xie, 2002). The reason for studying the ML decoder is two-fold. First, while the population vector decoder above

allows for mathematical analysis, we want to make sure that the observed difference in the bias was not limited to a particular type of decoder. Secondly, as opposed to the population vector, the ML decoder can be extended to decode both the center and surround orientations. This allows one to take into account the effect of surround modulation in decoding, which the population vector is naive to.

In contrast to the population vector decoder, it becomes important to include noise in our model as in the ML decoder biases can arise that are dependent on the noise level. Here we modeled the neural responses as Poisson processes with a rate given by the tuning functions $f()$. The observation time was initially set to 0.5s. The results shown below extend qualitatively to Gaussian additive and multiplicative noise models.

First we use an ML decoder that is naive to the presence of the surround, and assumes that the responses were generated by the center drive only (Methods). This decoder has a bias identical to the population vector decoder (compare Fig.4.4 A and B, this can be shown formally in the limit of low noise, see Methods). The naive ML decoder, as such, confirms our results for the population vector.

However, the naive ML decoder uses an incomplete model of the response likelihood, since it does not include how the response changes when a surround is present. Seriès et al. (2009) used a similar decoder for sequential stimuli, and provided the ML decoder the exact value of the contextual stimulus (their analog of the surround). However, a local population would not have direct access to this information. Instead our full ML decoder has knowledge of the effect of surround modulation, by knowing the full function and estimating both θ_c and θ_s. The estimated center and surround orientations $(\hat{\theta}_c, \hat{\theta}_s)$, are found by maximizing the likelihood with respect to both angles. The estimates have corresponding biases $b_c = \langle\hat{\theta}_c\rangle - \theta_c$ and $b_s = \langle\hat{\theta}_s\rangle - \theta_s$.

While the center dependent model remained bias-free, the fixed model was again biased, although reduced compared to the population vector decoder, Fig.4.4C. Where the bias in the fixed case comes from was more fully explored in Chapter 3.

The average bias across surround orientations is shown as a function of measurement time in Fig.4.5, for both ML decoders and both types of surround modulation. When the surround modulation is center dependent, both the naive and full ML decoders consistently have zero bias. A naive ML decoder of a fixed modulation model has a consistently large bias. However, the bias of a full ML decoder depends on measurement time. In the limit of very long measurement time (or zero noise), the log likelihood landscape becomes very sharp and is always maximized when $(\hat{\theta}_c, \hat{\theta}_s) = (\theta_c, \theta_s)$, i.e. the estimate equals the true value and the bias is zero. This re-

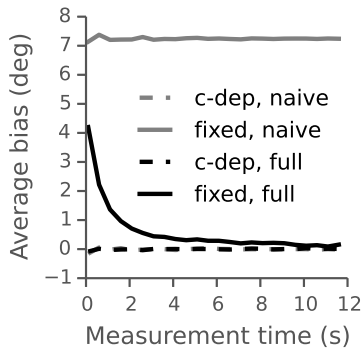

Figure 4.5: The dependence of the average bias on the measurement time, for the two ML decoders and both surround modulation types. C-dep refers to center dependent modulation, and fixed to fixed modulation. The average bias for each measurement time was found over 500 trials. Center dependent modulation never results in a bias (the gray dashed curve fully overlaps with the black dashed curve). Fixed modulation always leads to a bias, however, when using the full ML decoder it decreases with longer measurement times and disappears in the limit.

flects that every center-surround combination yields a unique population response; the ML decoder finds exactly the stimulus pair responsible for the response. In this limiting case decoding is bias-free for both fixed and center dependent models. However when noise is present, the estimate is biased.

In summary, when decoding using a ML decoder center dependent modulation results in zero decoding bias as opposed to fixed surround modulation. For a decoder with full knowledge of the underlying encoding models, the bias can be reduced using long observation times or other ways to reduce the noise level (i.e. more neurons or higher firing rates).

4.3.2.3 Decoding from a mixed population

As mentioned in the Introduction, in recordings the surround modulation in V1 is neither always center dependent nor always fixed, but is a mixture of both (with some neurons showing an intermediate tuning). To test the consequences of decoding from a mixed population we use a population of 3200 neurons and varied the percentage of the center dependent ones. We decoded the center orientation using the population vector, Fig.4.6. The amount of bias strength is proportional to the percentage of neurons with fixed modulation. This also holds when using the full ML decoder (not shown).

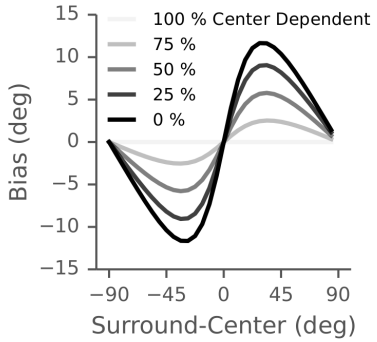

Figure 4.6: The bias in the population vector decoder pooling from a population mixing both types of surround modulation. For 100 populations of neurons with either center dependent or fixed modulation, the percentage gives the number of neurons that are center dependent. A percentage of 75% center dependent neurons would correspond approximately to psychophysical tilt illusion data.

4.3.3 Saliency signals

Surround modulation is furthermore thought to underlie saliency (Sillito et al., 1995; Zhaoping, 1999; Petrov and McKee, 2006), whereby a stimulus differing from its surround pops out. While saliency computation likely involves feedback from higher areas, it has been proposed that part of the saliency is computed from the V1 response in a feed-forward manner, via so called bottom-up saliency (Zhaoping, 1999, 2002). The basic idea is that as surround suppression reduces the response of neurons to parallel stimuli, deviations from the global orientation will stand out. In other words, one would expect a stimulus configuration in which center and surround are aligned to have a lower response than a non-aligned configuration.

We examine how center dependent and fixed surround modulation affect the population responses to the center and the resulting visual saliency. Hereto we extend our model to have multiple populations. Each bar is associated with a population which has that bar as the center stimulus and the neighboring bars as the surround. We compare the population response at the center location to the response in the background. Because the background is mainly surrounded by bars of identical orientation, we neglect the influence of the center bar in these populations. Thus the populations corresponding to the surround bars receive co-linear center-surround stimuli ($\theta_c = \theta_s$). The response of the populations is independent and calculated as in equations 4.4 and 4.5.

Psychophysically saliency is often investigated by search time (how long does it take to find some differing element?)(e.g. Treisman and Gormican, 1988), or the pro-

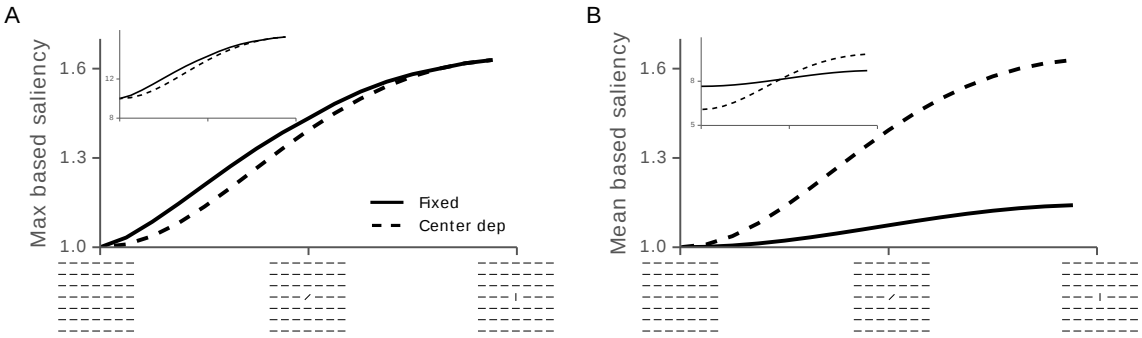

Figure 4.7: Comparison of saliency computation in the center dependent and the fixed surround modulation models, for a deviant bar (center) in a homogeneous background (the surround). (A) Saliency values based on maximum responses, as the center orientation changes. Because the response of neurons with fixed modulation is always lower than for center dependent modulation for the intermediate orientations, saliency is slightly higher for fixed modulation for these orientations. The inset shows the non-normalized maximum response rate in the center. (E) Saliency values based on mean responses. Because the responses to the center bar relative to the background are larger for center dependent modulation, saliency is stronger. The inset shows the mean response in the center.

portion of correct indications of the differing element location (e.g. Field et al., 1993). Such specific decisions, however, likely involve some processing well beyond V1. In the computational literature that tries to explain saliency from just V1 it some measure of the difference in neural response is used (Zhaoping, 1999; Spratling, 2012). Similarly, we define two measures of saliency. First we assume the saliency of the center bar is given by the maximum response within the corresponding population, relative to the maximum response in a surround population (Zhaoping, 1999),

$$s_{max} = \frac{\max(\mathbf{r}^t)}{\max(\mathbf{r}^b)},$$

where $\mathbf{r}^t$ and $\mathbf{r}^b$ are the population responses to the center and surround locations respectively, and $\max(\mathbf{r})$ is the maximum of a population response. If the center bar's maximum response is the same as in the rest of the image the saliency equals 1, reflecting no saliency.

The second saliency measure is the mean response within the target population compared to the mean response in a background population

$$s_{mean} = \frac{\text{mean}(\mathbf{r}^t)}{\text{mean}(\mathbf{r}^b)}.$$

For either saliency readout and either modulation model the response to the center increases relative to the surround as the center bar is rotated away from the surround orientation (Fig.4.7A and B, insets). Consequently, the saliency also increases with the rotated center for either model (Fig.4.7A and B). This is similar to the experimentally observed increase in neural response and perceived contrast of the center (Cannon and Fullenkamp, 1990; Shushruth et al., 2013).

Although the results are the same qualitatively for either model, the quantitative effects are different. The maximum-based saliency s_{max} is similar for both types of surround modulation, with only a slight difference for the orientations between 0 and 90 degrees (compare the solid and dashed curves), with fixed surround modulation resulting in slightly higher saliency, Fig.4.7A. The reason for the saliency difference can be understood intuitively from the center drive and surround modulation curves in Fig.4.3. As the population response curve shifts away from the center orientation for the fixed modulation model, its maximum is higher than for the center dependent model.

The effect of center dependent modulation on mean-based saliency, s_{mean}, is quite different. Although saliency increases for more deviant centers for both modulation types, the increase is stronger in the center dependent model, being over 50% higher in the most salient case, Fig.4.7B. The s_{mean} saliency differences can be explained by the population responses, Fig.4.3. In a non-salient situation (iso-orientation), every neuron in the center dependent model receives as much modulation as the strongest modulated neuron in the fixed model (compare the dashed and solid blue curves). However, in the most salient situation (e.g. orthogonal center and surround), every neuron in the center dependent model is modulated as little as the least modulated neuron in the fixed model. This results in a larger difference in total population response, Fig.4.7B, inset. The saliency is stronger with center dependent modulation across all target orientations.

While we have illustrated this effect using experimentally constrained parameters, these results can be shown to hold more generally (Appendix): for s_{mean} the difference in mean response between iso-orientation and orthogonal stimuli is always larger for center dependent modulation, while (under some broad assumptions) the difference in maximum response is always lower.

In summary, both modulation types lead naturally to a saliency signal. Shushruth

et al. (2012) suggested that a center dependent population will have a larger difference in total response between salient and non-salient locations than a fixed population, resulting in better saliency detection. Our results show that this depends strongly on the saliency computation: Under center dependent surround modulation saliency is actually slightly lower compared to fixed modulation if saliency is computed from the maximum response. However, it is indeed strongly increased compared to fixed modulation if based on the mean response.

4.3.4 Saliency and orientation decoding of complex stimuli

So far we have considered only simple scenes. This raises the question how surround modulation affects perception of more complicated scenes. To study this we need a generalized model that is not limited to a single surround orientation but allows a constellation of stimuli in the surround. We use an existing model for encoding scenes of oriented bars, where the modulation from flankers depends the smoothness of the curve connecting a flanker to a reference bar (Keemink and van Rossum, 2016, Chapter 3). This model reproduces both the tilt illusion and several pop-out effects.

We present the fixed and center dependent modulation variants of this model with various scenes, Fig. 4.8 left column. The decoded scenes, middle and right columns, are based on the model responses. The bar orientations are given by the population vector of each bar's population, and bar opacity is proportional to the mean based saliency. Either type shows basic saliency effects such as pop-out of single features (top and second row), contours (third row) and boundaries (bottom row). However, the pop-out is stronger in each case for center dependent modulation, and there are clear biases in the decoded orientations from fixed modulation. With max-based saliency the results are similar to Fig. 4.7, but instead center dependent modulation then still leads to generally stronger pop-out (albeit with a less dramatic effect than with mean based saliency).

4.4 Discussion

We compared two types of surround modulation in V1: 1) fixed modulation, which is strongest when the surround orientation equals the neuron's preferred orientation, and 2) center dependent modulation, which is strongest when the center and surround stimuli are aligned, and we asked how saliency and orientation coding depend on these

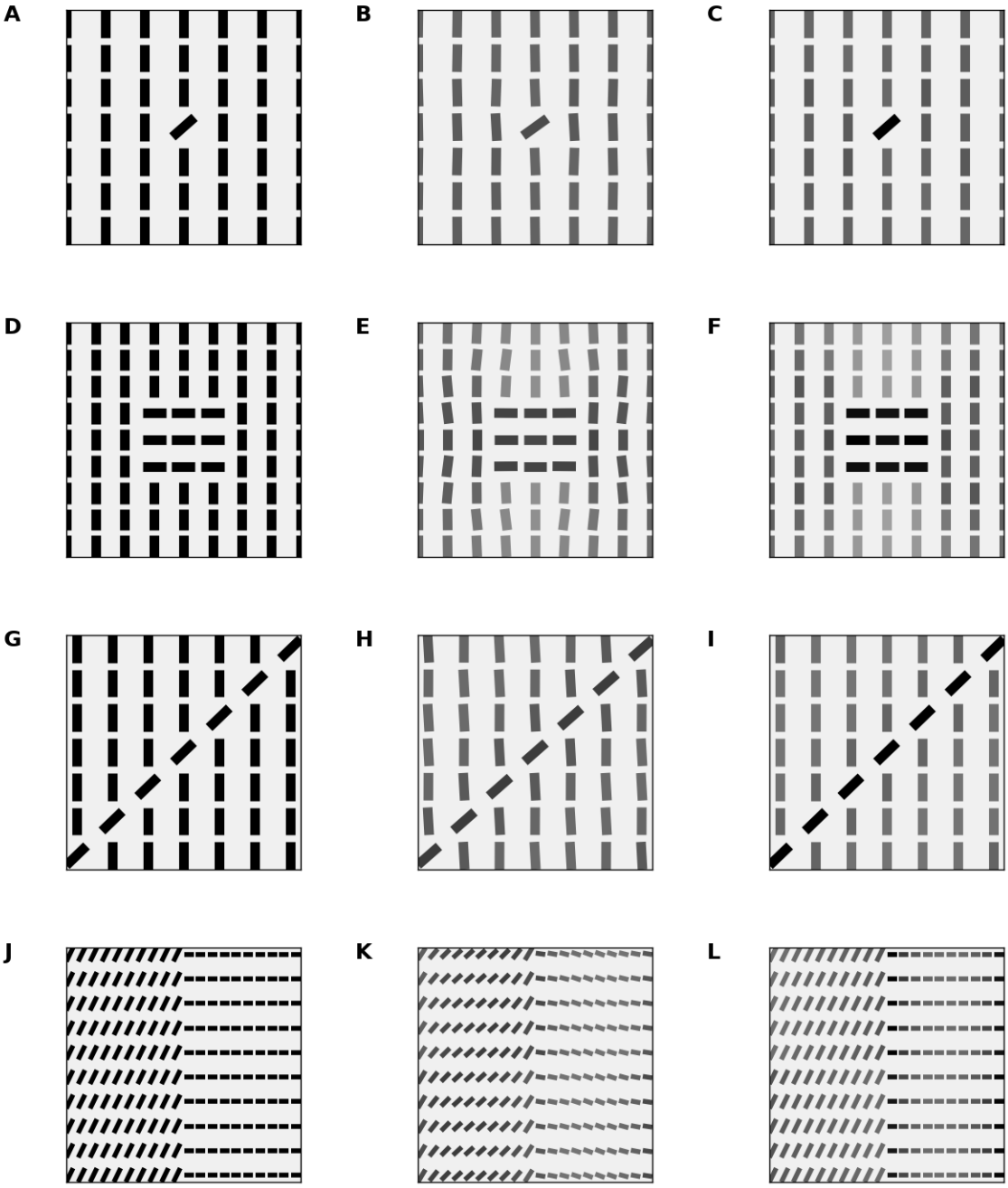

Figure 4.8: Four examples scenes encoded by either a fully center dependent or fully fixed population model. (Left Column) Encoded scene. (Middle Column) Decoded scene with fixed modulation. (Right Column) Decoded scene with center dependent modulation. The orientation of the bars in the decoded scenes is given by the orientation vector of the local neural population. The opacity of the bars is proportional to the mean-based saliency of that bar.

surround modulation variants.

We first examined orientation coding biases arising from surround modulation by decoding the center orientation from the population response. Under fixed surround modulation, due to each neuron being modulated differently, a population vector decoder has a strong repulsive tilt bias, irrespective of the observation time or number of neurons in the population, as is well known from existing tilt illusion models (Blakemore et al., 1970; Clifford et al., 2000; Schwartz et al., 2009). In contrast, for a center dependent model, where the modulation is identical across the population, the bias is completely absent.

For maximum likelihood (ML) decoders the situation is more subtle for fixed modulation. The bias depends on whether the decoder decodes the center orientation only (naive decoder) or both the center and surround (full decoder). For the naive ML decoder we showed that the bias equals the population vector bias. For the full ML decoder the bias depends on the neuronal noise level and observation time. The origin of the bias of the full ML decoder was more fully explored in Chapter 3.

4.4.1 Explaining tilt illusion data

What do our results mean for the origin and magnitude of the tilt illusion? *Qualitatively* the repulsive tilt illusion is explained well by a fixed modulation model, while in a population modeled after the majority of V1 neurons, i.e. with center dependent modulation, the illusion disappears.

We propose that the tilt illusion stems from a mix of center dependent and fixed modulated neurons, Fig.4.6. In our model, fitted to monkey V1 neurons, and assuming a population vector decoder, a population with only fixed modulation neurons leads to a repulsive bias of maximally 12 degrees, compared to about 3 degrees psychophysically (Clifford, 2014). Therefore a maximum repulsive bias of 3 degrees would correspond to a population with about 75% center dependent and 25% fixed modulation. Both types of modulation have been observed electro-physiologically, the first is more prevalent, although a specific ratio is hard to infer from current literature (Sillito et al., 1995; Cavanaugh et al., 2002a; Shushruth et al., 2012). While extracting a quantitative match between neural tuning curve properties and human psychophysics is full of pitfalls, and depends on the decoder used, this ratio is reasonable.

The bias in the full decoder, 4.4C, is also in the right ballpark, and could therefore also explain the observed tilt illusion magnitude. However, in contrast to the popula-

tion vector and naive ML decoders, the bias in the full ML decoder depends on the noise level and disappears at lower noise levels, Fig.4.5 In principle these results imply that psychophysical data could help to determine which decoder best describes the down-stream neural processing involved in orientation estimation. While several studies have reported effects of the presentation time on the bias magnitude (Corbett et al., 2009), the results are conflicting as to the exact nature of the dependence. Across all studies the illusion persisted when using long observation times, the illusion persisted as well as in studies where subjects were free to rotate a test grating until it matched the perceived vertical (Wenderoth and Johnston, 1988), arguing against a full ML decoder.

While to our knowledge no other study has examined the decoding bias in center-surround stimuli using a maximum likelihood decoder, nor with center dependent modulation, similar effects have been studied for the tilt after-effect (Seriès et al., 2009). In the tilt after-effect the orientation of a full field grating is misjudged if presented after a differently oriented adapter grating (Schwartz et al., 2007). Mathematically, the adapter stimulus is the temporal analogue to the surround stimulus in our model. The equivalent of center dependent modulation corresponds to the case where adaptation depends on the difference in orientation between the adapter and test grating. A center dependent naive ML decoder with no knowledge of the adapter was also bias-free, compared to a strong bias for the fixed equivalent. When an aware decoder was used this bias disappeared (Seriès et al., 2009). However, in that study the decoder had perfect knowledge of the adapter orientation which is unlikely physiologically. Instead our decoder infers both center and surround orientations from the population response. With fixed modulation, in principle one could decode the center accurately but needs access to direct information about the context. Under center dependent modulation this is not necessary, significantly simplifying the decoding problem.

4.4.2 Saliency

To examine the effect of modulation on saliency computations we first tested the saliency of a rotated target bar among a field of homogeneously oriented background bars. Assuming a saliency measure based on either the mean or maximum responses to the bars, either surround modulation type provides a bottom-up saliency signal by enhancing the responses to the target relative to the background. For saliency based on the maximum response there is only a small difference between the two surround modulation types. For saliency based on the average population response, the larger response

differences between salient and non-salient locations lead to a stronger saliency signal for center dependent modulation. These results confirm a suggestion from Shushruth et al. (2012) that center dependent modulation might lead to better saliency detection, but we show that this only holds if saliency computation is based on the mean response. To further strengthen the assumption that center dependent modulation increases saliency, when encoding more complex scenes both max and mean based saliency are enhanced when using center dependent modulation, Fig. 4.8.

4.4.3 Normative models of center dependent modulation

Recent work has argued that center dependent modulation emerges from a normative perspective. Lochmann and Deneve (2011) and Lochmann et al. (2012) built a spiking network model with connections that are loosely derived from a Bayesian probability model. Similarly, Coen-Cagli et al. (2012) based the connectivity in a neural model on the image statistics at several surround locations. In both models center dependency seems to arise automatically, from which one could argue that that center dependence follows from these theories. One reason their models are center dependent could be that in their estimation of the parameters, they implicitly assumed an unbiased decoder; the only way to fit parameters under such a decoder, is for the surround modulations to be center dependent. We predict that in both these models the tilt illusion is absent. This does not necessarily argue against the validity of those models, but it does indicate that such normative principles might not underly whichever computation or function done by the visual cortex that causes tilt illusion might not.

In summary, seemingly minor changes in the modulation of neurons, might have important functional consequences. In particular explanations of the tilt illusion are strongly dependent on it. It would be of interest to know whether these conclusions translate to other cortical areas of sensory processing (more on this in the general discussion).

4.5 Appendix

4.5.1 Maximum-based saliency

Here we show that under some assumptions center dependent modulation leads to a lower saliency signal for saliency based on the maximum response. In other words, the

solid line in Fig.4.7D always lies above the dashed line.

We denote the response of a neuron with fixed modulation by

$$f^{fix}(\phi,\theta_s) = g(\phi,0)[1-h(\theta_s,\phi)],$$

and a center dependent neuron by

$$f^{dep}(\phi,\theta_s) = g(\phi,0)[1-h(\theta_s,0)],$$

where we treat the preferred orientation ϕ as a continuous variable (i.e. in the limit of infinite neurons) and assumed $\theta_c = 0$. In other words, θ_s can be considered the center-surround orientation difference. The saliences of the target (see results) are then

$$s^{dep}(\theta_s) = \frac{\max_\phi f^{dep}(\phi,\theta_s)}{\max_\phi f^{dep}(\phi,0)} \qquad \text{center-dependent}$$

and

$$s^{fix}(\theta_s) = \frac{\max_\phi f^{fix}(\phi,\theta_s)}{\max_\phi f^{fix}(\phi,0)} \qquad \text{fixed.}$$

First we note that across the population activity f^{dep} is maximum when $\phi = 0$, i.e. when the center stimulus matches the preferred orientation, independent of the surround orientation. This maximum is given by

$$\max_\phi f^{dep}(\phi,\theta_s) = g(0,0)[1-h(\theta_s,0)].$$

Meanwhile, in the fixed model the activity of the neuron with preferred orientation θ_c is

$$f^{fix}(0,\theta_s) = g(0,0)[1-h(\theta_s,0)] = \max_\phi f^{dep}(\phi,\theta_s),$$

given that $h(\theta_s) = h(-\theta_s)$. In other words, the neuron with the highest activity in the center dependent model always has the same activity as the corresponding neuron in the fixed model, as can be observed from the intersection of solid and dashed curves at $\phi = 0$ in the middle column of Fig.4.3.

However, in the fixed model the corresponding neuron is not the neuron with the highest activity. To see this, consider the derivative of the activity with respect to the preferred orientation

$$\frac{d}{d\phi} f^{fix}(\phi,\theta_s) = -k_c \sin(\phi) f^{ind}(\phi,\theta_s) + k_s \sin(\phi-\theta_s) g(\phi,0) h(\theta_s,\phi)$$

which at $\phi = 0$ equals

$$\frac{d}{d\phi} f^{fix}(0,\theta_s) = -k_s \sin(\theta_s) g(0,0) h(\theta_s,0),$$

This is only zero if $\theta_s = 0 + n\pi$, showing that $f^{fix}(0)$ is neither a minimum nor a maximum for any surround stimulus for which $\theta_s \neq 0 + n\pi$. There must therefore be a point where $f^{fix}(\phi) > \max f^{dep}$. Under the reasonable assumption that in the background condition the maximum response $f^{fix}(\phi)$ occurs when $\phi = 0$, center dependent modulation thus leads to a lower maximum-based saliency signal.

4.5.2 Mean-based saliency

Here we demonstrate that our saliency measure based on the mean responses at different locations always gives better detection under center dependent modulation, compared to fixed modulation.

Assuming $\theta_c = 0$, the mean response to a center stimulus only is given by

$$\frac{1}{\pi}\int_0^{\pi} g(\phi,0) d\phi = 2A_c I_0(k_c).$$

For a center-surround stimulus the mean response to a background bar is

$$\text{mean}(\mathbf{r}^b) = \frac{1}{\pi}\int_0^{\pi} g(\phi,0)(1 - h(\theta_{ref},\theta_s)) d\phi.$$

Substituting θ_c and ϕ_i for θ_{ref} we find for the both modulation types

$$\text{mean}(\mathbf{r}^b) = 2A_c I_0(k_c)\,[1 - h(0,0)] \qquad \text{center-dependent}$$

and

$$\text{mean}(\mathbf{r}^b) = 2A_c\,[I_0(k_c) - A_s I_0(|(k_c + k_s)\mathbf{v}_c|)] \qquad \text{fixed.}$$

The mean response to the target bar meanwhile is

$$\text{mean}(\mathbf{r}^t) = 2A_c I_0(k_c)\,[1 - h(0,\theta_s)] \qquad \text{center-dependent}$$

and

$$\text{mean}(\mathbf{r}^t) = 2A_c\,[I_0(k_c) - A_s I_0(|k_c\mathbf{v}_c + k_s\mathbf{v}_s|)] \qquad \text{fixed.}$$

The saliency follows as

$$s^{dep}(\theta_s) = \frac{1 - h(0,\theta_s)}{1 - h(0,0)} \qquad \text{center-dependent}$$

and

$$s^{fix}(\theta_s) = \frac{I_0(k_c) - A_s I_0(|k_c \mathbf{v}_c + k_s \mathbf{v}_s|)}{I_0(k_c) - A_s I_0(k_c + k_s)} \qquad \text{fixed.}$$

While we found no proof that $s^{fix} < s^{dep}$ for all orientations θ_s, we will compare the non-salient to the most salient condition. In the non-salient collinear condition , i.e. $\theta_s = 0$ the saliency is the same for both modulation types:

$$s^{fix}(0) = s^{dep}(0) = 1.$$

In the most salient condition of orthogonality, i.e. $\theta_s = \pi/2$, the saliences are given by

$$s^{dep}(\frac{\pi}{2}) = \frac{1 - A_c \exp(-k_s)}{1 - A_c \exp(k_s)} \qquad \text{center-dependent}$$

and

$$s^{fix}(\frac{\pi}{2}) = \frac{I_0(k_c) - A_s I_0(k_c - k_s)}{I_0(k_c) - A_s I_0(k_c + k_s)} \qquad \text{fixed.}$$

Since $1 - A_c \exp(-k_s) > 1 - \frac{A_s I_0(k_c - k_s)}{I_0(k_c)}$ and $1 - A_c \exp(k_s) < 1 - \frac{A_s I_0(k_c + k_s)}{I_0(k_c)}$, it follows that

$$s^{dep}(\pi/2) > s^{fix}(\pi/2).$$

In other words, the relative range in Fig.4.7E is larger for center dependent modulation than for fixed modulation.

4.5.3 The attractive tilt illusion

In the main text our model only accounts for the repulsive tilt illusion. A small but consistent attractive effect has been observed for larger orientation differences. The attractive effect can be achieved in our model by adding fixed facilitatory modulation. If these are center dependent, there would again be no illusion (not shown). However, our model was never meant to replicate the tilt illusion explicitly, and it is not clear how well our results generalize to models specifically designed to explain the tilt illusion.

We study a recent model that successfully accounts for both the attractive and repulsive effects, by including an estimate of how likely it is that the center and surround are part of the same image, as informed by image statistics (Schwartz et al. (2009), see Fig.4.9A, solid curve). This model was also recently shown to predict newer psychophysics experiments based on contrast (Qiu et al., 2013). Similarly to our model, the Schwartz model is of the form $f_i(\phi_i, \theta_c, \theta_s) = g(\phi_i, \theta_c) h(\phi_i, \theta_c, \theta_s)$, but with a non-linear function $h()$ dependent on both the preferred, center and surround orientations.

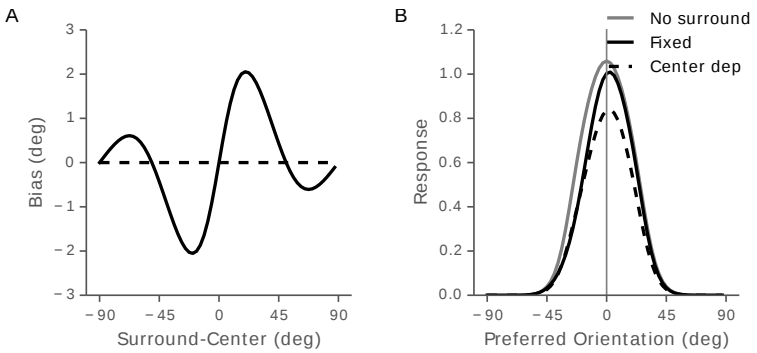

Figure 4.9: Bias results for the Schwartz model. (A) Bias illustrated in Schwartz model. The original model reproduces the real tilt illusion exceptionally well (black line), accounting for both the repulsive and attractive effects. However, the bias completely disappears if we adjust the model so that the surround modulation is center dependent rather than fixed (dashed line). (B) Population responses with surround orientation at -30 degrees. The population response without a surround is illustrated by the gray line. If the modulation is fixed the population response shifts resulting in a bias (black line). But, as in our model, for center dependent modulation the population response is modulated, but not shifted (dashed line), resulting in zero bias in panel A. The parameters used were $k = 0.125$, $\sigma = 16$, $\lambda = \sqrt{4000}$, $N = 64$ (see Appendix).

While our model parameters were fit to neural responses, the Schwartz model parameters are based on image statistics (Appendix).

While the Schwartz model accounts for the psychophysics well, it is not clear if this is solely due to fixed modulation as in our model. When we investigate the surround modulation tuning in detail, we find that it is always strongest when the surround and preferred orientations of a neuron are aligned, Fig.4.10B and C, indicating that the surround modulation is fixed. If the model is changed to have center dependent surround modulation (see below, and Fig.4.10E and F), the bias disappears, as in our model, Fig.4.9A.

As to the underlying cause of the tilt illusion, or lack thereof, the same reasoning as in our model applies here, as is illustrated in Fig.4.9B. The gray curve shows the population activity in the absence of a surround, and the solid and dashed black curves correspond to the center dependent and fixed versions of the Schwartz model respectively, with a surround at -30 degrees. Both population responses are modulated, but only the solid curve shifts away from the surround orientation due to the fixed modulation. This shifted population activity leads to a bias for the fixed model, while the

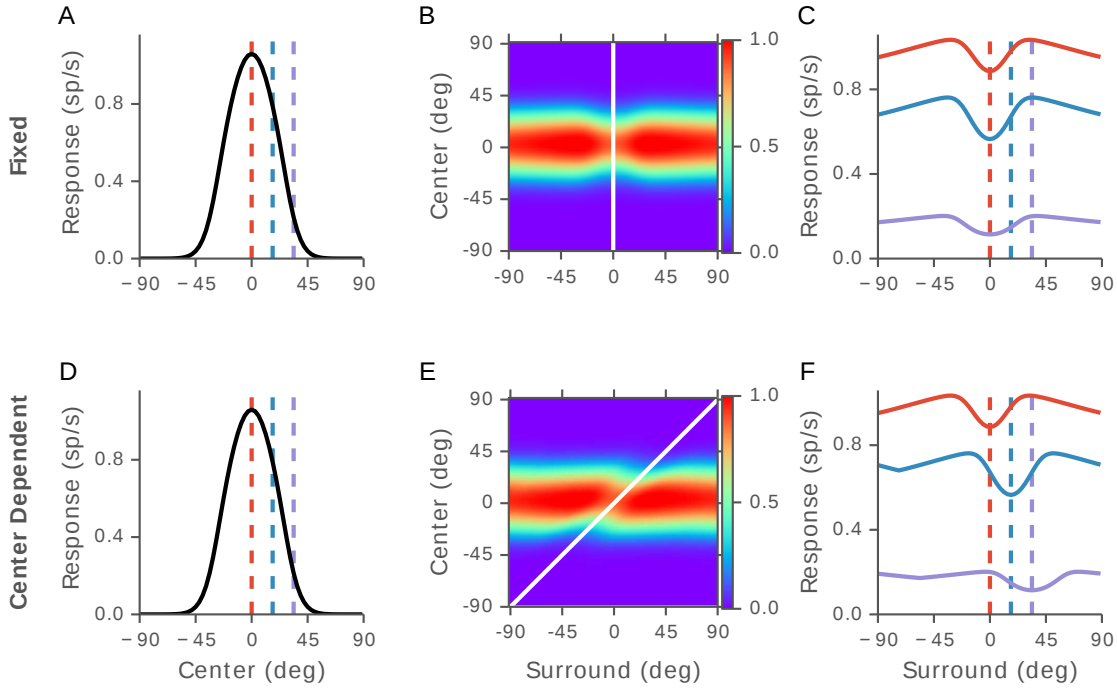

Figure 4.10: Illustration of the Schwartz models. Panels A to C describe the original (fixed), model, and panels D to F the modified (center dependent) model. (A/D) Tuning curve of a neuron in the absence of surround. Identical for both models. The colored dashed lines indicate at which center orientations the surround tuning curves are plotted in panels C and F. (B/E) Response of a neuron to all center-surround combinations. The white lines indicate where the lowest responses are for each center orientation. In the fixed model this results in a vertical line (since the are always oriented the same as the preferred orientation), while for the center dependent model it is tilted 45 degrees (since the minima are always oriented the same as the center orientation). (C/F) Surround tuning of a neuron, for different center orientations. The center orientations are indicated by the dashed lines and their respective colors. For every center orientation the response is shown for all relative surround orientations by the solid lines. Note that the curves remain stationary for the fixed model, but shift with the center orientation for the center dependent model.

stationary population response under center dependent modulation results in no bias at all.

We will now detail how the Schwartz model was made center dependent. In the original model the neural responses are given by a function of the form

$$f_i(\phi_i, \theta_c, \theta_s) = g(\phi_i, \theta_c) h(\phi_i, \theta_c, \theta_s),$$

where $g(\phi_i, \theta_c) = sign(l_c) l_c$ and $h(\phi_i, \theta_c, \theta_s) = p\sqrt{\frac{1}{l_1}} \frac{B(\frac{n}{2} - \frac{1}{2}, l_1)}{B(\frac{n}{2} - 1, l_1)} + (1-p)\sqrt{\frac{1}{l_2}} \frac{B(0, l_2)}{B(-\frac{1}{2}, l_2)}$ (for details see Schwartz et al., 2009). Here B is the Modified Bessel Function of the Second kind, and l_c, l_s, l_1 and l_2 are given by

$$l_c(\phi_i, \theta_c) = \exp\left(-\frac{|\phi_i - \theta_c|^2}{2\sigma^2}\right)$$

$$l_s(\theta_{ref}, \theta_s) = \exp\left(-\frac{|\theta_{ref} - \theta_s|^2}{2\sigma^2}\right)$$

$$l_1(\phi_i, \theta_c, \theta_s) = \sqrt{l_c^2 + l_s^2 + k}$$

$$l_2(\phi_i, \theta_c) = \sqrt{l_c^2 + k}$$

where k is an additive constant. The probability p describes whether a given center and surround are part of the same object (as informed by image statistics) and is given by

$$p(\theta_{ref}, \theta_s) = \exp\left(\frac{-|\theta_{ref} - \theta_s|^2}{2\lambda^2}\right).$$

In the original paper θ_{ref} was set as ϕ_i, and as a result the model shows fixed surround modulation.

By setting $\theta_{ref} = \theta_c$ we can make the model center dependent. By setting $\theta_{ref} = \theta_c$ instead of ϕ_i, two parts of the model change. Importantly, these changes can be made without changing the model assumptions, as we will argue below. First the drive from the surround becomes

$$l_s(\theta_{ref} = \theta_c, \theta_s) = \exp\left(-\frac{|\theta_c - \theta_s|^2}{2\sigma^2}\right)$$

This is acceptable because θ_{ref} was chosen to be ϕ_i assuming neurons experience fixed surround modulation in the first place. The second change is a neuron's guess of the probability that the center and surround are part of the same object:

$$p(\theta_{ref} = \theta_c, \theta_s) = \exp\left(\frac{-|\theta_c - \theta_s|^2}{2\lambda^2}\right).$$

This was originally assumed to depend on the preferred orientation of a neuron, rather than the actually presented center orientation. But there is no inherent reason why this

probability should be center dependent or fixed, except to make the neural responses conform to the experimental responses.

Chapter 5

Contextual modulation in mouse visual cortex and calcium-imaging signal separation

This chapter consists of a paper in preparation, which describes a toolbox nearing release, developed as part of a collaboration with the Rochefort lab (University of Edinburgh). The experiments where designed by N. Rochefort and J. Pakan. All the recordings were collected by E. Dylda and J. Pakan. The data was extracted and prepared using code co-written by me, J. Pakan and S. Lowe. The algorithms and their implementations described in this chapter were primarily developed by me, with some of the code written in collaboration with Scott Lowe. The first application of this work can be seen in Pakan et al. (2016).

5.1 Introduction

Recently, the mouse has become the model of choice for visual cortex research, mainly due to the richness of available genetic tools, part of which is the use of 2-photon imaging in awake animals using fluorescent calcium indicators. Although classically the mouse primary visual cortex was not considered to have strong contextual modulations, being weaker and only with large stimuli (Van den Bergh et al., 2010), this is now regarded to be due to the use of anesthetized animals. Several studies have now shown that receptive fields in awake animals are smaller, and contextual effects (at least with center-surround stimuli) are present (Haider et al., 2013; Vaiceliunaite et al., 2013b; Self et al., 2014). While a number of classic experiments performed in

other species have been reproduced (mainly that surround modulation grows stronger with bigger surrounds, and that it is surround orientation dependent (Self et al., 2014)), several questions still remain. In Chapter 2 we discussed center dependent modulation. Is this also present in the mouse visual cortex? In chapter 3 we looked at the elastica model, and how it reproduces both experimental and psychophysical results. Is the surround modulation in mice similarly context dependent?

One challenge is that the difference in neural responses between center dependent or fixed modulation, or elastica-like tuning and general association fields, can be subtle. For this reason contextual effects beyond center-surround modulation have predominantly been studied using single-unit electrophysiology, which has a high signal to noise ratio (SNR) and temporal resolution. However, the number of neurons recorded with this approach is typically low, and electrophysiological recordings are biased towards overly active cells (Olshausen and Field, 2005a). Now, genetically encoded calcium indicators, such as GCaMP6, combined with 2-photon imaging, have enabled monitoring the activity of many cells simultaneously with single-cell resolution and reasonably good temporal resolution (although still low compared with electrophysiology) (Chen et al., 2013b). Although 2-photon calcium imaging has seen continuing improvements in both the temporal and spatial resolution a given cell's signal is often contaminated by signals from background noise and nearby sources such as neighboring cells, overlapping dendrites and axons. This contamination is collectively referred to as neuropil contamination, Fig. 5.1 red area around cell and red curve in B and C. It is essential to remove neuropil contamination in order to accurately estimate the true signal (Harris et al., 2016), which is especially important when measuring subtle contextual effects. Additionally, in our experience the contamination can be strongly correlated to the stimulus, leading to an overestimation of the number of tuned cells.

Generally there have been two approaches to removing contamination and detecting cells (Harris et al., 2016). First, and more classically, region of interest (ROI) based methods define ROIs for each cell body (usually by hand labeling, sometimes by automated detection algorithms). Then, some neuropil region is defined, of which the average signal is subtracted from the measured average ROI signal (as in Chen et al., 2013b). Subtraction has the advantage that it is fast and intuitive, and can be used on predefined cells. However, it often leads to artifacts (such as negative signals relative to baseline), reduces the signal-to-noise ratio, and is strongly user dependent, as hand-tuned parameters chosen for each cell are needed for best results. Secondly, and more recently, pixel based methods use blind source separation (such as Indepen-

dent Component Analysis (ICA) (Stetter et al., 2000; Mukamel et al., 2009) or Non-Negative Matrix Factorization (NMF) (Maruyama et al., 2014; Pnevmatikakis et al., 2015)), which simultaneously extract cell locations and signals using source separation techniques. The resulting signals typically also have much reduced contamination. Pixel-based methods are principled, generally work well, and are reproducible. However, they are computationally intensive and rarely able to detect all cells. To get the majority of cells labeled (especially the more sparsely active ones), hand labeling is often still necessary. For those cases the neuropil contamination is not automatically corrected, and a better method than subtraction is needed.

As an alternative we developed the Fast Image Source Separation Analysis (FISSA) toolbox. Whereas existing pixel-based methods take every individual pixel in the image as signal sources (sometimes using PCA to reduce the dimensionality), FISSA avoids this by exploiting the fact that contamination generally comes from nearby signal sources. We define a set of neuropil regions around a pre-defined ROI, from which a set of signals is extracted (the raw somatic signal, plus the surrounding neuropil signals), all of which are a mixture of the underlying true signals; one of which is the somatic signal we want to extract. FISSA then uses blind source separation to extract the underlying signals, and finds the signal most strongly represented in the somatic ROI. Since only a few signals need to be separated, this process is orders of magnitude faster than the pixel-based methods.

We applied this method to both experimental and simulated data. We found that the performance is similar to pixel-based methods, whilst computationally two orders of magnitude faster. FISSA thus offers the best of both worlds: it is fast and easy to use like subtraction, but is as accurate as other signal separation methods. We provide both MATLAB and Python implementations at: `https://github.com/rochefort-lab/fissa` and `https://github.com/rochefort-lab/fissa-matlab`. These repositories are currently private; publication and online access are planned for 2017.

5.2 FISSA workflow

FISSA assumes ROIs have been drawn in advance, using either manual detection or some detection algorithm. For all the experimental results in this chapter, the ROIs were drawn by J. Pakan. We now first give a quick overview of the FISSA workflow, before discussing the steps in more detail. Briefly, given a known somatic ROI, FISSA first sets the extent of a neuropil region by growing the somatic ROI a fixed amount,

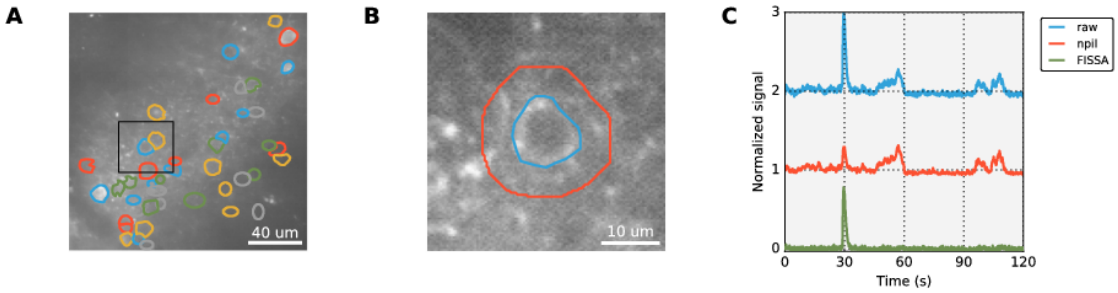

Figure 5.1: Example calcium data and extraction. (A) Two-photon image of layer2/3 neurons in the mouse primary visual cortex. The neurons are labeled with GCamP6f. The background image is an average across 2,400 frames, with several ROIs drawn around neural bodies. Black square shows region shown in B. (B) Zoomed in on a specific cell, with a ROI around a cell body (blue) and neuropil region in red (this only extends to the ROI). (C) The raw ROI trace (blue), the surrounding neuropil trace (red), and the extracted signal by our method, FISSA (green). For plotting purposes, all signals first had their median subtracted, and were divided by the global maximum across the four signals. Each trace was plotted offset by 1 relative to the next trace.

Fig. 5.2A. The final neuropil area is then defined as the grown shape, excluding the original ROI. Next, the neuropil region is divided in subregions of equal area, Fig. 5.2B. The signals for each region are then extracted, Fig. 5.2C. These signals will include both the somatic signal of interest, as well as contaminating signals. To reduce the high frequency noise, the signals are low-pass filtered at 5 Hz, Fig. 5.2D. Although this may at first seem like a low frequency to choose, the calcium indicator itself already acts as a low-pass filter (a convolution of the spiking signal with an exponential filter, see simulated data generation, Fig. 5.6), therefore 5 Hz is reasonable. These signals are then separated using blind source separation, Fig. 5.2E. For every separated signal, blind source separation also determines how strongly it is present in the originally measured signals from the results of blind source separation. FISSA chooses the signal most strongly present in the somatic ROI relative to the neuropil regions as the estimated somatic signal, Fig. 5.5F green curve.

5.2.1 Neuropil definition

To decontaminate the somatic signal, we first define a region which we believe will also contain the contaminating neuropil signals. Typically, this is done by generating a circle around a ROI, and adjusting the size by hand (e.g. Chen et al., 2013b), with

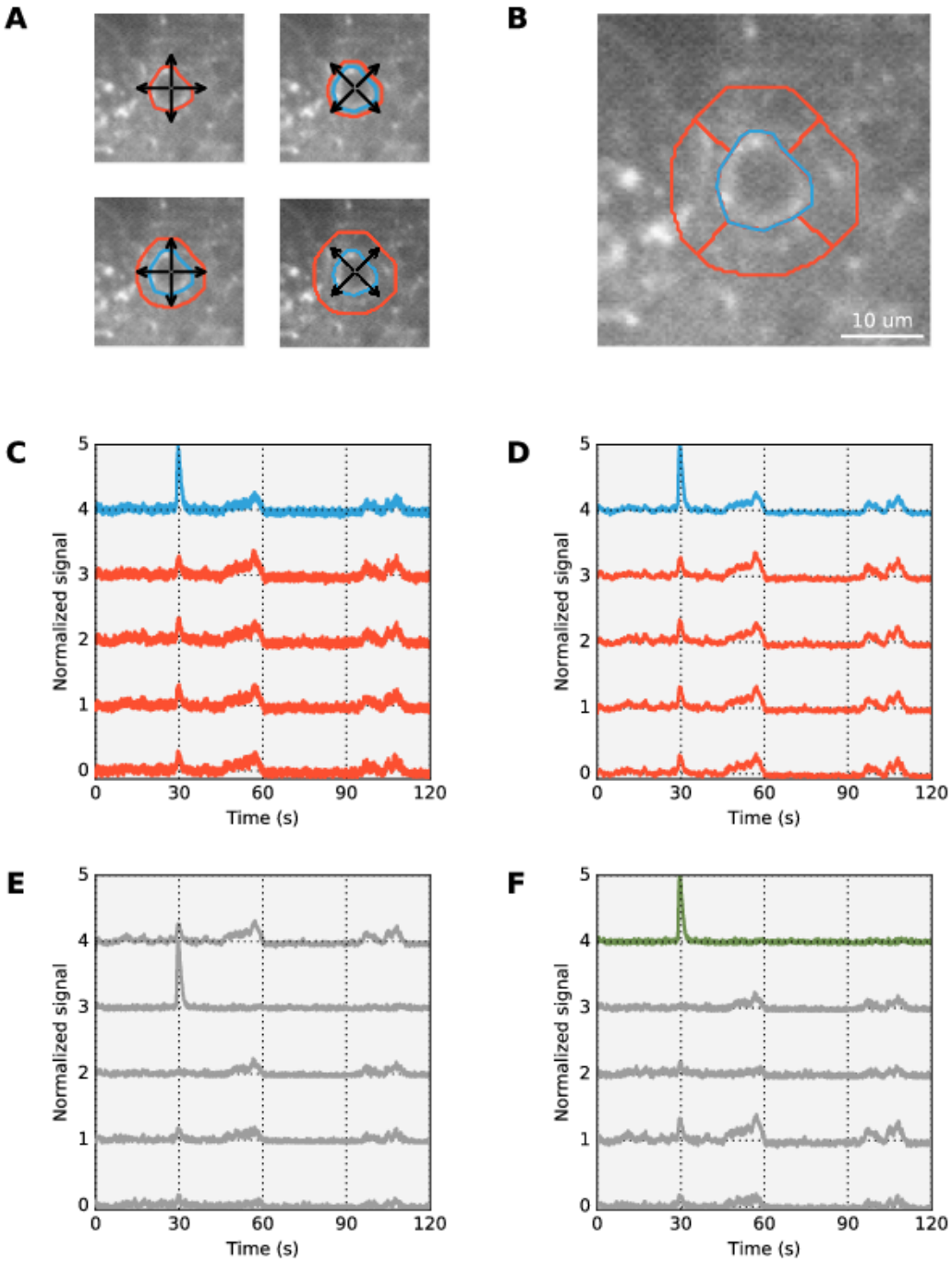

Figure 5.2: Basic FISSA workflow. In each panel, the median was subtracted from each signal, and the traces were divided by the global maximum. Each trace was offset by 1 relative to the next trace. (A) First, a neuropil region (red) is defined around the ROI (blue). The basic ROI shape is used as a base, which is grown over a fixed number of iterations (see main text for details). (B) The resulting neuropil region is then split up into a set of equal area subregions. (C) The signals extracted from the ROI and neuropil regions, and (D) passed through a low-pass filter. (E) The signals are then separated by NMF, at which point it is unknown which of the separated signals is the true somatic signal. (F) The NMF-extracted signal that is most strongly present in the ROI, relative to the neuropil regions, is chosen as the 'true' signal (green curve). Here the NMF-separated signals are scaled and ordered by how much they are present in the central ROI.

some other additions such as leaving out correlated pixels (Peron et al., 2015), or manually, by inspection, choosing per cell whether decontamination is necessary (Chen et al., 2013b). This hand-tuning is necessary, since each cell has different contaminating signals. In FISSA, the surrounding region is defined by 'growing' the ROI, Fig. 5.2A, by moving the neuropil shape alternately in the cardinal and diagonal directions, starting with the basic ROI shape. This process is repeated a fixed number of steps. We used 15 steps which approximately corresponds to a neuropil region thickness of 15 pixels (or 10μm, depending on the imaging resolution). Afterwards, the Neuropil region is split up into N equal area subregions, by taking the polar coordinates relative to the Cell ROI center and taking the $\frac{1}{N}'$th percentile for each subregion. For the examples in this paper we set $N = 4$, but a higher N would be more appropriate for datasets with more densely packed neurons.

5.2.2 Decontamination algorithms

5.2.2.1 Neuropil subtraction

Given a somatic ROI and a surrounding neuropil region, a common decontamination method is subtraction of the neuropil signal. This rests on the assumption that the measured signal in a ROI is the true signal, $f_{\texttt{true}}(t)$, plus some contaminating signal $f_{\texttt{cont}}(t)$ as follows:

$$f_{\texttt{ROI}}(t) = a\, f_{\texttt{true}}(t) + b\, f_{\texttt{cont}}(t),$$

where a and b are constants. A surrounding neuropil region is therefore defined and subtracted, which should get rid of some of the contamination. The estimated signal is then

$$f_{\texttt{est}}(t) = f_{\texttt{ROI}}(t) - k\, f_{\texttt{npil}}(t),$$

where k is a constant, usually found by hand tuning per dataset to find the best value to remove most contamination (e.g. Chen et al., 2013b). We set $k = 1$. For our purposes, we define $f_{\texttt{npil}}(t)$ to be the average signal across all neuropil subregions.

Although widely used, the subtraction method leads to many problems, such as negative events when too much signal is removed, or signal loss when the somatic signal bleeds into the neuropil. This is partly due to the assumption that the contaminating signal originates from a single source removable by subtraction, and that the

neuropil signal is exactly proportional to the contaminating signal. However, contamination generally comes from several sources. For example, there can be a global background contamination, an active nearby cell, or an overlapping axon or dendrite. Consequently, a more accurate model of the measured signal is

$$f_{\mathtt{ROI}}(t) = a\, f_{\mathtt{true}}(t) + \sum b_i\, f_{\mathtt{cont},i}(t),$$

where $f_{\mathtt{cont},i}(t)$ are the different contaminating signals, and b_i describes how strongly each signal is present in the ROI. If we now define some nearby region (such as a surrounding neuropil) the signal for that region would be

$$f_{\mathtt{npil}}(t) = \sum_i c_i\, f_{\mathtt{cont},i}(t) + d\, f_{\mathtt{true}}(t).$$

where c_i and d are constants. Crucially, across the different signal sources i, each c_i does not necessarily equally relate to each b_i, so that when one attempts a subtraction of the form $f_{\mathtt{ROI}}(t) - k\, f_{\mathtt{npil}}(t)$, there is not always a k which will result in the complete removal of all contamination. Instead, one might over-correct (resulting in negative signals), or reduce the SNR when d is large enough (i.e. when the true signal is strongly present in the neuropil). Normally, to adjust for this, a certain degree of hand-tuning is necessary for each cell. A more principled approach, given sets of mixed signals such as we have now described, is to use blind source separation.

5.2.2.2 Blind Source Separation

As described in section 2.2.1, in calcium imaging recordings there is set of underlying true signals, and measurements which are mixed versions of the true signals. In blind source separation, given a set of measured signals, $\mathbf{f}_{\mathtt{measured}}(t)$, one tries to estimate the underlying true sources $\mathbf{f}_{\mathtt{true}}(t)$. This approach relies on the assumption that the measured signals are a linear mix of the true signals

$$\mathbf{f}_{\mathtt{measured}}(t) = W\, \mathbf{f}_{\mathtt{true}}(t),$$

where W is the mixing matrix (which would contain the variables a and b_i, etc. from the subtraction method). Blind source separation allows us to estimate a mixing matrix V and separated sources $\mathbf{f}_{\mathtt{sep}}(t)$ such that

$$\mathbf{f}_{\mathtt{measured}}(t) = V\, \mathbf{f}_{\mathtt{sep}}(t),$$

under some constraints on the signals $\mathbf{f}_{\texttt{sep}}(t)$. The basic process of blind source separation is illustrated in Fig. 5.3. The biggest advantage compared with the subtraction method is that no per-cell hand-tuning is necessary, as the underlying sources are estimated as well as their relative contributions, instead of attempting to remove a single signal.

We define $\mathbf{f}_{\texttt{measured}}(t)$ as the average signals from the central ROI and the N neuropil regions, resulting in $N+1$ signals. FISSA provides the option to perform blind source separation with either Independent Component Analysis (ICA) or Non-Negative Matrix Factorization (NMF), both as implemented in the scikit-learn toolbox (Pedregosa et al., 2011). ICA relies on the fact that the distribution of the sum of random variables will be more Gaussian than the individual components. Thus by finding the most non-Gaussian projections, the sources can be found. ICA is the faster of the two options, but allows that signals can be negative, which is not possible with calcium imaging. This allows for separated signals to be present positively in one measured signal, but negatively in another measured signal. In practice this can lead to negative signal artifacts similar to those found with subtraction. NMF makes the strong assumption that all signals are strictly non-negative, a property which is necessarily true for calcium imaging data. We therefore recommend to use the NMF method, which was used for all the results in this chapter. Although ICA is more principled, and it gives much quicker and generally comparable results, in our experience it too often has negative events similar to those found with neuropil subtraction.

The signals over time $\mathbf{f}_{\texttt{measured}}(t)$ and $\mathbf{f}_{\texttt{sep}}(t)$ can be written as matrices with components

$$F_{i,t,\texttt{type}} = f_{i,\texttt{type}}(t).$$

The NMF algorithm then minimizes an objective E in alternating steps with respect to V and $F_{\texttt{sep}}$, until a target threshold is reached (Lin, 2007; Cichocki and Anh-Huy, 2009). The objective is the total squared difference between the measured signals and the estimated signals $F_{\texttt{measured}}$, plus some additional norms:

$$\begin{aligned} E = \tfrac{1}{2}\,||F_{\texttt{measured}} - V\,F_{\texttt{sep}}]^2||_{\texttt{Fro}^2} + \alpha l_{1\texttt{ratio}}\,||\texttt{vec}(V)||_1 + \alpha l_{1\texttt{ratio}}\,||\texttt{vec}(F_{\texttt{sep}})||_1 \\ + \tfrac{1}{2}\,\alpha(1-l_{1\texttt{ratio}})\,||V||_{\texttt{Fro}^2} + \tfrac{1}{2}\,\alpha(1-l_{1\texttt{ratio}})\,||F_{\texttt{sep}}||_{\texttt{Fro}^2} \end{aligned},$$

where the Frobernius norm is given by $||A||_{\texttt{Fro}^2} = \sum_{i,j} A_{ij}^2$ and the element-wise L1 norm is given by $||\texttt{vec}(A)||_1 = \sum_{i,j} \texttt{abs}(A_{ij})$. We set $= 0.1$ and $l_{1\texttt{ratio}} = 0.5$. $l_{1\texttt{ratio}}$

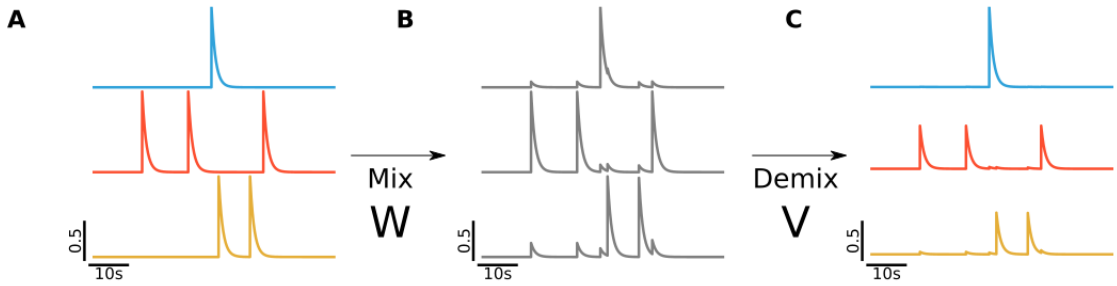

Figure 5.3: Basic principle of blind source separation. (A) Three true underlying signals. (B) Mixed versions of the the signals in A. (C) Signals separated and matched using NMF.

determines the ration between the Frobernius norm and the element-wise norm. α roughly determines the degree of sparsity on both the mixing matrix V and the separated signals F_{sep}. Finally, both the separated signals and the mixing matrix are constrained to be fully positive.

5.2.2.3 Signal selection

An important step in the FISSA algorithm is choosing which output signal corresponds to the underlying somatic signal. Blind source separation returns a set of signals, but it is not obvious which one is the true signal, and which ones are contaminating signals. However, we do know the weight with which each identified source signal contributes to the original input signals, characterized by the estimated mixing matrix V. For an example of three measured signals, the mixing matrix is

$$V = \begin{bmatrix} v_{00} & v_{01} & v_{02} \\ v_{10} & v_{11} & v_{12} \\ v_{20} & v_{21} & v_{22} \end{bmatrix},$$

where v_{ij} is the weight of the j-th separated signal in the i-th measured signal, with each row representing how strongly each of the three estimated underlying signals are present in the measured signal, as illustrated in Fig. 5.4. Each column shows how strongly a given underlying signal is represented across all three measured signals. To choose the underlying signal which corresponds to the true somatic signal, we make a single assumption: the somatic signal is most strongly present in the central ROI, compared to the neuropil regions. We therefore rate each signal by how strongly it is present in the ROI compared to the neuropil regions. This is done by normalizing the

weights in each column so that their sum is 1:

$$v'_{ij} = \frac{v_{ij}}{\sum_i v_{ij}}.$$

The values v'_{ij} now represent how strongly each signal is present in each region, relative to the other regions. If $v'_{ij} = 1$, signal j is only present in the corresponding region. When $v'_{ij} = 0$ signal j is absent. Values between 0 and 1 indicate mixing across the measured signals. Each underlying signal can now be ranked according to v'_{0j}, the relative presence in the originally measured ROI signal. The signal that we take as the estimated true somatic signal is given by the signal for which v'_{0j} is the highest, multiplied by its contribution to the measured ROI signal

$$j_{\mathtt{max}} = \arg\max_j v'_{0j}.$$

Now the final estimated signal follows as

$$f_{\mathtt{est}}(t) = v_{0j_{\mathtt{max}}} f_{\mathtt{sep}}^{j_{\mathtt{max}}}(t),$$

where $f_{sep}^{j}(t)$ is the j-th signal as separated by blind source separation.

5.2.2.4 Number of signals

For given number of neuropil regions, it is still possible that the number of local signals is much lower than the number of regions. This could lead to a degradation of the estimated signal in extreme situations. This is mainly an issue when using the ICA separation method; our experience has shown single signals can be split apart due to the enforced independence between extracted signals. To prevent this, FISSA estimates the number of true signals present using Principal Component Analysis (PCA). The target number of signals extracted is set to the number of components which explain at least 1% of the variance. Using the input parameters, it also possible to manually define how many signals will be extracted.

5.2.2.5 Multiple trials

Most experimental paradigms involve recording multiple trials for a given ROI. FISSA concatenates the traces for each trial, and applies blind source separation to the concatenated traces. This improves the separation quality, and ensures separated signals match up across trials.

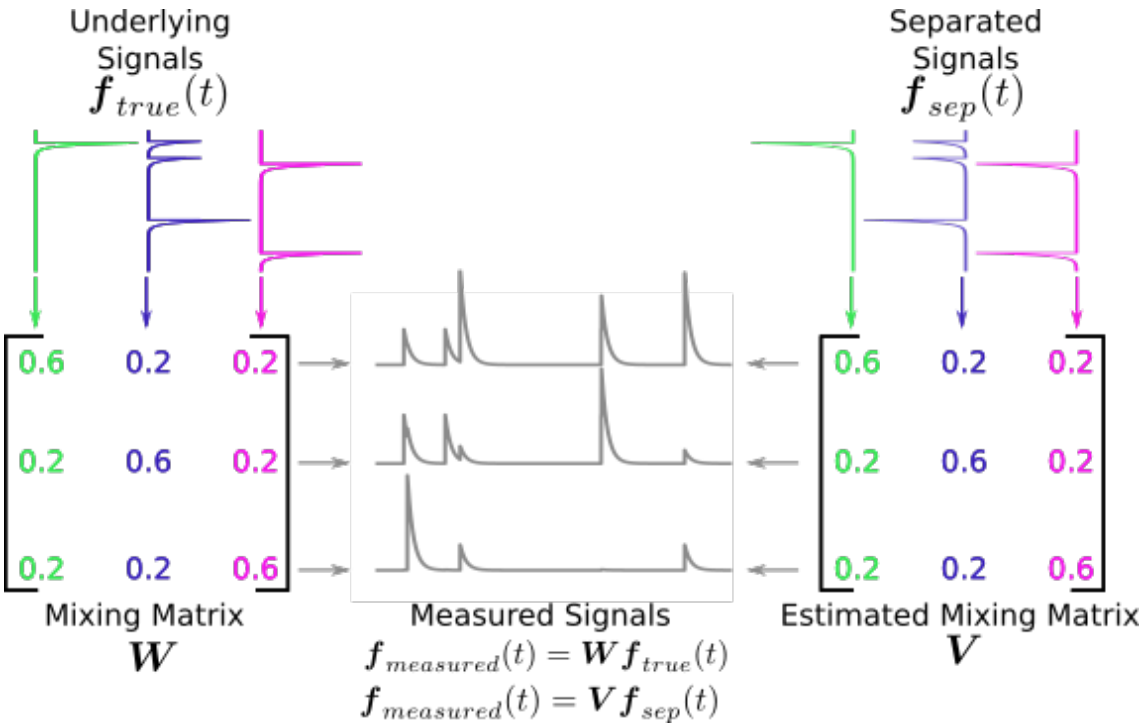

Figure 5.4: Cartoon illustrating the relationship between the true signals, measured signals, separated signals, the mixing matrices. The underlying signals are mixed by W, the mixing matrix. Each column in the mixing matrix represents how strongly a given true signal is represented across the measured signals. Each row shows how strongly each underlying signal is present in a given measured signal. The same applies to blind source separated signals, and estimated mixing matrix V.

5.3 Method comparisons

5.3.1 Experimental data

We will now compare the results of applying FISSA and subtraction on some example data. A neuropil region is defined, as described in section 2.2.1, the signal(s) of which are then either subtracted, or separated using FISSA. When a cell is particularly clear, with relatively small background signal, both subtraction and FISSA keep this signal intact, Fig. 5.5A+B. When there is a substantial neuropil signal contaminating the cell, as in Fig. 5.5C+D, both subtraction and FISSA appear to be able remove much of the background fluctuations, although subtraction occasionally removes too much (yellow curve). It is in this setting that subtraction is most often used. However, when there are particularly bright signals nearby, as in Fig. 5.5E+F, subtraction can lead to serious problems. Here, after subtracting the neuropil, the extracted signal (yellow) becomes mostly senseless due to the negative signals, whereas FISSA's extraction is more sensible.

From these examples we can tell that at least FISSA does not share the neuropil subtraction problems. But does it do a good job of extracting the true signal? Unfortunately, with our current data set we do not know the ground truth, so we next turn to simulated data to perform further empirical evaluation of FISSA.

5.3.2 Simulated data

Simulated data generation

To better estimate the efficiency of the different methods in FISSA, we generate simulated calcium imaging data. In previous work generally Poisson spikes were generated and convolved with an exponential kernel, for which simple spatial kernels are defined. Then, Gaussian noise was added to every pixel (Mukamel et al., 2009; Pnevmatikakis et al., 2015). For our purposes, this simulated data does not explicitly model neuropil contamination beyond possibly overlapping cells. We therefore expand on the simulated data protocols by including general background fluctuations and have each cell's signal spread beyond its structure, Fig.5.6. We will now describe this process in more detail.

Each neuron i's spike train, $s_i = [0,0,0,1,1,....]$, is generated by a Poisson process at a given rate (with the data simulated at 40Hz, such that each time bin covers $1/40$s. To model the effects of stimulus presentation, and induce correlations between neurons,

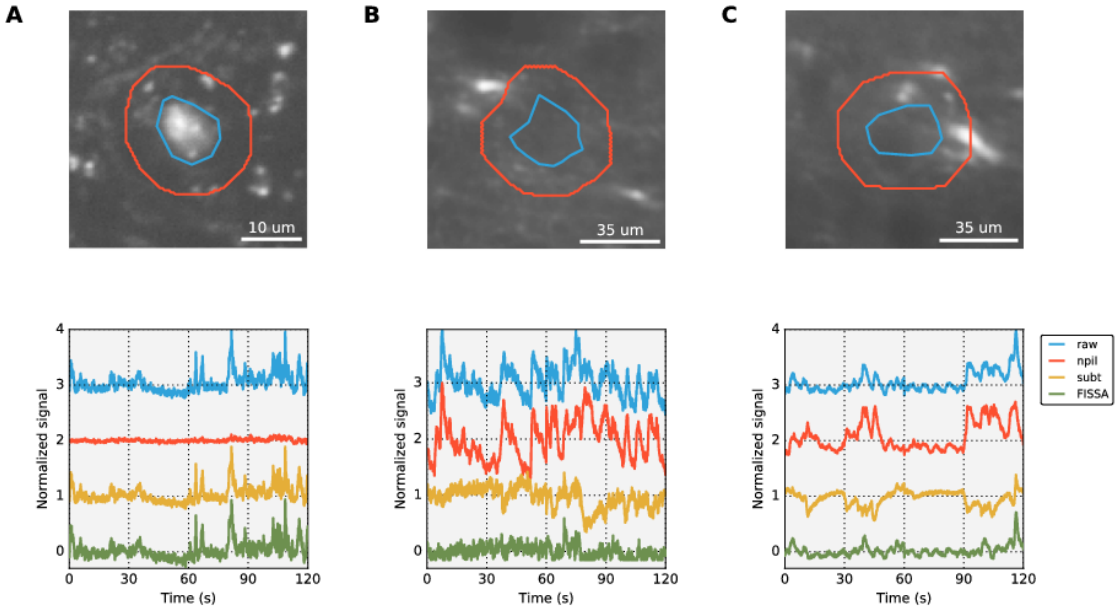

Figure 5.5: Example extractions, comparing FISSA and subtraction of the neuropil signal. (A, top) The mean image for a cell averaged across 2400 frames, with a particularly quiet background. The cell ROI outline is shown in blue, and the non-divided neuropil outline in red. (A, bottom) The extracted traces; raw (blue), the surrounding neuropil (red), with the neuropil subtracted (yellow) and the FISSA extracted signal (green). The subtraction and FISSA perform very similarly in this case. For plotting purposes, each signal first had its median subtracted, and was divided by the global maximum across the four signals. Each trace was plotted offset by 1 relative to the next trace. (B) A different cell with a strong contaminating background. Subtraction removes too much here, while FISSA seems to extract a good signal. (C) Same as A and B, but with a particularly bright signal nearby. This bright signal leads to a negative signals relative to baseline in the final signal when subtracted (yellow curve), whereas FISSA does not have this problem (green curve).

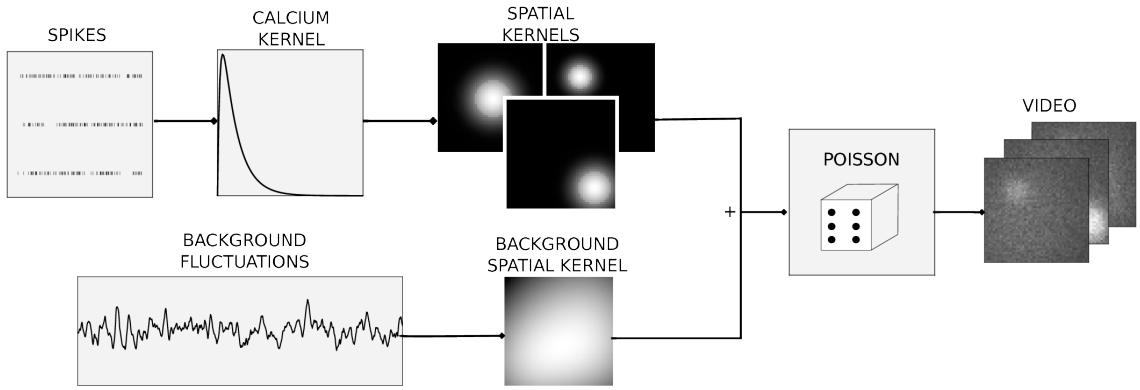

Figure 5.6: Illustrating basic simulated data setup. For each simulated neuron, a Poisson spike train is generated, which is then convolved with a difference of exponentials kernel. Next, a spatial kernel is generated consisting of a circular mask (to model the cell's structure) and Gaussian to simulate signal spread. Additionally, a slow noise process generates background fluctuations, which is convolved with a spatial background kernel. All resulting signals are summed, before being passed through a Poisson generation process to simulate photon emission, resulting in the final video.

the rates were periodically doubled for a fixed duration. To model the calcium indicator response, the spike train is convolved with a difference-of-exponential kernel

$$k(t) = a\left[\exp\left(-\frac{t}{\tau_d}\right) - \exp\left(-\frac{t}{\tau_r}\right)\right],$$

where a is a scale constant set to 0.5, and the rise and decay time constants τ_r and τ_d given by 0.05 and 0.3 seconds respectively (such that the kernel roughly matches that of the GCaMP6f indicator, as plotted in Chen et al. (2013b)). This is used to generate the underlying calcium response $f_{\texttt{true}}(t)$ for that neuron as

$$f_{\texttt{true},i}(t) = (k * s)(t).$$

Each resulting signal is then convolved with a two dimensional spatial kernel given by a 2D Gaussian plus a step function

$$K_i(x,y) = \begin{cases} \exp\left(-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma^2}\right) + c, & \texttt{if } \exp\left(-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma^2}\right) < \texttt{T} \\ \exp\left(-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma^2}\right), & \texttt{otherwise}, \end{cases}$$

where $\sigma^2 \in [25, 50]$ is the spatial spread, and μ_x, μ_y give the mean x and y positions, which were randomly drawn within the image limits (50 by 50 pixels). The offset

$c = 0.1$ models the physical structure of a cell, the threshold $T = 0.5$ models its extend, and the Gaussian models the spread of a cell's calcium signal beyond its structure to model cross contamination between nearby structures.

A background signal is generated from Gaussian noise convolved with the calcium response kernel, $k_{\mathtt{bg}}(t)$, with a longer time constant ($\tau_r = 3$s) which is subsequently added to every pixel,

$$f_{\mathtt{bg}}(t) = (k_{\mathtt{bg}} * n)(t),$$

where n is Gaussian noise. The background signal is convolved with a spatial kernel given by the sum of two Gaussians

$$K_{\mathtt{bg}}(x,y) = \exp\left(-\frac{(x-\mu_{x1})^2 + (y-\mu_{y1})^2}{2\sigma_1^2}\right) + \exp\left(-\frac{(x-\mu_{x2})^2 + (y-\mu_{y2})^2}{2\sigma_2^2}\right).$$

Finally, we have the full calcium response at a given pixel at position x, y and time t given by

$$F(x,y,t) = \sum_i K_i(x,y) f_{\mathtt{true},i}(t) + K_{\mathtt{bg}}(x,y) f_{\mathtt{bg}}(t),$$

where the sum over i is across structures. This gives the final true calcium indicator signal, which would then be observed experimentally by photon count. To simulate photon emission, F is used as the rate in a Poisson process to get generate the simulated response. Signals for each structure are estimated with its ROI mask

$$M(x,y) = \begin{cases} 1, & \mathtt{if}\ \exp(-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma^2}) < \mathtt{T} \\ 0, & \mathtt{else.} \end{cases}$$

To estimate the quality of the extracted signal, we compare the measured trace $f_{\mathtt{measured},i}(t)$ for structure i to its true trace $f_{\mathtt{true},i}(t)$ by taking their Pearson correlation.

Simulated data results

We now generate simulated data for cells similar to those in Fig. 5.5, and compare the correlation of the extracted signals to the true signals. First, we consider a single cell contaminated only by general background fluctuations, Fig. 5.7A+B. The true calcium signal is plotted by the gray curve, with clear transients for every underlying spike. Although the raw trace shows the same transients (blue curve), the neuropil

signal indicates the slow fluctuations might be background signal (red curve). Indeed, subtracting the neuropil removes much of the contamination (yellow curve), as does FISSA (green curve), both of which now match the true signal. When comparing the correlation between the extracted signals and the true signal, both subtraction and FISSA appear successful across a wide range of firing rates for the central cell, Fig. 5.7C.

Next, we add a neighboring neuron whose spatial extent overlaps with our target neuron and has a stronger calcium response, Fig. 5.7D+E. Now the raw curve has newly added transients, some of which are not from the neuron of interest (i.e. do not correspond to the true transients, gray curve). As a result the correlation curve drops down compared to the lonely cell, the blue curves in 5.7F to C. Whilst subtracting the neuropil again succeeds in removing the background fluctuations, this does not fully remove the 'faulty' transients. Accordingly, the correlation curve drops down for subtraction, yellow curve, but not as much as the raw. When using FISSA, both the background fluctuations and the contaminating spikes are removed successfully, while keeping the true transients intact. As a result, the correlation curve when using FISSA is roughly the same as for the isolated neuron.

Next we add a second nearby source, which has a high calcium response but is localized, resulting in no contamination of the central ROI, Fig. 5.7G+H. Although the neuropil signal (red) now shows many new transients, these do not show up in the raw signal (blue), and thus there is no further drop of the correlation curve, Fig. 5.7I. Subtracting the neuropil now leads to negative transients (yellow), and the corresponding correlation curve drops closer to that of raw correlation. This is avoided when using FISSA, with both the extracted example signal (green) and the correlation curve being very roughly the same to the previous two test cases.

Finally, we halve the calcium response of the target cell without changing the surrounding cells to investigate whether the signal could still be recovered, Fig. 5.7J+K. The true transients are now barely visible by eye, neither in the raw trace nor in the trace with neuropil subtracted. While FISSA is not able to fully recover the signal, it performs better than either the raw or subtraction. The correlation traces all drop down (as expected with a lower SNR), but the relative advantage of FISSA remains clear.

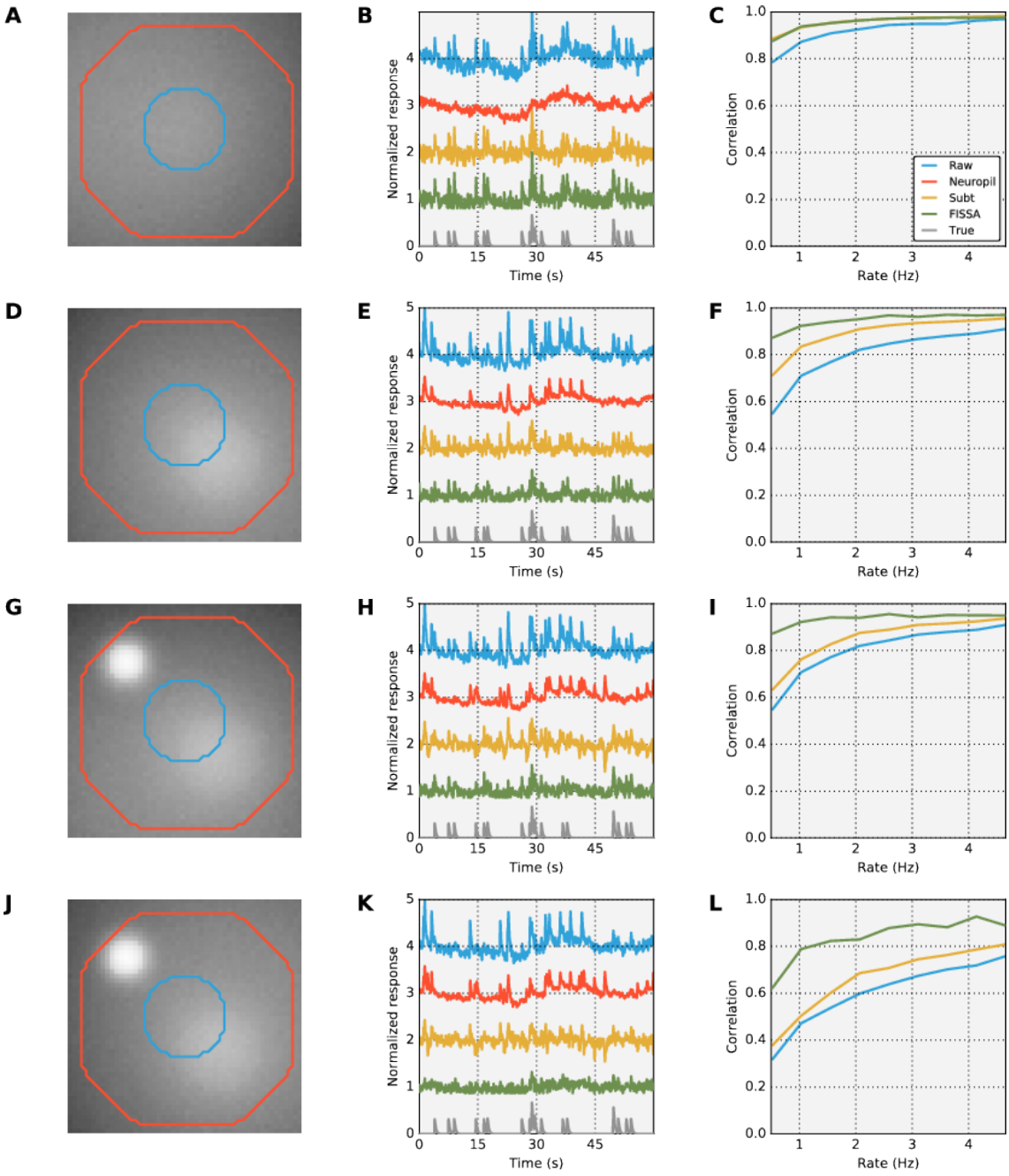

Figure 5.7: Simulated data results. Each example in the left and middle columns has the same background noise and spikes, with all firing rates set to 0.2 Hz. Simulated data was generated for 150 seconds, 60 seconds are plotted for clarity. (A) The mean image across all frames of a single cell amongst a quiet background. (B) The raw extracted signal (blue), the average surrounding neuropil signal (red), with the mean neuropil subtracted (yellow), FISSA extraction (green) and the true underlying signal (gray). For illustrative purposes, each signal has the median subtracted, and was divided by the global maximum. Signals are offset by one relative to the signal below. (C) The correlations of the extracted signals with the true underlying signals, across different firing rates for the central cell. The initial conditions were randomly drawn for each data point. (D-F) Same as A-C, but with a bright overlapping cell contaminating both the central cell and the neuropil. (G-I) Same as D-F, but with a bright localized signal contaminating the neuropil without contaminating the ROI. (J-L) Same as G-I, but with a less bright central cell, such that its response barely shows up in the raw trace (blue).

5.3.3 Pixel-based method comparison

Here we compare FISSA's extraction results to those produced by constrained NMF (cNMF) (Pnevmatikakis et al., 2015), which is one of the most recent and best performing pixel based methods. To directly compare the cNMF method to FISSA, we analyzed experimental data, focusing on a square twice the width of a ROI of a known cell (roughly 80x80 pixels depending on the ROI), over 4800 frames. Then, regarding the necessary inputs for the cNMF method, we set the number of cells to be detected to be 5, and the size of these cells to be equal to the width of the predrawn ROI; we used the default values for other parameters. From the detected cells we then handpicked which was the central detected cell, and compared its signal to the FISSA extracted signal. We then ran each algorithm 4 times, and noted the average computation time of the signal separation.

When we revisit the cell from Fig. 5.1, we see that the two methods perform similarly with FISSA removing slightly more contamination, Fig. 5.8A+B. It is of course possible that the small signals at around 60 and 100 seconds are in fact cell activity, but that would have to be compared with the underlying real signal, which in this case is not known. The two methods perform almost identically for a more active cell in terms of signal extraction, although the detected ROI is slightly larger than the actual cell (most likely due to the cell signal bleeding into the surround), Fig. 5.8D+E.

Thanks to the simplicity of the FISSA method it vastly outperforms cNMF in terms of computation time, taking only a small fraction of the time, Fig. 5.8C+F. Of course this does not take into account the time spent hand labeling the ROIs, but as noted in the introduction this is still sometimes preferable for accuracy. ROI labeling is also a one-time cost, and although both cNMF and FISSA are fairly automated, an experimenter is likely to want to rerun the analysis several times, which is far more feasible with FISSA.

5.4 Other FISSA features

Baseline detection Although not used in this chapter, in the 2-photon literature responses are most usually reported as $\frac{f-f_0}{f_0}$, where f_0 is the baseline of the signal f. We used different normalizations for plotting purposes (see figure captions), as $\frac{f-f_0}{f_0}$ lead to bad comparisons due to the negative signals in with neuropil subtraction. Nevertheless, in FISSA $\frac{f-f_0}{f_0}$ can be calculated as follows. A simple but effective baseline

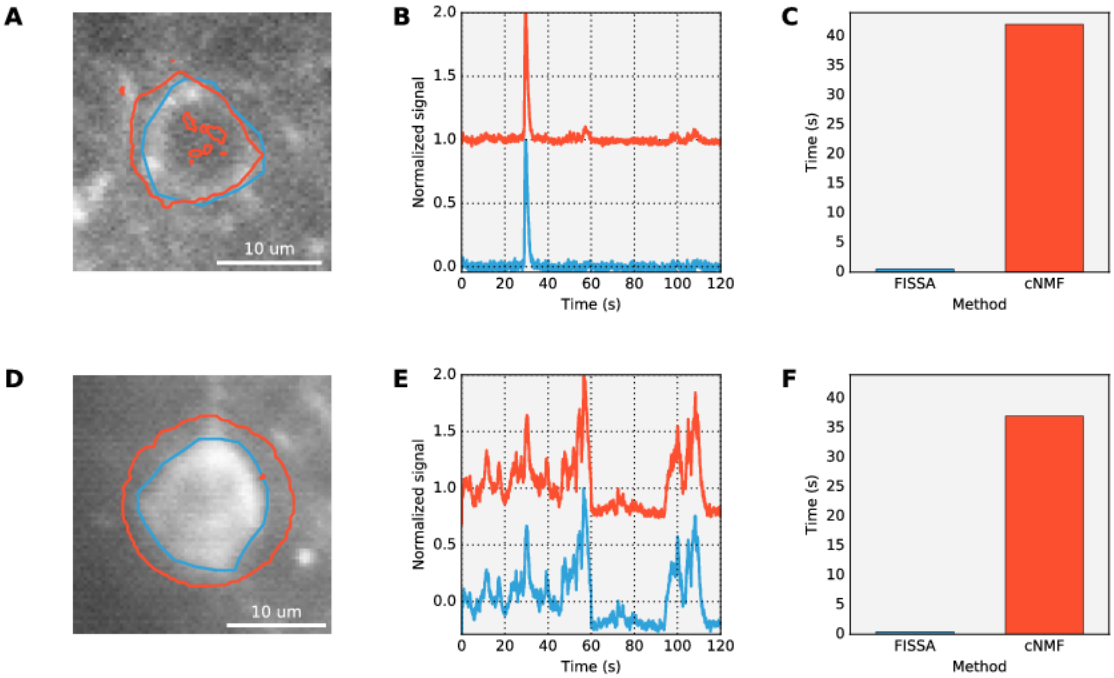

Figure 5.8: Comparing FISSA to the cNMF method (Pnevmatikakis et al., 2015). (A) The cell analyzed with both methods. In blue, the hand drawn ROI and in red the detected ROI. This is the same cell as in Fig. 5.1. The background is the recorded image averaged across 2400 frames. (B) Extracted signals obtained using FISSA (blue) and cNMF (red). Each signal had their median subtracted, and was divided by its own maximum for plotting purposes. The cNMF extracted signal was offset by 1. (C) Computation wall time for each method. FISSA on average took 0.5 seconds, and cNMF 42 seconds. (D-F) Same as A-C but for a different cell. FISSA took on average 0.45 seconds, and cNMF 37 seconds.

calculation is achieved by estimating f_0 as the 5th percentile of the 1Hz low-pass filtered trace. As the baseline often varies per trial due to changes in recording sensitivity or background activity, a baseline is calculated and subtracted per trial. Then, the average baseline across trials is used for the division step. This method was developed and implemented by S. Lowe.

Tiff data extraction The most common data format in 2-photon imaging is the TIFF image format. FISSA has fast TIFF reading scripts, using the open source `tifffile` and `Pillow` packages, and can readily be implemented in analysis workflows.

ImageJ ROIs ImageJ is an image processing software package especially useful for studying multi-frame images and defining ROIs, and is the most commonly used tool for drawing ROI. The FISSA analysis suite can optionally start from a set of ImageJ ROI files, define the neuropil regions, and extract their signals.

Plotting tools In analyzing and understanding extracted calcium signals, one often has to study both the frame-by-frame video of a cell, as well as the extracted trace over time. To aid with such analysis, the Python implementation of FISSA comes with a collection of plotting tools which extend the HoloViews toolbox (Stevens et al., 2015). These tools provide intuitive exploration of both an interactive frame-by-frame video of a ROI, as well as the different types of extracted traces (raw, separated, etc.). The plotting tools are completely optional, and FISSA can be installed and used without.

5.5 FISSA usage

FISSA can be applied to an existing dataset, and the resulting data explored, in just a few lines of code, Fig. 5.9. First, the user defines the data paths and some experiment metadata (only the frame-rate is necessary), Fig. 5.9A. After this setup, only two lines are necessary to run the full FISSA analysis pipeline: from neuropil region definition to signal separation and selection. Finally, if the optional plotting tools are installed, the extracted data can be intuitively explored through plotting HoloViews objects, Fig. 5.9B+C. The MATLAB analysis would look similar, but would not include the plotting capabilities.

Thanks to its simple implementation, FISSA can be easily slotted into existing workflows provided the ROIs to analyze are already known.

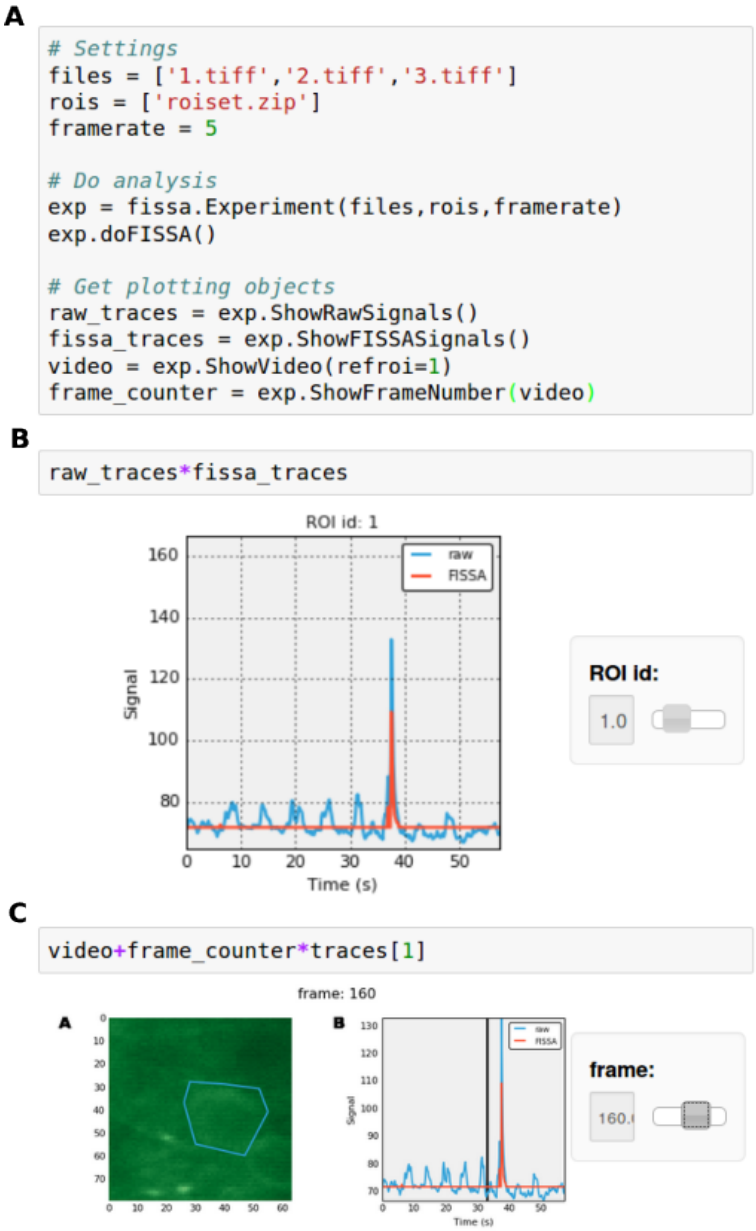

Figure 5.9: Example workflow. (A) Example code for setup, analysis and plotting objects. Here, data is defined as a list of tiff files, but one could also provide a set of Python arrays (this is to be implemented). ROIs are defined as a zip of ROIs, as made with ImageJ, but one could also provide a set of arrays (also to be implemented). (B) Exploring the traces for a given trial across ROIs. (C) Exploring the data for given ROI across time. Subfigure A here depicts a video the frames of which can be changed with the slider. The black line in subfigure B indicates what the trace looks like for the current frame. The plotting objects are generated using the HoloViews toolbox (Stevens et al., 2015), for more information on how to use and combine these objects see `https://holoviews.org/`.

5.6 Discussion

This chapter presented a fast and easy to use toolbox for signal decontamination in calcium imaging data. It was shown to provide similar accuracy to more involved pixel-based methods, while being two orders of magnitude faster, without the negative signal artifacts from neuropil subtraction. Thus, FISSA is a great option for experiments where pixel-based methods are too slow or not yet accurate enough in terms of cell detection.

Although FISSA offers increased reproducibility and less need for hand-tuning than subtraction, there are still several parameters that need to be chosen by the experimenter. First, there is a set number of neuropil regions. Whilst for our dataset four regions were adequate, dataset with a higher density of recorded neurons would be well advised to use a higher number. The other highly important parameter is the α term in the objective function for NMF. Roughly, the higher α is, the more sparse the extracted signal will be (although there is no one-to-one relationship between α and sparsity). We have found that setting α too high can lead to a higher cut-off of the extracted data, especially when there is high noise. When setting it too low (especially when $\alpha = 0$), we found that NMF performs much closer to neuropil subtraction. The effect of α is furthermore dependent on both the frame-rate and the magnitude of the signal. The signals are currently normalized by their median before separation to counteract this, but future work is needed to develop a more detailed understanding of how much tuning of these parameters is required, and when. The effect of both these parameters could be studied by using the simulated data and varying the neuron density, signal strength, frame rate, etc. Finally, pixel-based methods have similar fine tuning needs (number of sources to estimate, timescale of the signal, etc.), but it is more feasible to explore parameters in FISSA thanks to its relative speed.

Although this chapter showed several experimental data results, the ground truth was unknown. Several datasets are now available online consisting of both electrophysiological and calcium imaging data, thus providing some form of ground truth. Future work would include an analysis comparing the different extraction methods, followed by some form of spike detection, and testing the extracted spikes against those observed in the electrophysiological data.

So far, of the pixel-based methods, we have only compared FISSA to cNMF (which itself compares favorably against older methods (Pnevmatikakis et al., 2015)). Very recently another pixel-based method called Suit2P was uploaded to bioRxiv which

claims better and faster results, and also explicitly models neuropil contamination (Pachitariu et al., 2016). In future work, comparisons should be made between FISSA and Suit2P, as both make similar claims.

FISSA has been implemented and is used routinely in the Rochefort lab. The first application of this work can be seen in *Pakan et al. (2016).*

5.7 Methods

5.7.1 Data details

Our specific dataset is roughly 600 by 600 pixels (the exact values varies based on motion correction), at 40 frames per second. For the experimental details see Pakan et al. (2016).

5.7.2 Motion correction

To motion correct the imaging data we used the discrete Fourier transform method from the SIMA toolbox Kaifosh et al. (2014).

5.7.3 Software

The Python implementation is implemented in Python 2.7, using the Numpy 1.7, SciPy 0.12, Matplotlib 1.2 and HoloViews 1.6 toolboxes. We implemented ICA with the FastICA implementation in the Python scikit-learn toolbox, and the MATLAB FastICA 2.5 toolbox. We implemented NMF with the python NIMFA toolbox (Žitnik and Zupan, 2012) and the scikit-learn NMF function, and in MATLAB with the Non-Negative Matrix Factorization Toolbox (Li and Ngom, 2013).

5.7.4 Hardware

For the purpose of speed comparisons FISSA was run on a computer running Ubuntu 16.04 with an Intel® Core™ i5-4590T CPU @ 2.00GHz.

Chapter 6

General Discussion

In this thesis I attempted to explain a range of psychophysical and physiological results relating to contextual modulations. Many of these results had been explained individually or as a small set before, but not under a single unifying principle.

First, in Chapter 2, I used a smooth-contour principle (as specified by elastica theory) to describe how neurons respond to bars outside their receptive field. This model explained several psychophysical and biophysical phenomena. While the theory of elastica had been used before, it had not been used before to explain such a wide array of both psychophysical and physiological phenomena, further strengthening the case for smooth contours, specifically in the form of elastica, as an underlying principle.

Secondly, in Chapter 3, I approached a set of V1 neurons with contextual modulation (temporal or spatial) as a multivariate encoding system, with a maximum likelihood decoder. This is by itself quite novel, as most theoretical coding/decoding studies have used one dimensional encoding models, and most coding treatments of contextual modulation have used a one-dimensional decoder. We found that encoding models with contextual effects with an ML decoder are inherently biased, due to ambiguities in the coding model that get brought out by the presence of noise.

Thirdly, in Chapter 4, I updated known models for the tilt illusion and saliency effects to incorporate center dependent modulation (which is strongest whenever center and surround are aligned). Whereas the classic assumption of fixed modulation (which is strongest when the surround is aligned with a neuron's preferred orientation) is known to account for both saliency and tilt illusion effects, it turns out that center dependent modulation allows for unbiased representation of the center grating, and a stronger saliency effect. This was not known in the context of surround modulation before, and gives a possible function for center dependent modulation.

Finally, in Chapter 5, I attempted to improve signal extraction methods for 2-photon calcium imaging. This was done by developing the FISSA toolbox which used non negative matrix factorization to extract the true signal from contaminated measurements. This is both a general contribution to the field of 2-photon imaging, where accurate signal extraction is still a big problem, and the field of contextual interactions, as the more subtle effects are hard to measure without very accurate signal extraction.

Each chapter contained an extensive discussion about their individual results, so I will here discuss comparisons and conclusions that can be drawn beyond the individual chapters.

6.1 Beyond the individual chapters

The overall hypothesis was:

> The psychophysical observations can be explained from known V1 responses and contextual interactions through an encoding/decoding approach, and those contextual interactions from a single unifying principle.

Overall we can confirm that indeed, much of the psychophysical results and contextual modulations can be explained from low-level V1 responses, and simple individual modulations from contextual elements. While this is not a new conclusion as such (see also 'Comparison to previous work' below), each chapter extended our understanding of this overall hypothesis. Chapter 2 defined such a unifying principle in the form of elastica, which is a particularly elegant and well defined principle, Chapter 3 worked out the encoding/decoding problem in more detail, Chapter 4 implemented the latest physiological results, and Chapter 5 laid the groundwork for studying these effect in the mouse V1 using calcium imaging.

How do the results about center dependent modulation translate to the other chapters? In Chapter 4 we already adjusted the elastica model from Chapter 2 to be center dependent, and found that the general results held: there was no bias and the saliency effects were stronger. We also revisited the tilt illusion from using a maximum likelihood decoder from Chapter 3 in Chapter 4, and found that fixed modulation still lead to biases, and that center dependent modulation had zero bias. In Chapter 3 we showed how the bias in the fixed model comes from bimodal decoding distributions, which in turn was caused by ambiguities in the encoding model. Why did the bias disappear with center dependent modulation? Since center dependent modulation, unlike fixed modulation, does not cause the population response to shift (Fig. 4.3), there is also no

resulting encoding ambiguity, and thus no bimodality in the decoding distribution, and thus no decoding bias.

In chapters 2 to 4 the coding/decoding framework was used, often with a maximum likelihood decoder. This decoder was used as it is considered both optimal and unbiased (Kay, 1993), and is furthermore in theory implementable by a neural network (Deneve et al., 1999; Jazayeri and Movshon, 2006b). How relevant is the concept of 'optimal coding' to the brain to begin with? Certainly the maximum likelihood decoder has had some success in explaining (and predicting) psychophysics (Jazayeri and Movshon, 2006a, 2007), but one could probably also find many examples of suboptimal behavior. Even in the case where behavior might not be optimal, it is still useful to understand the upper-bound on performance. Indeed, it is only when we can put a behavior in such a perspective, that we can truly understand the underlying mechanisms and strategies. For example, the tilt illusion as explained in many previous models, relies on using a decoder which is not aware of the changes effected by contextual modulation. Chapter 3 shows that even if we use a more knowledgeable decoder biases will still persist.

6.2 Comparison to previous work

While the set of illusions and saliency effects we presented in Chapter 2 have not been combined in one model before, nor has center dependent modulation been implemented widely, several previous studies have used binding principles to explain at least a subset of psychophysical effects.

Bednar (2012) used simple plasticity rules and stimulus presentation to learn a V1 model which portrays much of the contextual modulation effects (in terms of gratings) as well as the tilt-after effect. Since it appears to have fixed surround modulation, it likely also reproduces at least the repulsive part of the tilt illusion. It is not known how well it reproduces the bar-based illusions however, nor the saliency results. Bednar's model thus reaches similar conclusions as our elastica model from a plasticity perspective, while our model gives a good account of the computations done by the visual cortex (thus giving the principle the plasticity should possibly aim to reach). Future work marrying the two models would be of interest. A problem there would be that there is no evidence of location variable modulation in their model.

Another model in which the connectivity is learned based on image statistics (specifically the relationships between different orientation/location combinations) (Coen-

Cagli et al., 2012) explains much of saliency and contextual modulation (including location dependent modulation). Although they don't explicitly test any orientation illusions with this model, the precursor to this model was explicitly used to reproduce the tilt illusion (Schwartz et al., 2009), which would suggest the new model also should be able to explain at least some orientation illusions. However, the 2012 model appears to inherently have center-dependent modulation, which our Chapter 4 would predict to lead to no illusion at all. Thus this could possibly be something missing from this model. They predict strong inhibitory effects for aligned gratings, and less inhibition for parallel gratings on the side, which does not correspond to our elastica model (but does correspond to studies on higher contrast gratings, more on that later).

Another range of studies established the saliency-map hypothesis for V1, which states that its interactions exist to aid in saliency extraction (Zhaoping, 1998, 1999, 2002). These models rely on a butterfly-like connectivity structure which depends on both the relative position and orientation, which is qualitatively quite similar to our elastica principle. However, where elastica is based on rigorous mathematical rules, the butterfly connectivity they use is built up more arbitrarily. Predictive coding is another possible principle that reproduces the results from the saliency-map hypothesis (Spratling, 2012). Both the Zhaoping and Spratling studies reach similar conclusions as us in terms of saliency effects, and show a larger scala of effects in that regard. Although, they have not tested any visual illusions themselves, early testing indicates these models has fixed contextual modulation, and at least portrays the repulsive tilt illusion. Further investigations would be necessary to find out if there are further points of overlap.

An illusion very similar to the tilt illusion also exists for color, temporal context, and motion (Clifford et al., 2000). Chapter 4 showed how the tilt illusion would disappear in this type of model, if based on orientation (and the same can be shown for the tilt after effect, Seriès et al., 2009). The important thing for the lack of illusion is the dependence of the contextual modulation on the RF stimulus and the contextual stimulus, not orientation dependence necessarily. Chapter 4 results should therefore fully extent to other modalities with similar illusions, as long as the physiological equivalent of center dependent modulation exists.

6.3 Limitations

All the models in this thesis have been highly phenomenological, and included neither biophysical mechanisms nor any dynamics. This allowed for the use of the coding decoding framework and for mathematical analysis and greater understanding of the models themselves, but in reality the systems studied are part of a dynamical system. Given the passive models that we have used, our results might not necessarily translate to dynamical models. It might be, for example, that the biophysical mechanisms determining whether surround modulation more center dependent or fixed, includes some kind of trade-off which increases bias magnitudes (this could be studied with the model for center dependent modulation presented in Shushruth et al. 2012). In the elastica model, currently the contextual modulation is determined directly from the flanking bars. However, a more realistic model would have this modulation mediated through other neurons which are ultimately influenced by the cells representing each flanking bar. This might in particular have subtle influences on the illusion results presented in Chapter 2. Future work would have to extend the models in this thesis to take all this into account.

Initial work on this was done by University of Edinburgh honours students in their final research projects, under my supervision. Susan Lechelt investigated the effect of implementing center dependent modulation in a dynamical model, (Lechelt, 2015), by implementing the model from Shushruth et al. 2012. We found that in this model, it was hard to relate the illusion strength directly to the level of center dependency, as changing this also meant changing the strength of the overall contextual modulations. Further research would be needed to expand on this, and to find out of this is a fundamental feature of the circuit, or just of this particular model. Martin Asenov investigated the impact of dynamics on the Elastica model (Asenov, 2016), by building a dynamically interacting version of the model from Chapter 2. He found that the results generally held, but that some details were slightly different (most prominently, the attractive illusion for the full center-surround disappeared). It was not clear yet however, if this was an essential result from the dynamics, or a result of parameter choice, as the attractive illusion can also be reduced in the phenomenological model in Chapter 2 by parameter choice, Fig. 2.8. Further research would be needed in both of these cases to make more specific claims.

Most of the contextual modulation modeled in this thesis was multiplicative, linear in the combination of many elements. While this allowed for easy mathematical

analysis and understanding, it does have a sever limitation; as more and more elements are added, this could potentially lead a fully inhibited response, which is not often the case experimentally. In our model investigations this was not a direct problem as we never used enough elements to do this, and furthermore, in the elastica model the influence from bars diminished as they were further away. Even with 16 surrounding bars at different distance, Fig. 2.6, the response was not close to disappearing. However, to extent this some kind of nonlinearity in the strength of the modulation after many elements might be necessary.

In all models in this thesis we have kept stimulus contrast constant. This allowed for easier mathematical treatment and interpretation, but is clearly far from realistic. Both tilt illusion (e.g. Qiu et al., 2013), and the physiological contextual interactions (Kapadia et al., 2000b) can change strongly based on stimulus contrast. These results are sometimes somewhat contradictory, such as aligned bars or gratings either facilitating or inhibiting for respectively low or high contrast. While some simple parameter changes can mimic some of the qualitative effects, Fig. 2.8 top and bottom rows, it is not clear yet if these are realistic changes. Future investigations would be needed in this.

While we partly confirmed our hypothesis that a single principle could explain low-level V1 contextual interactions and link these to psychophysical results, especially in terms of the contextual interactions there are some limitations. The basic orientation dependence of modulation arises quite naturally from elastica, but this dependence being fixed or center dependent is something set somewhat arbitrarily by the modeler.

It is unlikely that either center dependent modulation or contrast dependent changes can be reasoned out purely from the elastica principle, and thus this does add some more assumptions to the model beyond elastica.

6.4 Future work

Throughout the thesis we made several testable predictions. In Chapter 2 we found that the magnitude of the tilt illusion could depend on the ratio of neurons with center dependent versus fixed surround modulation. The tilt illusion magnitude varies strongly among subjects, so a strong test for our theory would be to show a direct link between varying illusion magnitude and the ratio of neurons center dependent versus fixed surround modulation. Since these experiments would realistically not be possible in human subjects, the first step would be to find an analogue for the tilt illusion in

animals. Reproducing visual illusions in animals would be a good goal in general, as models explaining human perceptual illusions are generally based animal neural measurements. Yet not many studies exist studying those same illusions in animals, except for a few notable exceptions have been found to persist across species (e.g. Fujita, 1997; Huang et al., 2002; Petkov et al., 2003). The tilt illusion has unfortunately not been reported outside humans, but it existing across species (and depending on neural properties of V1) would be a good test of the idea that the tilt illusion arises in V1 in the first place. While the exact tilt bias might be hard to measure, it might be possible to make a binary decision task based on the perceived orientation of the central grating.

In Chapter 4 we proposed an experiment based on overlapping random dot motion patters, each with a different motion direction. The subject's task is to estimate the difference between the two motion directions, Fig. 6.1A. The prediction would be that the perception of the direction difference is bimodal similarly to Chapter 4's results. We have started implementing the experiment, and initial self tests suggest decoding distributions similar to Fig. 3.3A, but the experiment is still in its early phases. Additionally, one could imagine an unbiased decoder, but which has takes into account a prior for a slight motion direction difference, which could also lead to a bi-modal decoding distribution. We have yet to determine a good test to distinguish distributions resulting from the maximum likelihood decoder, from a such a decoder with a prior .

The FISSA toolbox from Chapter 5 was partly developed to allow for experiments exploring subtle contextual effects using 2-photon calcium imaging. We would like to test two things in the mouse brain: presence of center dependent modulation, and association-field like experiments. The reasons for this are two-fold. First, neither of these phenomena have been reported in the mouse as of yet. Secondly, if they do indeed exist, the genetic tools available for the mouse would allow for a detailed study of the underlying mechanisms.

Most existing contextual modulation studies have been done through electrophysiology. In such experiments one identifies a neuron first, and optimizes the stimulus for this neuron (position, spatial frequency, etc.). Although with 2-photon imaging many neurons are measured at once, one does not have the luxury of stimulus optimization. We therefore propose the following protocol. Four different gratings can be shown, in different orientations and in combinations of two, Fig. 6.1A. Additionally, the size, distance between gratings, spatial frequency, etc. would be informed by their most common values in mouse V1. This should increase the number of responsive neurons, and one of the gratings would be in the RF of some cells, and function simultaneously

as the context for other cells. Early pilot experiments with this protocol have started, but the data is yet to be analyzed.

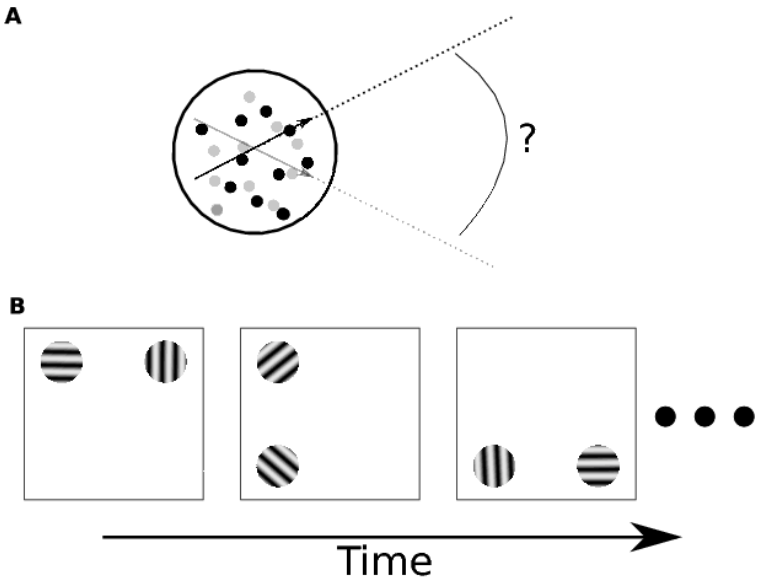

Figure 6.1: Cartoons of proposed experiments. (A) Two overlapping random dot motion patterns are briefly presented to the subject. The subject is then asked to estimate the angle between the two motion directions. The dots are only colored differently to illustrate the two groups, in actual experiments the two groups can only be identified by their motion direction. (B) Contextual modulation experiment for mouse V1. Gratings can be shown in combinations of two in four positions, and four orientations.

Bibliography

A. Adesnik, W. Bruns, H. Taniguchi, Z. Huang, and M. Scanziani. A neural circuit for spatial summation in visual cortex. *Nature*, 490:226–231, 2012.

H. Alitto and W. Usrey. Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron*, 57:135–146, 2008.

J. Alonso and Y. Chen. Receptive field. *Scholarpedia*, 4(1):5393 (doi: 10.4249/scholarpedia.5393), 2009.

M. Asenov. *Dynamic Model of Interactions between Orientation Selective Neurons in Primary Visual Cortex*. PhD thesis, University of Edinburgh, 2016.

D. Azzimonti and D. Ginsbourger. Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation. *arXiv preprint*, arXiv:1603.05031, 2016.

W. Bair, J. Cavanaugh, and J. Movshon. Time course and time-distance relationships for surround suppression in macaque v1 Neurons. *The Journal of Neuroscience*, 23 (20):7690–7701, 2003.

J. Barlow and P. Földiák. Adaptation and decorrelation in the cortex. *The computing neuron*, (ed. Durbin R. and Miall, C. and Mitchison, G.J.):54–72, 1988.

R. Bauer and S. Heinze. Contour integration in striate cortex. *Experimental Brain Research*, 147(2):145–152, November 2002.

G. Baumgartner. Indirekte Größenbestimmung der rezeptiven Felder der Retina beim Menschen mittels der Hermannschen gittertäuschung. *Pflügers Archiv für die gesamte Physiologie*, 272:21–22, 1960.

J. Bednar. Building a mechanistic model of the development and function of the primary visual cortex. *Journal of Physiology*, 106:194–211, 2012.

H. Beh, P. Wenderoth, and A. Purcell. The angular function of a rod-and-frame illusion. *Perception & Psychophysics*, 9(4):353–355, 1971.

C. Blakemore and E. Tobin. Lateral inhibition between orientation detectors in the cat's visual cortex. *Exp Brain Res*, 15:439–440, 1972.

C. Blakemore, R. H. Carpenter, and M. A. Georgeson. Lateral inhibition between orientation detectors in the human visual system. *Nature*, 228(5266):37–39, October 1970.

W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, 17(6):2112–2127, 1997.

O. J. Braddick, K. A. Wishart, and W. Curran. Directional performance in motion transparency. *Vision Res*, 42(10):1237–1248, May 2002.

A. Bruckstein and A. Netravali. On minimal energy trajectories. *Computer Vision, Graphics, and Image Processing*, 49(3):283–296, 1990.

M. Cannon and S. Fullenkamp. Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. *Vision Research*, 31(11):1985–1998, 1990.

M. Carandini and D. L. Ringach. Predictions of a recurrent model of orientation selectivity. *Vision research*, 37(21):3061–3071, 1997.

J. R. Cavanaugh, W. Bair, and J. Movshon. Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 Neurons. *Journal of Neurophysiology*, 88:2547–2556, 2002a.

J. R. Cavanaugh, W. Bair, and J. Movshon. Nature and interaction of signals from the receptive field center and surround in macaque V1 Neurons. *Journal of Neurophysiology*, 88:2530–2546, 2002b.

T. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458): 295–300, 2013b.

A. Cichocki and P. H. A. N. Anh-Huy. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 92(3):708–721, 2009.

C. W. G. Clifford, P. Wenderoth, and B. Spehar. A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society B: Biological Sciences*, 267(1454):1705–1710, September 2000.

C. Clifford. The tilt illusion: Phenomenology and functional implications. *Vision Research*, In press, 2014.

R. Coen-Cagli, P. Dayan, and O. Schwartz. Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Comput Biol*, 8(3):e1002405, March 2012.

J. Corbett, T. Handy, and J. Enns. When do we know which way is up? The time course of orientation perception. *Vision Research*, 49:28–37, 2009.

J. M. Cortes, D. Marinazzo, P. Series, M. W. Oram, T. J. Sejnowski, and M. C. W. van Rossum. The effect of neural adaptation on population coding accuracy. *J Comput Neurosci*, 32(3):387–402, June 2012.

T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley, New York, 1991.

P. Dayan and L. F. Abbott. *Theoretical Neuroscience*. MIT Press, Cambridge, MA, 2001.

S. Deneve, P. E. Latham, and A. Pouget. Reading population codes: a neural implementation of ideal observers. *Nat. Neuro.*, 2:740–745, 1999.

A. S. Ecker, P. Berens, A. S. Tolias, and M. Bethge. The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci*, 31(40):14272–14283, October 2011.

M. Edwards and J. A. Greenwood. The perception of motion transparency: A signal-to-noise limit. *Vision Research*, 45(14):1877–1884, 2005.

U. Ernst, S. Mandon, N. Schinkel-Bielefeld, S. Neitzel, A. Kreiter, and K. Pawelzik. Optimality of human contour integration. *PLoS Computational Biology*, 8(5): e1002520, 2012.

D. Field, A. Hayes, and R. Hess. Contour integration by the human visual system evidence for a local "association field". *Vision Research*, 33(2):1117–1131, 1993.

R. Freeman, I. Ohzawa, and G. Walker. Beyond the classical receptive field in the visual cortex. *Prog. Brain Res.*, 134:157–170, 2001.

W. Fries, K. Albus, and O. Creutzfeldt. Effects of interacting visual patterns on single cell responses in cat's striate cortex. *Vision Research*, 17(9):1001–1008, 1977.

K. Fujita. Perception of the ponzo illusion by rhesus monkeys, chimpanzees, and humans: Similarity and difference in the three primate species. *Perception & Psychophysics*, 59(2):284–292, 1997.

T. J. Gawne and J. M. Martin. Responses of Primate Visual Cortical V4 Neurons to Simultaneously Presented Stimuli. *J. Neurophysiol.*, 88:1128–1135, 2002.

J. Geier, L. Sera, and L. Bernath. Stopping the hermann grid illusion by simple sine distortion. *Perception*, 33(53):Supplement, 2004.

W. Geisler, J. Perry, B. Super, and D. Gallogly. Edge co-occurrence in natural images predicts contour grouping and performance. *Vision Research*, 41:711–724, 2001.

A. Genz. Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.*, 1(2):141–149, 1992.

A. Genz. MVNDST: Software for the numerical computation of multivariate normal probabilities, available from web page at http://www. sci. wsu. edu/math/faculty/genz/homepage. 1998.

A. Georgopoulos, A. Schwartz, and R. Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, September 1986a.

J. Gibson and M. Radner. Adaptation, aftereffect and contrast in the perception of tilted lines: I. quantitative studies. *Journal of Experimental Psychology*, 20:453–467, 1937.

C. Gilbert and T. Wiesel. The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, 30:1689–1701, 1990.

S. Girman, Y. Sauvé, and D. Raymond. Receptive field properties of single neurons in rat primary visual cortex. *J Neurophysiol*, 82:301–311, 1999.

B. Haider, M. Häusser, and M. Carandini. Inhibition dominates sensory responses in the awake cortex. *Nature*, 493:97–100, 2013.

T. Hansen and N. Neumann. A recurrent model of contour integration in primary visual cortex. *J Vision*, 8:1–25, 2008.

K. Harris, R. Quiroga, J. Freeman, and S. Smith. Improving data quality in neuronal population recordings. *Nature Neuroscience*, 19(9):1165–1174, 2016.

H. Hartline. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, 121:400–415, 1938.

X. Huang, S. MacEvoy, and M. Paradiso. Perception of brightness and brightness illusions in the macaque monkey. *The Journal of Neuroscience*, 22(21):9618–9625, 2002.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106–154, January 1962.

D. Hubel. *Eye, Brain, and Vision*. New York : Scientific American Library, 1995.

D. Hubel and T. Wiesel. Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, 28:229–289, 1965.

D. Hubel and T. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243, 1968.

D. Hubel and T. Wiesel. *Brain and Visual Perception: The Story of a 25-year Collaboration*. Oxford University Press, Inc., 2005.

L. Itti. Visual salience. *Scholarpedia*, 2(9):3327 (doi: 10.4249/scholarpedia.3327), 2007.

M. Jazayeri and J. A. Movshon. Optimal representation of sensory information by neural populations. *Nature neuroscience*, 9(5):690–696, 2006a.

M. Jazayeri and J. A. Movshon. Optimal representation of sensory information by neural populations. *Nature neuroscience*, 9(5):690–696, 2006b.

M. Jazayeri and J. A. Movshon. A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446(7138):912–915, 2007.

H. Jones, K. Grieve, W. Wang, and A. Sillito. Surround suppression in primate V1. *J Neurophysiol*, 86:2011–2028, 2001.

P. Kaifosh, J. D. Zaremba, N. B. Danielson, and A. Losonczy. Sima: Python software for analysis of dynamic fluorescence imaging data. *Frontiers in neuroinformatics*, 8, 2014.

M. Kapadia, M. Ito, C. Gilbert, and G. Westheimer. Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15:843–856, 1995.

M. Kapadia, G. Westheimer, and C. Gilbert. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, 84(4):2048–2062, 2000b.

S. Kastner, H. Nothdurft, and I. Pigarev. Neuronal responses to orientation and motion contrast in cat striate cortex. *Visual Neuroscience*, 16:587–600, 1999.

S. Kay. *Fundamentals of statistical signal processing: Estimation theory*. Prentice-Hall, NJ, 1993.

S. Keemink and M. van Rossum. A unified account of tilt illusions, association fields, and contour detection based on elastica. *Vision Research*, 126:164–173, 2016.

S. Keemink and M. van Rossum. Biases in multivariate neural population codes. *bioRxiv*, doi: http://dx.doi.org/10.1101/113803, 2017.

B. Kimia, I. Frankel, and A. Popescu. Euler spiral for shape completion. *International Journal of Computer Vision*, 54(1/2/3):159–182, 2003.

M. Kinoshita, C. Gilbert, and A. Das. Optical imaging of contextual interactions in V1 of the behaving monkey. *J Neurophysiol*, 102:1930–1944, 2009.

J. Knierim and van Essen D.C. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 1992.

H. Ko, S. Hofer, B. Pichler, K. Buchanan, P. Sjöström, and T. Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345): 87–91, 2011.

K. Koffka. *Principles of gestalt psychology*. Lund Humphries, London, 1935.

S. Lechelt. *Origins of biases in biological vision from network interactions*. PhD thesis, University of Edinburgh, 2015.

T. Leung and J. Malik. Contour continuity in region based image segmentation. *International Journal of Computer Vision*, 43(1):7–27, 2001.

R. Levien. The elastica: a mathematical history. Technical report UCB/EECS-2008-103. Berkeley: EECS Department, 2008.

W. Li, V. Piëch, and C. Gilbert. Contour saliency in primary visual cortex. *Neuron*, 50:951–962, 2006.

Y. Li and A. Ngom. The non-negative matrix factorization toolbox for biological data mining. *BMC Source Code for Biology and Medicine*, 8:10, 2013.

C.-J. Lin. Projected gradient methods for non-negative matrix factorization. *Neural Computation*, 19:2756–2779, 2007.

T. Lochmann and S. Deneve. Optimal cue combination predict contextual effects on sensory neural responses. *Sensory Cue Integration*, Chapter 22:393–405, 2011.

T. Lochmann, U. A. Ernst, and S. Deneve. Perceptual inference predicts contextual modulations of sensory responses. *Journal of Neuroscience*, 32(12):4179–4195, March 2012.

L. Maffei and A. Fiorentini. The unresponsive regions of visual cortical receptive fields. *Vision Research*, 16:1131–1139, 1976.

W. Marshak and R. Sekuler. Mutual repulsion between moving visual targets. *Science*, 205(4413):1399–1401, 1979.

R. Maruyama, K. Maeda, H. Moroda, I. Kato, M. Inoue, H. Miyakawa, and T. Aonishi. Detecting cells using non-negative matrix factorization on calcium imaging data. *Neural Networks*, 55:11–19, 2014.

R. Moreno-Bote, J. Beck, I. Kanitscheider, X. Pitkow, P. Latham, and A. Pouget. Information-limiting correlations. *Nature neuroscience*, 17(10):1410–1417, 2014.

E. A. Mukamel, A. Nimmerjahn, and M. J. Schnitzer. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron*, 63(6):747–760, 2009.

D. Mumford. Elastica and computer vision. In C. Bajaj, editor, *Algebraic Geometry and Its Applications*, page 491. Springer, 1994.

J. J. Nassi, S. G. Lomber, and R. T. Born. Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *Journal of Neuroscience*, 33(19): 8504–8517, 2013.

J. Nelson and B. Frost. Orientation selective inhibition from beyond the classical receptive field. *Brain Research*, 139:359–365, 1978.

H. Nothdurft, J. Gallant, and D. van Essen. Response modulation by texture surround in primate area V1: Correlates of "popout" under anesthesia. *Visual Neuroscience*, 1999.

H. Nothdurft. The conspicuousness of orientation and motion contrast. *Spatial Vision*, 7:341–363, 1993.

A. Oleksiak, M. Manko, A. Postma, I. J. M. van der Ham, A. V. van den Berg, and R. J. A. van Wezel. Distance estimation is influenced by encoding conditions. *PLoS One*, 5(3):e9918, 2010.

B. Olshausen and D. Field. How close are we to understanding and V1? *Neural Computation*, 17:1665–1699, 2005a.

A. E. Orhan and W. J. Ma. Neural population coding of multiple stimuli. *The Journal of Neuroscience*, 35(9):3825–3841, 2015.

B. O'Toole and P. Wenderoth. The tilt illusion: repulsion and attraction eeffect in the oblique meridian. *Vision Research*, 17:367–374, 1977.

M. Pachitariu, C. Stringer, S. Schröder, M. Dipoppa, L. Rossi, M. Carandini, and K. Harris. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*, http://dx.doi.org/10.1101/061507, 2016.

J. Pakan, S. Lowe, E. Dylda, S. Keemink, S. Currie, C. Coutts, and N. Rochefort. Behavioural state modulation of inhibition is context-dependent and cell-type specific in mouse V1. *Elife*, 5:e14985, 2016.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

S. Peron, T.-W. Chen, and K. Svoboda. Comprehensive imaging of cortical networks. *Current opinion in neurobiology*, 32:115–123, 2015.

C. Petkov, K. O'Connor, and M. Sutter. Illusory sound perception in macaque monkeys. *The Journal of Neuroscience*, 23(27):9155–9161, 2003.

Y. Petrov and S. McKee. The effect of spatial configuration on surround and suppression of contrast sensitivity. *Journal of Vision*, 6:224–238, 2006.

S. Pilarski and O. Pokora. On the cramér–Rao bound applicability and the role of fisher information in computational neuroscience. *Biosystems*, 136:11–22, 2015.

E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, M. Ahrens, R. Bruno, T. M. Jessell, D. S. Peterka, R. Yuste, and L. Paninski. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron*, pages 285–299, 2015.

U. Polat, K. Mizobe, M. Pettet, T. Kasamatsu, and A. Norcla. Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 391:580–584, 1998.

C. Qiu, D. Kersten, and C. A. Olman. Segmentation decreases the magnitude of the tilt illusion. *Journal of Vision*, 13(13):1–17, November 2013.

H.-J. Rauber and S. Treue. Reference repulsion when judging the direction of visual motion. *Perception*, 27(4):393–402, 1998.

H.-J. Rauber and S. Treue. Revisiting motion repulsion: evidence for a general phenomenon? *Vision research*, 39(19):3187–3196, 1999.

M. P. Sceniak, D. L. Ringach, M. J. Hawken, and R. Shapley. Contrast's effect on spatial summation by macaque V1 neurons. *Nat Neurosci*, 2(8):733–739, August 1999.

L. Schwabe, J. M. Ichida, S. Shushruth, P. Mangapathy, and A. Angelucci. Contrast-dependence of surround suppression in macaque v1: experimental testing of a recurrent network model. *Neuroimage*, 52(3):777–792, September 2010.

O. Schwartz, T. Sejnowski, and P. Dayan. A Bayesian framework for tilt perception and confidence. *Advances in Neural Information Processing Systems*, 18:2680–2718, 2006a.

O. Schwartz, T. J. Sejnowski, and P. Dayan. Perceptual organization in the tilt illusion. *Journal of Vision*, 9(4):1–20, April 2009.

O. Schwartz, A. Hsu, and P. Dayan. Space and time in visual context. *Nat Rev Neurosci*, 8(7):522–535, July 2007.

G. Sclar, J. H. R. Maunsell, and P. Lennie. Coding of image contrast in the central visual pathways of the macaque monkey. *Vis. Res.*, 30:1–10, 1990.

M. Self, J. Lorteije, J. Vangeneugden, E. van Beest, M. Grigore, C. Levelt, J. Heimel, and P. Roelfsema. Orientation-tuned surround suppression in mouse visual cortex. *the Journal of Neuroscience*, 34(28):9290–9304, 2014.

P. Seriès, A. Stocker, and E. Simoncelli. Is the homunculus "aware" of sensory adaptation? *Neural Computation*, 21:3271–3304, 2009.

P. Seriès, J. Lorenceau, and Y. Frégnac. The "silent" surround of V1 receptive fields: theory and experiments. *Journal of Physiology-Paris*, 97(4-6):453–474, July 2003.

M. Shamir and H. Sompolinsky. Implications of neuronal diversity on population coding. *Neural Comput*, 18(8):1951–1986, August 2006.

M. Shamir. Emerging principles of population coding: in search for the neural code. *Curr Opin Neurobiol*, 25:140–148, April 2014.

E. Sharon, A. Brandt, and R. Basri. Completion energies and scale. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(10):1117–1131, 1997.

C. Sherrington. *The integrative action of the nervous system*. C Scribner and Sons, New York., 1906.

S. Shushruth, P. Mangapathy, J. M. Ichida, P. C. Bressloff, L. Schwabe, and A. Angelucci. Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. *Journal of Neuroscience*, 32(1):308–321, January 2012.

S. Shushruth, L. Nurminen, M. Bijanzadeh, J. M. Ichida, S. Vanni, and A. Angelucci. Different orientation tuning of near- and far-surround suppression in macaque primary visual cortex mirrors their tuning in human perception. *Journal of Neuroscience*, 33(1):106–119, January 2013.

M. Sigman, G. Cecchi, C. Gilbert, and M. Magnasco. On a common circle: Natural scenes and gestalt rules. *PNAS*, 98(4):1935–1940, 2001.

A. Sillito and H. Jones. Context-dependent interactions and visual processing in V1. *J Physiology*, 90:205–209, 1996.

A. Sillito, K. L. Grieve, H. Jones, J. Cudeiro, and J. Davis. Visual cortical mechanisms detecting focal orientation discontinuities. *Letters to Nature*, 378:492–496, 1995.

H. P. Snippe. Parameter extraction from population codes: a critical assessment. *Neural Comp.*, 8:511–529, 1996.

H. Sompolinsky, H. Yoon, K. Kang, and M. Shamir. Population coding in neuronal systems with correlated noise. *Phys. Rev E*, 64:51904, 2002.

M. Spratling. Predictive coding as a model of the V1 saliency map hypothesis. *Neural Networks*, 26:7–28, 2012.

M. Stetter, I. Schiessl, T. Otto, F. Sengpiel, M. Hübener, T. Bonhoeffer, and K. Obermayer. Principal component analysis and blind separation of sources for optical imaging of intrinsic signals. *NeuroImage*, 11:482–490, 2000.

J. Stevens, P. Rudiger, and J. Bednar. HoloViews: Building complex visualizations easily for reproducible science. *PROC. OF THE 14th PYTHON IN SCIENCE CONF. (SCIPY 2015)*, 2015.

Q. Tang, N. Sang, and T. Zhang. Extraction of salient contours from cluttered scenes. *Pattern Recognition*, 40:3100–3109, 2007.

A. Treisman and S. Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1):15–48, 1988.

S. Treue, K. Hol, and H. J. Rauber. Seeing multiple directions of motion-physiology and psychophysics. *Nat Neurosci*, 3(3):270–276, March 2000.

A. Vaiceliunaite, S. Erisken, F. Franzen, S. Katzner, and L. Busse. Spatial integration in mouse primary visual cortex. *J Neurophsysiol*, 110:964–972, 2013b.

G. Van den Bergh, B. Zhang, L. Arckens, and Y. Chino. Receptive-field properties of V1 and V2 Neurons in mice and macaque monkeys. *The Journal of Comparative Neurology*, 518:2051–2070, 2010.

R. J. van Wezel, M. J. Lankheet, F. A. Verstraten, A. F. Marée, and W. A. van de Grind. Responses of complex cells in area 17 of the cat to bi-vectorial transparent motion. *Vision research*, 36(18):2805–2813, 1996.

von Mises. Uber die "ganzzahligkeit" der atomgewicht und verwandte fragen. *Phys. Z.*, 19:490–500, 1918.

P. Wenderoth and S. Johnston. The different mechanisms of the direct and indirect tilt illusions. *Vision Research*, 28(2):301–312, 1987.

P. Wenderoth and S. Johnston. The different mechanisms of the direct and indirect tilt illusions. *Vision Research*, 28(2):301–312, 1988.

M. Wertheimer. Untersuchungen zur Lehre von der Gestalt: II. *Psycheologische Forschung*, 4:301–350, 1923.

M. Westheimer. Simultaneous orientation contrast for lines in the human fovea. *Vision Research*, 30:1913–1921, 1990.

L. Williams and D. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9(4):837–858, 1997.

L. Williams and K. Thornber. Orientation, scale, and discontinuity as emergent properties of illusory contour shape. *Neural Computation*, 13:1683–1711, 2001.

X. Xie. Threshold behaviour of the maximum likelihood method in population decoding. *Network: Computation in Neural Systems*, 13:447–456, 2002.

R. S. Zemel, P. Dayan, and A. Pouget. Probabilistic interpretation of population codes. *Neural Comput*, 10(2):403–430, February 1998.

R. S. Zemel and P. Dayan. Distributional population codes and multiple motion models. *Advances in neural information processing systems*, pages 174–182, 1999.

K. Zhang and T. J. Sejnowski. Neuronal Tuning: to sharpen or to broaden? *Neural Comp.*, 11:75–84, 1999.

L. Zhaoping. A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10:903–940, 1998.

L. Zhaoping. Contextual influences in V1 as a basis for pop out asymmetry in visual search. *Natl. Acad. Sci. USA*, 96:10530–10535, 1999.

L. Zhaoping. A saliency map in primary visual cortex. *TRENDS in Cognitive Sciences*, 6(1):9–16, 2002.

H. Zhou, J. Zheng, and X. Yang. Euler arc splines for curve completion. *Computers & Graphics*, 36:642–650, 2012.

M. Žitnik and B. Zupan. NIMFA: a Python library for nonnegative matrix factorization. *The Journal of Machine Learning Research*, 13(1):849–853, 2012.

D. Zoccolan, D. D. Cox, and J. J. DiCarlo. Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci*, 25(36):8150–8164, September 2005.