



Contextual Modulation of Early Visual Processing and the Role of Cortico-Cortical Feedback

Citation

Trott, Alexander. 2016. Contextual Modulation of Early Visual Processing and the Role of Cortico-Cortical Feedback. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493254>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)

Contextual modulation of early visual processing and the role of cortico-cortical feedback

A dissertation presented

by

Alexander Richard Trott

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Neurobiology

Harvard University

Cambridge, Massachusetts

April 2016

©2016 Alexander Richard Trott

All rights reserved.

Contextual modulation of early visual processing and the role of cortico-cortical feedback

Abstract

How does the brain form a useful representation of its external environment from the astoundingly complex patterns of sensory signals it receives? This question has motivated the field of systems neuroscience for decades. Our current understanding stems from the discovery that sensory cortex is a hierarchical system, where the behaviorally relevant information that is only implicit in sensory inputs is made explicit through sequential stages of processing. The connections between cortical areas that create this processing sequence are abundantly reciprocal, such that information flows from lower- to higher-order areas through feedforward connections and in the reverse direction through feedback connections. Despite the anatomical abundance of feedback, its role in computation is poorly understood. Recent evidence suggests that the role of feedback relates to contextual modulation, a process by which neural responses to sensory information are shaped by the behavioral or stimulus setting in which that information arrives. A canonical example of contextual modulation is surround suppression in primary visual cortex (V1). Here, we demonstrate the richness and specificity with which the stimulus surround tailors the representations of context-embedded stimuli in V1 of awake macaque monkeys. In addition, we reversibly inactivated cortical areas V2/V3 and their feedback projections to V1 to identify a possible role of feedback in mediating these sophisticated forms of modulation. The effects of feedback inactivation were observed to differ considerably between subjects, but our overall findings suggest that feedback contributes to divisive modulation in a manner that, as shown by previous theoretical work, reflects optimal adaptation to the statistics of natural stimuli. In addition, we observed that noise correlations in V1 were restructured from day to day in a stimulus-dependent and feedback-dependent manner. We use these findings to propose future

experiments that will allow us to better understand the diversity of inactivation effects across subjects as well as how the precise network-level signatures of contextual modulation depend on feedback. We expect that these details will provide critical insights towards an understanding of the role of cortico-cortical feedback in sensory processing and cortical computation.

Table of contents

Abstract	iii
List of figures	viii
Acknowledgments	ix
1 Introduction	1
1.1 Information pathways of the cortical hierarchy	1
1.2 Theoretical perspectives on feedback	4
1.2.1 Instructive processing	4
1.2.2 Efficient coding	7
1.3 Inactivation of feedback	10
1.4 Contextual modulation of early visual processing	12
1.5 Summary of results	16
2 Input-gain control produces feature-specific surround suppression . . .	19
2.1 Introduction	20
2.2 Experimental Procedure	22
2.2.1 Multiunit activity recordings	22
2.2.2 Experimental design	22
2.2.3 Center/Surround stimuli	23
2.2.4 Plaid/Surround stimuli	23
2.2.5 Inclusion criteria	24
2.2.6 Analysis	25
2.2.7 MUA fitting	25
2.2.8 Model parameter fitting	26
2.3 Results	26
2.3.1 Surround suppression <i>scales</i> population responses to single-grating center stimuli	27

2.3.2	Surround suppression <i>shifts</i> population responses to plaid center stimuli . . .	33
2.3.3	Suppression as feature-selective input-gain	38
2.3.4	Eye movement controls	40
2.4	Discussion	42
3	Effects of feedback inactivation on feature-specific surround suppression in V1	47
3.1	Reversible feedback inactivation by cortical cooling	50
3.1.1	Surgical preparation	50
3.1.2	Cryoloops	50
3.1.3	Design of an inactivation experiment	51
3.1.4	Epileptiform activity during feedback inactivation	52
3.2	Experimental procedures	58
3.2.0.1	Monkey R: measuring match suppression	59
3.2.0.2	Inclusion criteria (center-only/center+surround data)	60
3.2.0.3	Monkey R: measuring feature-specific suppression	61
3.2.0.4	Inclusion criteria (plaid-only/plaid+surround data)	62
3.2.1	Experimental procedure: Monkey V	63
3.2.2	Analysis	64
3.3	Monkey R: Effects of feedback inactivation on match suppression	66
3.4	Monkey R: Effects of feedback inactivation on feature-specific suppression	73
3.5	Monkey V: Effects of inactivation on match suppression	78
3.6	Monkey V: Effects of feedback inactivation on feature-specific suppression.	87
3.7	Monkey V: Configuration-specific divisive modulation as a general explanation. . .	93
3.8	Discussion	97
4	Feedback-dependent correlation structure during contextual modulation	105
4.1	Experimental procedures	108
4.1.1	Receptive field localization	108
4.1.2	Measurement of noise correlations	110
4.1.3	Using stimulus location to characterize correlations	112
4.2	Monkey R	114

4.3 Visualizing stimulus-dependent correlations and effects of inactivation	115
4.4 Quantifying stimulus-dependent correlations and effects of inactivation	121
4.5 Discussion	128
References	146

Listing of figures

2.1	Tuning of surround suppression is sensitive to stimulus context.	28
2.1	(Continued)	29
2.2	Scaling of population responses.	31
2.3	Component-specific suppression with plaid center stimuli.	35
2.4	Population responses predicted by input gain model.	41
3.1	Experimental setup.	51
3.2	Epileptiform activity observed during feedback inactivation.	54
3.2	(Continued)	55
3.3	Firing rate changes during control, inactivation, and re-warming: Monkey R.	67
3.4	Time course of average responses measured in the center-only/center+surround data set using a Gaussian sliding window: Monkey R.	70
3.4	(Continued)	71
3.5	Suppression of responses by matching surrounds across three response epochs: Monkey R.	72
3.6	Population responses to plaid-only and plaid+surround stimuli during control (top) and inactivation (bottom): Monkey R.	77
3.7	Best-fitting weights applied to the population responses to plaid stimuli in monkey R and the effect of inactivation.	78
3.8	Effects of the surround and feedback inactivation on basic response dynamics of individual units in monkey V.	80
3.9	Firing rate changes during control, inactivation, and re-warming: Monkey V.	82
3.10	Time course of average responses measured in the center-only/center+surround data set using a Gaussian sliding window: Monkey V.	84
3.10	(Continued)	85
3.11	Suppression of responses by matching surrounds across three response epochs: Monkey V.	87

3.12 Population responses to plaid-only and plaid+surround stimuli during control and inactivation: Monkey V.	89
3.13 Best-fitting weights applied to the population responses to plaid stimuli in monkey V and the effect of inactivation.	90
3.14 Non-specific changes in component weights fitting plaid+surround data after feedback inactivation.	91
3.15 Plaid+surround population responses fit with the scaled-weights modification.	93
3.16 Monkey V: Configuration-dependent responses during the sustained period of activity for center+surround and plaid+surround data.	95
3.17 Monkey V: Time course of average responses measured across both surround modulation data sets using a Gaussian sliding window.	96
3.18 Responses, surround modulations, and inactivation effects observed with the center-only/center+surround data in each monkey.	101
4.1 Estimates of receptive fields.	111
4.2 Total distance as a metric for organizing and describing correlations in terms of RF and stimulus location.	114
4.3 Control correlation matrices visualized with the default electrode ordering.	116
4.4 Control correlation matrices visualized with the sorting determined by total distance.	117
4.5 Sorted center-only correlation matrices.	118
4.6 Sorted center+surround correlation matrices.	118
4.7 Averages of the sorted correlation matrices across the 8 days	119
4.8 Averages of the correlation matrices across the 8 days for iso-oriented and orthogonal center+surround stimuli.	120
4.9 The relationship between variables used to describe electrode pairs and correlations.	123
4.10 Noise correlations reflected in the interactions between total distance, geometric mean, and Fano factor.	124
4.11 Examining the relationship between geometric mean, Fano factor, and noise correlations along the range of total distances.	127

Acknowledgments

The thanks that are deserved cannot be given briefly. There are many people reflected in this work and the endurance that has made it possible.

First, I want to thank my adviser, Rick, without whom this work never would have been possible. Rick, it's been an adventure. I want to thank you for the years of encouragement, motivation, honesty, and, when I needed it, space that have made me feel up to the challenges of being a scientist and, even more daunting, a graduate student. I only wish I had more frequently availed myself of your many wisdoms, which you communicated strictly in the form of hilarious idioms. Very Yoda-esque. That's high praise.

To Till, my "post-doc." Till, I do not know how you put up with being the person I brought all of my technical questions to, but I never could have done it without you. You made grad school at times easier and at others much harder, and I mean both in the best way. If I could ever convince you of something, that's when I knew I was on the right track. (I'm not sure that I ever did.) Thanks, Till. Sharing our corner of the office sure was... *cool*!

No lab would be complete without its token Canadian, and my experience of grad school would have been so much less rich if it did not feature Camille. Camille, you are one of the absolute nicest people I have ever met. Your ability to consider so many points of view while still adding your own is something I look up to tremendously. There were many times when I sought to emulate your

thoughtful and exact approach to science, and I think I am far the better for it. Camille, I mean it when I say “Sorry.” (Which I believe is Canadian for “Thanks.”)

And one mustn’t forget the wise, battle-hardened Vladimir, keeper of all of life’s secrets. People often throw around the advice to be true to yourself, but I’m guessing no one ever felt the need to remind Vladimir. Thank you, Vladimir, for being a completely reliable, honest, encouraging, and endlessly helpful part of my time in the lab. You add such character to the lab and I hope you know how valued it is.

I also owe a huge debt of gratitude to the many people outside the Born lab that have contributed in countless ways. I want to thank my DAC, Marge Livingstone, Chris Harvey, and Gabriel Kreiman, for years of sanity checks and good advice. I want to also thank the brave souls who have agreed to serve on the exam committee for this dissertation, John Assad, Mark Andermann, Gabriel Kreiman, and David Sheinberg. If this work meets with your approval, I consider that a true mark of success. Friends have been an essential part of making this whole experience so rewarding and manageable. To all my fellow classmates, if you think that maybe, just maybe, I want to thank you, I do. I’d like to especially thank Mr. Ari Morcos, my partner in crime throughout it all. Wherever life takes us, Ari, you can’t imagine the positive influence that you’ve had on me through our years of being lunch buddies and great friends.

And what success story would be complete without the fearless, *amazing* single mother character? Certainly not mine. Mom, I’ve never once questioned the support I could find in you. Whether we managed to talk every week or every month when things got busy, you were always the foundation of my support structure and I can’t thank you enough for that. I am who I am because of you, and I am very proud of that.

If there is one person who I would like to pay special tribute to, it is my loving partner, Leah Kandel. As it just so happens, today is our ten year anniversary. Leah, if I ever convince you to read this document, it will bring me some joy to know that you will see this section and know what a difference you have made throughout it all. You spark joy, creativity, passion, and pure determination in every conversation we have. You are my role model in countless ways and my best friend. Thank you.

Chapter 1

Introduction

Constructing a behaviorally relevant and useful representation of the external world given the signals received from the sensory periphery is a difficult challenge. The generally fluid, robust, and unambiguous experience of conscious perception belies the abundant ambiguities in the sensory signals from which it is created. Understanding how the cerebral cortex approaches these challenges (and, frequently, overcomes them) is a fundamental goal of systems neuroscience. The last few decades have seen many advances towards this understanding as research continues to reveal the computational principals of hierarchical systems. However, accounting for the function(s) of the many feedback pathways in the brain has been a significant hurdle. The overarching goal of this thesis project has been to address this gap by studying the role of cortico-cortical feedback in early sensory cortex, with an emphasis on understanding its role in contextual modulation.

1.1 Information pathways of the cortical hierarchy

External stimuli are received by the sensory periphery. Signals from the periphery are routed to the brain, where they are relayed through a series of processing stages specific to the relevant modality. These processing stages refer to the distinct cortical areas involved in analyzing and representing the

information from the outside world. Distinct cortical areas can be distinguished from one another by their functional differences. In early visual cortex, for example, separate areas each represent a retinotopic “map” of the visual world in terms of a particular class of features (Albus & Beckmann, 1980; Shipp et al., 1995; Blasdel & Campbell, 2001; Schuett et al., 2002). Functional borders such as these correspond to a more general indicator of cortical area boundaries: axonal projection patterns.

Axonal terminations between cortical areas usually follow one of two laminar patterns, either originating in superficial layers and targeting layer IV or originating in (primarily) deeper layers and targeting layers I, II/III, V, and VI, though not always to the same extent (Rockland & Pandya, 1979). Importantly, connections between cortical areas are typically reciprocal (Rockland & Pandya, 1981; Felleman & Van Essen, 1991; Markov et al., 2014), but differ in projection pattern depending on the direction of connectivity (Rockland & Pandya, 1979). These termination patterns allow one to order cortical areas into a self-consistent hierarchy (Felleman & Van Essen, 1991) in which (feedforward) connections from lower- to higher-order areas originate in superficial layers and target layer IV and (feedback) connections from higher- to lower-order areas originate in deeper layers and target layers I, II/III, V, and VI.

A hierarchical organization is consistent with observations that cells in progressively higher areas extract increasingly more sophisticated stimulus features by building off of the features encoded by their antecedent inputs (Nassi & Callaway, 2009). This notion dates back to the pioneering work of Hubel and Wiesel, who first suggested that center-surround cells are combined by simple cells to create phase-sensitive orientation filters (Hubel & Wiesel, 1959) and that simple cell outputs are combined by complex cells so as to create phase-insensitive orientation filters (Hubel

& Wiesel, 1962). In fact, this framework remains applicable well beyond early visual cortex: for example, when considering the way representations of faces and object identity seem to be built and maintained towards the “apex” of the ventral visual pathway (Li & DiCarlo, 2008; Freiwald & Tsao, 2010).

An examination of the features represented along the visual hierarchy suggests that the goal of the visual cortex is to make explicit the behaviorally relevant information contained within a scene that is only implicit in sensory inputs. Given this, a hierarchical framework accounts well for the observed patterns of feedforward connectivity between cortical regions. Interestingly, computational models of vision trained towards a similar goal (recognizing the abstract contents of an image in a manner invariant to their low-level presentation) can perform impressively well when implemented as a purely feedforward hierarchy (Krizhevsky et al., 2012; Szegedy et al., 2013, 2015). The success of these computational models underscores the power of feedforward processing in extracting and representing complex, abstract features of the incoming information.

However, the brain is not composed of only feedforward connections. With few exceptions, connections between cortical areas are abundantly reciprocal (Felleman & Van Essen, 1991; Markov et al., 2014), such that feedback represents a massive anatomical investment. From this standpoint, it seems very likely that feedback plays a crucial role in cortical information processing. But, while the hierarchical organization discussed above is useful for understanding the difference between feedforward and feedback connections and even provides a compelling account of feedforward processing, it largely fails to identify the challenges of sensory processing that feedback helps to solve. As such, the defining question concerning feedback is that of its computational role. How could feedback allow the brain to better handle the complexities of the

sensory world? There are many feasible answers to this question (Mumford, 1992) but a dearth of conclusive empirical evidence resolving the possibilities. In the following sections, I will briefly review some theoretical perspectives on the role of feedback and then cover experimental findings.

1.2 Theoretical perspectives on feedback

There are two broad perspectives concerning the function of feedback that I will focus on. The first, ‘instructive processing,’ explains feedback as a mechanism through which lower areas exploit potentially disambiguating information that only becomes available at higher areas; the second, ‘efficient coding’ (Barlow, 1959), explains feedback as a means to maximize the efficiency of information representation/transfer within and between cortical areas. The critical difference between the two comes from their predictions regarding the conditions under which feedback should excite and suppress its targets. According to the instructive processing framework, a given representation in a higher area should, through feedback, bias a lower area towards the same representation. According to the efficient coding framework, representations in higher areas should suppress representations of the same information in lower areas.

1.2.1 Instructive processing

Attention studies provide some evidence for instructive processing. The effect of spatial attention on orientation tuning in V4 is well explained as an increase in the response gain of the neurons whose receptive fields cover the attended location (McAdams & Maunsell, 1999). The effects of feature attention on firing rates in MT were also well explained by gain changes; however, the magnitude and sign of the gain change were no longer fixed but instead depended on the attended

feature (Martinez-Trujillo & Treue, 2004). These findings are important in the context of the feature-similarity gain model (Treue & Martínez Trujillo, 1999) and when considering the possibility that spatial and feature attention need not use separate mechanisms. Such ideas are consistent with the notion that the degree of neuronal gain modulation caused by attention depends on the match between the neurons' preferred features (including location, direction-of-motion, etc.) and the features captured by the attentional state (Maunsell & Treue, 2006). In short, gain is increased when the preferred and attended features agree and decreased when they disagree.

While attention is undeniably a top-down process, it reflects the outcome of a high-level behavior (specifically, the behavior that determines how to allocate attention). As such, it has an unclear relationship with the form of top-down processing implemented through the projection patterns that we used previously to define feedback. For example, the way in which a given state of V2 affects the processing of information in V1 via feedback may be quite distinct from the way in which volitional allocation of top-down resources does so. Therefore, when discussing feedback it is useful to be aware of the computational strategies captured by attention, but it is also important to note that these insights may not generalize to all instances of feedback processing.

Some of these difficulties in interpretation are avoided by considering studies that measure choice-related activity during psychophysical tasks that do not explicitly direct attention. A consistent finding amongst these studies is that neurons exhibit positive “choice probability,” a metric describing the extent to which trial-to-trial fluctuations in neural activity track the reported percept of a noisy stimulus during a psychophysical task (Britten et al., 1992; Dodd et al., 2001; Williams et al., 2003; Nienborg & Cumming, 2006). Traditionally, choice probability has been

given a causal interpretation—that the fluctuations actually influence perception (Shadlen et al., 1996)—but more recent work has challenged this notion by providing a top-down explanation (Nienborg & Cumming, 2009; Wimmer et al., 2015; Cumming & Nienborg, 2016). In addition, Cohen and Newsome (2008) observe that such fluctuations are correlated between neighboring neurons in a manner depending on the demands of the task. Recent theoretical work (Haefner et al., 2014) offers a compelling explanation of this body of findings in terms of feedback. The key insight is that learned understanding of the task structure enables inference through an ‘internal model’ of the noisy stimulus and the generative process from which it comes. This inference process integrates prior knowledge with sensory information and feedback mediates this integration.

The work of Haefner and colleagues (2014) serves as a particularly useful example model for exploring an instructive role of feedback. Their model is primarily constructed around a behavioral task that requires a subject to report the orientation of a Gabor that is corrupted by visual noise, where the underlying orientation takes on one of two values. The experimenter-controlled generation of each stimulus is hierarchical, beginning with the highest-level description (underlying orientation) and ending with the noisy stimulus. The job of the subject is to “invert” the generative process by observing the stimulus and inferring the highest-level, task-relevant parameter (underlying orientation). Correctly inferring the task-relevant parameter requires inferring all the parameters describing the generative process; therefore, the subject must perform hierarchical inference. This amounts to identifying the most probable set of parameters that generated each image.

In this model setting, there is a natural analogy between the stages of hierarchical inference and

the processing at each stage of the cortical hierarchy, where neural activations at each level of the hierarchy represent the inferred parameters at each level of the generative process. Insofar as this analogy reflects the computations performed by cortex, the role of both feedforward and feedback connections must be to ensure that the entire set of generative parameters represented across the cortical areas is of high probability. To meet this goal, feedback connections should bias the activity of neurons in lower-order areas towards the activity patterns that would be predicted given the activity in the higher area. The details of that prediction reflect the brain's 'internal model' of the external, generative process. Therefore, one advantage of this feedback mechanism is that it will promote representations that are maximally consistent with the external variables that produced the sensory data (assuming the internal model is accurate).

Such a framework for understanding feedback is consistent with the general notion that feedback is instructive. The elaboration of feature selectivity achieved by a hierarchical processing strategy allows the extraction of abstract information; however, this also means that much information only becomes available at higher stages. In the case where such information is useful to lower areas, feedback potentially provides a mechanism to coordinate the representation of various levels of information across cortical areas.

1.2.2 Efficient coding

In contrast to the first framework, the second posits that higher representations of a given stimulus feature should suppress matching (i.e. predictable) representations in lower areas. The most direct evidence for this notion comes from a pair of studies examining both the architecture and function of feedback in early visual areas. Feedback projections from V1 cells to the lateral geniculate nucleus (LGN) tend to terminate either parallel or perpendicular to the elongated on- and

off-regions that define their orientation selectivity (Murphy et al., 1999). Concerning those with parallel terminations, the functional alignment of V1 feedback effects and LGN input zones is inverted. That is, feedback from V1 cells onto LGN on-regions (regions that, when activated, excite V1 cells) tended to suppress the on-regions; the opposite being true for terminations onto off-regions (Wang et al., 2006).

The results of the above pair of studies are exceptional, as most evidence for this framework tends to be indirect. However, the overall body of relevant work demonstrates some interesting trends in its support. Some of the most compelling examples are found in the patterns of surround suppression/modulation in V1, especially with respect to its orientation tuning. Surround modulation refers to the ability of a stimulus that is otherwise ineffective at driving a cell to modulate its response to a separate stimulus that would be effective alone. Whether a given stimulus drives the cell or modulates it depends, by definition, on whether it falls in the cell's classical receptive field (CRF) or non-classical receptive field (nCRF)—also termed the surround—respectively. As revealed by the tuning of many V1 cells to optimally oriented stimuli of varying sizes, surround stimulation tends to be suppressive (Sceniak et al., 1999, 2001; Cavanaugh et al., 2002a). Interestingly, however, suppression evoked by such stimuli tends to decrease as the surround orientation is varied while keeping the center orientation at the optimal orientation (Jones et al., 2002). Even more interesting is the fact that this pattern does not derive from a fixed suppressive surround orientation; instead, surround suppression is maximal when the nCRF and CRF stimuli contain the same orientation (Sillito et al., 1995). Moreover, the effectiveness of this pattern of suppression depends on the CRF and nCRF stimuli being coaxially aligned in the case where the surround stimulus is not annular (Cavanaugh et al., 2002b).

This complex sensitivity of surround tuning speaks to the potential involvement of feedback, as a possible determinant of the suppression strength seems to be the extent to which the overall stimulus configuration drives orientation-tuned areas beyond V1. Consider the functional connectivity from V1 to LGN discussed above. A LGN cell may be more suppressed when its center stimulus and its surround form a contour because this is what drives V1. In the same way, a V1 cell may be more suppressed when its center and surround stimuli are coaxial and iso-oriented because this is what would drive higher, orientation-selective areas with larger receptive fields (such as V2, V3, or V4). This viewpoint would seem to agree with the results of a mechanistic model that describes the role of feedback in shaping surround suppression (Schwabe et al., 2006).

Two intriguing theoretical studies examine coding schemes designed to leverage the inherent statistics of the sensory world in order to make information coding maximally efficient (Rao & Ballard, 1999; Schwartz & Simoncelli, 2001). The models are capable of predicting how various stimuli yield previously observed patterns of surround modulation, as well as the functional connectivity from V1 to LGN described above (Jehee et al., 2006). How then might this coding system be implemented in the brain? While the answer remains unclear, there is evidence suggesting that interactions among neighboring neurons are a likely vehicle for surround modulation (Sillito et al., 1995; Ozeki et al., 2009; Hashemi-Nezhad & Lyon, 2012; Shushruth et al., 2012). It is intriguing to speculate that the main target of feedback may be interneuronal processing (Gilbert et al., 2009; Ramalingam et al., 2013).

In this framework, feedback is thought to enhance cortical information processing by making responses of lower areas sparser, less correlated, and generally more informative. Though it is presently unclear whether such effects are mediated by feedback, horizontal connections, or both,

all three of those effects occur as a result of increasing stimulation size (from one to several times the CRF) when presenting V1 neurons small patches of natural movies (Vinje & Gallant, 2000, 2002; Haider et al., 2010). This type of coding strategy may also be useful for the biased representation of the most informative components of a scene, such as feature discontinuities introduced by object borders. Indeed, early visual cortex is highly sensitive to discontinuities across a variety of stimulus features (Lamme, 1995; Jones et al., 2002; Lee, 2003; Shen et al., 2007). This sensitivity may play a role in figure-ground segregation and perceptual grouping, processes proposed to depend on feedback (Roelfsema, 2006; van Kerkoerle et al., 2014).

1.3 Inactivation of feedback

None of the studies described above directly manipulated cortico-cortical feedback. While their results are essential for a complete understanding of the topic, any clues that they provide about the role of feedback are a matter of interpretation. Perhaps the most direct method of examining the role of feedback in cortical processing comes from directly manipulating its availability, and, over the past few decades, numerous studies have attempted exactly this. While the basic experimental approach has remained fairly consistent, the specifics—such as inactivation method, areas of inactivation and recording, and model organism—differ amongst groups. To a large degree so do their results.

The majority of feedback inactivation studies demonstrate predominantly decreased responsiveness to CRF-confined stimuli (Sandell & Schiller, 1982; Bullier et al., 1996; Hupé et al., 1998, 2001; Bardy et al., 2009), full-field stimuli (Galuske et al., 2002; Schmidt et al., 2011), or both (Wang et al., 2010). However, mixed increased and decreased (Alonso et al.,

1993a) and predominantly increased (Alonso et al., 1993b; Martinez-Conde et al., 1999) responses have also been observed. In many of these studies, significant effects were observed only rarely. All of the above studies examined possible functional roles of feedback, but the combined results are equivocal. For example, studies disagree on whether the influence of feedback is contrast dependent (Hupé et al., 1998; Wang et al., 2010). Furthermore, even within the same author groups there is evidence both supporting (Bullier et al., 1996; Bardy et al., 2009) and refuting (Hupé et al., 2001) a role of feedback in surround modulation.

While the current body of inactivation studies effectively demonstrates that feedback exerts an appreciable effect on cortical processing, their collective results are not easily synthesized into a coherent description of what feedback is doing. One potential reason for this is that all of the studies described in this section were done using anesthetized animals. It is commonly suggested that feedback acts as a modulator rather than a driver of activity (Zipser et al., 1996; De Pasquale & Sherman, 2013; Zagha et al., 2013). A prominent electrophysiological and computational study has identified the importance of the conductance state from background synaptic inputs in the control of neuronal gain modulation (Chance et al., 2002). While the mechanisms of anesthesia are unknown, it has clearly been shown to exert an effect on the state of network conductance in vivo (Destexhe et al., 2003). As such, it is very unclear exactly how the various types of anesthesia used in the above studies contribute to the factors that may determine the impact of feedback signals and/or their removal. To illustrate this concern, the finding that anesthesia disrupts the representation of figure-ground borders (Lamme & Spekreijse, 1998) casts significant doubt on the finding rejecting an involvement of V2 to V1 feedback in surround modulation of static textures (Hupé et al., 2001).

Recently, our lab tested the effects of reversible inactivation of V2/V3 (and their feedback

projections) on surround modulation of V1 responses in awake monkeys. In contrast to results from anesthetized monkeys, inactivation effects were seen to be common for V1 neurons. Furthermore, the average effect of inactivation was to reduce the suppressive influence of stimuli that were larger than the CRF (Nassi et al., 2013). The effects of inactivation were clearly visible for both low- and high-contrast stimuli, in contrast to the trend observed in previous studies for effects to be more common at low contrasts. Therefore, the results of Nassi et al. (2013) contradict the suggestion developed by studies in anesthetized animals that feedback provides a weak, excitatory input that modulates low-salience, CRF-confined stimuli. Instead, their results identify feedback as providing a predominantly suppressive influence that acts as a source of contextual modulation.

1.4 Contextual modulation of early visual processing

Interpreting the results of Nassi et al. (2013) through the lens of the efficient coding framework would seem relatively straightforward: large, homogenous gratings are easily represented by higher-order areas such as V2/V3 and feedback suppresses the distributed representation of these stimuli in V1 to favor their more compact representation in V2/V3, allowing V1 to instead represent the components of the stimulus that higher-order areas do not explain (Mumford, 1992; Rao & Ballard, 1999). Could this viewpoint similarly make some sense of the observation that suppression from the surround generally depends on its feature similarity with the center stimulus (Sillito et al., 1995; Shen et al., 2007; Shushruth et al., 2009)? It would seem that it could, if we assume that the suppressive nature of large, feature-homogenous stimuli stems from their tendency to activate V2/V3.

However, this assumption neglects the issue of feature-specific surround modulation and potentially introduces a subtle but important misinterpretation regarding the influence of feedback. The hypothetical viewpoint from above treats suppression as a relatively non-specific influence whose magnitude depends on some set of conditions (in this case, the appropriateness of the stimulus for recruiting feedback). An alternative explanation is that suppressive signals are highly feature-specific such that the reason the response to a center stimulus is not suppressed by, say, an orthogonal surround is that suppression from said surround targets features that are not contained within the center. In this case, suppression could be appropriately described as a form of effective input-gain control. This notion was previously only addressed in theory (Spratling, 2008, 2010), but we provide experimental confirmation that suppression from the surround is coordinated in this feature-specific manner (Trott & Born, 2015).

As such, the natural question to follow the findings of Nassi et al. (2013) is whether feedback from V2/V3 orchestrates this form of suppression—or, more generally, context-dependent input-gain control. Achieving an answer to this question (expressed more broadly as how stimulus context shapes responses in early visual cortex and the extent to which feedback mediates these effects) forms the motivation for this thesis project as a whole.

Addressing the possible role of feedback in directing input-gain control is critical not only because it would add to our understanding of feedback but also because input gain offers a potential bridge between the two theoretical perspectives laid out in Section 1.2. In addition to the fact that attention has been shown to interact with surround suppression (Bair et al., 2003; Roberts et al., 2007; Sundberg et al., 2009), many parallels can be drawn between the form of input-gain control produced through surround suppression and the manner in which attention modulates

neuronal responses to one or more stimuli (McAdams & Maunsell, 1999; Reynolds et al., 1999; Reynolds & Desimone, 2003; Lee & Maunsell, 2009, 2010; Ni et al., 2012). As such, establishing a more concrete understanding of the forms of contextual modulation that feedback controls is an attractive avenue towards the goal of developing a unified framework for thinking about top-down processing.

On the other hand, only a subset of theoretical work invokes input-gain control as the mechanism underlying surround modulation. In fact, a large body of theoretical work, which also frames surround modulation as a manifestation of efficient coding, draws explanatory power by treating divisive normalization (Carandini & Heeger, 1994; Carandini et al., 1997; Busse et al., 2009) as the underlying mechanism. These studies capture efficient coding in that they describe strategies that the brain could use to minimize avoidable redundancies in the outputs of neurons whose response properties are modeled after those in V1. Importantly, these redundancies stem from the complex dependencies produced by the structure of natural visual stimuli. Modeling approaches that seek to find compact linear encodings of natural images that maximize statistical independence yield receptive fields that resemble those of V1 neurons (Olshausen & Field, 1996), but retain characteristic redundancies. One solution is to adapt normalization to the dependencies produced by natural images; model outputs produced by this normalization scheme replicate some of the contextual effects observed in V1 and, when adapted to natural sounds, similarly replicate those observed in auditory cortex (Schwartz & Simoncelli, 2001). In this approach, normalization weights between model neurons are tuned so as to maximize independence when presented with natural stimuli. Continuations of this work introduced the idea that statistical dependencies between neighboring image regions are themselves variable (Schwartz et al., 2006)

and models that include an inference step that first assigns the input image to the appropriate normalization pool capture some of the impressive flexibility of surround modulation in V1 (Coen-Cagli et al., 2009, 2012) and provide among the current best fits to responses of V1 neurons presented with natural images (Coen-Cagli et al., 2015).

Intriguingly, the inference procedure modeled in these studies is inherently hierarchical, such that the highest representation provides some compact description of how the activations of the units at the lower layer are distributed (Karklin & Lewicki, 2005; Schwartz et al., 2006; Coen-Cagli et al., 2009, 2012, 2015). The idea that visual neurons encode the higher-order moments of their inputs can explain complex cells in V1 (Karklin & Lewicki, 2008) and provides a useful model for activity in V2 (Freeman & Simoncelli, 2011), raising the possibility that V2 can represent patterns of activity distributions in V1 and, through feedback, flexibly gate its normalization accordingly (Coen-Cagli et al., 2015).

The empirical characterization of surround modulation as a form of input-gain control contradicts, in a subtle but important way, the theoretical models just described. That is, these models describe the surround using normalization, which constitutes a form of output-gain control. Therefore, the feature-specificity that defines input-gain control is very much not realized in the current state-of-the-art model that relates natural image statistics to surround modulation. It is imaginable that future instantiations of this model will resolve this contradiction, but, for the present work, we considered the hypothesis that feedback flexibly gates normalization signals as separate from the hypothesis that feedback produces input-gain control through feature-specific suppression. Indeed, this separation served as part of the motivation for studying whether feedback contributes to the feature-specificity of suppression. We set out to answer this question with the hopes that it would

not only help to establish a more exact picture of the computational repertoire underlying contextual modulation but also would present a description of these computations in terms of the cortical hierarchy.

The volume of theoretical perspectives and speculative interpretations regarding the role of feedback far outweighs the amount of available empirical evidence. At this point, how to adjudicate between these theories depends critically on questions that we must answer experimentally. Here, we choose to address the question of whether the feature-specificity that characterizes surround modulation in V1 is a function of feedback from V2/V3.

1.5 Summary of results

Chapter 2 covers the work reported in Trott and Born (2015). This work provides experimental validation of the proposal that surround modulation in V1 acts as a form of effective input-gain control through feature-specific suppression. To demonstrate this, we recorded V1 responses to stimuli composed of CRF-confined center stimuli either with or without an oriented, annular surround grating. The effect of the surround appeared to differ between cases when the center stimulus contained either a single grating or when the center was composed of a superposition of two independently oriented gratings; however, we showed that both types of effect could be explained with a single mechanism: feature-specific surround modulation. This work identified a previously unappreciated level of specificity to modulation and suggests a fundamental mechanism of contextual modulation, whether context is defined in terms of behavioral demands, stimulus settings, or both.

In Chapter 3, I introduce the method used throughout the remainder of the project to interrogate

the role of feedback: specifically, reversible inactivation of areas V2/V3 while recording multiunit activity (MUA) in V1 of awake macaques using multi-electrode arrays (MEAs). Even though our method of reversible inactivation (cortical cooling) has been successfully employed for a long time, we are the first group, to our knowledge, to combine this technology with MEA recording. We find that, to our surprise, this combination can produce epileptiform activity under the recording array during inactivation. Fortunately, this problem diminished with time from array implantation, allowing a limited data set to be collected from each of two monkeys. While the epileptiform activity was an obstacle to the experiments analyzing center-surround interactions, the fact that it could be reliably produced by feedback inactivation (observed in three monkeys) is also revealing about the nature of feedback.

The main focus of Chapter 3 is the effect of feedback inactivation on the feature-specificity of surround modulation in V1. We were able to collect a limited dataset in each of two monkeys. These data replicate the previously published findings of diminished surround suppression following inactivation (Nassi et al., 2013), but show that inactivation primarily leaves the specificity of the surround intact.

In the remainder of my dissertation, I present two other interesting findings we observed within the inactivation effects. While interesting, they differ between the two monkeys that provided inactivation data, limiting the conclusions we can hope to draw but allowing room for useful speculation. The first such finding concerns the effect of inactivation on the tendency for surround suppression to track the similarity between the center and surround stimuli. In one monkey (R), we report evidence that feedback inactivation weakens this tendency. This effect was not present for the other monkey (V), for whom inactivation made the behavior of the

surround, if anything, *closer* to what is reported in the literature. Our second interesting but uncorroborated finding was that the correlated fluctuations in the MUA of monkey V were structured in a stimulus-dependent manner (Chapter 4). These (noise) correlations tracked the center of the stimulus, even when a surround was present (such that the whole array was stimulated). As a result, pairs of multiunits showed highest correlations when they were nearby one another and nearby the stimulus center. This latter dependency was diminished with feedback inactivation.

This work identifies feature-specificity as a key mechanism of contextual modulation, adding to its already impressive flexibility in shaping responses to complex stimuli. Moreover, this work hones our understanding of feedback by showing that its role is likely distinct from input-gain control. While we still cannot claim a concrete understanding of the computations performed by feedback, the steps taken here serve to narrow the scope of possibilities, bringing us closer to an answer. In particular, the data from monkey V provide a clue that feedback modulates the influence of the surround in a feature-*nonspecific* but stimulus-dependent manner, offering tentative support for its role in flexible normalization. Future work exploring the role of feedback in contextual modulation through the use of natural stimuli holds particular promise. Continued improvements in the quality and stability of methods to record populations of neurons while manipulating feedback will allow us to probe this function in increasingly naturalistic settings. Hopefully, the work presented here will stand as an important precursor to such efforts.

Chapter 2

Input-gain control produces feature-specific surround suppression

In primary visual cortex (V1), neuronal responses are sensitive to context. For example, responses to stimuli presented within the receptive field (RF) center are often suppressed by stimuli within the RF surround, and this suppression tends to be strongest when the center and surround stimuli match. We sought to identify the mechanism that gives rise to these properties of surround modulation. To do so, we exploited the stability of implanted multi-electrode arrays to record from neurons in V1 of alert monkeys with multiple stimulus sets that more exhaustively probed center-surround interactions. We first replicated previous results concerning center-surround similarity using gratings representing all combinations of center and surround orientation. With this stimulus set, the surround simply scaled population responses to the center, such that the overall population tuning curve had the same shape and peak response. However, when the center contained two superimposed gratings (i.e. a visual “plaid”), one component of which always matched the surround orientation, suppression selectively affected the portion of the response driven by the matching center component, thereby producing shifts in the peak of the population orientation tuning curve. In effect, the surround caused neurons to respond

predominantly to the component grating of the center plaid that was unmatched to the surround grating, as if by reducing the effective strength of whichever stimulus attributes were matched to the surround. These results provide key physiological support for theoretical models that propose feature-specific, input-gain control as the mechanism underlying surround suppression.

2.1 Introduction

Neuronal responses to stimuli confined to the receptive field (RF) center are modulated by the simultaneous presentation of stimuli in the RF surround (McIlwain, 1964; Hubel & Wiesel, 1965; Allman et al., 1985). In primary visual cortex (V1), one common manifestation of such spatial modulation is a decreased neuronal response to stimuli extending beyond the RF center (Sceniak et al., 1999, 2001; Cavanaugh et al., 2002a), a phenomenon generically referred to as “surround suppression.” This operation is thought to enhance the efficiency of visual information processing by reducing the redundancy inherent in natural images (Barlow, 1959; Mumford, 1992; Rao & Ballard, 1999; Vinje & Gallant, 2000; Schwartz & Simoncelli, 2001; Vinje & Gallant, 2002; Haider et al., 2010).

The surround, like the center, exhibits selectivity for orientation (Bair et al., 2003; Ozeki et al., 2009; Hashemi-Nezhad & Lyon, 2012), but, interestingly, the surround tuning does not appear to be fixed. Instead, suppression is usually strongest when the center and surround stimuli match (Sillito et al., 1995; Cavanaugh et al., 2002b; Jones et al., 2002; Shushruth et al., 2012). We replicated these results by recording V1 responses to gratings representing all combinations of center and surround orientation. We show that, under these conditions, the surround effectively scaled population responses to the center according to its orientation similarity to the surround.

We further sought to identify the type of mechanism that gives rise to this behavior of the surround. One possibility suggested by the above result is that the surround simply modulates neural outputs according to the similarity between the center and the surround. Alternatively, recent theoretical studies have proposed that the mechanism of surround suppression is a form of input-gain control (Spratling, 2010; Lochmann & Denève, 2011; Spratling, 2011; Lochmann et al., 2012). This mechanism produces efficient coding by allowing the surround to suppress predictable inputs, and unifies a number of observations regarding surround modulation (Lochmann et al., 2012).

In order to evaluate these different possibilities, we measured the effect of the surround on center stimuli composed of visual plaids. The orientations of each center component varied independently but the orientation of the surround grating was matched to one of the center components. We found that suppression from the surround selectively affected the portion of the responses driven by the matched center component, such that the peak of the population tuning curve shifted towards the orientation of the unmatched center component, consistent with the predictions of input-gain control (Lochmann et al., 2012). These findings demonstrate that the surround is capable of modulating the representation of central stimuli in a highly selective way to better represent informative stimulus features at the expense of spatially redundant ones.

We offer a simple, quantitative model to account for our results, showing that the full range of surround behaviors observed by us and others can be explained by an input-gain control mechanism in which suppression selectively targets the effective strength of central features that match the surround.

2.2 Experimental Procedure

2.2.1 Multiunit activity recordings

Two male macaque monkeys (*Macaca mulatta*; P and R) were each implanted with a 10x10 electrode array (400 μm spacing) in the right hemisphere of V1 as well as a head post on the anterior portion of the skull. Monkeys were trained to fixate a 1.5° window for approximately 3 s while stimuli were presented. Completed fixations resulted in a liquid reward. Multiunit activity (MUA) was sampled using a Cerebus 128-channel system (Blackrock, Utah). At the start of each recording day, activity thresholds were set for each electrode as -3.6 RMS background noise; MUA events were logged as threshold crossings.

2.2.2 Experimental design

All stimuli used for the present experiments consisted of stationary, sinusoidal gratings and were presented on a CRT monitor (100 Hz refresh rate), using mean luminance as background. Stimuli included a central grating/plaid, an annular surround, or both. To simplify pooling the data, the diameter of the central stimulus and the inner and outer diameters of the annular surround were fixed across sessions (at 0.3°, 0.6° and 2.0°, respectively), as was the spatial frequency of the gratings (at 3.33 Hz). Orientation was sampled at a spacing of 30°. For each session, the stimulus location was approximately aligned to the RF center(s) of one or more multiunit sites. Whether a given site was included for analysis depended on how well its RF boundaries aligned with the stimulus boundaries, as assessed using the criteria described below. Once selected, experimental sessions consisted of 3 distinct portions. First, surround-only activity was recorded by measuring responses to all orientations of the full-contrast (100% Michelson contrast) surround with the center contrast

set to 0%. Second, we recorded responses to the plaid-only/plaid+surround stimulus set (2.2.4). Lastly, we recorded responses to the center-only/center+surround stimulus set (2.2.3). Within each portion, stimuli were presented in a random block interleaved design. For each portion, the goal was to repeat each stimulus 20 times. Occasionally, this goal was not met for the final portion; data were included provided that at least 10 blocks had been completed. A trial began when the monkey achieved fixation; 300-500 ms after fixation began the stimulus appeared and remained on screen for another 500 ms, after which the trial ended. If the monkey broke fixation before the end of the trial, the trial was aborted without reward.

2.2.3 Center/Surround stimuli

For the center-only/center+surround portion of the experiments, we measured responses to all combinations of center and surround orientations, presented in pseudorandom order. The center was always presented at 100% contrast. In half the trials, the center was presented in isolation (giving center-only data) and, in the other half, a full-contrast surround was included (giving center+surround data).

2.2.4 Plaid/Surround stimuli

For the plaid-only/plaid+surround portion, we presented the central stimulus as a “plaid,” drawn by summing two component gratings (termed Center_1 and Center_2), each at 50% contrast. We measured responses to all $\text{Center}_1/\text{Center}_2$ orientation combinations. As with the center-only/center+surround stimuli, half of the trials presented the center in isolation (giving plaid-only data). The other half of trials included a 100% contrast surround (giving plaid+surround data) with the added manipulation that the surround orientation was always set

equal to that of Center₂.

2.2.5 Inclusion criteria

We wanted to restrict our analysis to data from multiunits whose RFs overlapped the central stimulus and only minimally overlapped the annular surround. To that end, we applied a set of inclusion criteria to ensure the interpretability of our analyses. For the plaid-only/plaid+surround data set, inclusion required that: 1) surround-only responses showed no significant tuning to surround orientation; 2) responses to plaid-only stimuli significantly exceeded spontaneous activity; 3) responses to plaid-only stimuli significantly exceeded responses to plaid+surround stimuli; 4) evoked activity was significantly tuned to center orientation (assessed using plaid-only trials where component orientations were identical). For the center-only/center+surround data set, multiunits that met all 4 plaid-only/plaid+surround criteria were included, provided that center-only responses were also significantly tuned to the orientation of the center. Significance of tuning was tested using a permutation test (1000 iterations, $\alpha = 0.05$). This yielded a total of 97 and 61 multiunits for the plaid-only/plaid+surround and center-only/center+surround data sets, respectively. However, some sites met all criteria on more than one session. To prevent any potential miscounting of data, we determined that the most conservative approach was to only allow each site to enter the analysis once. Therefore, if a given site met criteria for multiple sessions its data was only included for the session in which its RF most overlapped the center. We used the ratio of the average center-only (plaid-only) to the average center+surround (plaid+surround) activity as a benchmark for this overlap. (Because the sites included for the plaid-only/plaid+surround and center-only/center+surround experiments only partly overlapped, we performed this final exclusion step separately for the two sets of data). These steps resulted in

plaid-only/plaid+surround and center-only/center+surround sample sizes of 71 and 47 multiunit sites, respectively. The qualitative outcome of our analyses remained the same if this final exclusion step was omitted.

2.2.6 Analysis

Responses were analyzed over the final 350 ms of stimulus presentation (chosen to include only the sustained portion of the response). Before data were pooled, spontaneous activity was first subtracted from responses, which were then normalized to their grand mean response to all stimuli in the no-surround conditions. Where applicable, stimuli orientations are reported relative to the preferred orientations of the measured multiunits. Preferred orientation was calculated as the vector average of the responses to the no-surround stimuli (for the plaid-only data, specifically the conditions where the orientations of Center_1 and Center_2 were identical). For simplicity, preferred orientations were binned before analysis (bin width: 30°).

2.2.7 MUA fitting

Population responses (2.3.1 and 2.3.2) were fit according to the equations and approaches described in the Results. The underlying templates used in these fits were themselves acquired by fitting the population response with a circular Gaussian (two free parameters: width and height). Importantly, these template fits were acquired before attempting to fit the remaining data and were therefore not optimized for the entire data set. Where different models were compared, the significance associated with the improved fit of the full model was determined using Sequential F-testing. Any data used to fit the underlying templates were excluded from such significance calculations.

2.2.8 Model parameter fitting

Our entire data set (the center-only, center+surround, plaid-only, and plaid+surround data) was fit to a single model with 5 free parameters (Equations 2.3.3 and 2.3.3). During fitting, these parameters were chosen to minimize the squared adjusted error between the observed population responses and those produced by the model. The adjusted error refers to an additional step where the error on each data point was normalized by the grand mean of the data set it belonged to. (For example, the errors in the fit to the plaid+surround data were normalized by the grand mean of the actual plaid+surround data.) We observed that the standard least squared error fitting approach (where this normalization does not occur) underfit the center+surround and plaid+surround data, likely because some model parameters only applied to the half of the data where a surround was present and this half of the data contained less variance. Normalizing the error as we described ameliorated this bias and produced better qualitative fits.

2.3 Results

We measured MUA in primary visual cortex (V1) of two awake male macaque monkeys during a fixation task involving the presentation of stimuli designed to probe center/surround interactions (see Methods). Responses were recorded using chronically implanted 10 x 10 microelectrode arrays (MEAs). Our experimental procedure involved two distinct stimulus sets (see 2.2): data collected using a single, full-contrast grating in the center are referred to as ‘center-only’ and ‘center+surround’; data collected using two superimposed half-contrast gratings are referred to as ‘plaid-only’ and ‘plaid+surround’. For each stimulus set, a given electrode was only allowed to enter analysis on at most one session, yielding sample sizes of 47 multiunit sites for the

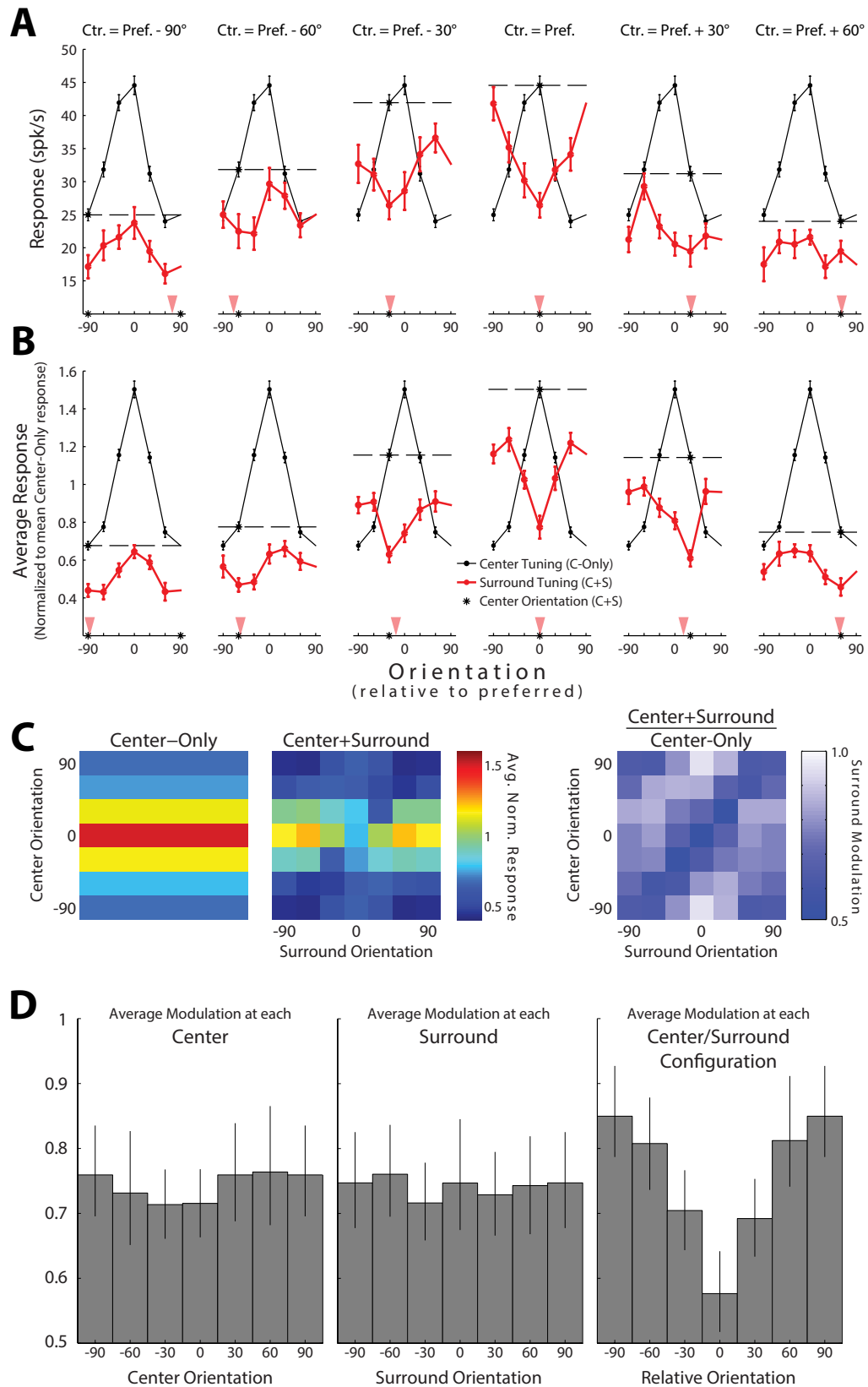
center-only/center+surround data and 71 for the plaid-only/plaid+surround data.

2.3.1 Surround suppression *scales* population responses to single-grating center stimuli

Previous studies characterizing center/surround interactions in V1 have shown that surround suppression is greatest when surround features (i.e. orientation, direction, color) match those of the center (Sillito et al., 1995; Zipser et al., 1996; Cavanaugh et al., 2002b; Shen et al., 2007). With the goal of identifying the surround mechanism underlying this form of context-dependent response modulation, we first sought to replicate these findings by measuring V1 responses to stationary sinusoidal gratings representing all combinations of center and surround orientation. Consistent with previous reports, we found that tuned surround suppression was a common property of the multiunits recorded: in the subset of trials with the center at the preferred orientation, suppression was significantly tuned to surround orientation ($p < 0.05$, permutation test) in a majority of the analyzed units (25/47) and the average suppression from the preferred surround (46%) significantly exceeded the average suppression from the anti-preferred surround (19%). In addition, the orientation tuning of the surround depended on the orientation contained in the center. This trend was often visible in the responses of individual multiunits (Figure 2.1A) and was clearly visible in the combined population data (Figure 2.1B). Each of these plots shows the tuning to the orientation of the center when no surround was present (plotted in black) as well as the tuning to the orientation of the surround (plotted in red) measured when the center was held at a particular orientation. The six subplots show the surround tuning measured for each of the fixed center orientations, which is indicated by the title of the plot and by the black asterisk along the abscissa. For each center condition, the red triangle indicates the maximally suppressive surround orientation, which clearly tracks the orientation in the center.

Figure 2.1 (following page) | Tuning of surround suppression is sensitive to stimulus context. **A**, tuning of an example MU to center orientation with no surround (black curve) and to surround orientation (red curves). Data points are represented as mean $\pm$ SEM. The data plotted in black indicate the orientation tuning of the example MU to center orientation. These data serve as a visual reference and are identical across the six subplots. The data plotted in red indicate the orientation tuning of the MU to the orientation of the surround (with the center orientation fixed at a particular orientation). Each plot depicts the surround tuning measured using a specific center orientation (indicated by the title and again as a black asterisk along the abscissa). The dashed line indicates the response to the relevant center when presented without a surround. Red triangles indicate the orientation of the maximally suppressive surround, calculated from the mean vector of the difference between the center-only response (dashed line) and the surround tuning. **B**, same as A, for population-averaged data. **C**, (left) same data in B, represented as response maps; (right) surround modulation map obtained by dividing (element-wise) the center+surround response map by the center-only response map. Smaller values indicate greater suppression. **D**, collapsed representations of the modulation map, showing the average relationships between surround modulation and center (left), surround (middle), and relative center/surround orientations (right). 95% confidence intervals (black) were obtained using bootstrapping.

Figure 2.1 | (Continued)



We compared responses to center-only stimuli and responses to center+surround stimuli after aligning responses by preferred orientation and pooling across units (Figure 2.1C; $n=47$; see 2.2 for details). From there, we estimated the condition-by-condition effect of the surround by normalizing the response to each center+surround stimulus by the response to the center-only stimulus with the same center (Figure 2.1C). Collapsing the resulting suppression profile across surround orientations failed to reveal any systematic relationship between center orientation and degree of modulation and the same was true for surround orientation when collapsing across center orientations (Figure 2.1D). Instead, the pattern of modulation was almost entirely determined by the relative orientation (that is, the absolute orientation difference) between the center and surround (Figure 2.1D, right). Therefore, while the surround may have exhibited orientation tuning at a given center, the full set of center+surround responses suggested that suppression was not concentrated toward any particular orientation of the surround itself (nor of the center itself); rather, the key variable for describing suppression at the population level appeared to be the degree of orientation similarity between the center and surround (i.e. the relative center/surround configuration).

We re-represented the combined center-only/center+surround data as a series of population responses under each of the relative center/surround configurations (Figure 2.2). This illustrates how the presence of a surround affected the representation of center orientation, demonstrating that, for the center-only/center+surround data set, the addition of a surround resulted in a scaling of the center-only population response (black curves). To confirm this intuition, we fit the center+surround population response in each configuration (red curves) with a scaled version of the center-only population response, which itself was fit as a circular Gaussian centered on 0° (the

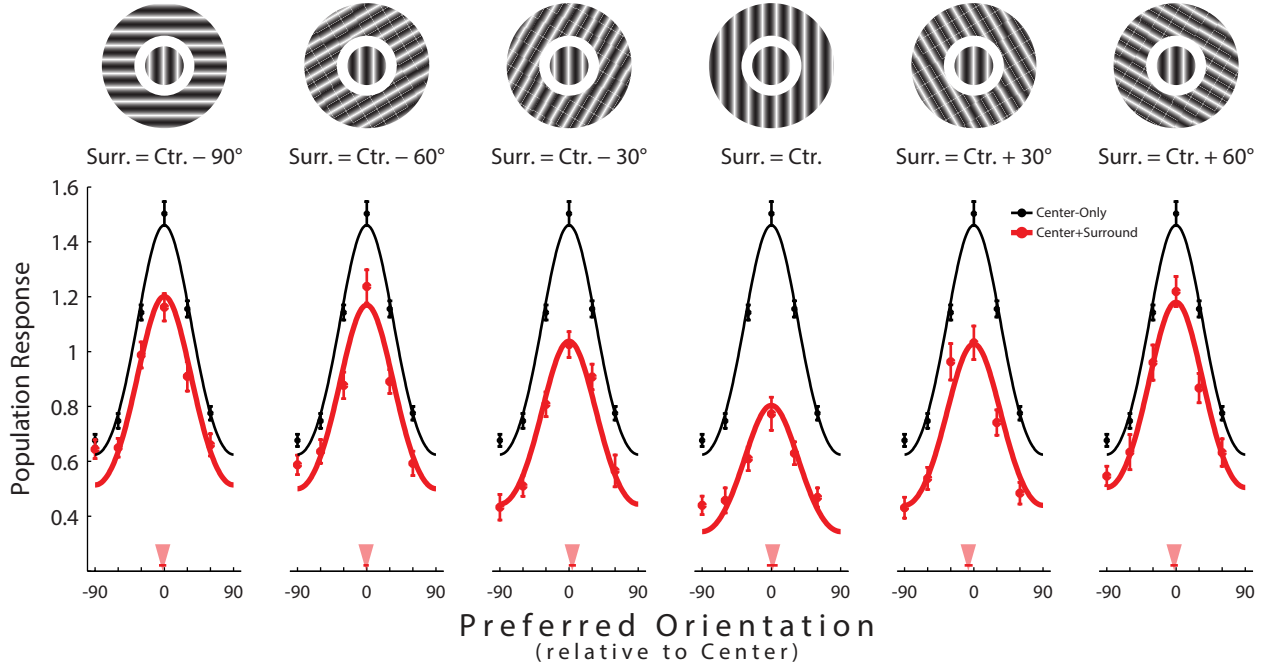


Figure 2.2 | Scaling of population responses. Population responses under each of the 6 relative center/surround configurations (a stimulus exemplifying the relevant configuration is illustrated above each subplot). Data points are represented as mean $\pm$ SEM. Smooth black curves depict the circular Gaussian fit to the center-only population response. Smooth red curves represent the center-only Gaussian, scaled to best fit the center+surround data depicted in red. Red triangles point to the orientations of the mean vectors calculated from the center+surround population responses. The thin red lines superimposed on the triangles indicate the 95% confidence intervals obtained from the bootstrapping procedure described in the main text. Stimuli are shown for illustration and are not to scale.

preferred orientation). As such, the center+surround population responses were expressed as:

$$R_{\theta} = w_{\theta} \cdot T \quad (2.1)$$

where R_{θ} is the center+surround population response for relative center/surround configuration θ , w_{θ} is the scaling factor for that configuration, and T is the Gaussian template fit to the center-only population response. This analysis resulted in 6 unique weights, one for each of the 6 relative center/surround configurations and their respective population responses. The best fitting weights very closely approximate the average surround modulation measured at each relative configuration (Figure 2.1D, right). We compared this reduced model (Eq. 2.3.1) to a full model in which T was allowed to shift (to accommodate potential shifts in the peak orientation) and found that the fit

improvements offered by the full model were not significant ($r^2_{\text{full}} = 0.96$; $r^2_{\text{reduced}} = 0.94$; $p = 0.15$, Sequential F test), suggesting that, within the center+surround data, the surround did not shift the center tuning curves. One corollary of the scaling effect is that the orientations of the population responses' mean vectors continue to be 0° . We might fail to detect shifting behavior because the orientations in fact remain at 0° or because the data was too noisy. To estimate our confidence in the mean vector orientations measured for each of the 6 center+surround population responses, we iteratively re-sampled (with replacement) the underlying units in the complete data set and re-measured the mean vectors at each iteration ($n = 10,000$) to determine the 95% confidence intervals. These intervals are plotted as red horizontal bars in Figure 2.2 and were quite small. As such, any shifting effects of the surround that we failed to detect using the Sequential F test were unlikely to be of more than a few degrees in magnitude.

Lastly, we adapted a fitting procedure used by Benucci et al. (2013; their Figure 4) in order to quantitatively compare how center/surround orientation difference and neuronal orientation preference each contribute to the suppression patterns we measured. In essence, this approach fits surround modulation (Figure 2.1C, right) as the product of two Gaussian-shaped gain factors, one tuned to the difference between the center and surround orientations (“stimulus” gain) and one tuned to the difference between the preferred orientation and the surround orientation (“neuronal” gain). The data were fit by adjusting the magnitude of each gain factor, which revealed that 84% of the observed suppression was attributable to stimulus gain (data not shown). Importantly, stimulus gain should scale population responses and neuronal gain should shift them (Benucci et al., 2013). As such, this analysis supports the conclusion that, when the center consists of a single grating, the surround acted primarily to scale population responses to the center.

2.3.2 Surround suppression *shifts* population responses to plaid center stimuli

The goal of this study was to understand the surround mechanism that gave rise to the suppression patterns illustrated in Figures 2.1 and 2.2 and reported previously (Sillito et al., 1995; Jones et al., 2002; Cavanaugh et al., 2002b; Shushruth et al., 2012). In particular, we wanted to distinguish between two alternative classes of models. We first considered the possibility that the surround acts as a global modulator of activity, and suppression similarly affects the output of all neurons whose receptive fields cover the “center,” regardless of each neuron’s feature selectivity, thereby scaling the population response. This mechanism assumes that the pool of neurons controlling suppression is most active when the center and surround stimuli match. The second possibility we considered was that the surround acts to modulate the input strengths of specific features within the center stimulus, specifically those features that match the surround (Spratling, 2010; Lochmann et al., 2012). Accordingly, this mechanism would cause neurons to respond predominantly to the stimulus attributes that were unmatched to the surround stimulus. For example, a vertical surround stimulus would act to suppress whatever portion of the neurons’ response was driven by vertical components of the center stimulus. Comparing these potential mechanisms more generally amounted to asking whether surround suppression is best understood as a form of output- or input-gain control.

The data presented thus far are useful for motivating the two hypothetical mechanisms described above but do not provide any evidence for one over the other. To distinguish between them, we performed an additional experiment using a “plaid” center stimulus, created by superimposing two 50%-contrast component gratings (termed Center₁ & Center₂, abbreviated in figures as C₁ & C₂), and recorded V1 responses to all Center₁/Center₂ orientation combinations. On half of the trials the plaid was presented by itself (‘plaid-only’), and in the other half a surround was added whose

orientation always matched that of Center₂ ('plaid+surround'). In this way, the center+surround and plaid+surround stimulus sets were quite similar; the difference was that, for the plaid+surround stimuli, the orientation presented in the surround also contributed to the center. This allowed us to measure responses under a similar range of center/surround configurations but introduced the manipulation that, at all times, at least one component orientation would match the surround.

Critically, this experiment allowed us to directly compare our two hypotheses, since they make distinct predictions for how the surround should affect responses to the plaid centers. According to the first hypothesis, the surround mechanism is sensitive to center/surround similarity and globally modulates neuronal responses in proportion to this similarity, which would predict scaling of population responses similar to what was seen in the previous experiment (Figure 2.2). According to the second hypothesis, the surround mechanism modulates the input strength of each feature of the center stimulus according to how much that feature matches the surround. Importantly, this predicts that the surround should shift population responses toward the orientation of the unmatched component of the center.

When examining the combined population data ($n=71$) to the plaid-only stimulus set, we found that, on average, responses appeared equally tuned to the two component orientations (Figure 2.3A-B, left). Unsurprisingly, the average plaid-only response curves to the two components were nearly identical since, without a surround, the distinction between Center₁ and Center₂ was arbitrary. Interestingly, the complementary response profile for the plaid+surround data set revealed that the responses were more tuned to the orientation of Center₁ than to the orientation of Center₂, indicating that the surround had disproportionately suppressed the responses driven by the matched Center₂ component (Figure 2.3A-B, right).

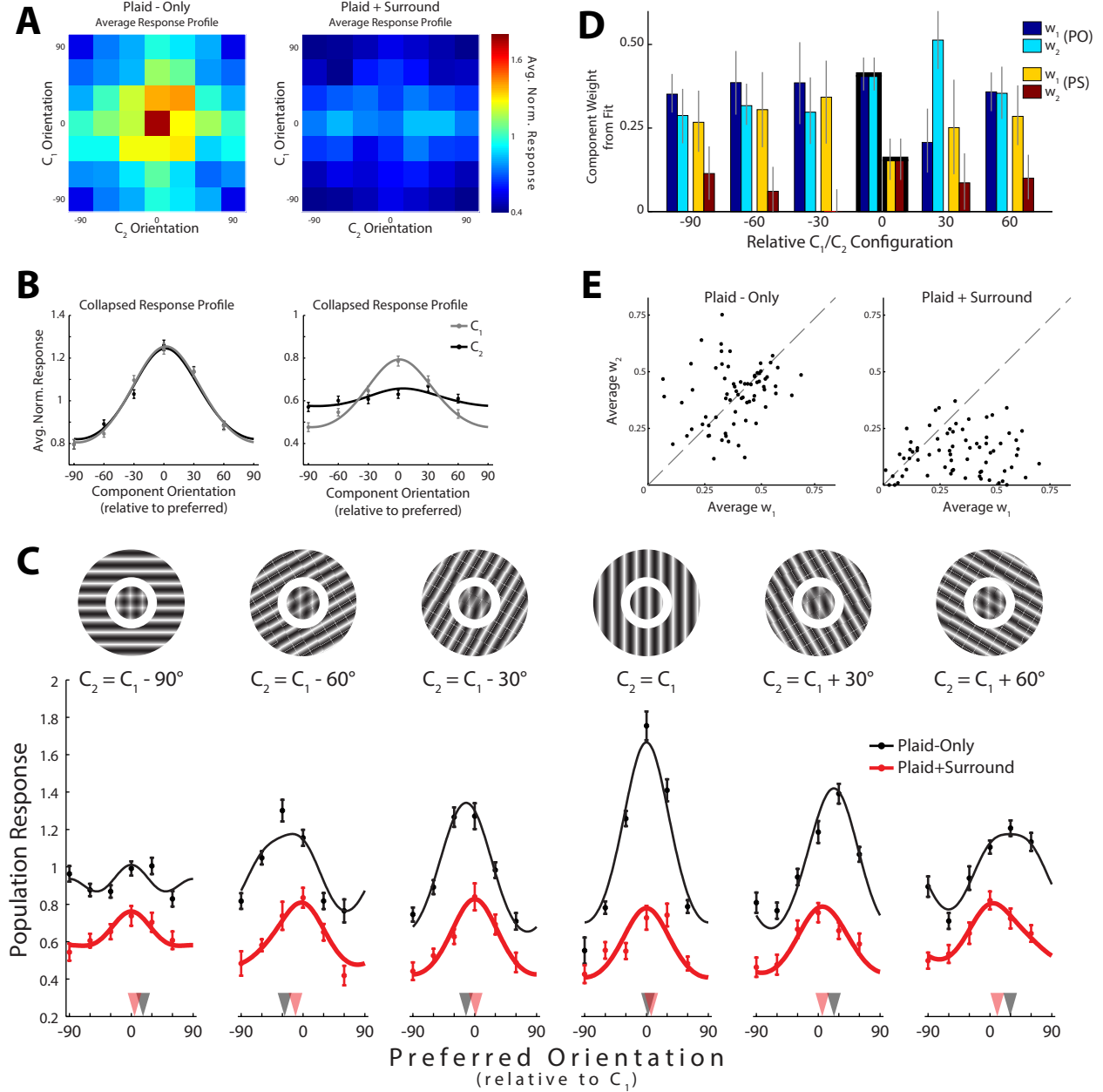


Figure 2.3 | Component-specific suppression with plaid center stimuli. **A**, average response maps measured under the two surround conditions. **B**, response maps in **A**, collapsed across stimulus dimensions, showing the average tuning to each of the center components with (right) and without (left) a surround. **C**, population responses under each of the 6 relative configurations (a stimulus exemplifying the relevant configuration is illustrated above each subplot). Data points are represented as mean $\pm$ SEM. Smooth curves represent the results of the fitting procedure described in the main text (Eq. 2.3.2). Black (red) triangles point to the orientations of the mean vectors calculated from the plaid-only (plaid+surround) population responses. Stimuli are shown for illustration and are not to scale. **D**, best-fitting component weights (Eq. 2.3.2). For fitting the two $C_1=C_2$ population responses, w_1 and w_2 were constrained to be equal. 95% confidence intervals (black) were obtained using bootstrapping. **E**, scatter plots comparing the mean weight assigned to each component for fits to the plaid-only data (left) and fits to the plaid+surround data (right). Each data point corresponds to an individual multiunit.

To investigate this possibility more closely, we examined population responses measured for the plaid-only/plaid+surround stimuli under each of the relative configurations (Figure 2.3C). We reasoned that the measured population responses should be describable as a weighted sum of the responses to each center component alone (Busse et al., 2009; MacEvoy et al., 2009). To approximate this, we created a template by fitting a circular Gaussian to the plaid-only population response measured when the orientations of Center₁ and Center₂ were the same. For each of the remaining 11 curves, we fit the population response as a weighted sum of two such templates, one centered on the orientation of Center₁ and the other on the orientation of Center₂. When applied to the population responses, each curve could be expressed as:

$$R_{\theta} = w_1 \cdot T_0 + w_2 \cdot T_{\theta} + k \quad (2.2)$$

where R_{θ} is the population response for a given configuration (identified by the Center₁/Center₂ orientation difference, θ), T_0 is the Gaussian template centered on 0° (representing the response to Center₁), T_{θ} is the Gaussian template shifted by θ° (representing the response to Center₂), and w_1 and w_2 are the weights assigned to Center₁ and Center₂, respectively. To reduce the tradeoff between fitting the mean vs. the shape of the data, we allowed a constant offset, k . This model was then fit simultaneously to all 12 of the population response curves shown in Figure 2.3C (to ensure that the term k was the same across fits).

As stated above, our goal was to identify the class of mechanism underlying surround modulation, and we focused on two possibilities. The first was that the surround globally scales neuronal outputs in proportion to center/surround similarity. The second was that the surround effectively modulates the input strength of distinct center features depending on whether they match the

surround. The most obvious difference between these two hypotheses is that they explain the surround mechanism as a form of output-gain control and as a form of input-gain control, respectively. That is, they differ in whether the effect of the surround occurs after or before inputs are combined into a neuronal response. In the context of Eq. 2.3.2, this difference could be expressed as whether the surround modulates the response itself (R_θ ; output-gain control) or whether the surround modulates the specific weight given to each component (w_1/w_2 ; input-gain control). The advantage of fitting our data with Eq. 2.3.2 was that it allowed us to compare these two mechanisms simply by asking whether the surround differentially affected the weights associated with each center component.

The set of weights obtained by fitting the data to Eq. 2.3.2 is plotted in Figure 2.3D. The fits to the plaid-only data produced roughly equal weights, on average, between the two components, as expected. The fits to the plaid+surround data show that the surround disproportionately reduced the contribution of the matched center component to the population responses: w_1 exceeded w_2 for all plaid+surround fits where they were not constrained to be equal. We additionally performed this fitting procedure for each multiunit individually. The plots in Figure 2.3E compare, for each unit, average w_1 to average w_2 for the plaid-only data (left) and for the plaid+surround data (right). (For averaging, we ignored data from conditions where the two center components had the same orientation.) As one would expect, there was not any systematic difference between the average component weights in the plaid-only data ($p > 0.5$, paired t-test). In contrast, the average weight given to the unmatched Center₁ component consistently exceeded that given to the matched Center₂ component, and this trend was highly significant ($p \ll 0.00001$, paired t-test).

These analyses show that the surround effectively suppressed the portion of the response

specifically driven by the matched component. This effect is evident not only in the combined data (Figure 2.3C-D) but also in the responses of individual multiunits (Figure 2.3E). The specificity of this suppression is further visible in Figure 2.3C, where plaid-only population responses appear to represent the approximate average orientation of the two plaid components but plaid+surround population responses instead appear to represent almost exclusively the orientation of Center_1 . These representational shifts, caused by a selective suppression of the response to Center_1 , directly contradict the hypothesis that the surround performs population scaling as a function of center/surround similarity. Instead, our findings provide strong physiological evidence in support of input-gain control as the mechanism underlying surround suppression (Spratling, 2010; Lochmann & Denève, 2011).

2.3.3 Suppression as feature-selective input-gain

From this perspective, our entire data set should be easily describable by a model with two simple stages: one in which the surround modulates the weighting of specific inputs and another in which inputs are combined into a response according to their weighting. Here, we formalize such a model in order to demonstrate how an input-gain control mechanism can recapitulate the modulation patterns we observe.

The first stage, implementing input-gain control, can be written as:

$$w_{\theta} = c_{\theta} \cdot [1 - \beta - \alpha \cdot G_{\text{surround}}(\theta - \theta_s)] \quad (2.3)$$

where c_{θ} is the stimulus contrast of the center component with orientation θ . The term inside the brackets represents the effects of the surround. Within this term, $G_{\text{surround}}(\theta - \theta_s)$ is a circular Gaussian function tuned to the difference between the orientation of the center component, θ , and

the orientation of the surround, θ_s . $G_{surround}$ implements feature-selective input-gain control by modulating the component contrast according to how much the orientation of that component matches the surround. The magnitude of this feature-selective input gain is determined by a scaling factor α ; additionally, the term β is a constant offset that allows the surround to control input gain in a feature-nonselective manner. The result of this modulation is represented as w_θ , which can intuitively be thought of as the effective strength or “weight” of the center component with orientation θ . To model responses when no surround is present, α and β are simply set to 0.

The second stage of the model is then the neuron’s response function, where each input is combined (summed) into a net output according to its weighting and the neuron’s orientation preference:

$$R = \frac{\sum_\theta w_\theta^n \cdot G_{tuning}(\theta)}{\sigma^n + w_{rms}^n} \quad (2.4)$$

where w_θ is the weight of the center component with orientation θ (from Eq. 2.3.3), and $G_{tuning}(\theta)$ is a circular Gaussian representing the orientation tuning of the neuron. The weighting of each center component is multiplied by the neuron’s preference for the orientation of that component; the neuron’s response, R , is taken as the sum of these products, normalized by a constant, σ , plus the root mean square of the weights of all center components. The exponent, n , is a constant. We chose this response function for our model because it has been shown to account for population responses to plaids as a function of their component contrasts (Busse et al., 2009). While our Eq. 2.3.3 similarly expresses population responses as a function of the components’ “weight,” we do not intend to equate these weights to contrast, per se. That is, we cannot say with certainty that the effect of the surround on a given component is the same as changing that component’s contrast (see 2.4).

The model given by Eqs. 2.3.3 and 2.3.3 is described by 6 free parameters: the width of $G_{surround}$, the width of G_{tuning} , α , β , σ , and n . We used this model to simulate the full range of experiments we performed. We chose parameter values based on what produced the best agreement between the population responses returned by the model and those observed experimentally (see 2.2.8). This model, which relies on feature-specific input-gain control, provided an excellent fit to the data and produced population response patterns qualitatively identical to what we observed (Figure 2.4, original data re-plotted as dots for comparison). When simulating the center-only/center+surround data, the effect of the surround was to scale population responses (Figure 2.4A; compare to Figure 2.2), while, when simulating the plaid-only/plaid+surround data, the effect of the surround was to shift population responses away from the orientation of the surround (Figure 2.4B; compare to Figure 2.3C). Therefore, the full range of surround effects that we observed experimentally is explained by a mechanism that functions to reduce the effective input strength of specifically those stimulus features that are matched to their surroundings.

2.3.4 Eye movement controls

Given that peri-foveal V1 receptive fields are of similar spatial scale to fixational eye movements, we wanted to be certain that our results were not affected by such eye movements. First, although the fixation window we used was on the generous side (1.5 deg. in diameter), both animals maintained fixation over a much smaller range: the median within-trial variability in eye position was 0.07 degrees (RMS), well within the scale of the center stimulus (generally a diameter of 0.3 degrees). Second, we re-analyzed the plaid data after removing half of the trials with the largest variance in eye position and obtained results that were qualitatively identical to those from the full data set. Finally, we note that any deviations in eye position would have the effect of moving the

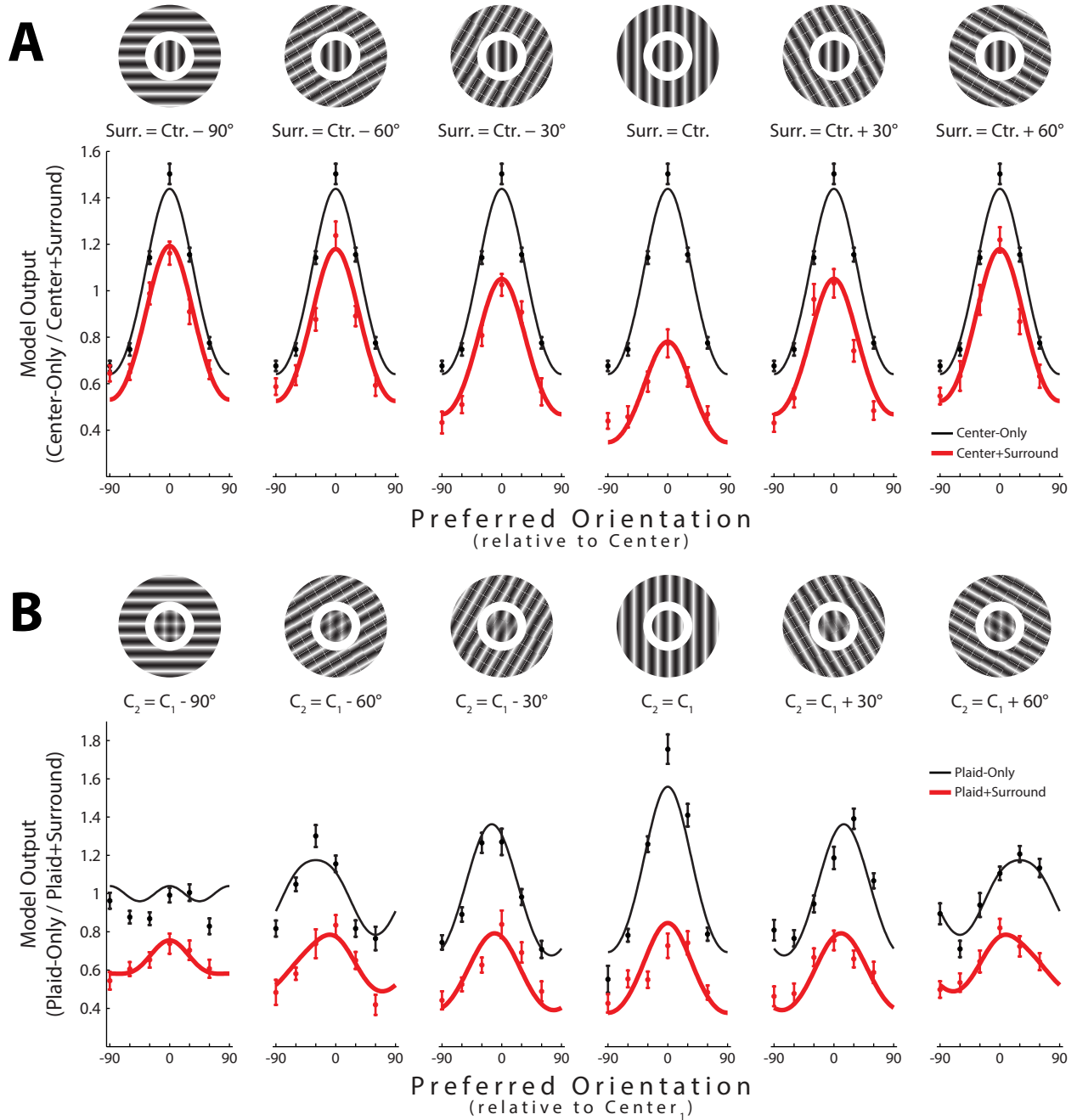


Figure 2.4 | Population responses predicted by input gain model (Eqs. 2.3.3 and 2.3.3). Continuous lines represent model responses and dots represent actual data. **A**, population responses under each of the 6 relative center/surround configurations, shown for the center-only (black) and center+surround (red) data. Conventions are identical to Figure 2.2. **B**, same as A, for the plaid-only (black) and plaid+surround (red) data. Conventions are identical to Figure 2.3C.

stimulus surround into the RF center, a manipulation that would produce responses favoring the orientation of the surround stimulus—opposite to the effects we observed. We are thus confident that small fixational eye movements did not influence our results.

2.4 Discussion

We investigated the effect of surround suppression on population coding in V1 of alert macaque monkeys under a range of stimulus configurations. Previous studies have found that the effect of a surround stimulus on neuronal responses depends on the center stimulus with which it is presented (Sillito et al., 1995; Cavanaugh et al., 2002b; Shen et al., 2007; Shushruth et al., 2012). The common result throughout these studies is that, in general, suppression follows the degree of feature-similarity between the center and surround. This tendency was readily visible in the V1 responses we measured to all combinations of center and surround orientation (Figures 2.1 and 2.2). In addition, we measured V1 center/surround interactions using stimuli in which the center consisted of plaids. The effect of the surround in this data set was to specifically reduce the portion of the response driven by the central plaid component whose orientation matched that of the surround (Figure 2.3). These seemingly distinct suppression patterns were both consistent with a single input-gain control mechanism that reduces the effective strength of center features that match the surround (Figure 2.4).

The response function used to model our data (Eq. 2.3.3) is adapted from the normalization model previously shown to account for population responses to plaids as a function of their component contrasts (Busse et al., 2009). Whereas Busse and colleagues (2009) systematically varied the contrast of the plaid components, in our experiments it was kept constant. As such, we do not

really know whether the surround's effect on the center components behaves exactly like contrast (i.e. shows the same characteristic nonlinearity). While it is intriguing to interpret our result in the context of the contrast-weighted normalization model (Busse et al., 2009) by saying that the addition of the surround lowers the effective contrast of the matched plaid component, the experiments needed to make that claim have not been done. We thus emphasize the common feature of the two models that we think is necessary for explaining our results, which is the relative weighting of the inputs—that is, input-gain control. However, we do note that normalization (the denominator of Eq. 2.3.3) improves the performance of our model; this mechanism exaggerates the shifting effect of the surround when responses are driven by multiple orientations with unequal weights, producing an output that is closer to “winner-take-all,” with the unmatched component winning.

While the results obtained using the plaid stimuli are, to our knowledge, novel, the observation that surround tuning depends on the contents of the center is well established. Despite this, relatively little is known about the biophysical mechanisms directing this form of contextual modulation. The most successful biophysical model to date (Shushruth et al., 2012) derives its explanatory power from the assumption that the feedforward activity evoked by a given center stimulus engages recurrent activity in an orientation-nonspecific manner; each neuron receives the strongest suppression when the surround contains its preferred orientation but only when the center and surround stimuli match does the suppression inhibit the primary source of recurrent drive to the network. While this model is capable of replicating the non-fixed suppression patterns observed previously and in our center+surround data, it is not intuitively obvious whether it could reproduce the results of our plaid experiment. Thus our data provide an important

benchmark for future biophysical models.

The impact of future research directed towards understanding the circuit mechanisms that control input gain will likely extend beyond the scope of surround suppression. For one thing, the firing rate changes observed in studies of selective attention are well captured when attention is modeled as changes in the effective strength of inputs into a normalization mechanism (Lee & Maunsell, 2009; Ni et al., 2012). This description reconciles the finding that attention performs multiplicative scaling with a single center stimulus (McAdams & Maunsell, 1999) with the finding that attention often has non-linear effects with multiple stimuli in the RF center (Reynolds et al., 1999; Reynolds & Desimone, 2003). Interestingly, the effects of a surround on the population responses we observed with either one or two orientations in the RF center resemble the effects of attention under similar conditions (Lee & Maunsell, 2010). This similarity is further evidenced by the ability of the same basic model (where attentional or surround context controls a feature-selective input-gain mechanism) to explain both the effects of attention and the full range of surround behavior reported above. While this descriptive overlap may hint at a common mechanism, one has yet to be identified experimentally. Nevertheless, a unified explanation of attentional and surround modulation (based on context-dependent input-gain control) seems within reach (Spatling, 2008, 2010).

The common link between surround suppression and selective attention may be that both mechanisms serve to enhance the representation of relevant information. In the case of attention, relevance is behaviorally determined, whereas, in the case of surround suppression, relevance appears to be determined by the statistics of natural stimuli. That is, a given surround may establish predictions regarding the contents of the center based on learned statistics of naturally

occurring stimuli (Schwartz & Simoncelli, 2001; Coen-Cagli et al., 2012). Elements of the center that violate the predictions established by the surround are more relevant by virtue of the added information that they carry (Rao & Ballard, 1999). This fact appears to be exploited by the surround mechanism, as evidenced by the observation that, when the center contains both redundant and informative elements, (i.e. orientations that do and do not, respectively, match the surround), representation is specifically biased towards the more informative (i.e. less spatially redundant) element. This feature of the surround likely explains its observed tendency to decorrelate and “sparsify,” and thereby increase the information content of representations of natural stimuli in V1 (Vinje & Gallant, 2000, 2002; Haider et al., 2010). Importantly, our conclusions support those of theoretical studies that approached surround suppression as a form of input-gain control, especially with regard to their implications for how the surround enhances efficient coding (Spatling, 2010; Lochmann et al., 2012).

The idea that the surround promotes efficient processing is not new (Barlow, 1959; Mumford, 1992), nor is the idea that top-down mechanisms interact with surround suppression (Bair et al., 2003; Roberts et al., 2007; Sundberg et al., 2009; Nassi et al., 2013, 2014). However, these ideas bear further investigation because they underscore the notion that contextual modulation (in the form of context-dependent, feature-selective, input gain) is a solution to the inherent challenges of representing natural stimuli. With this in mind, our findings constitute key physiological evidence supporting a framework for understanding contextual modulation and its role in information processing. According to this framework “context” (in the case of passive sensing) is determined by the surround (rather than by its continuity with the center); modulation occurs through a form of feature-specific input-gain control that can fundamentally tailor the representation of

context-embedded stimuli; and, this mechanism prioritizes the efficient representation of the most informative features of the sensory input. Identifying the circuit mechanisms that perform this type of input-gain control remains a crucial step towards understanding how the cortex implements contextual modulation.

Chapter 3

Effects of feedback inactivation on feature-specific surround suppression in V1

The observation that feature-specific surround suppression acts as a form of effective input-gain control and the possibility that this reflects a fundamental function of top-down processing raises the question of how feedback inactivation affects this mechanism. Nassi et al. (2013) demonstrated that the most common effect of reversible inactivation of V2/V3 (and their feedback projections) was a reduction in the suppressive influence of the receptive field surround. This effect was demonstrated by measuring responses to optimally oriented gratings of various sizes and therefore leaves unanswered the question of whether more sophisticated behaviors of the surround depend on feedback. Forming an answer to this question was the general aim of this project. We reasoned that the results presented in Chapter 2 illustrate one of the most sophisticated behaviors of the surround that can be readily characterized with parametric stimuli. Moreover, we reasoned that use of this particular lens was most likely to yield results that would inform multiple potential viewpoints on the role of feedback (see Section 1.4).

Our approach involved reversible inactivation of areas V2/V3 by cortical cooling while recording (primarily) multiunit activity (MUA) in V1 with chronically implanted multi-electrode arrays

(MEAs). This chapter describes the methodology of the inactivation experiments and provides an overview of the main obstacle faced during the project: the tendency for feedback inactivation to produce epileptiform activity under the MEA. This tendency strongly limited the pace of data collection but is also noteworthy because of the implications it holds for the function of feedback.

In the remainder of the chapter, I describe the effect of feedback inactivation on the forms of surround modulation characterized through the two classes of stimuli used in Chapter 2 (center-only/center+surround and plaid-only/plaid+surround). We were able to collect data from 2 male macaque monkeys (monkey R and monkey V) across a subset of experimental sessions in which inactivation did not cause any observable epileptiform activity. Overall, evaluation of the data set as a whole is difficult, the main reason being some baseline differences and considerable inactivation effect differences between the two monkeys. In addition, the timeline of the experiments with each monkey were different with respect to the overall timeline of the project, leading to some differences in experimental design choices, complicating direct comparison at times. Lastly, experiments were conducted as soon as possible after array implantation for monkey V but several years after implantation for monkey R. We believe that the data from both monkeys are worth considering but we also point out that the signal quality of the data from monkey V was considerably higher. One simple, quantitative, example of that is the threshold settings that we were able to use to determine spike times in each monkey: -3.1 RMS for monkey R and -4.7 for monkey V. This setting reflects the discriminability of signal waveforms from those of background noise and that discriminability was quite low in monkey R. Therefore, we place more emphasis on interpreting the data from monkey V in the cases when the data disagree. We believe that the nature of the effect differences between the two monkeys will be made clearer

as experiments move to a third monkey, but we are presently limited to speculation. For the reasons listed above, the data from each monkey are presented separately. Since this separation results in a presentation of the data that is somewhat harder to digest, we provide a condensed summary of the baseline response properties and effects of inactivation measured for each of the two classes of stimuli in Table 3.1.

	Monkey R	Monkey V
Single-grating center	Suppression maximal for similarly oriented CS stimuli. Later responses show weaker dependency on center/surround configurations. Reduced overall suppression. Largest increase in responses for similarly oriented CS stimuli.	Maximal suppression for similarly oriented CS stimuli during the response transient. Later responses suppressed uniformly by all center/surround configurations. Reduced overall suppression. Largest increase in responses for orthogonal CS stimuli.
Plaid center	Clear feature-specific suppression of sustained responses. Reduced overall suppression. Slight tendency for increase to be feature-specific, but not statistically significant.	Clear feature-specific suppression of sustained responses. Reduced overall suppression. Reduction clearly not feature-specific. Suggests that feedback divisively scales population responses.

Table 3.1 | Suppression patterns and inactivation effects throughout the stimuli used in Chapter 3. Each column provides a rough summary of the suppression patterns for the labeled monkey throughout the control data (red) and inactivation data (blue).

3.1 Reversible feedback inactivation by cortical cooling

3.1.1 Surgical preparation

Two male macaque monkeys (*Macaca mulatta*; R and V) were each implanted with a 10x10 electrode array (400 μm spacing) in the right hemisphere of V1 and cryoloops in the lunate sulcus of the same hemisphere as well as a head post on the anterior portion of the skull. The head post was implanted as an initial surgery to allow for familiarizing the monkeys with the experimental setting. Once the monkeys were sufficiently accustomed to the experimental procedure, a second surgery was performed to implant the array and the cryoloops. A craniotomy was performed to expose the parafoveal region of V1 corresponding to the lower left visual field. Skull along the anterior portion of the craniotomy was removed so as to expose the lunate sulcus. A durotomy was performed to further expose the cortical surface and lunate sulcus. Three cryoloops were implanted into the lunate with the technical assistance of Dr. Stephen G. Lomber. The electrode array was then implanted in V1 (Figure 3.1A). The dura and skull fragment were replaced before finally fortifying the implanted equipment with dental acrylic. All surgeries were performed under isoflourane anesthesia and were in compliance with approved IACUC protocols.

3.1.2 Cryoloops

Cryoloops are constructed of hollow metal tubes attached to a small thermoprobe. The cryoloops we used were bent to form loops such that, when surgically implanted in the lunate sulcus, they rested against the cortical surfaces of areas V2 and V3 corresponding to the lower visual hemifield. The inlet and outlet of each cryoloop protruded out of the skull, allowing the experimenter to flow chilled methanol through them (Figure 3.1B). The thermoprobes were

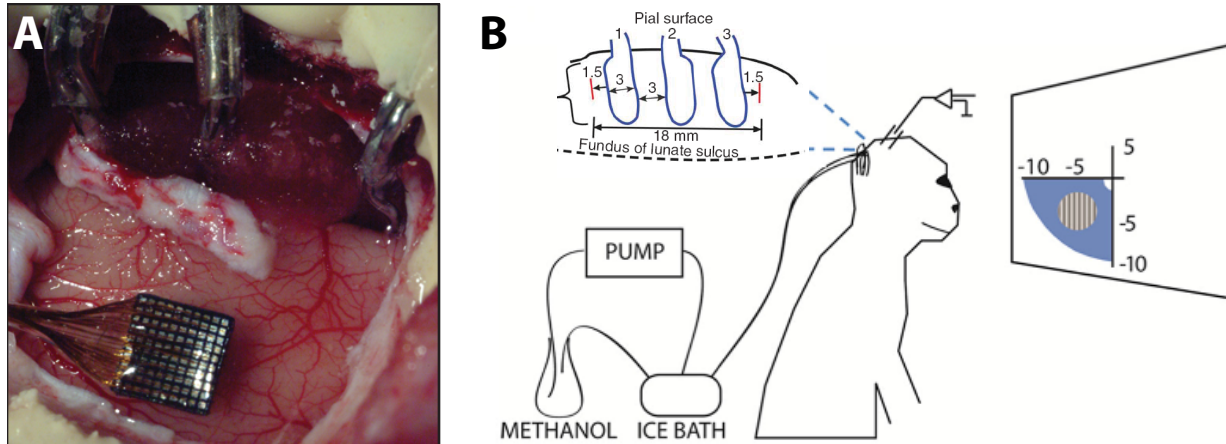


Figure 3.1 | Experimental setup. **A**, a photograph from the surgery to implant monkey R. This picture was taken after the cryoloops and array had been implanted. **B**, cartoon of the full experimental setup, including cryoloops, the electrophysiology equipment (an array), the awake monkey, stimulus presentation, and cooling equipment.

situated to measure the temperature of the portion of the loop inside the brain, allowing the experimenter to control the flow of methanol in order to achieve a desired temperature range for inactivating nearby cortex. Due to a malfunction with the thermoprobe of the lateral cryoloop in monkey R, we were unable to use it for cooling.

This technique has been successfully employed by the Born lab for many years to study the computational design of the visual hierarchy surrounding V2/V3 (Ponce et al., 2008, 2011; Nassi et al., 2013, 2014; Smolyanskaya et al., 2015). The present work benefits from the numerous validations and controls done in these studies demonstrating the effectiveness with which the cryoloops inactivate V2/V3, the reversibility of this inactivation, and the lack of direct-cooling effects in V1.

3.1.3 Design of an inactivation experiment

Determining the effects of feedback inactivation required measuring the activity of the same neuron (or multiunit) in response to a given set of stimuli both before and during inactivation of V2/V3. When possible, activity was again measured after inactivation had ended, providing an additional

control. Following an initial characterization of basic receptive field properties, responses to one full run-through of the experimental stimuli were collected (with each run-through consisting of 20 repetitions per stimulus). Next, the methanol pumps were turned on to allow the cryoloops to reach the target inactivation temperature of 10 degrees Celsius. This temperature was maintained throughout the duration of a second run-through. Once these responses were collected, the pumps were turned off and the brain was allowed to return to physiological temperatures, after which responses to a third and final run-through were collected. The data collected during each of these experimental phases are referred to as “control,” “inactivation” and “re-warming,” respectively. In some cases, we were unable to collect re-warming data when the monkey had reached satiety before the end of the experiment. While it is not entirely clear what the duration of continuous cooling is that produces lasting physiological changes in V2/V3, we find that no such apparent damage occurs within an hour and designed experiments such that a single run-through would only ever minimally exceed this arbitrary time limit.

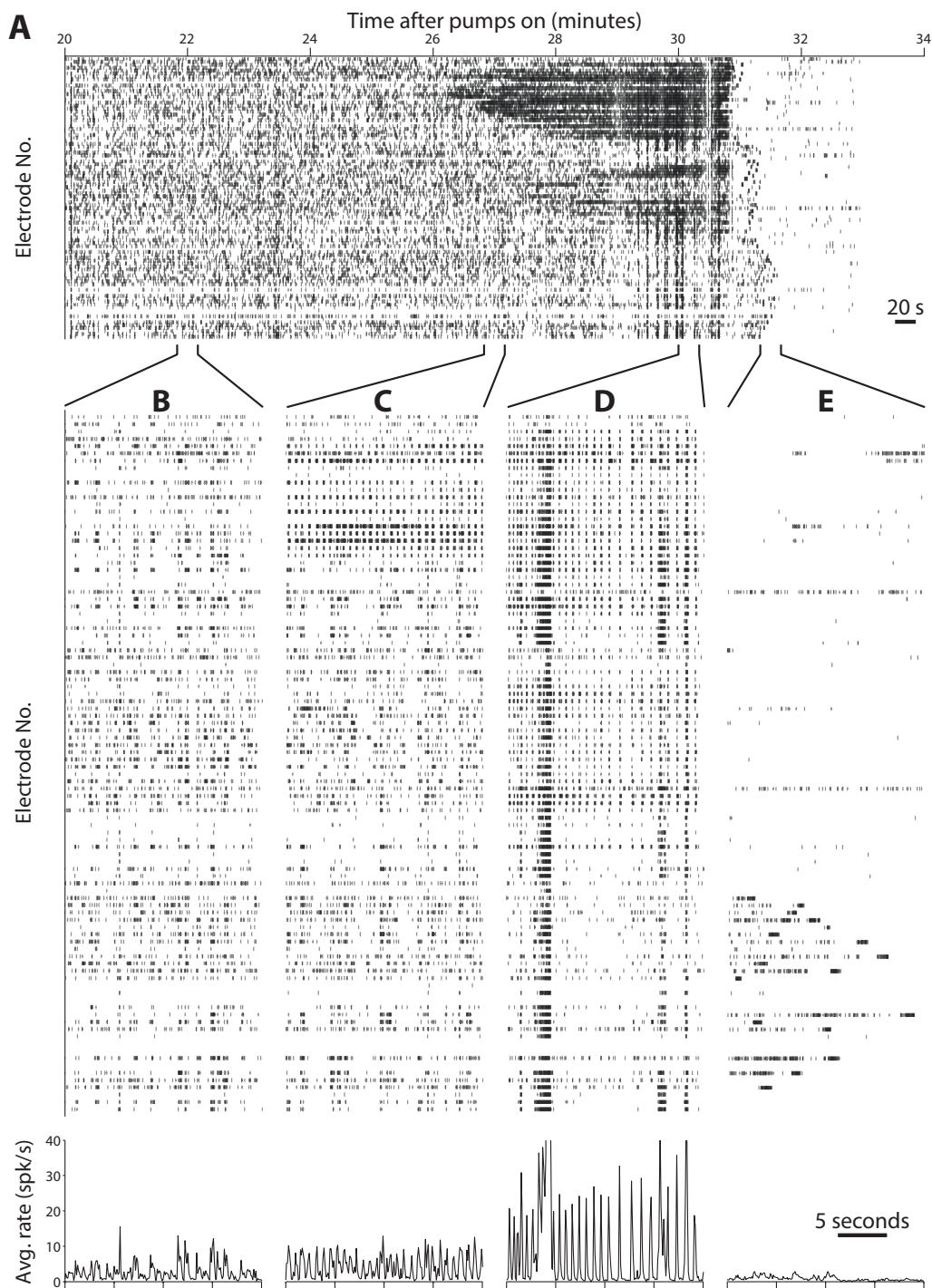
3.1.4 Epileptiform activity during feedback inactivation

The largest obstacle facing this technique is the creation of epileptiform activity under the MEA following inactivation of V2/V3. By “epileptiform activity” I mean spontaneous, rhythmic discharges that are unrelated to a visual stimulus and that appear synchronously on many, and eventually all, of the electrodes in the MEA. As mentioned above, the Born lab has successfully used cryoloops to inactivate V2/V3 for many years, but only relatively recently have we attempted to combine this inactivation method with chronically implanted arrays in V1. We have attempted this combination in 3 monkeys to date (R, V and U) and have observed feedback inactivation to have this outcome in all 3, at least at early stages of the project. In two of the

monkeys (R and V), the frequency and severity of the abnormal activity declined with time. Figure 3.2 provides a demonstration of the array-wide activity measured over the time course of such an event. Panel A represents the activity observed over the entire example event. Activity was fairly normal for the first 25 minutes of cooling and was determined mostly by stimulus presentation (Figure 3.2B). The epileptiform activity in this instance, like in most, emerged first on only a fraction of electrodes (Figure 3.2C; electrodes near top of plot; see also the emergence of rhythmic activity illustrated in the bottom panel) but eventually spread across the full array (Figure 3.2D). The likely pathological (that is, epileptiform) nature of this activity is evidenced by its utterly extreme behavior: first, throughout its duration the activity was, as far as one could tell, propagated intrinsically (stimuli—and methanol pumps—were stopped as soon as this was observed (at approximately 29 minutes) but doing so did not appear to have an effect); second, once this activity was fully realized, the array fired rhythmically in complete unison and the duty cycle of this rhythm tended to smoothly vacillate around a frequency of 1-2 bursts per second; third, the waveforms measured during each burst appeared quite distinct from those normally associated with multiunit activity, likely because the signal was overwhelmed by the coordinated activity of thousands of nearby neurons; and, lastly, the cessation of this activity was always followed by near-total silence across the array for one to several minutes (Figure 3.2E). This general description applies to the observed activity patterns across all three monkeys, though emerging evidence suggests that time and the exact cooling approach may alter the balance of the bursting patterns and the silencing patterns (Figure 3.2 D and E, respectively).

Figure 3.2 (following page) | Epileptiform activity observed during feedback inactivation. **A**, each row represents an individual electrode of the array and each tick indicates the timing of a waveform crossing on that electrode. Time since the beginning of cooling (pumps on) is represented along the abscissa. **B-E**, the activity is broken up into several sections to provide a closer view of the array-wide activity over the progression of the epileptiform activity. Each section represents a window of 20 seconds and the average firing rate over the entire array at each point in time is plotted along the bottom. **B**, activity measured during inactivation several minutes before the onset of epileptiform activity. Responses across the array reflect normal activity during stimulus presentation. **C**, epileptiform activity first becomes noticeable on only a few electrodes (near the top), and rhythmic activity begins to take over the array-wide response pattern (bottom panel). **D**, a full minute after the stimuli and pumps were turned off, rhythmic bursts of activity continue to dominate most electrodes and the overall activity pattern (bottom panel). **E**, the epileptiform activity terminates with a near-silencing of the entire array. Many units transition into silence with a prolonged burst of firing.

Figure 3.2 | (Continued)



Fortunately, the epileptiform activity is an extremely salient, all-or-none event, which is easily recognized by the experimenter. Therefore, we were able to collect data from inactivation sessions and focus on those sessions where no such extreme behavior was observed. But we were understandably conservative with the number of attempts made to inactivate, following the general concern that the integrity of the cortex under the array may suffer under such extreme conditions.

Interestingly, this tendency seems to require the combination of the MEA and feedback inactivation; either alone will not produce these effects. This point is particularly well made in monkey R, from whom Nassi et al. (2013) recorded V1 activity with single, tungsten electrodes during inactivation and never observed any pathological activity. One might be concerned that a single electrode recording would prevent the detection of what is so easily recognized in the activity across the MEA (Figure 3.2). However, we expect that the repetitive bursts of activity associated with this phenomenon would be quite visible with a single electrode because they are unrelated to the visual stimulus and display highly abnormal spike waveforms. Furthermore, no overtly abnormal activity was observed in the second monkey from that study, from whom activity was only ever recorded with single electrodes. This suggests that the epileptiform activity was *only* in the nearby vicinity of the MEA. When it was seen on the MEA during cooling, monkey R was observed to occasionally make repeated saccades to the location of the visual field represented by the neurons under his array. This is exactly the behavior one might expect if the coordinated bursts of activity produced retinotopically fixed phosphenes and monkey R was attempting to foveate them. While circumstantial, this further speaks to the array as the locus of this activity. The additional requirement of feedback inactivation is obvious; in the substantial

amount of time various members of the Born lab have recorded from the arrays of monkeys R, V and U, epileptiform activity has only been observed following inactivation.

How then is this finding to be explained? We characterize this effect of feedback inactivation as adding “insult to injury,” where “insult” refers to the removal of a suppressive influence and “injury” to the traumatic insertion of the MEA. There is clinical evidence that focal trauma to the surface of the cortex increases the risk of epileptiform activity (Annegers et al., 1998; Ferguson et al., 2010). The insertion of 100 rigid electrodes across 16 mm² of cortex is very traumatic. However, the Utah array and its insertion method are standards in the field and, while we are aware of reports of similar post-operative epileptiform activity in monkeys (JH Maunsell, personal communication), the effects of this trauma are well managed with a regimen of peri-operative dexamethasone (an anti-epileptic). With this added step we observe no obvious pathology associated with the arrays themselves. The explanation may rest on the nature of the changes to the system when feedback is temporarily removed. Importantly, as demonstrated previously (Nassi et al., 2013) and confirmed here (see below), the most common effect of feedback inactivation is to weaken the suppressive influence of the surround, such that inactivation amounts to a net decrease in inhibition within retinotopically matched V1. Over the last several years, a number of studies have examined the question of how excitation and inhibition are coordinated and typically report a characteristic balance between the two (Mariño et al., 2005; Okun & Lampl, 2008; Sun et al., 2010). One study (that coincidentally also accounts for the effect of the surround on excitatory and inhibitory drive) explains this balance as evidence that visual cortex operates as an inhibitory stabilized network (Ozeki et al., 2009). An inhibitory stabilized network is defined by a network regime in which recurrent excitation is strong enough

to be inherently unstable but the network itself is stable through the influence of inhibition (Tsodyks et al., 1997). As such, the instability of V1 during inactivation underscores the observed inhibitory influence of feedback. It is tempting to further speculate that this instability reflects the loss of a more detailed organization of inhibition but making this claim would require a much more detailed understanding of the involved circuit motifs and how network stability depends on the availability of each.

While the epileptiform activity we encountered impeded our ability to characterize the role of feedback in contextual modulation, it provided a characterization of feedback nonetheless. Unfortunately, the immediate insight provided by this characterization is limited. However, we believe that, as continued work further informs our understanding of feedback and models to describe its function, these models should predict a contribution of feedback to network stability. Given that array implantation is a factor in the epileptiform activity, such models would not need to predict that feedback removal lead to complete instability but should be able to show an effect in that direction.

3.2 Experimental procedures

The experiments and results from this point onward are described one monkey at a time. The purpose of this separation is made quite clear when examining the effects of inactivation for each monkey and observing the occasionally striking differences—especially where those differences relate to interpreting the role of feedback. In terms of stimuli and their associated responses, I will organize the results from each monkey similar to as in Chapter 2: first, I will present the effect of feedback inactivation on responses measured to all combinations of center and surround

orientations in the case where the center contains a single grating (a stimulus setting that typically elicits ‘match’ suppression); next, I will present data from the “plaid” case (a stimulus setting that we find to elicit ‘feature-specific’ suppression). I will then provide an overview of the differences between the data from each monkey. Here and below, the patterns of suppression seen with the two stimulus sets are referred to as ‘match’ suppression and ‘feature-specific’ suppression, respectively, with full acknowledgement of the possibility that they in fact reflect the same underlying mechanism.

3.2.0.1 Monkey R: measuring match suppression

Monkey R was trained to fixate a 1.5° window for approximately 3 s while stimuli were presented. Completed fixations resulted in a liquid reward. Multiunit activity (MUA) was sampled using a Cerebus 128-channel system (Blackrock, Utah). At the start of each recording day, activity thresholds were set for each electrode as -3.6 RMS background noise; MUA events were logged as threshold crossings.

All stimuli used for the present experiments consisted of stationary, sinusoidal gratings and were presented on a CRT monitor (100 Hz refresh rate), using mean luminance as background. Depending on the experiment, stimuli included a central grating/plaid, an annular surround, or both (experiments to characterize match suppression and feature-specific suppression were done separately for this monkey). To simplify pooling the data, the diameter of the central stimulus and the inner and outer diameters of the annular surround were fixed across sessions (at 0.3° , 0.7° and 3.0° , respectively), as was the spatial frequency of the gratings (at 5 Hz). Orientation was sampled at a spacing of 30° . For each session, the stimulus location was approximately aligned to the RF center(s) of one or more multiunit sites. Whether a given site was included for analysis

depended on how well its RF boundaries aligned with the stimulus boundaries, as assessed using the criteria described below.

Once a location was selected, experiments followed the general framework outlined in Section 3.1.3: for determining inclusion off-line, the responses to the surround-only stimuli were measured; afterwards, responses to center-only and center+surround stimuli were measured for each portion of the experiment—control, inactivation, and re-warming. Responses to an isolated stimulus component were measured by setting all other components to have a (Michelson) contrast of 0% while keeping the relevant component of the stimulus at 100% contrast and without changing any other stimulus parameters (except orientation, which was varied to measure tuning). Center+surround stimuli represent all of the 6x6 orientation combinations between the center and surround components.

Within each stimulus group, stimuli were presented in a random block interleaved design. For each portion, the goal was to repeat each stimulus 20 times. A trial began when the monkey achieved fixation; 300-500 ms after fixation began the stimulus appeared and remained on screen for another 800 ms, after which the trial ended. If the monkey broke fixation before the end of the trial, the trial was aborted without reward.

3.2.0.2 Inclusion criteria (center-only/center+surround data)

We wanted to restrict our analysis to data from multiunits whose RFs overlapped the central stimulus and only minimally overlapped the annular surround. To that end, we applied a set of inclusion criteria to ensure the interpretability of our analyses. For the data set characterizing match suppression, inclusion required that: 1) responses to the surround-only stimuli were either

unturned to the surround orientation or did not show above-baseline responses to the surround alone; 2) responses to center-only stimuli exceeded baseline firing on average and (3) exceeded responses to center+surround stimuli on average. These criteria were based on data measured during initial characterization or during the control portion. In addition, to avoid analysis of units that were driven by the surround, we further required that responses to the center+surround were not significantly tuned to the surround orientation itself during control or inactivation. (Given the form of surround modulation that serves as our lens, applying this final criterion to both control and inactivation is important to avoid biased sampling.) Significance of response level differences was determined with a 2-sample t-test, and significance of tuning was tested using a permutation test (1000 iterations). Significance was assessed using $\alpha = 0.05$, applying Bonferroni correction where appropriate. This yielded a total of 25 multiunits from monkey R for this data set, measured across 7 inactivation sessions, with 2 of the sessions excluded based on the observation of epileptiform activity. For the purposes of analyzing match suppression, it is not required that the units are tuned to the center orientation, so that did not serve as a criterion here. Responses measured during the “sustained” portion of the trial were used to determine inclusion. The sustained portion refers to the time interval of the trial over which activity is relatively stable (i.e. after the onset transient and ends at the stimulus offset); for monkey R, the sustained activity describes responses after 150 ms from the onset of the stimulus.

3.2.0.3 Monkey R: measuring feature-specific suppression

The experiments to measure feature-specific suppression and the effects of feedback inactivation in monkey R followed the basic outline described for measuring match suppression, with a few differences. Specifically, the diameter of the central stimulus and the inner and outer diameters

of the annular surround were set to 0.3° , 0.6° and 2.0° , respectively; stimuli were drawn with a spatial frequency of 10 Hz; surround-only stimuli (for determining inclusion) were presented during control, inactivation, and re-warming; and, lastly, stimuli were presented for 500 ms.

The stimulus set was designed and presented in the same way as the “plaid” stimuli described in Chapter 2 (Section 2.2.4). However, in the present experiment, the two center components did not have the same contrast. Instead, the “unmatched” (C1) and “matched” (C2) components of the center were drawn with Michelson contrasts of 33% and 66%, respectively. The rationale for this setting was that, if the effect of the surround under control settings is to modulate the effective strength of the centers, V1 would represent the two components as having similar contrasts when a surround was present and that this would provide the best baseline for seeing the direction in which inactivation had an effect. While the surround roughly equated the effective contrast of the two center components (Section 3.4), the use of unequal contrasts ended up not being ideal for reasons discussed below. In later experiments with monkey V, we returned to 50/50 contrast.

3.2.0.4 Inclusion criteria (plaid-only/plaid+surround data)

The goal of the inclusion criteria were, as before, to restrict our analysis to data from multiunits whose RFs overlapped the central stimulus and only minimally overlapped the annular surround. For this experiment, we apply the same criteria described in Section 2.2.5, excluding the last criterion which is not applicable. Here, however, we must slightly adapt this method to deal with the multiple portions of the experiment (i.e. control, inactivation, and re-warming). For exclusive criteria (in this case, tuning to the orientation of the surround-only stimulus), which require a non-significant p-value at each of the portions of the experiment that are considered, applying Bonferroni correction to the significance threshold is sufficient. For inclusive criteria, which

require a significant p-value at each of the considered phases, we chose to simply relax the significance threshold such that the probability of incorrectly rejecting the null hypothesis for all the required tests matched some desired alpha level. We found the best setting for reasonable inclusion and exclusion was $\alpha = 0.01$. We based inclusion only on the data recorded during control and inactivation in order to (1) be consistent with the approach taken with the center-only/center+surround data and (2) allow the re-warming to serve as a control against the inclusion procedure itself. This provided a sample of 31 multiunits measured across 5 inactivation sessions. Responses measured during the “sustained” portion of the trial were used to determine inclusion.

3.2.1 Experimental procedure: Monkey V

Monkey V was trained to fixate a 1.0° window for approximately 3 s while stimuli were presented. Completed fixations resulted in a liquid reward. Multiunit activity (MUA) was sampled using a Cerebus 128-channel system (Blackrock, Utah). At the start of each recording day, activity thresholds were set for each electrode as -4.7 RMS background noise; MUA events were logged as threshold crossings.

The main difference in the approach used for monkey V was to present the 4 classes of stimuli (center-only, center+surround, plaid-only, and plaid+surround; defined in Sections 2.2.3 and 2.2.4) as a single pseudo-random interleaved block. In addition, the size dimensions of the stimuli were slightly more tailored to the units targeted on a given day. The range of spatial frequencies used was 3.33-5 Hz, center diameters were chosen from between 0.3° to 0.5° , and surround inner diameters were chosen from between 0.6° to 0.8° depending on the smallest diameter of the surround-only stimulus that generated above-baseline responses, as observed during online characterizations of

RF properties.

We attempted (unsuccessfully, I believe) to curb the occurrence of epileptiform activity by providing a less rhythmic stimulus schedule. We experimented with several timing schedules, such that, for the data reported, stimulus durations were either 400 ms, 500 ms, or, for one session, 1200 ms. Unless otherwise specified, analysis was based on the responses measured during the sustained portion of the trial, defined in monkey V as beginning 250 ms after stimulus onset and lasting until stimulus offset. The sustained portion in monkey V begins later because units in this monkey displayed, to varying degrees, what could be described as multiple transients (Section 3.5). Inclusion was determined based on sustained responses, using the same method as just described for monkey V, yielding a total of 39 units measured across 8 inactivation sessions (11 sessions, with 3 excluded based on aberrant, array-wide activity during inactivation).

3.2.2 Analysis

Unless otherwise specified, responses were analyzed over the sustained portions of the responses (defined above for each monkey). Before data were pooled, spontaneous activity was first subtracted from responses, which were then normalized to their grand mean response to all stimuli in the no-surround conditions. Where applicable, stimuli orientations are reported relative to the preferred orientations of the measured MUA. Preferred orientation was calculated as the vector average of the responses to the no-surround stimuli but, for the plaid-only data, ignored conditions where the two center components had different orientations. For simplicity, preferred orientations were binned before analysis (bin width: 30°).

A number of analyses in this Chapter examine surround/feedback effects over time. In these

cases, activity at each time point was measured as the activity density within a Gaussian window, centered on the relevant time point. This simply amounted to a time-weighted spike count, such that firing rate was expressed as the weighted sum of all the spike times divided by the area under the Gaussian window. The effects of time were studied by simply sliding the center of the Gaussian window in time. The time-varying activity is expressed in normalized units. For monkeys R and V, respectively, the rate used to normalize a given unit was its average center-only response from 150 ms to stimulus offset (monkey R) or 250 ms to stimulus offset (monkey V).

Sections 3.4 and 3.6 explore the effects of feedback inactivation on responses measured with the plaid stimulus set. The data associated with these responses are studied using the same type of fitting approach described in Chapter 2, where population responses are fit as weighted sums of two response templates. In this Chapter, control and inactivation data are fit separately; to facilitate comparison of the fitting results, we omit the offset term, k , used in Chapter 2. For monkey V, we apply this general method twice, the second time with a modification to address the possibility that feedback inactivation has a scaling effect (Section 3.6). We include the 95% credible intervals of the posterior predictive distribution, which is essentially a function that takes the observed data and an additional orientation value as inputs and returns a probability distribution describing the range of expected responses to that orientation, given the noise in the data and the quality of the fits. These intervals are useful because they capture the uncertainty in the fits and do so based on the variability of the responses themselves plus the uncertainty produced by the positive correlation between the response templates themselves. We add them as a visual demonstration that the errors produced by the modified fitting procedure are reasonably small.

3.3 Monkey R: Effects of feedback inactivation on match suppression

The tendency for responses in V1 to be suppressed according to the feature-similarity of the center and surround has been reported a number of times (Sillito et al., 1995; Cavanaugh et al., 2002b; Jones et al., 2002; Shushruth et al., 2012) as well as here (Chapter 2). The simplest way to demonstrate this is to compare responses to stimuli containing only a center, stimuli with an iso-oriented surround, and stimuli with an orthogonal surround. For this level of analysis, we can simply average over the various center orientations to take the mean firing rate in each of the 3 conditions listed above.

These average responses are plotted for the 25 units from monkey R that contributed to this data set in Figure 3.3. The top row of plots compares responses to the various stimulus configurations separately for control (red points) and inactivation (blue points). Overall, responses are noticeably decreased in the presence of an iso-oriented surround compared to the center-only responses (Figure 3.3A), indicated by the strong tendency for points to fall below the line of unity. Importantly, this same level of suppression is not observed in the presence of an orthogonal surround (Figure 3.3B), as evidenced by the observation that average responses to orthogonal surround stimuli are almost always larger than average responses to iso-oriented surround stimuli (Figure 3.3C).

The middle row of Figure 3.3 illustrates the effect of feedback inactivation on the responses within each of these 3 conditions. Feedback inactivation does not systematically affect the overall responsiveness to center-only stimuli; the average response differences between control and inactivation are not significantly different from 0 ($p > 0.5$, Figure 3.3D). Interestingly, there appears to be a systematic increase in responses to center+surround stimuli (Figure 3.3E-F, see

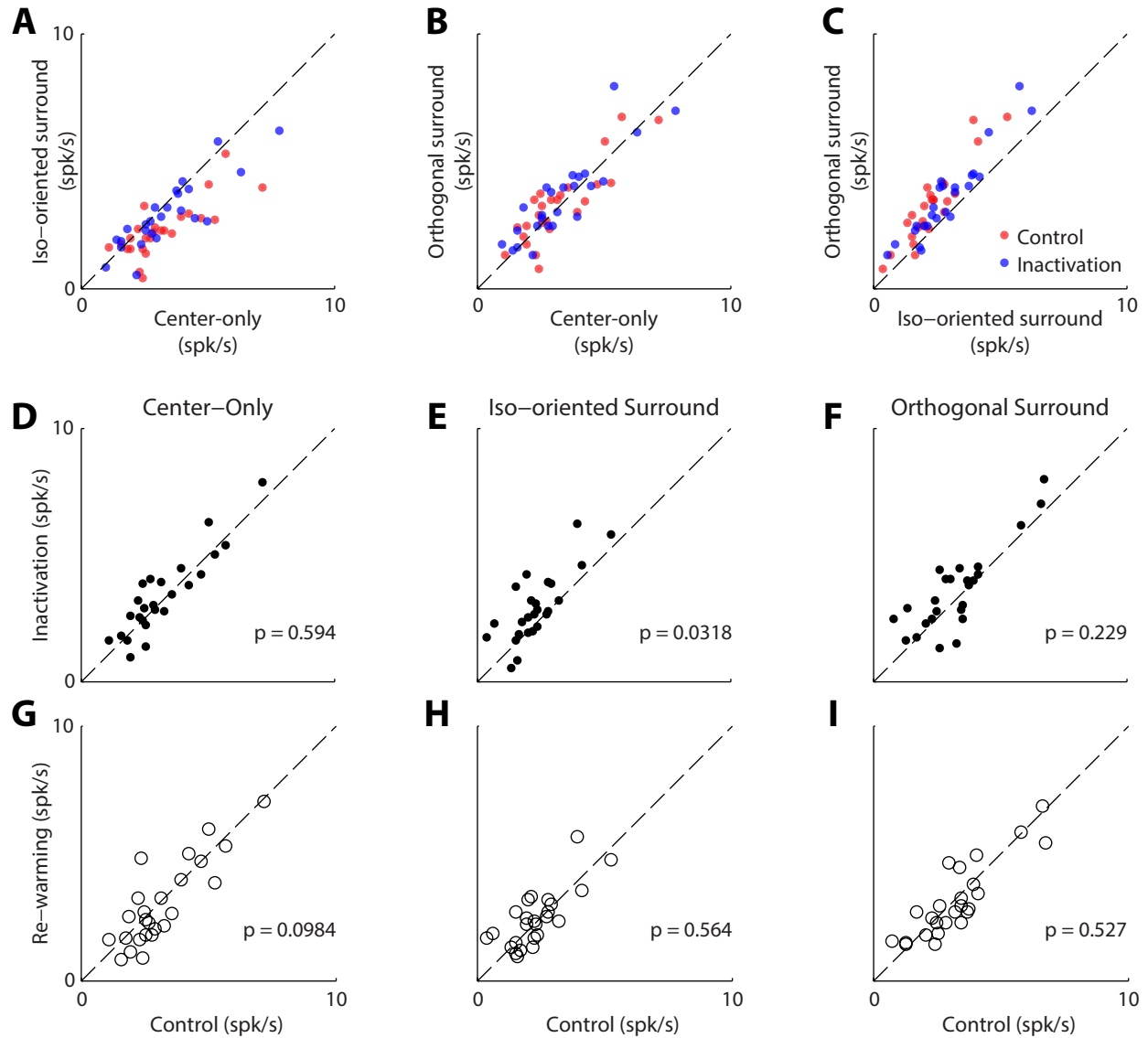


Figure 3.3 | Firing rate changes during control, inactivation, and re-warming: Monkey R. **A-B**, modulation of center-only responses by an iso-oriented surround (A) and an orthogonal surround (B). **C**, comparison of responses to iso-oriented and orthogonal center+surround stimuli. **D-F**, effect of inactivation on responses to center-only stimuli (D), iso-oriented stimuli (E), and orthogonal stimuli (F). **G-I**, same as D-F but comparing control and re-warming activity. All firing rates were measured over the sustained portion of the trial, as defined for monkey R. To generate p-values, we represented rate changes using the difference over the sum (which prevents high-firing units from biasing the variance) and performed a t-test according to the null hypothesis the mean of the resulting distribution was 0.

also Figure 3.4), corroborating previous reports from our lab that feedback inactivation weakens surround suppression. The data described here, however, provides an unprecedented view into the effects of inactivation by measuring these effects under a variety of center/surround configurations. Indeed, the effect of inactivation on responses to iso-oriented stimuli ($p = 0.032$, Figure 3.3E) appears to be more pronounced than the effect on orthogonal surround ($p = 0.23$, Figure 3.3F). This observation suggests that feedback may play a role in generating this property of surround suppression. Importantly, these effects are not seen during the re-warming portion of the experiment (Fig 3.3G-I), further supporting such a role for feedback.

To further explore this possibility, we examined how the average responses within each of the center/surround configurations evolved over time. To visualize this evolution, we measured spike counts across each of the 6 center/surround configurations within a sliding Gaussian window (see 3.2.2 for details). Within each window, spike counts for each unit were normalized to the average center-only response of that unit during the entire sustained period. Figure 3.4A-B illustrates the average time course of activity for the control and inactivation, separated into center-only and center+surround data. The reduced suppression during inactivation is illustrated by the consistently higher normalized rate during the sustained portion of the response (Figure 3.4B, right). When looking at the time course of responses for each of the 6 configurations (Figure 3.4C, bottom), one can see that match suppression (that is, suppression that comes from the center and surround having no orientation difference) is visible very early on in the response, especially for the control data. Though somewhat difficult to see because the color range of these plots is clipped to highlight the sustained response, even the transient responses show match suppression.

The configuration-specific effect of the surround over time is further illustrated by Figure 3.4D,

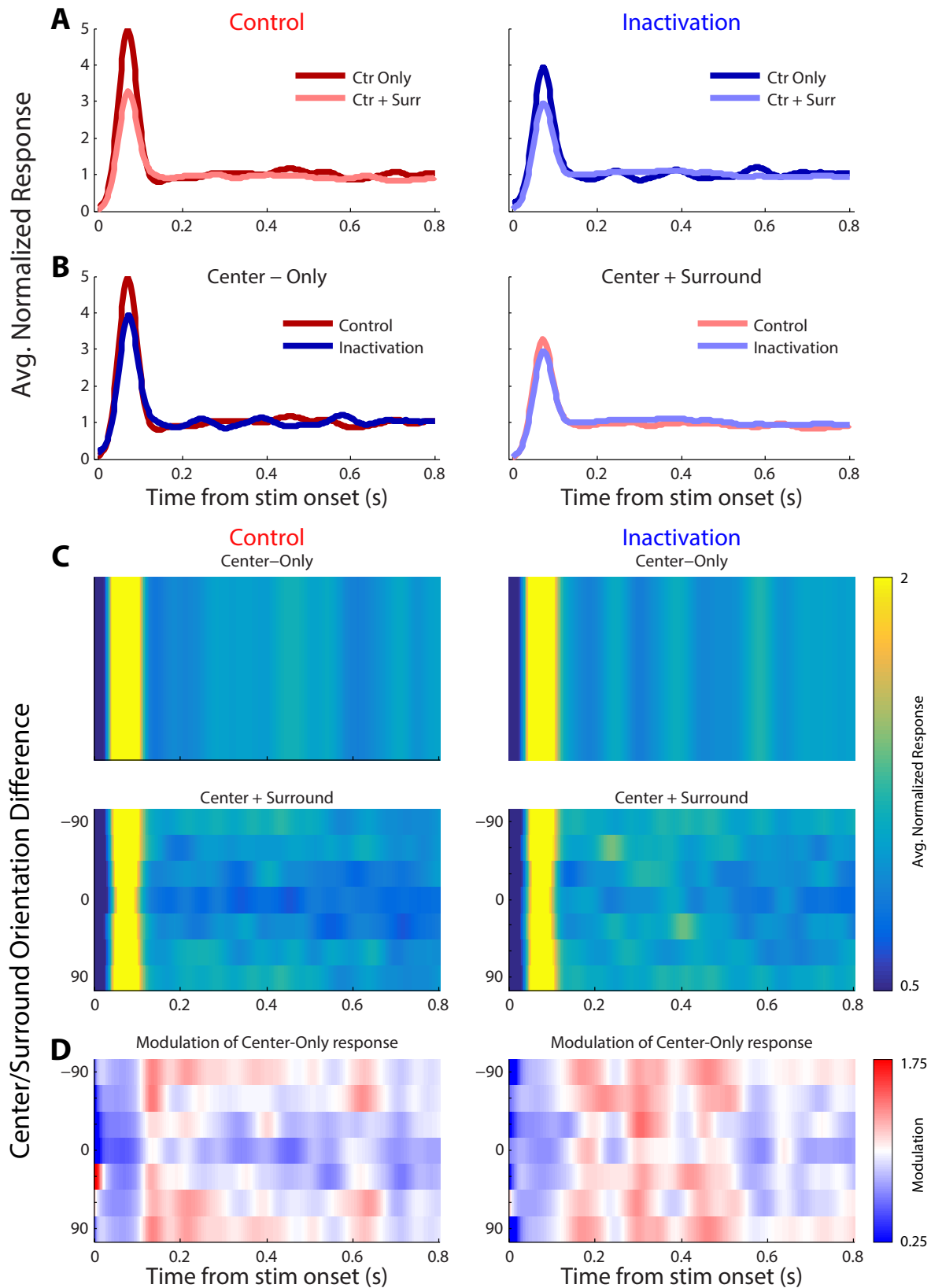
which represents the effect of the surround as a modulation of the center-only responses. This visualization is obtained simply by dividing the time-varying center+surround responses (C, bottom) by the time-varying center-only response (C, top). Here, blue represents a suppressive influence of the surround, red facilitation, and white indicates no effect. The surround modulation of the control responses clearly depends on the relative orientation between the center and the surround. Even as the overall influence of the surround changes (especially in the earliest portion of the response), suppression (or the lack of facilitation) is concentrated around configurations with small center/surround orientation differences (Figure 3.4D, left).

In contrast, this pattern is harder to identify within the inactivation data (Figure 3.4D, right). This may be due, in part, to dividing the center+surround responses by the fluctuating center-only response, but close visual inspection reveals a more significant change. These modulation maps show how the influence of the surround depends on the center/surround orientation difference at each point in time. Qualitatively, this dependency appears less consistent across time for the inactivation data compared to control.

Furthermore, this visualization reveals what appear to be slow changes in the strength of match suppression in the control data. To confirm this intuition, we separated the trial into the 3 epochs: the onset transient, early sustained, and late sustained. We wanted to ask whether/how the relationship between center/surround configuration varied across these epochs and whether the effects of inactivation displayed any similar variability. To minimize the influence of random variations in the center-only responses, we normalized both sustained portions by the center-only responses measured across the entire sustained period; we normalized the transient responses by its time-matched center-only data.

Figure 3.4 (following page) | Time course of average responses measured in the center-only/center+surround data set using a Gaussian sliding window: Monkey R. **A**, population response profiles averaged across all center-only (dark curve) and center+surround (light curve) stimuli, for control (left) and inactivation (right). **B**, same data as in A, organized to compare control and inactivation for each of the two stimulus subsets. **C**, bottom: population response profiles separated according to center/surround configuration (i.e. orientation difference); top: center-only response profile (same as in A) expanded into a false-color map to illustrate calculation of surround modulation in D. **D**, configuration-specific modulation of center-only responses over time, calculating by dividing the center+surround responses (C, bottom) by the center-only responses (C, top). This analysis used a Gaussian window with a standard deviation of 20 ms. All units' responses are pre-normalized to their average firing rate over the sustained period before being pooled for analysis.

Figure 3.4 | (Continued)



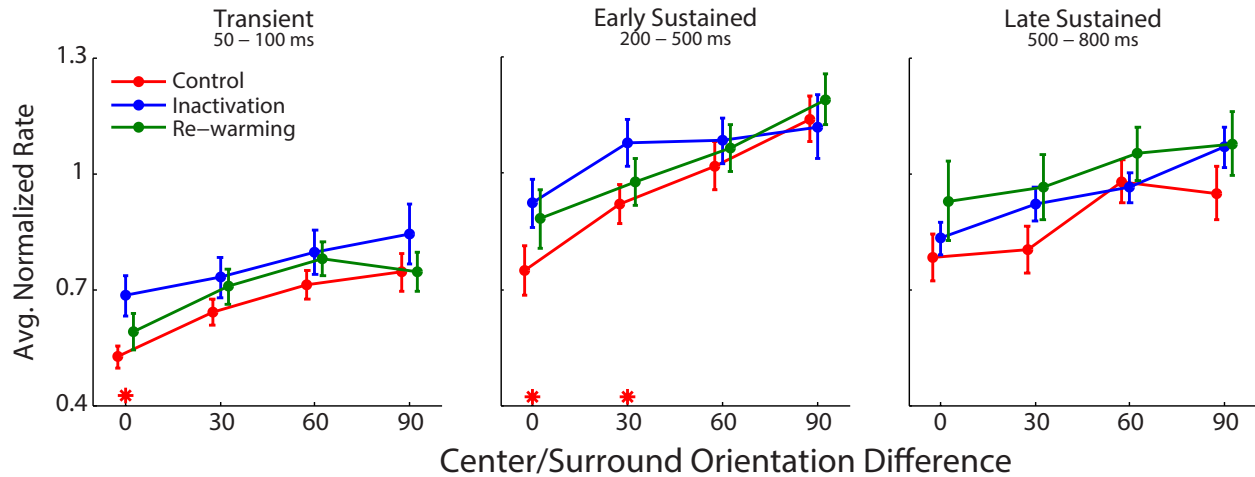


Figure 3.5 | Suppression of responses by matching surrounds across three response epochs: Monkey R.

Responses were separated according to the absolute orientation difference between the center and surround and averaged within each condition after pre-normalizing. Transient responses (left) were pre-normalized by the center-only responses within the same window. Early (middle) and late sustained (right) responses were both pre-normalized by the center-only responses across the combined window (200–800 ms). Error bars represent SEM. Asterisks along the abscissa indicate a significant difference ($p < 0.05$, paired t-test). The outcome of control vs. inactivation is plotted in red, re-warming vs. inactivation in green, and control vs. re-warming in black. The latter two comparisons were never significant.

Figure 3.5 shows the average normalized responses within each of these epochs as a function of the absolute orientation difference between the center and the surround. The depth of modulation (that is, how much responses change as a function of orientation difference) varies over time in the control data and, to a lesser extent, the re-warming data. For both control and re-warming data, the depth of modulation is greatest during the early sustained period (Figure 3.5, middle). For the control data, the most notable change over the course of the center+surround responses is a loss of facilitation when the surround is orthogonal to the center (Figure 3.4; Figure 3.5, compare early to late sustained). The resulting decrease in modulation depth can also be seen for the re-warming data. In contrast, the relationship between center/surround orientation difference and surround modulation is relatively stable during inactivation, only exhibiting changes in offset.

These data suggest that the richness of surround modulation decreases during inactivation. Feedback inactivation results in a clear overall decrease in the amount of surround suppression

but also manifests as changes in the way that more sophisticated modulatory mechanisms, like match suppression, are deployed. This is most evident when considering how the effect of feedback inactivation depends on center/surround configuration. This interaction is most visible during the early sustained portion of the response (defined as 200-500 ms after stimulus onset; Figure 3.5, middle), where feedback inactivation does not simply result in a global increase in responses to center+surround stimuli but instead diminishes the suppression from surrounds that are oriented similar to the center. This is highlighted by the statistically significant changes between control and inactivation responses to iso-oriented stimuli during the transient and early sustained periods (Figure 3.5).

The responses observed during re-warming are perhaps best described as an intermediate between control and inactivation. This may reflect that the reversibility of inactivation is not entirely immediate but also suggests that the effects of inactivation are not artifacts created by recording the inactivation data later in the experiment. When analyzing the data from monkey V, we see a much clearer reversal of inactivation effects during re-warming, speaking to the fact that inactivation itself does not introduce lasting physiological changes.

3.4 Monkey R: Effects of feedback inactivation on feature-specific suppression

In Chapter 2, we provided evidence that a feature-specific form of modulation can actually explain match suppression. According to this explanation, it is not the matching, per se, that produces suppression; instead, suppression effectively targets the features of the center that are also in the surround. When the center stimulus contains a single orientation the influence of the surround is

only revealed when the center orientation is similar to that in the surround. On the other hand, a consistent, feature-specific influence of the surround is seen when the center contains a plaid, drawn by superimposing two oriented gratings where the orientation of one of the gratings is always matched to the surround. This influence is better understood as suppressing the representation of the matched component rather than recruiting suppression depending on the overall similarity between the center and surround. Feedback has been shown to play a role in generating surround suppression and the preliminary data from monkey R suggests that feedback also contributes to match suppression. Therefore, a natural next question is whether this effect of feedback inactivation reflects a role of feedback in controlling input gain through feature-specific modulation.

To address this question, we collected a different set of preliminary data in monkey R using the plaid stimuli introduced in Chapter 2, which are instrumental in characterizing the specificity of suppression. We employed a variation on the plaid stimuli where the unmatched center component (C_1) was drawn with a Michelson contrast of 33% and the matched component (C_2) was drawn with 66% contrast. The rationale for this choice was that it would result in the two components having roughly equal effective contrasts in the presence of a surround, providing an ideal baseline for observing the direction of influence of inactivation. As more experiments were done to characterize this form of suppression, it became clear that interpreting the data was more straightforward when starting from equal contrasts. However, we did not have the chance to repeat these experiments in this animal with 50/50 contrast plaids. Still, it is worthwhile to explore this data where possible and acknowledge caveats where they arise.

The fitting procedure introduced in Chapter 2 provides an ideal tool for quantifying the patterns of suppression associated with this set of stimuli. To recap, the plaid-only and plaid+surround

data are each grouped according the orientation difference between the two center components. Within each group, responses are organized according to each unit's preferred orientation relative to the unmatched, C_1 orientation. Arranging the data this way creates 12 population responses, one for each of the 6 C_1/C_2 configurations for both the plaid-only and plaid+surround data. Our fitting procedure imagines these population responses as the weighted sum of 2 underlying responses, one driven by C_1 and the other driven by C_2 . We describe each underlying response as a "template" that is aligned to the orientation of its respective component. The template itself is created by initially fitting the scale, width, and offset of a circular Gaussian to the plaid-only population response representing the conditions where C_1 and C_2 are identical. From here, the process of fitting each of the remaining population responses is to simply choose the best fitting weights of the C_1 -aligned and C_2 -aligned templates. The resulting weights provide a condensed description of the data and, as such, the influence of the surround. It should be noted that, unlike in Chapter 2, we do not include an offset term during the fits. This choice is made to simplify the comparison of weights across conditions and units.

The population responses and their associated fits are illustrated in Figure 3.6 for the control (top) and inactivation data (bottom). The plaid-only responses are clearly biased towards the C_2 component, as expected given its stronger contrast. The control data display feature-specific suppression, as the main effect of the surround is to suppress responses driven by the matching C_2 component, shifting population responses towards the C_1 orientation (which is always aligned to 0 in the plots). The fits to the control responses provide a quantitative basis for this claim: the weights assigned to the unmatched component were only slightly affected by the surround (a small increase on average), whereas the weights assigned to the matched component were, on

average, roughly halved (Figure 3.7A).

It is readily apparent that the overall influence of the surround has not drastically changed in the inactivation data (Figure 3.6, bottom; Figure 3.7B). By the same criteria used with the control data to identify feature-specific suppression, it is also present during inactivation. However, we can attempt to address the direction of change, if any, caused by inactivating feedback. The specificity of suppression is captured by the differential effect of the surround on the matched and unmatched weights. Similarly, evidence for the involvement of feedback can be evaluated based on whether the effect of its inactivation also shows component specificity. Figure 3.7C shows the average change in the unmatched and matched weights from plaid-only to plaid+surround for the control and inactivation data. For both data sets the surround contributes to an increase in the unmatched weights and a decrease in the matched weights. For the inactivation data, the decrease in the matched weights is weaker than seen in the control data, suggesting a loss of suppression, but there does not appear to be an accompanying increase in the unmatched weights, suggesting a feature-specific loss of suppression. Taken at face value, this is the direction of effect one would expect to see if feedback contributes to this form of modulation. However, the effects of the surround on each component are not significantly different between control and inactivation (Figure 3.7C).

Therefore, these data are mostly inconclusive. This is due in large part to the difficulty in quantifying small changes in feature-specific suppression within the responses elicited by plaids configured with this contrast setting. Unfortunately, this caveat was only obvious in hindsight, after the opportunity to recollect this data using better-optimized stimulus parameters was lost. This same difficulty renders more detailed analyses, like those examining the time course of these

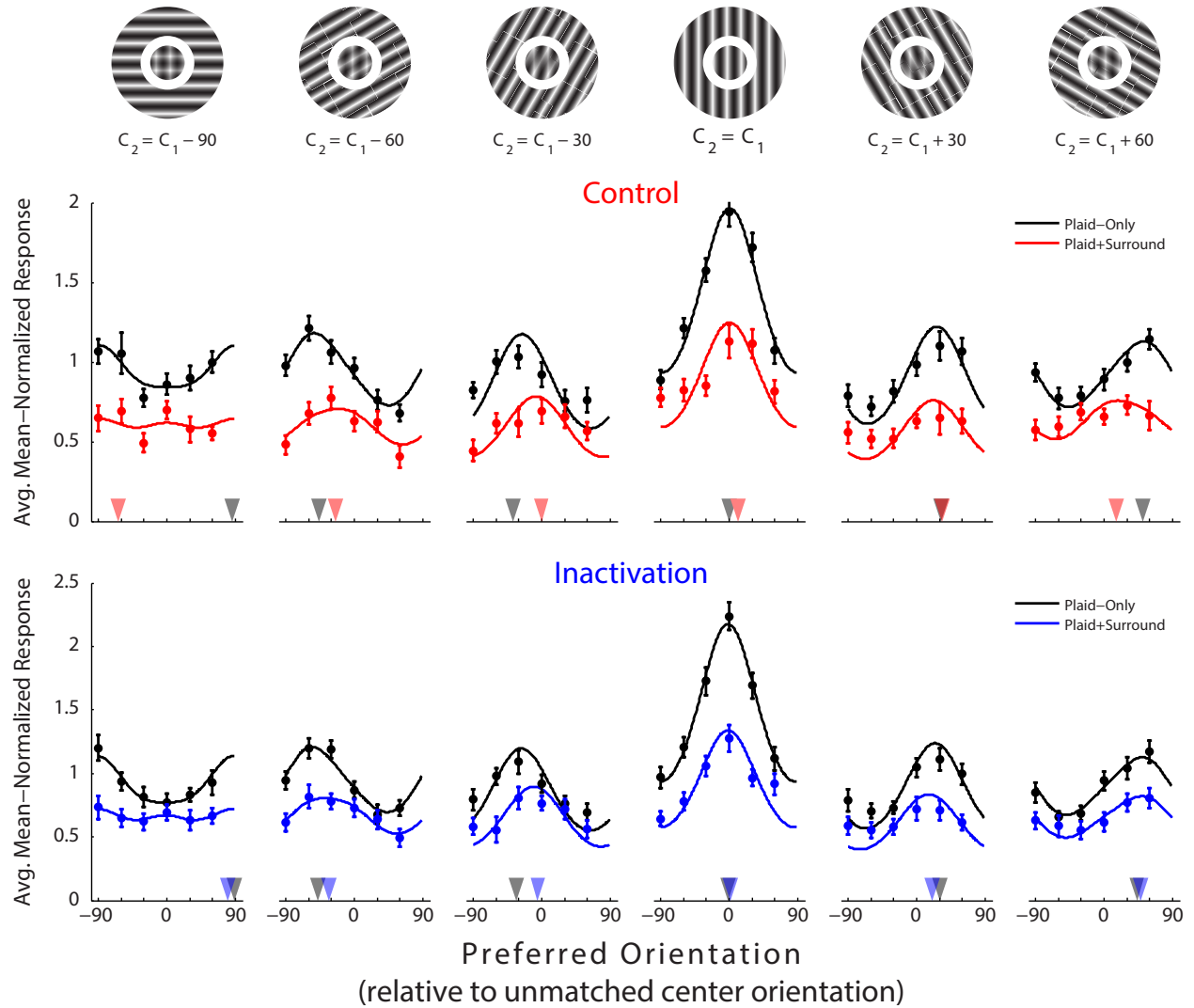


Figure 3.6 | Population responses to plaid-only and plaid+surround stimuli during control (top) and inactivation (bottom): Monkey R. Conventions adopted from Figure 2.3. Each panel represents the pair of population responses measured in response to stimuli with the labeled C_1/C_2 configuration. Plaid-only responses and their associated fits are plotted in black and the plaid+surround data are represented in color. Individual points indicate actual data and error bars indicate SEM. Solid lines describe the fits produced by the best-fitting weights (see main text). Triangles along the abscissa point to the orientation of the population response mean vector.

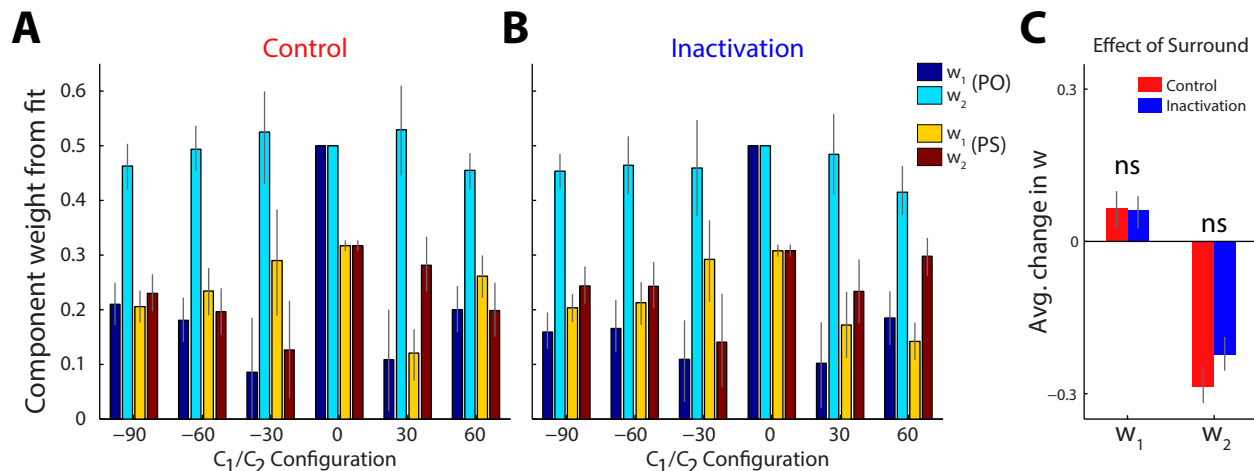


Figure 3.7 | Best-fitting weights applied to the population responses to plaid stimuli in monkey R and the effect of inactivation. **A**, the pairs of weights used to fit the control plaid-only and plaid+surround data are depicted for each of the C_1/C_2 configurations, indicated along the abscissa. Error bars represent bootstrapped SEM, calculated by repeating the fitting procedure many times on pseudopopulations representing random resamplings (with replacement) from the underlying units. **B**, same as in **A** for inactivation data. **C**, each bar represents the change in average C_1 or C_2 weight caused by the addition of a surround during control and inactivation. Error bars represent bootstrapped SEM. Significance was assessed by a permutation test performed by randomly swapping each unit's control and inactivation data.

effects, beyond useful interpretation. These data do provide a useful reference point when considering the results from monkey V (below). In addition, they provide some weak evidence that feedback is involved in feature-specific suppression in monkey R. The question of whether the putative contribution of feedback to match suppression (Section 3.3) reflects a role in generating feature-specific modulation cannot be answered conclusively with this data.

3.5 Monkey V: Effects of inactivation on match suppression

Based on the number of studies reporting what we refer to as match suppression and the ease with which we have observed it in the past (Sections 2.3.1 and 3.3), we expected it to be readily observable in the V1 activity of monkey V. Instead, match suppression in monkey V was often surprisingly weak, at least during sustained responses. In contrast, responses measured during the onset transient were more consistent with previous observations. Indeed, the neural responses, their

modulation by the surround, and the influence of feedback inactivation are very much characterized by how they evolved over the time course of stimulus presentation. This temporal richness exceeds what we were able to observe in monkey R, reflecting perhaps a difference in overall signal quality or perhaps a difference in the source (i.e. layer) of the signal. Connecting empirical observations to such a difference is a topic of discussion revisited later on (Section 3.8). First, it is necessary to characterize the time-varying responses measured in monkey V, with a focus on how (and when) they exhibit match suppression.

One of the more striking features of the responses measured in monkey V is plainly visible with a simple spike raster plot (Figure 3.8A). The three panels in Figure 3.8A correspond to three example units recorded on separate days. The trials (rows) are color coded according to the portion of the experiment during which they were measured. One recognizes the transition from spontaneous activity to the onset transient roughly 50 ms after stimulus onset. The following transition to sustained activity commonly involves a second, intermediate peak in activity around 100 ms after stimulus onset. Interestingly, this transition appears noticeably altered during feedback inactivation but is clearly visible again during re-warming (Figure 3.8A). The structure and ubiquity of the changes affected by feedback inactivation are further evident when examining the normalized PSTH plots of all 39 units contributing to this data set, separated according to the presence of a surround and inactivation state (Figure 3.8B). These plots reveal a consistent interaction between the surround and feedback inactivation in shaping responses as activity settles towards a sustained rate. The responses measured in monkey R, in contrast, do not show any obvious phasic dynamics beyond the typical onset transient (Figure 3.8C).

Because these responses were so consistent in monkey V, we defined the sustained period of

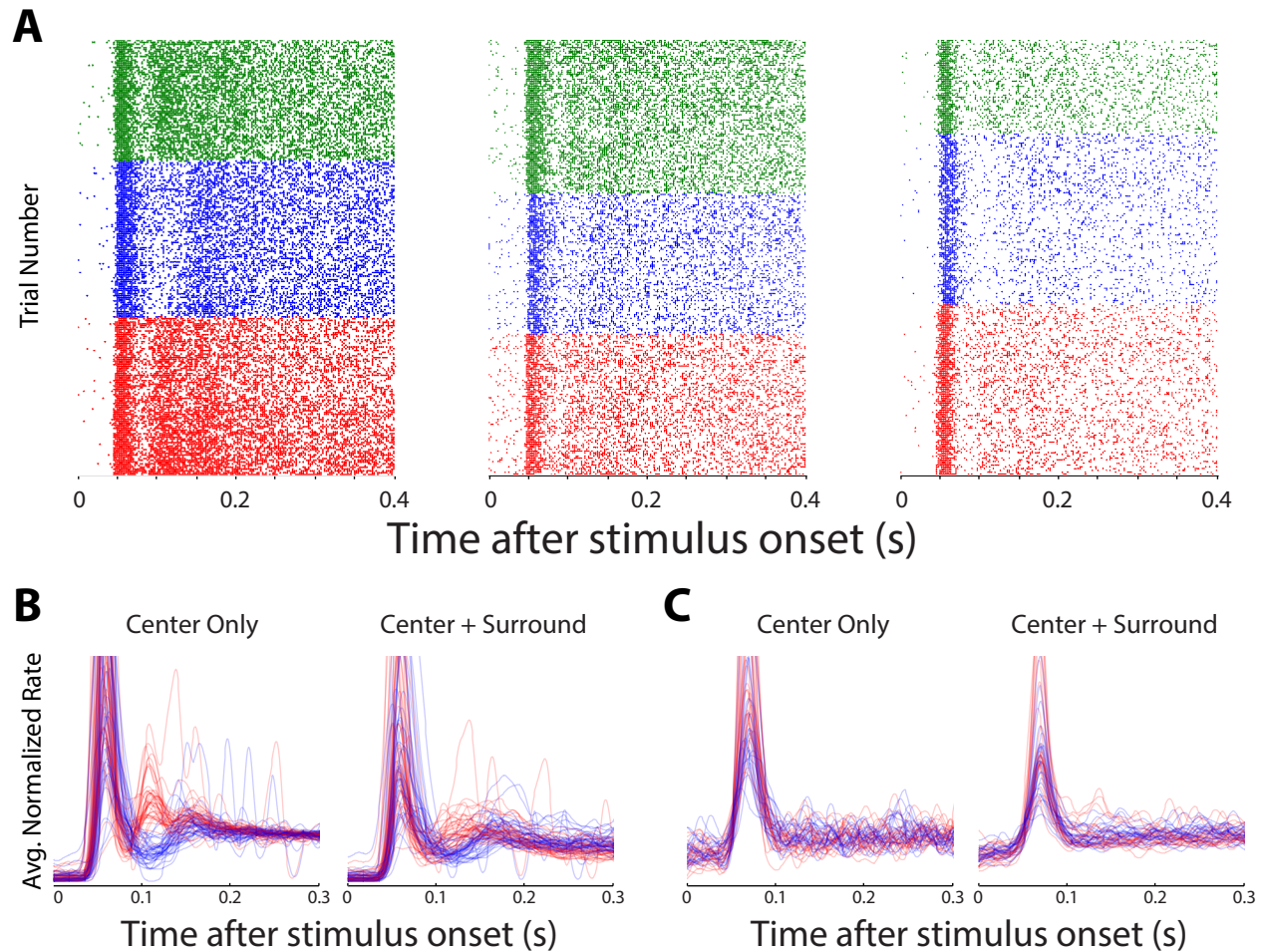


Figure 3.8 | Effects of the surround and feedback inactivation on basic response dynamics of individual units in monkey V. **A**, peri-stimulus spike raster plots for 3 example units, measured on different days. Each row represents a completed trial and each dot indicates the timing of a spike. Control, inactivation, and re-warming trials are concatenated (not reflecting timing gaps) and color-coded using the established convention. **B**, center-only and center+surround PSTH plots for each unit recorded with monkey V during control (red) and cooling (blue). Responses are normalized by their average center-only activity from 250–400 ms. The last 100 ms of response are omitted to make the early responses clearer. **C**, same analysis as in B for monkey R. Spike times were smoothed with a Gaussian sliding window (SD = 5 ms) for B and C.

activity as beginning at 250 ms after stimulus onset. To characterize match suppression once the activity had reached this relatively steady rate, we compared firing rates evoked by center-only stimuli, iso-oriented surrounds, and orthogonal surrounds during the sustained epoch. As expected, responses to iso-oriented surround stimuli were suppressed relative to center-only responses (Figure 3.9A). However, control responses exhibited a nearly identical amount of suppression to orthogonal stimuli (Figure 3.9B-C), which was not expected. As a further departure from our expectations, feedback inactivation led to a clear and consistent reduction in center-only responses ($p < 0.001$; Figure 3.9D). Inactivation also resulted in a consistent decrease in the response to iso-oriented center+surround stimuli ($p = 0.024$) and had little effect on orthogonal stimuli (Figure 3.9E-F). However, the decrease in responses to iso-oriented stimuli should not be construed as a change in the modulation produced by these stimuli, as the relationship between responses to center-only and iso-oriented stimuli is almost indistinguishable between control and inactivation (Figure 3.9A). In contrast, there is a very noticeable difference between this relationship when considering orthogonal stimuli (Figure 3.9B). The comparison between control and re-warming firing rates across these conditions speaks to the reversibility of inactivation. During re-warming, center-only and iso-oriented center+surround responses returned to normal (Figure 3.9G-H). There did appear to be a small increase in responses to orthogonal surrounds during re-warming, compared to control (Figure 3.9I), but there was a strong qualitative agreement between the modulation patterns seen during control and re-warming in monkey V (see below), suggesting that this minor change was insubstantial.

These data would seem to suggest that, within the sustained period of activity, control responses are not affected by match suppression. This does not stem from a lack of suppression but an

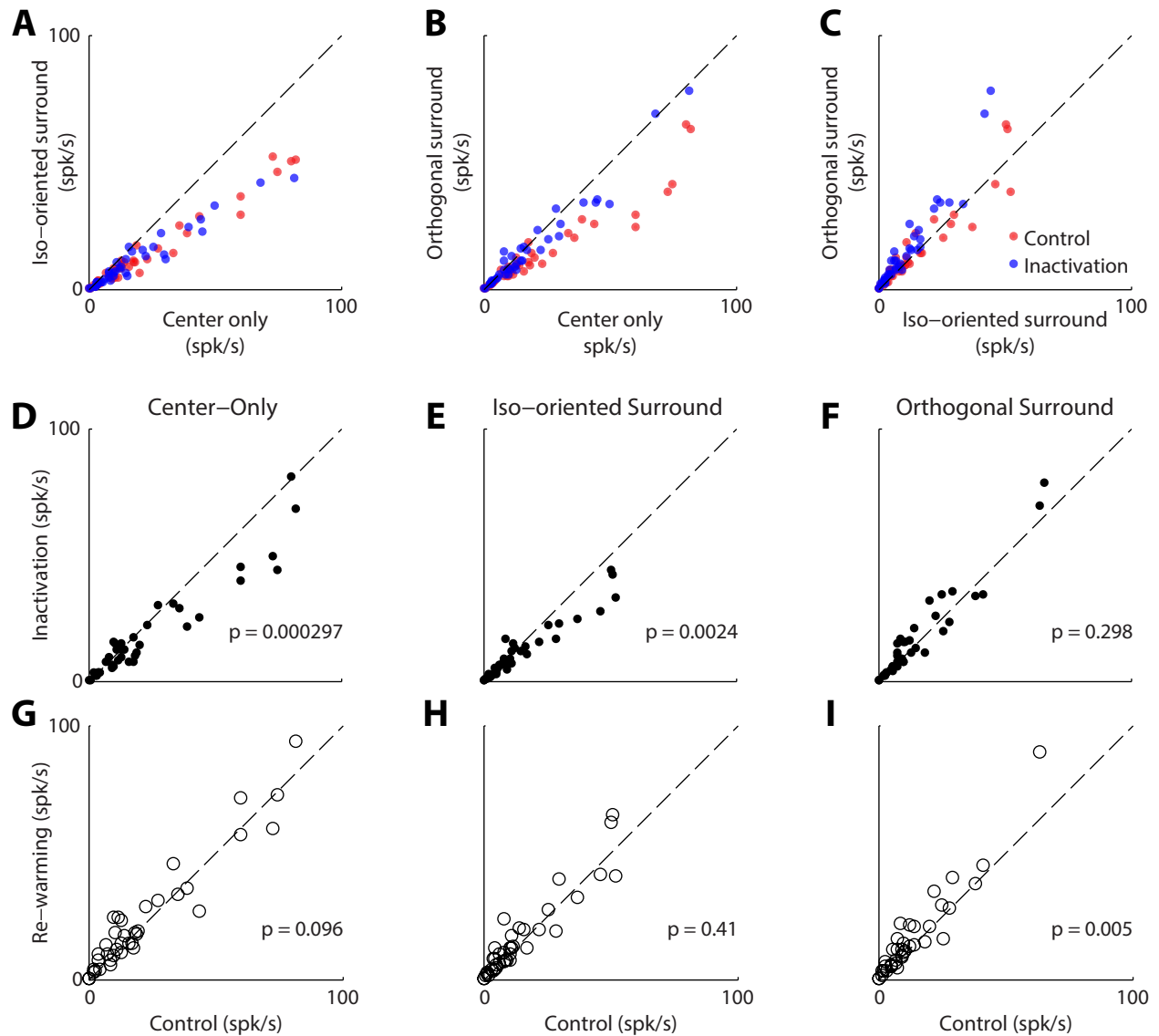


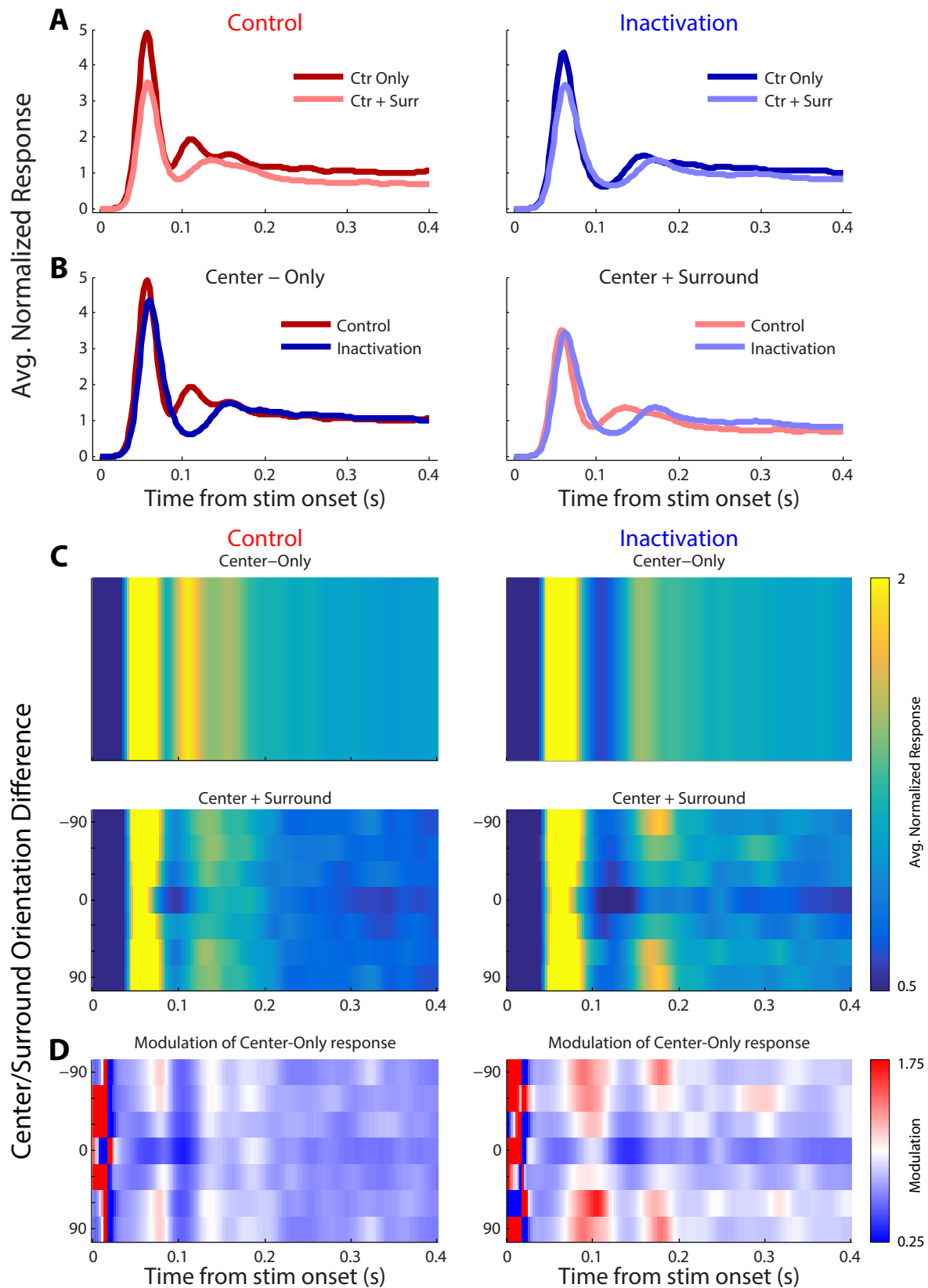
Figure 3.9 | Firing rate changes during control, inactivation, and re-warming: Monkey V. Conventions are identical to Figure 3.3. **A-C**, component-specific modulation of responses by the surround. **D-F**, effects of inactivation on responses to each group of stimuli. To generate p-values, we represented rate changes using the difference over the sum (which prevents high-firing units from biasing the variance) and performed a t-test according to the null hypothesis the mean of the resulting distribution was 0. **G-I**, same as D-F, comparing control to re-warming data.

abundance of it—the atypical feature of the control data is the amount of suppression associated with orthogonal surrounds. From that standpoint, feedback inactivation produces more “typical” responses. Normalizing to center-only responses within each condition, feedback inactivation has the effect of reducing the overall amount of surround suppression and that reduction tends to be configuration-specific (Figure 3.9D-F). In this particular regard, the data from monkey R and monkey V agree. However, the details differ considerably.

This description of the center-only/center+surround data (typically used to characterize match suppression) in monkey V pertains to the sustained activity. We were interested in whether the same description would apply throughout the time course of the response. To get at this question, we applied a sliding window analysis, similar to that used in Section 3.3 to this subset of the data. All recording sessions used a stimulus duration of at least 400 ms, so we limit the present analysis to responses measured within that window. Figure 3.10A-B illustrates the average time course of activity for the control and inactivation, separated into center-only and center+surround data. This representation nicely illustrates the stimulus- and feedback-dependent response dynamics during the transition into sustained activity and also that suppression during inactivation is reduced, as shown by the consistently higher normalized rate during the sustained portion of the response (Figure 3.10B, right).

Figure 3.10 (following page) | Time course of average responses measured in the center-only/center+surround data set using a Gaussian sliding window: Monkey V. Conventions are identical to Figure 3.4. **A**, population response profiles averaged across all center-only (dark curve) and center+surround (light curve) stimuli, for control (left) and inactivation (right). **B**, same data as in A, organized to compare control and inactivation for each of the two stimulus subsets. **C**, bottom: population response profiles separated according to center/surround configuration (i.e. orientation difference); top: center-only response profile (same as in A). **D**, configuration-specific modulation of center-only responses over time. This analysis used a Gaussian window with a standard deviation of 5 ms.

Figure 3.10 | (Continued)



Separating the data according to center/surround configuration (that is, orientation difference) further reveals an interesting pattern of modulation over time (Figure 3.10C-D). Consistent with the above analysis of unnormalized firing rates, the pattern of activity that describes match suppression was quite faint over the last 150-200 ms of the stimulus presentation under control conditions but was plainly visible during inactivation. However, this pattern was present during the earlier responses in the control data (50-150 ms; Figure 3.10C-D, left).

We separated the trial into 3 epochs reflecting the qualitative response phases identified in the PSTH plots: onset transient (50-100 ms), post-transient (100-200 ms), and sustained (200-400 ms). Within each epoch, we compared average normalized responses as a function of the absolute center/surround orientation difference (here, we normalized responses to center-only responses within the overlapping time window). Under control conditions, the modulation depth of match suppression diminishes heavily; it is heavily configuration-dependent during the transient and nearly entirely uniform during the final epoch (Figure 3.11). This change reflects a buildup of suppression of responses to orthogonal configurations, and, to a lesser extent, loss of suppression of responses to iso-oriented stimuli (Figure 3.11). Interestingly, the modulatory influence of orthogonal surrounds, which change the most over the time course of control responses, create the largest contrast between control and inactivation (Figure 3.11). These patterns are also evident when examining the full time course of modulation (Figure 3.4D). This interaction between match suppression and feedback inactivation in monkey V stands in rather stark contrast to what was observed in monkey R (Figure 3.5). It is also worth pointing out that there is a near complete reversal of the inactivation effects during re-warming. The re-warming data agreed very closely with the control data (Figure 3.11), supporting our claim that the inactivation effects were a

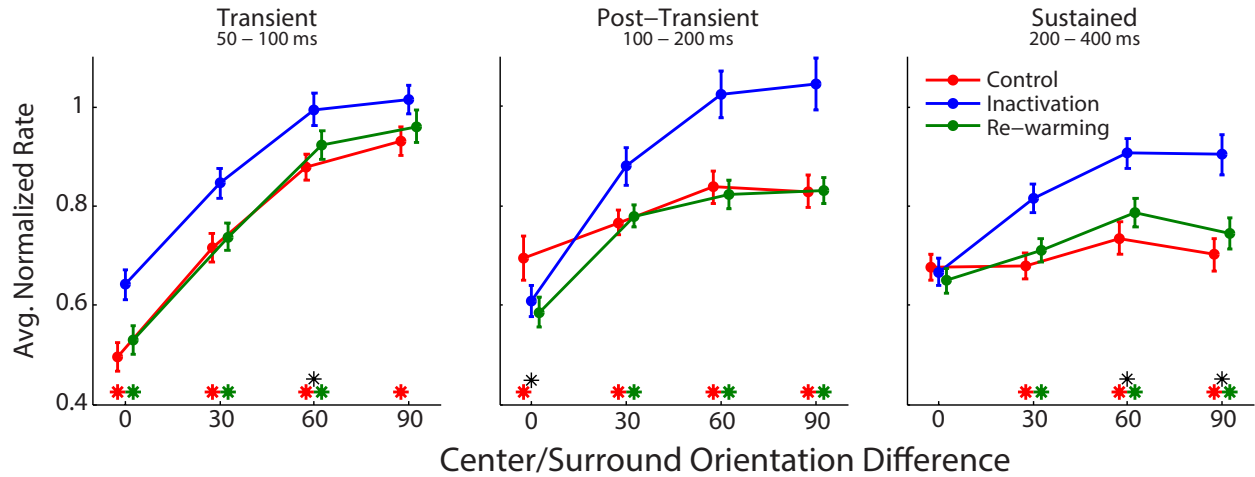


Figure 3.11 | Suppression of responses by matching surrounds across three response epochs: Monkey V. Responses were separated according to the absolute orientation difference between the center and surround and averaged within each condition after pre-normalizing. All responses were pre-normalized by the center-only responses within the same window. Error bars represent SEM. Asterisks along the abscissa indicate a significant difference ($p < 0.05$, paired t-test). The outcome of control vs. inactivation is plotted in red, re-warming vs. inactivation in green, and control vs. re-warming in black.

consequence of the temporary loss of feedback projections from V2/V3 to V1. Having demonstrated this reversibility here and for the sake of brevity, we focus only on the control and inactivation data for the remainder of the Chapter.

3.6 Monkey V: Effects of feedback inactivation on feature-specific suppression.

While collecting the data to measure match suppression, we also collected data to measure feature-specific suppression. This offers the advantage of being able to compare these forms of modulation (and their feedback dependency) within the same group of units. To examine feature-specific modulation, we isolated the subset of data recorded when the center stimulus was drawn as a visual plaid and apply the fitting procedure described above (Section 3.4). Figure 3.12 and Figure 3.13 illustrate the population responses and accompanying fits and the weights

describing those fits, respectively, for the data measured during the sustained period. As reported previously in Chapter 2 and in Section 3.4, the surround produces feature-specific suppression, as captured by the population responses shifting their peaks towards the unmatched orientation (aligned to 0 degrees) and the reduced weight assigned to the matched component in the plaid+surround data.

While the current chapter has thus far treated the modulation patterns measured with a single-grating center (“match suppression”) and with a plaid-center grating (“feature-specific suppression”) separately, they are described as stimulus-dependent manifestations of the same input-gain control mechanism (Chapter 2). With this idea in mind, the observation of feature-specific suppression during the sustained activity in the control data is somewhat surprising considering the nearly uniform suppression across stimulus configurations observed with the single-grating stimuli (Section 3.5). Reconciling these observations cannot be done within the framework we have already proposed, prompting the question of how best to adapt our explanation of the surround effects so as to capture this data.

We can begin to answer this question by first asking how the inactivation of feedback influences the responses to plaid stimuli and the effect of the surround. The hallmark of feature-specific suppression revealed by the analyses employed thus far is the disparate effect of the surround on the weights assigned to the unmatched and matched center components. As such, we are interested in whether this disparity changes when feedback is inactivated. To address this, we can examine the best-fitting weights for the unmatched and matched components (w_1 and w_2 , respectively) when representing the plaid+surround population responses (specifically, for the 5 configurations where the component orientations differ). Interestingly, the weights of both

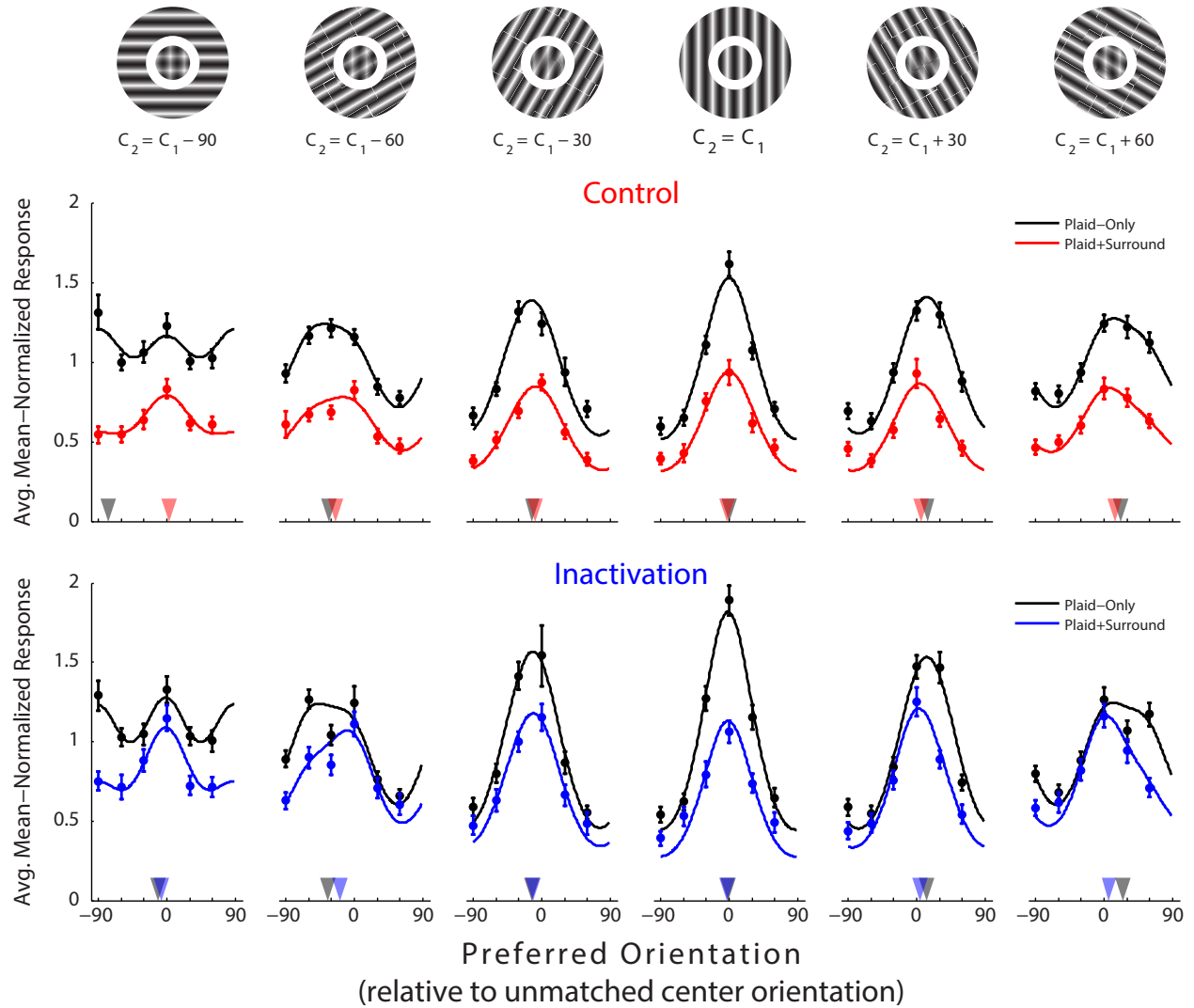


Figure 3.12 | Population responses to plaid-only and plaid+surround stimuli during control (top) and inactivation (bottom): Monkey V. Conventions adopted from Figure 2.3 and are identical to Figure 3.6. Each panel represents the pair of population responses measured in response to stimuli with the labeled C_1/C_2 configuration. Plaid-only responses and their associated fits are plotted in black and the plaid+surround data are represented in color. Individual points indicate actual data and error bars indicate SEM. Solid lines describe the fits produced by the best-fitting weights (see main text). Triangles along the abscissa point to the orientation of the population response mean vector.

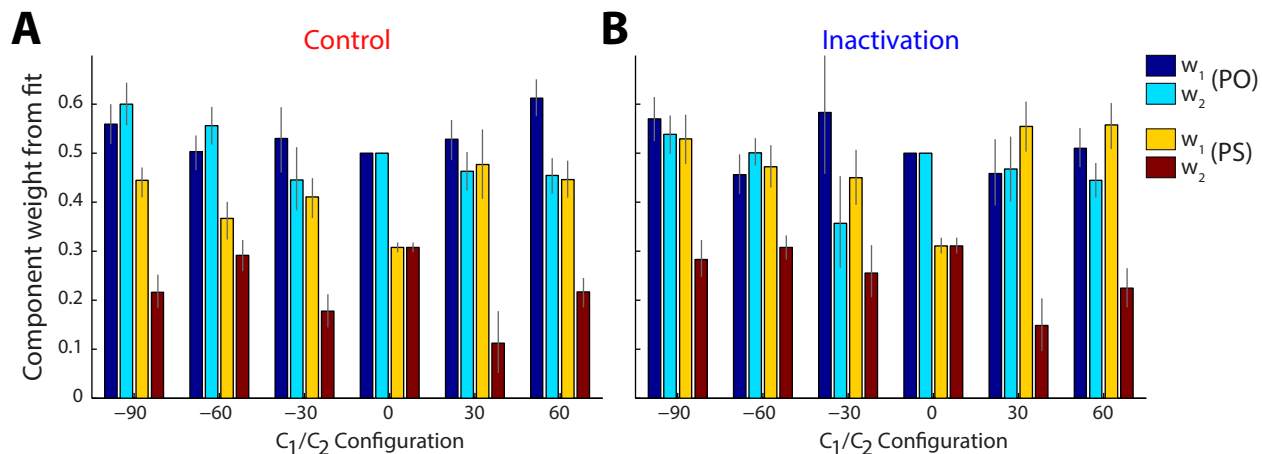


Figure 3.13 | Best-fitting weights applied to the population responses to plaid stimuli in monkey R and the effect of inactivation. Conventions are identical to Figure 3.7. **A**, the pairs of weights used to fit the control plaid-only and plaid+surround data are depicted for each of the C_1/C_2 configurations, indicated along the abscissa. Error bars represent bootstrapped SEM. **B**, same as in A for inactivation data.

components increased proportionately during inactivation. Figure 3.14A compares the best fitting plaid+surround weights during cooling and inactivation. Feedback inactivation seems to have a mostly multiplicative effect on the weights. Applying a simple linear regression model to quantify the scale and offset of these changes (pooling matched and unmatched weights) supports this intuition as it very nearly passes through the origin (scale = 1.138, offset = 0.0185; Figure 3.14A, cyan line). To further quantify this trend, we can take advantage of the relatively high signal quality in this data set and apply the fitting procedure to the each of the 39 units. Each unit is then represented by averaging its best-fitting unmatched weights and matched weights across the 5 relevant plaid+surround configurations (with averages denoted as $\langle w_1 \rangle$ and $\langle w_2 \rangle$, respectively), doing so separately for control and inactivation. Figure 3.14B shows the same type of comparison as panel A for the unit-by-unit analysis. Clearly, fits to individual units are much noisier but are consistent with the results from the pooled data, as they do not identify any systematically disproportionate change in one of the components during inactivation. To quantify this, we describe the unit-by-unit effect of inactivation on each center component using a weight

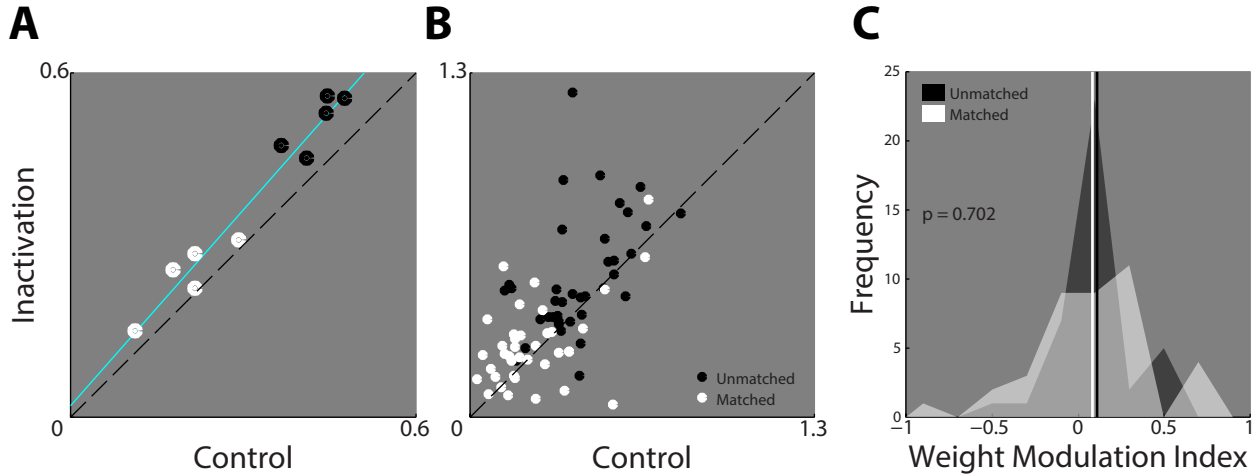


Figure 3.14 | Non-specific changes in component weights fitting plaid+surround data after feedback inactivation. **A**, Each data point indicates one of the ten weights used to fit the plaid+surround data (excluding $C_1=C_2$), showing the best-fitting value of that weight during control (abscissa) and inactivation (ordinate). The cyan line indicates the result of a linear regression fit to all ten points. The line of unity is plotted in black. **B**, similar to A but each data point represents the fits to the entire plaid+surround data for a particular unit. Each unit contributes 2 points (reflecting 4 average weights). The line of unity is plotted in black. **C**, WMI (see main text) distributions for the two plaid components. Vertical lines indicate distribution means. Significance assessed with a paired t-test.

modulation index (WMI), measured as the difference between $\langle w \rangle_{\text{inactivation}}$ and $\langle w \rangle_{\text{control}}$ divided by their sum, where $\langle w \rangle$ can refer to $\langle w_1 \rangle$ or $\langle w_2 \rangle$. The distributions of WMI values for the unmatched and matched weights are plotted in Figure 3.14C. WMI is a scale-invariant metric for quantifying change, resulting in nearly overlapping distributions that reflect proportionate modulations. The WMI for the matched component is more sensitive to noise, explaining its wider distribution. We performed a paired t-test and found no significant difference in the effect of inactivation on the unmatched and matched center weights ($p > 0.7$) and observe only the slightest of mean differences (Figure 3.14C).

These data do not reveal any feature-specificity to the effect of feedback inactivation in monkey V. However, feedback inactivation did weaken the suppressive influence of the surround (Figure 3.12-3.14), consistent with previous results. The fits to the plaid data suggest that this loss of suppression is better described as a multiplicative increase in response to plaid+surround stimuli.

As discussed in Chapter 1, a fundamental goal of this experiment is to identify the repertoire of contextual modulations to which feedback contributes. The data described here offer the crucial insight that the feature-specific modulatory mechanism producing input-gain control may be intrinsic to V1 and that feedback instead contributes to divisive inhibition. We wanted to explore the possibility that feedback mediates divisive modulation by asking how well this explanation could allow the responses to plaid+surround stimuli measured during control conditions to account for the responses to the same stimuli measured during inactivation.

To form an answer to this question, we introduced a modification to the procedure used to fit the population responses. Specifically, when fitting the plaid+surround population responses measured during inactivation, we no longer allowed the unmatched and matched weights to be fit directly but instead required the fits to take the form

$$[w_1, w_2]_{\text{inactivation}} = g \cdot [w_1, w_2]_{\text{control}}.$$

The bracketed weights on the right hand side indicate the pair of unmatched and matched weights used to fit the control data from a particular configuration. For each of the plaid+surround configurations, the unmatched and matched weights used to fit the inactivation data were obtained by fitting the scaling factor in the above equation, g , and were therefore expressed as scaled versions of the control weights. The control weights were optimized first and independently of the inactivation data. Figure 3.15 illustrates the fits produced by this modified procedure. The solid lines indicate the weighted sum of the two templates indicating the best-fit and the data points are the actual mean normalized responses. In general, scaled-weight fits provide excellent agreement with the actual inactivation data. To further illustrate this agreement, we included the

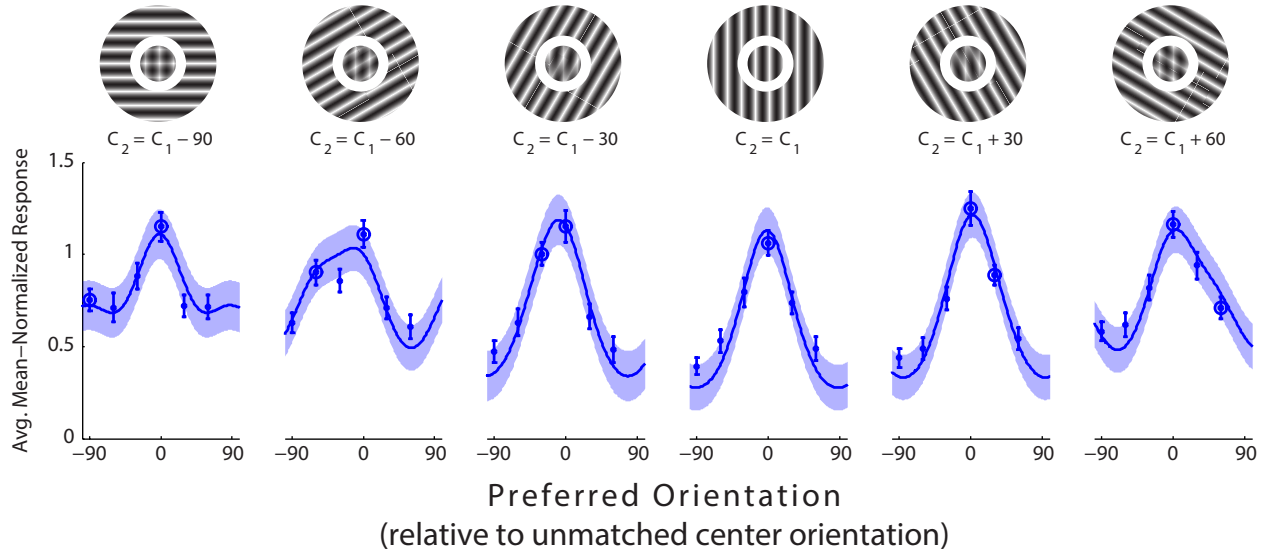


Figure 3.15 | Plaid+surround population responses fit with the scaled-weights modification. Conventions are similar to Figure Vince-Plaid-Pop. Population responses and fits of inactivation data, using the procedure that restricts weights to scalings of the configuration-matched weights in the control data (see main text for details). Shaded regions indicate 95% credible intervals. Solid lines indicate best fit. Actual data are plotted with error bars representing SEM. Open circles are used to identify the data points aligned to one of the component orientations to facilitate comparison of fit quality between these and other points.

95% credible intervals about the fits (shaded region), which were not adjusted to include the actual data but were simply determined by the variability of the normalized responses themselves (Section 3.2.2). These intervals, which include nearly all the data, demonstrate that the error produced by these fits is consistent with the response variability intrinsic to the data.

3.7 Monkey V: Configuration-specific divisive modulation as a general explanation.

It is useful at this point to step back and review the subtle differences that define the types of modulation discussed. In particular, feature-specific modulation and configuration-dependent modulation are quite different. Feature-specific modulation refers to the tendency of the surround to modulating the effective strength of specific features of the center; it refers more to the

fine-scale influence of the surround than to the stimuli that reveal it. On the other hand, configuration-dependent modulation refers to a form of modulation that is global (i.e. feature-nonspecific) and recruited in a stimulus-dependent manner. The plaid stimuli have primarily been used to diagnose feature-specific modulation but can also reveal configuration-dependent modulation when responses are not separated according to preferred orientation.

Overall, the analyses presented thus far for monkey V identify a multiplicative, feature-nonspecific effect of feedback inactivation on plaid+surround responses. As such, the data speak strongly against a possible role of feedback in generating feature-specific surround modulation. These observations await confirmation in a second monkey (since the data from monkey R cannot contribute much evidence with respect to specificity) but potentially constitute a key observation towards the goal of defining a computational role of feedback. However, some important interpretations of the data remain unresolved: in particular, the question posed above of whether the feature-specific suppression that shapes plaid+surround responses reflects a distinct mechanism from that producing match suppression. We have previously suggested that they are one in the same mechanism (Chapter 2) but must address the contradiction arising out of the lack of match suppression during the sustained activity of monkey V (Section 3.5) and the clear feature-specific suppression of plaid responses in these same units (previous Section).

As a reminder, the apparent lack of match suppression was evident through a number of analytical lenses, the simplest of which is shown in Figure 3.11 (and reproduced in Figure 3.16A), where the normalized center+surround rates show virtually no modulation as a function of center/surround configuration under control conditions during the sustained period but a clear dependency on this

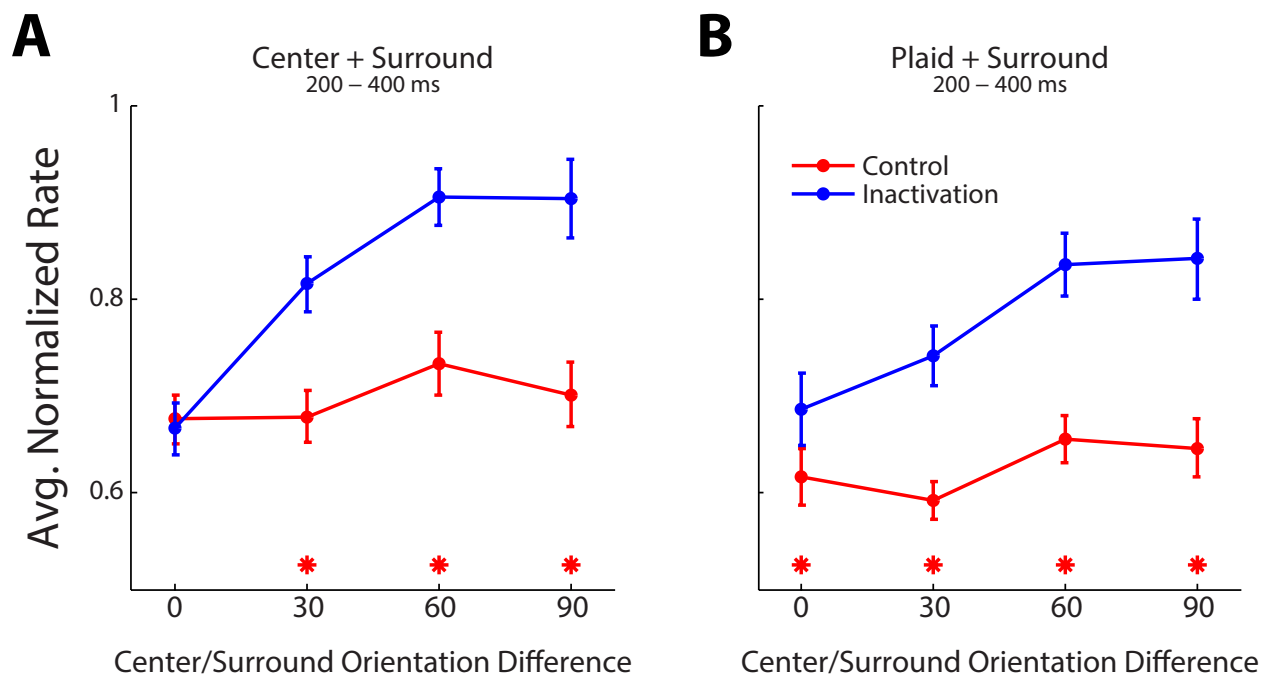


Figure 3.16 | Monkey V: Configuration-dependent responses during the sustained period of activity for center+surround and plaid+surround data. **A**, panel reproduced from Figure 3.11 (right), omitting re-warming. **B**, same convention as **A**, with analysis applied to the analogous data measured with a plaid stimulus.

variable during inactivation. Applying the same basic analysis to the plaid+surround reveals the same trend (Figure 3.16). We can take this a step further by applying the sliding window analysis to provide a glimpse into how configuration-dependent responses evolved throughout the response time course in the plaid data (Figure 3.17). This visualization reveals important similarities in the configuration-dependency of control responses across the trial, in particular the apparent loss of this dependency over time. Interestingly, for both subsets of the data, the effect of inactivation seems to be a relief of suppression that is strongest for orthogonal stimuli (Figure 3.17C).

This observation lends itself to a possible explanation for the seeming contradiction outlined above. According to this explanation, match suppression is a manifestation of feature-specific modulation, as are the responses to the stimuli containing plaids, and, importantly, this mechanism is intrinsic to V1. That is, it is unaffected by the inactivation of V2/V3 and their

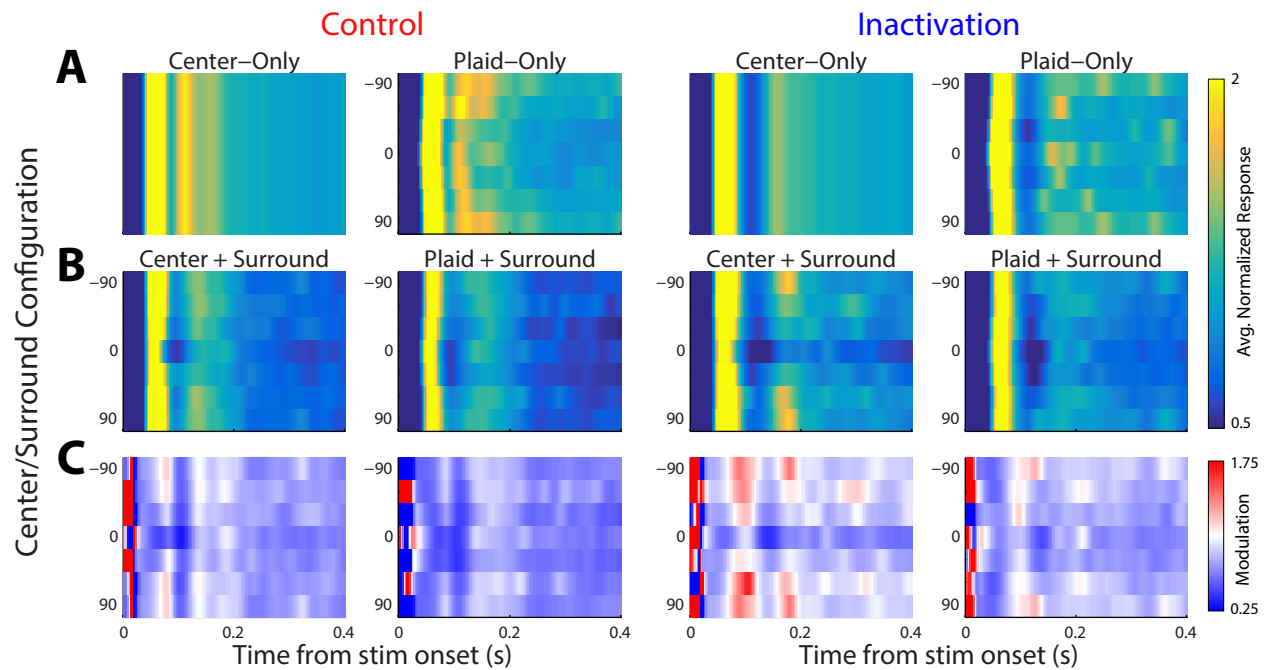


Figure 3.17 | Monkey V: Time course of average responses measured across both surround modulation data sets using a Gaussian sliding window. Conventions are similar to Figure 3.10. Here, “Stimulus Configuration” refers to either center/surround orientation difference or C_1/C_2 orientation difference, depending on the class of stimuli. **A**, population response profiles for center-only and plaid-only stimuli, separated by configuration where appropriate, for control data (left two panels) and inactivation data (right two panels). **B**, corresponding center+surround or plaid+surround data. **C**, surround modulation profiles. This analysis used a Gaussian window with a standard deviation of 5 ms.

feedback projections. This suffices to explain why these properties of contextual modulation are so plainly visible in the inactivation data from monkey V. In order to explain the control data in this monkey, we must invoke an additional mechanism: divisive modulation of responses evoked by stimuli with strong orientation-contrast between the center and their surroundings. Despite assuming a (arguably) novel mechanism, this explanation provides a rather parsimonious account of the data. It does not require a more nuanced explanation for the results reported in Chapter 2 and it provides a simple account of what has changed during inactivation. According to this interpretation, the mechanism underlying feature-specific modulation (and, as a corollary, match suppression) is active throughout control and inactivation. Under control conditions, the divisive mechanism proposed above is also active and the responses measured during this portion of the experiment reflect a superposition of these two mechanisms. This frames the inactivation data entirely in terms of established mechanisms and, importantly, suggests a role of feedback that includes configuration-dependent divisive modulation.

3.8 Discussion

This Chapter describes the effects of feedback inactivation on the more sophisticated forms of contextual modulation that one can probe with parametric stimuli. Specifically, we have attempted to characterize the extent to which feature-specific modulation is a function of feedback processing and took advantage of the rich and readily quantifiable modulatory behaviors evoked by our experimental stimuli.

As discussed in Section 3.1, our approach, which combines cortical cooling of V2/V3 with multi-electrode recording in V1, produced epileptiform activity locally in V1 when feedback was

inactivated. Because we did not set out to test this activity (rather, we very much tried to avoid it), it is difficult to analyze with any rigor or sense of experimental control. Nevertheless, its characterization is almost certainly relevant to understanding the nature of feedback. The simplest interpretation of this epileptiform activity is that feedback provides a net inhibitory—and therefore stabilizing—influence, which is entirely consistent with the effect of feedback inactivation on surround suppression (reported elsewhere and repeatedly here). It will be interesting to ask whether feedback-dependent stability reflects a more coordinated influence of feedback on network dynamics or simply reflects the global importance of sufficient inhibition. We expect that this question may gain tractability in network models of visual processing that encapsulate feedback, the development of which will require additional experimental work.

Fortunately, the problem of epileptiform activity (which required terminating the experimental session) diminished over time, though never vanished entirely. As such, we were able to use the described approach to collect a limited inactivation data set from each of two monkeys. The recording conditions and signal quality differed between the monkeys: monkey R, whose array was several years old at the time of data collection, provided preliminary data that were instrumental in optimizing experimental conditions for recording in monkey V, whose array was only several months old by the time we were able to successfully inactivate feedback. These differences carried over into the V1 activity patterns measured in each monkey as well as the effects of inactivation. Whether this reflected the particular experimental timelines and circumstances of each monkey is something that we have no way of answering. However, it is tempting to speculate that the differences reflected some functional heterogeneity within V1 that we did not set out to measure. When trying to synthesize the results from the two monkeys this

ends up being a very crucial point.

We can mostly summarize the responses to plaid stimuli and related effects of inactivation in monkey R (Section 3.4) as not saying much about the potential contribution of feedback to feature-specific suppression. Though they do not say much by themselves, they do importantly fail to contradict the suggested lack of contribution of feedback to this type of suppression in monkey V (Section 3.6).

Apart from an overall decrease in surround suppression, which has already been reported (Nassi et al., 2013), the noteworthy observations concerning feedback from each monkey are not corroborated in the other. Therefore, all of the results in this Chapter and their working interpretations must be considered preliminary. Even so, these results identify important features of contextual modulation that may reflect a contribution of feedback and therefore greatly inform the experiments and analysis we will perform with a third monkey. In particular, this work suggests that configuration-dependent scaling of population responses may reflect the mechanism by which feedback mediates surround modulation in V1.

With that in mind, it is worth revisiting the center-only/center+surround data from monkey R. We have already discussed how the above description of feedback provides a parsimonious explanation of the data from monkey V. Can it do the same for monkey R? It is possible that part of the suppression of responses to iso-oriented stimuli observed in monkey R reflects a divisive, configuration-dependent modulation. Framing feedback in this way would require us to slightly alter our explanation such that divisive modulation is recruited by iso-oriented stimuli in monkey R and orthogonal stimuli in monkey V. One prediction of this interpretation is that feedback inactivation should scale responses to plaid stimuli (i.e. unmatched and matched weights change

in proportion to one another) but with the largest increases for configurations with little orientation contrast. At face value, this prediction is not realized, but this failure is very difficult to interpret because, in monkey R, the matched component contrast was 66%, meaning that the overall orientation contrast between the center and surround was consistently small.

Developing an *ad hoc* explanation that will allow our full data set to be cleanly synthesized will likely prove a futile exercise. It makes more sense to simply acknowledge the very real differences between these two sets of results. The functional differences across both the layers and surface of V1 are well known and the extreme heterogeneity therein makes MEA placement a very likely source of sampling error. It is therefore possible that the MEAs in monkeys R and V provided access to different functional domains of V1. It is very difficult to find an analytical basis for this possibility; for example, the observed tendency for the signals from monkey V to be more commonly and more strongly orientation tuned is confounded by the poorer signal quality reflecting the age of monkey R's array. Therefore, we cannot address this possibility directly and instead stress the dissimilarity through a visual comparison. Figure 3.18 reproduces the sliding window analyses presented earlier (using the center-only/center+surround data) to facilitate a side-by-side comparison of the responses, surround modulations, and inactivation effects measured over time in each monkey. The left and right groups of panels represent the data from monkeys R and V, respectively. As the response progresses, the similarities across the monkeys become outnumbered by their differences, especially when considering the shape and time course of surround modulation (bottom panels). This visualization is not presented with the intention of identifying any particular aspect of feedback or surround modulation but simply to motivate the argument that the data from each monkey should remain separate until new evidence allows us to

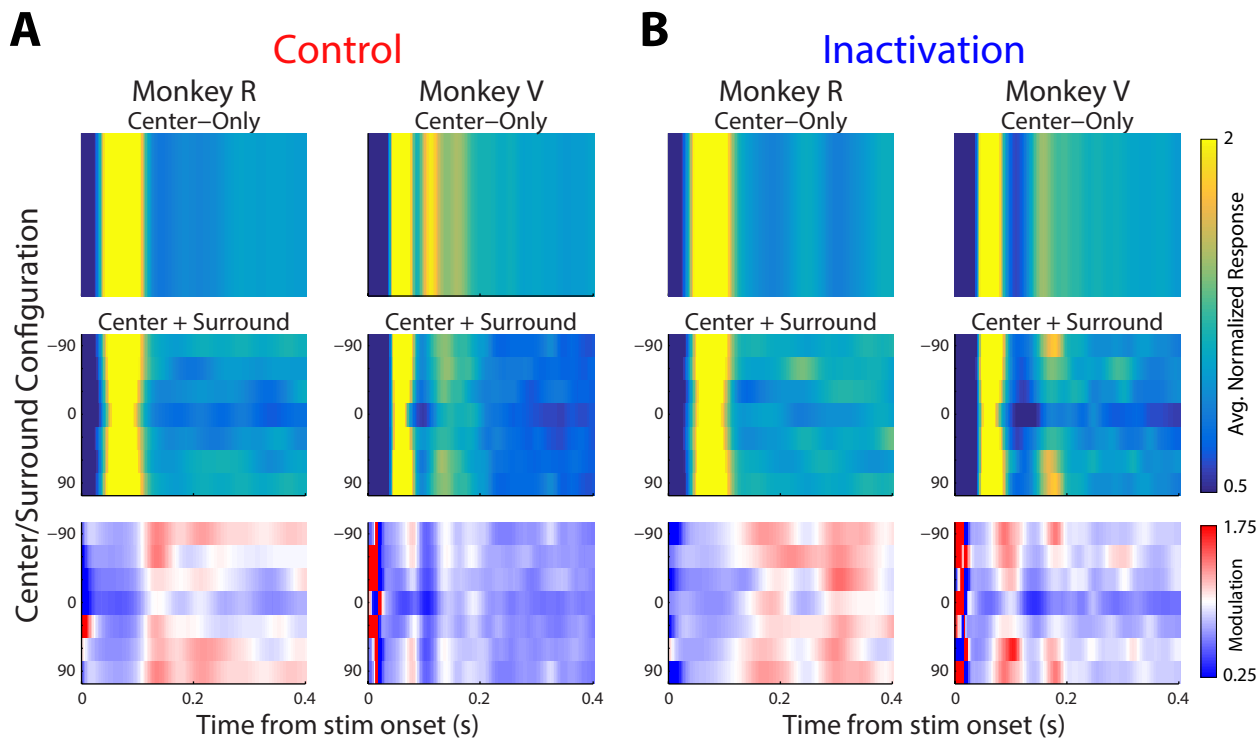


Figure 3.18 | Responses, surround modulations, and inactivation effects observed with the center-only/center+surround data in each monkey. The analysis window for monkey R is shown only over the first 400 ms, to align with the data available from monkey V. **A**, control response patterns in monkey R (left) and monkey V (right). Center-only responses, center+surround responses, and modulation of the center-only responses are shown in the top, middle, and bottom panels, respectively. For the top and middle panels, color indicates mean normalized firing rate, normalized to the responses measured during the entire sustained period. For the bottom panels, color indicates the effect of the surround, where blue indicates suppression, red indicates facilitation, and white indicates no effect. **B**, same as A, for inactivation data. This analysis used a Gaussian window with a standard deviation of 20 ms for monkey R and 5 ms for monkey V.

explain these obvious disparities.

If we are to understand the differences between the results from each monkey as sampling error (from, say, difficulty controlling array placement), we may expect that a coherent picture of feedback processing may not emerge until data from at least several more monkeys are available. Therefore, the best overall strategy for studying feedback through inactivation may be to develop parallel experiments with multi-electrode arrays (the recording stability of which is critical to the experiments performed here) and laminar probes, which would greatly accelerate the discovery of layer-specific inactivation effects but may require less elaborate stimuli. It seems certain that each

approach would inform the other considerably.

In any case, the next steps moving forward are clear. The data from monkey V speak strongly against a role of feedback in generating feature-specific suppression (Section 3.6-3.7). Importantly, the insight we gain from this observation is that input-gain control through this modulatory mechanism is likely intrinsic to V1. The stimuli used throughout this Chapter remain an important tool for characterizing suppression in terms of its feature-specificity and configuration-dependence. The full set of stimuli allows the experimenter to disentangle these modulatory influences and their changes during inactivation. Therefore, it is necessary to repeat the experiments performed with monkey V in a third monkey to form as clear as possible an answer to whether feedback produces feature-specific modulation or provides configuration-dependent, divisive modulation.

Match suppression has been reported repeatedly by studies characterizing surround modulation in V1 (Sillito et al., 1995; Cavanaugh et al., 2002b; Jones et al., 2002; Shushruth et al., 2012). The work documented in Chapter 2 identified feature-specific modulation and provided evidence that match suppression could be interpreted as a manifestation of it, raising the possibility of a unified mechanism. We proposed that the role of this modulation is to act as a form of effect input-gain control to flexibly gate particular streams of information in a context-dependent manner. However, match suppression has served as a benchmark for theoretical studies that frame contextual modulation as a result of optimal adaptation to the statistics of natural stimuli. These studies connect natural image processing to match suppression through a form of divisive normalization (Schwartz & Simoncelli, 2001; Coen-Cagli et al., 2009, 2012, 2015) and related work points out the hierarchical nature of such processing (Karklin & Lewicki, 2005; Schwartz

et al., 2006).

The proposal that feature-specific modulation encapsulates match suppression seemingly contradicts the theoretical explanation of match suppression through a divisive, feature-*nonspecific* mechanism. However, the interpretation we provide for monkey V resolves this contradiction by suggesting that both mechanisms are active in V1 (Section 3.7). Furthermore, it argues that feedback is specifically reflected in the mechanism that produces divisive modulation. Interestingly, this interpretation (that feature-specific and *–nonspecific* modulation is carried out through lateral connections within V1 and feedback connections, respectively) has also been used to explain the patterns of anatomical connections believed to mediate contextual interactions in V1 (Stettler et al., 2002).

Any evidence that the theoretical work mentioned above is in fact biologically accurate would support the ‘efficient coding’ hypothesis of feedback processing (discussed at length in Chapter 1). However, the observations that seem to generate this support also include the observation that feedback recruits suppression of stimuli with large orientation contrast. These are typically the kinds of stimuli that would not be targeted by suppression under the efficient coding hypothesis. Instead, suppression conditioned in this way would seem to support an instructive role for feedback, wherein feedback encourages V1 responses to “agree” with their surroundings. However, as previously discussed (Section 1.2.2), modulation deployed according to this strategy would be expected to have a feature-specific effect.

Piecing together the computations performed by feedback is no simple task. Fortunately, there is no shortage of theoretical perspectives on which to anchor the various observations encountered by the intrepid experimenter. The data we have collected thus far leave many questions unanswered

but suggest that we are close to being able to consolidate these theories into a deeper and more biologically motivated understanding of feedback processing.

Chapter 4

Feedback-dependent correlation structure during contextual modulation

Throughout this dissertation, the role of feedback has been discussed, tested, and analyzed through the lens of contextual modulation. In general, this has required us to summarize the overall effects of the stimulus surround by combining data across units and treating the descriptions that we can apply to this combined data as representative of the underlying computations. This approach has the advantage of distilling the characteristics of responses and their modulations that are most common across the sampled population. However, combining the data in this way comes at the cost of obscuring the potential heterogeneity throughout the population. Fortunately, the unit-by-unit analyses that we have been able to do provided little contradiction to the assumption that the descriptions we apply to the combined data are general—at least within an individual monkey. The other main disadvantage of analyzing a heavily pooled data set is that the detailed variations across the population’s responses to an individual stimulus presentation are entirely neglected.

These variations seem particularly worth considering when the responses represent contextual modulation. This point of view is motivated by the suggestion that modulation from the RF

surround is mediated by lateral processing (Stettler et al., 2002; Adesnik & Scanziani, 2010), the influence of which may be modulated by feedback (Ito & Gilbert, 1999; Ramalingam et al., 2013). From a more intuitive perspective, measuring the modulatory influence of the surround for one particular neuron necessarily involves providing a driving stimulus (i.e. the surround) to its neighbors. Therefore, it is natural to think of not only the neurons representing the stimulus center but the entire network involved in processing these modulatory stimuli.

In recent years, the topic of correlated variability within cortical networks has gained considerable attention (Kohn & Smith, 2005; Cohen & Newsome, 2008, 2009; Herrero et al., 2013; Ruff & Cohen, 2014; Snyder et al., 2014; Rabinowitz et al., 2015; Franke et al., 2016). Specifically, research has focused on “noise correlations”: correlated fluctuations in the responses to the multiple presentations of same stimulus (or, correlated fluctuations in the z-scored responses—when z-scoring for each unique stimulus). In part, this reflects the increasing availability of recording devices that allow for simultaneous recording. In addition, it reflects an enhanced appreciation for the implications of noise correlations with regard to information encoding/decoding (Cohen & Newsome, 2009; Herrero et al., 2013; Jeanne et al., 2013; Ruff & Cohen, 2014; Zylberberg et al., 2016) and also the growing set of observations that these correlations are dynamic and can be seen to vary according to behavioral and stimulus context (Kohn & Smith, 2005; Cohen & Newsome, 2008; Snyder et al., 2014; Franke et al., 2016).

Overall, it remains somewhat unclear how noise correlations are to be interpreted. This ambiguity is less of an issue for studies that can define some degree of optimal information representation—for example, when a monkey is performing an attention-demanding task. These studies mostly consider correlations from the standpoint of a hypothetical read-out network and

analyze how the putative success of such a downstream network is limited or enhanced by noise correlations and their particular structure (Cohen & Newsome, 2009; Ruff & Cohen, 2014; Franke et al., 2016). The lingering questions related to the circuit mechanisms and computations that give rise to such correlations and their flexibility, however, lacks an answer almost entirely. One possibility is that noise correlations reflect a common source of input that is not related to the stimulus itself (Rabinowitz et al., 2015). These inputs could be related to factors such as arousal, attentional focus, or any number of things that the experimenter cannot easily control or measure. It is conceivable that noise correlations also reflect local properties of the network, including direct and indirect influences from nearby neurons. Establishing direct influences (i.e. connectivity) is practically impossible with extracellular recording techniques. Indirect influences would reflect an emergent property of the network and would therefore be most readily modulated. A simple example would be disynaptic inhibition, where the influence of one neuron on another relies on an intermediate inhibitory neuron. The indirect influence of the first neuron onto the other could be modulated by changes in the gain of the intermediate inhibitory neuron. In this toy example, the various gain settings of this inhibitory neuron would lead to different patterns of noise correlations between the other two neurons. In summary, condition-specific patterns of noise correlations may reflect the condition-specific variability of common inputs and/or the effect of these conditions on the functional settings of the network itself. Whether noise correlations reflect functional adaptations to the demands of a particular behavioral or stimulus context is an open question of profound interest to systems neuroscience.

One of the key opportunities afforded by the use of multi-electrode arrays is the measurement of simultaneous activity across a local region of cortex (4 mm by 4 mm). Noise correlations are

implicitly captured in this simultaneous activity and we sought to identify potential signatures of contextual modulation within these correlations. To do so, we analyzed the neural activity collected during the experiments in Chapter 3 using this lens. The noise correlations observed in the multiunit activity of monkey V demonstrated a tendency to track the location of the stimulus center. This pattern was visible both when the stimulus included only the center as well as the center and the surround, suggesting that we were not simply measuring the difference between evoked and spontaneous correlations. Furthermore, this location-dependent correlation structure was reversibly weakened during feedback inactivation, specifically for stimuli that included a center and surround. Here, we report these findings and discuss how they might advance our understanding of feedback.

4.1 Experimental procedures

The experimental procedures for collecting the data have already been described in Chapter 3; here, we re-analyze this data with respect to noise correlations.

4.1.1 Receptive field localization

Our ability to measure the effects described below depended on our ability to assign a location (in visual coordinates) to the electrodes of the recording arrays. While receptive fields are a neural property, we can slightly abuse their definition to describe electrodes. This stems from the physical stability of the implanted arrays and the tight retinotopic organization of V1. As such, when we describe the RF of an electrode, we really mean the typical RF of the multiunit signals that the electrode provides access to.

This reflects the manner in which we estimated the RF centers of the signals we analyzed. That is, rather than attempting to estimate RFs for each day of recording, we devoted one large block of time to a thorough grid mapping procedure that yielded a large amount of data and produced excellent fits. Grid mapping was performed by presenting brief (200 ms), drifting gratings in each of 4 directions at each location in a grid tiling the horizontal and vertical span of the array in visual space (this span was estimated informally beforehand). The diameter of the grating, its wavelength, and the spacing of the sampling grid were 0.15° . Responses to each of the unique stimuli used were counted within the window from 50-200 ms after stimulus onset. Presentations were separated by a blank period of 100 ms. The monkey only received reward after approximately 3 seconds of successful fixation, using a fixation window of 1° . We ended data collection after a sufficient number of repetitions at each location/direction had been collected (30-50).

To estimate the RF location, we fit a 2D-Gaussian profile to the response map of each of the electrodes (MUA was combined with single-unit responses on the same electrode, where appropriate):

$$R(x, y) \propto N(x|\mu_x, \sigma_x) \cdot N(y|\mu_y, \sigma_y)$$

Where $R(x, y)$ represents the response measured to a stimulus centered at the coordinates indicated by x and y , the coordinates of the RF center are indicated by μ_x and μ_y , the width and height of the RF are indicated by σ_x and σ_y , respectively, and the function describing a normal distribution is given by N . Depending on the direction of the stimulus, we add a scaling coefficient to the right side of the equation to convert the product of Gaussians into a firing rate. A separate scaling coefficient was fit to each of the 4 directions. Therefore, the RF estimation involved optimizing

the 8 parameters used to fit responses to the grid mapping stimuli. The analyses described below made specific use of the parameters representing the coordinates of the RF center. At the start of each recording session, a shorter, more localized version of this approach was used to optimize the exact placement of the stimulus center (according to the experimental goals laid out in Chapter 3). We did not observe the accuracy of the RF estimates provided by the full grid mapping procedure to deteriorate over time and therefore, throughout the analyses below, used the RF centers fit to each electrode in this one batch of data as the RF centers for their associated MUA.

For the analyses described in this Chapter, we threw out data from electrodes that did not produce high-quality RF estimates (5 out of 96). The Gaussian RF estimates of the remaining 91 electrodes are plotted in Figure 4.1), which demonstrates the overall square shape of the combined RFs imposed by the shape of the array. In addition, though it is not especially obvious in this visualization, we observed a slight contraction in the size and spacing of RFs towards more foveal coordinates (a consequence of changing cortical magnification). Throughout the remaining analyses, no standard of quality was imposed to exclude activity from particular electrodes beyond the condition of reliable RF estimation at the end of the fitting procedure.

4.1.2 Measurement of noise correlations

Noise correlations are typically defined as correlated fluctuations about the mean response to multiple presentations of the same stimulus. An alternative definition is the correlation in residual activity, after accounting for the effect of the stimulus by subtracting the mean response. According to both, noise correlations must be the correlations in activity that are specifically not due to overlapping or antagonistic stimulus preferences. Often times it is desirable to measure noise correlations over a set of responses involving multiple stimuli.

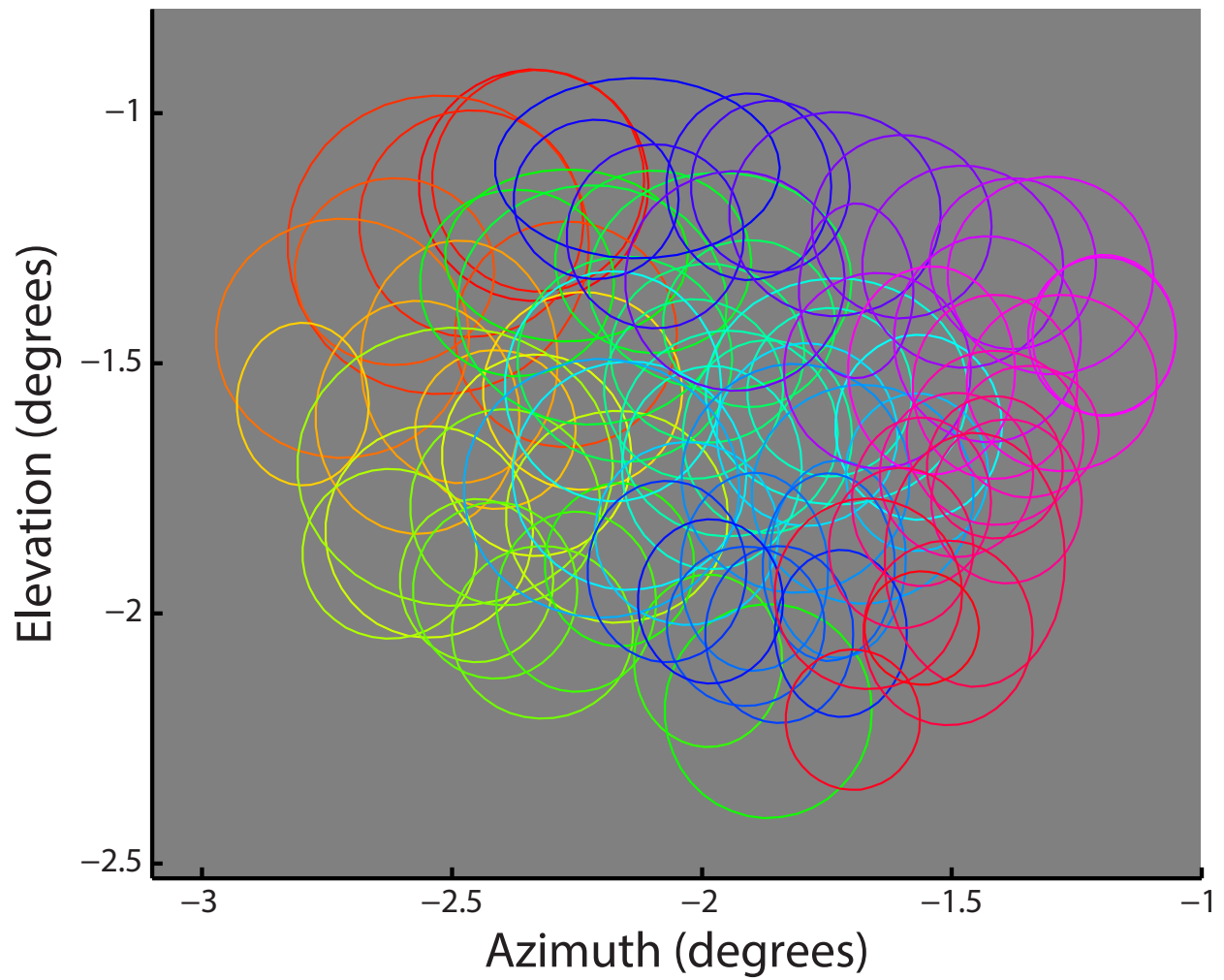


Figure 4.1 | Estimates of receptive fields. Each of the ellipses represents the 2D Gaussian estimated from the RF fitting procedure. Each ellipse defines the space of the visual field within one standard deviation of the RF center. The electrodes are colored according to their default numerical order, which is loosely related to their physical locations on the array.

There are two equivalent methods for doing this. One approach is to separate this data set into groups of responses according to the stimulus identity, calculate the correlation matrix within each subset of data, and then simply take the average of these matrices (assuming each stimulus was repeated the same number of times). The other approach, which is perhaps more intuitive, is to similarly separate the data according to stimulus identity, z-score the responses of each unit within each stimulus, re-combine the z-scored data and calculate the correlation matrix using this data. We employed the latter method, which allowed us to consider overall correlations within sets of trials comprised of multiple stimuli—for example, “center-only” and “center+surround” data.

Importantly, we limited ourselves to the half of the data where the stimulus center only contained a single grating. One reason was that it provides a slightly cleaner set of stimuli in a setting where the number of stimulus presentations should be as large as possible. The monkey sees fewer “versions” of the center portion of the stimulus when it is only comprised of a single grating (6) compared to when it is drawn as a plaid (36). Another reason that we chose to do this is to ensure that the stimuli were as similar as possible for the two monkeys. Recall that the plaid data in monkey R was recorded separately from the center-only/center+surround data. All analyses were performed using the “sustained” activity, which began 150 ms (monkey R) or 250 ms (monkey V) after stimulus onset and ended at the stimulus offset.

4.1.3 Using stimulus location to characterize correlations

Our main finding relates noise correlations to the location of the stimulus, which changed from day to day. We required a metric to describe electrodes in terms of the stimulus. An obvious choice for this is the Euclidean distance between the stimulus center and the RF center (RF-to-Stimulus distance). Since correlations are a second-order statistic (i.e. they describe a pair of units), any

metrics that we intend to relate to correlation values must themselves serve as a description of an electrode pair. Rather than simply using the sum of the RF-to-Stimulus distances of the two electrodes in a pair, we use this sum plus the RF-to-RF distance between the two RF centers of the pair. We refer to this sum of the three Euclidean distances as “total distance” (Figure 4.2A). Figure 4.2B represents the noise correlation matrix describing the control center-only responses during one of the 8 valid experimental sessions in monkey V, and panel C represents the map of total distances describing the same pairs. These matrices order the electrodes according to their default labels, which are somewhat arbitrary.

To illustrate the correlation structure revealed by the total distance metric, it is necessary to organize the data according to this signal. Ideally, we could sort the electrodes on a given day such that all pairs of electrodes that are ordered within a few indices of one another have fairly similar total distance values. An ordering such as this would result in a relatively “smooth” version of the total distance matrix in Figure 4.2C. The best method to approximate this ideal sorting is simply to iteratively move the ordering of each electrode into the location that most improves the “cost” associated with the current ordering. This cost encouraged orderings that concentrated the large total distance values towards the diagonal of the total distance matrix. Once this ordering was optimized, it was circularly shifted by 180° to give the final, relatively ideal electrode sorting. Sorting the electrodes in this way produced total distance matrixes with the desirable property of having the pairs with lowest total distance near the center (Figure 4.2D). An additional property of this sorting method not shown by the current visualization is that it puts pairs of electrodes that are nearby one another but far away from the stimulus in one of the 4 corners. Pairs that are neither in the center of the matrix nor in one of the corners represent pairs where one electrode

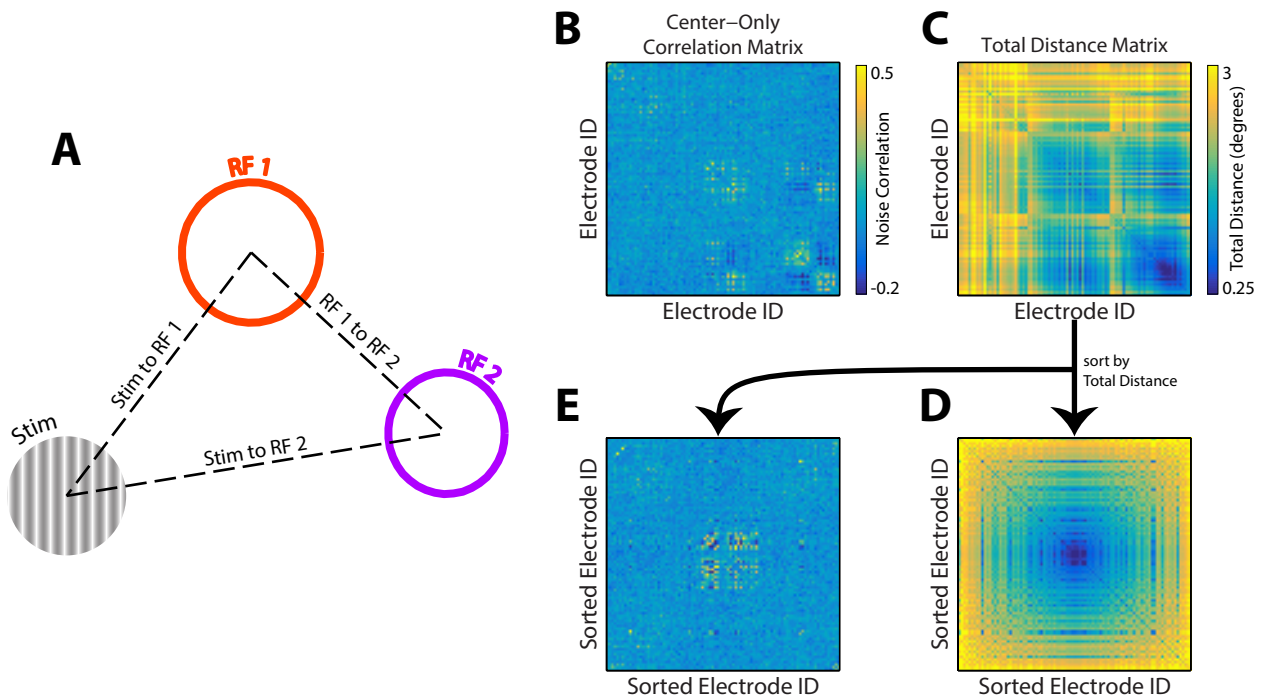


Figure 4.2 | Total distance as a metric for organizing and describing correlations in terms of RF and stimulus location. **A**, an illustration of how total distance is computed for a single electrode pair. Here, the grating represents the stimulus and the ellipses represent the RFs of the electrodes in the hypothetical pair. Total distance is the sum of the 3 distances between each of the centers—or, the perimeter of the triangle connecting the 3 centers. **B**, the noise correlation matrix measured from the control center-only data of one recording session. **C**, the matrix of total distance values describing the electrode pairs in **B**. **D**, the total distance matrix in **C** after sorting the electrodes. **E**, the correlation matrix in **B** after sorting the electrodes.

is near the stimulus and the other is relatively far away. The correlation matrix reproduced with this electrode sorting reveals a cluster of high correlation pairs corresponding to the electrode pairs with the smallest total distance (Figure 4.2E). This sorting method is used purely for visualization. None of the quantitative results depended on the outcome of this sorting procedure.

4.2 Monkey R

In keeping with the results from Chapter 3, monkeys R and V did not produce comparable results when their data was subject to the analyses we describe below. However, unlike in Chapter 3, where we were able to observe a number of surround/inactivation effects in monkey R, no such richness is

observed when measuring correlations. As such, we focus from here specifically on the data from monkey V. All such features of the correlations in this data that we suggest are interesting were not observed in monkey R. The reader should be aware of this partially negative result and understand that, despite the focus on data from monkey V, we performed these analyses in both monkeys and can only report noteworthy findings in one. I will provide a brief discussion of this disparity and its potential interpretations at the end of this Chapter.

4.3 Visualizing stimulus-dependent correlations and effects of inactivation

The first observation that stood out to us when attempting to analyze the correlation structure present in the data was the correlation matrices observed from day to day appeared to change considerably. We reasoned that this might reflect the fact that on different days we used different stimulus locations. According to this explanation, the apparent restructuring of noise correlations during the center-only data (Figure 4.3A) was perhaps not so surprising, since the small center-only stimulus only drives a fraction of the array, perhaps exposing a difference between evoked and spontaneous correlations. However, this explanation seemed less plausible in light of the observation that the correlation structure observed within the center-only responses was also observed within the center+surround responses (Figure 4.3B), which involved stimuli that largely covered the visual extent of the array, and which, on average, resulted in a decrease in the firing rate of neurons whose RFs were confined to the center.

To address the possibility that the day to day restructuring of the correlation matrices reflected the influence of the stimulus location, we defined a metric that relates each electrode pair to the stimulus in terms of the distances between the associated RFs and the stimulus (Section 4.1.3),

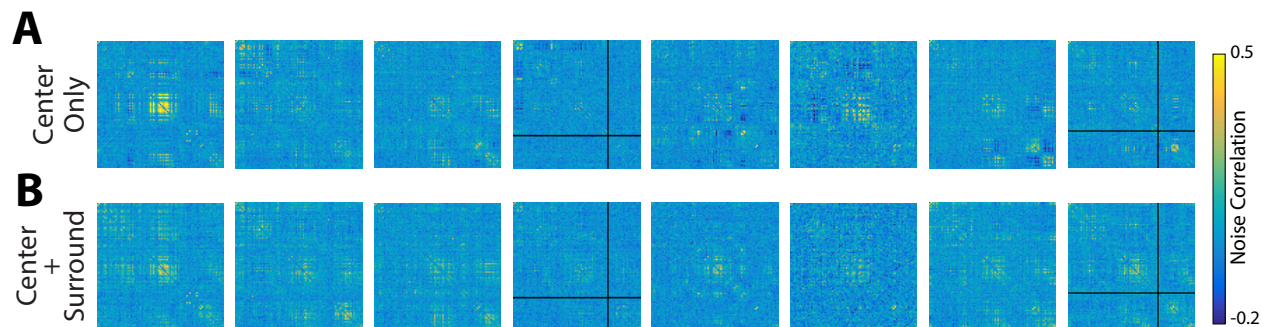


Figure 4.3 | Correlation matrices observed during the control portion on each of the 8 recording days, visualized with the default electrode ordering. A, correlation matrices describing the center-only responses. **B,** same as A, for center+surround data. Correlations of 1 along the diagonal are set to 0 for visualization.

which we term “total distance.” We can reorder the electrodes on each day such that the electrode pairs with the lowest total distance are placed at the center of the correlation matrices. Doing so offers a strong visual confirmation that the structure in the correlation matrices was in fact a reflection of the stimulus location (Figure 4.4A). The defining aspect of this structure is the tendency for more extreme correlation values (on average, larger) to be found amongst the pairs of electrodes who’s RFs mutually overlap with the stimulus center, at least approximately. In the center-only data, the units that are not driven by the stimulus are mostly represented in the more peripheral locations of the matrices. So, structure revealed by sorting the electrodes according to total distance could be a reflection of the different signals one is able to measure in evoked compared to spontaneous responses. This, however, would fail to explain the presence of the same stimulus-centered correlation structure in the center+surround data (Figure 4.4B). Instead, these data point to the intriguing possibility that the contextual modulations so commonly associated with this class of stimuli express some signature in the structure of noise correlations.

With this possibility in mind, we set out to ask whether feedback inactivation has any observable effect on this feature of the network activity. Figure 4.5 illustrates the total-distance-sorted correlation maps representing the center-only data measured during control (A), inactivation (B)

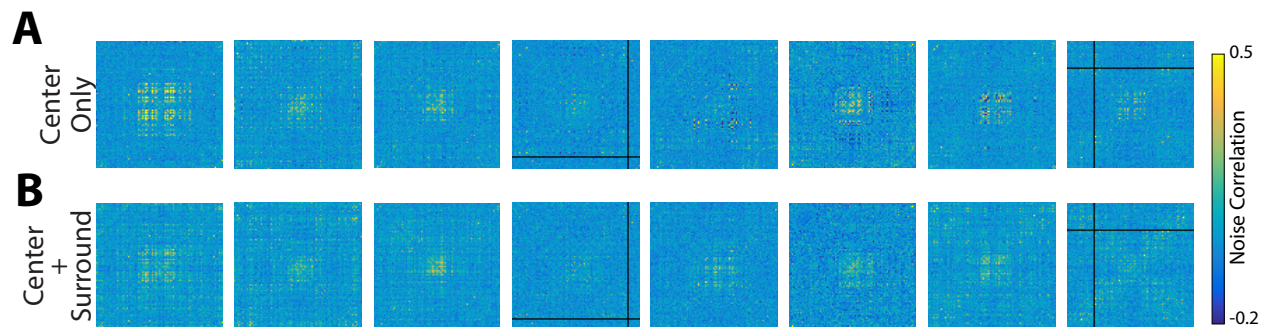


Figure 4.4 | Correlation matrices observed during the control portion on each of the 8 recording days, visualized with the sorting determined by total distance. All other conventions identical to Figure 4.3.

and re-warming (C). In the center-only data, the elevated correlations associated with electrode pairs near the stimulus center appear visible throughout the 3 portions of the experiment. However, feedback inactivation appears to introduce elevated correlations in the corners of the sorted correlation matrices, which represent pairs of electrodes that are nearby one another but distant from the stimulus center (Figure 4.5 B). This trend also appears in the re-warming data, though to a lesser extent (Figure 4.5 C).

A far more noticeable effect of inactivation is seen on the stimulus-centered correlation structure visible within the center+surround data (Figure 4.6). These data also display a slight elevation correlations observed between pairs that are distant from the stimulus center, but the defining change is clearly a reduction in the magnitude of correlations observed amongst the electrode pairs whose RFs represent the stimulus center (Figure 4.6). The effect produced by inactivation is apparently reversed during re-warming.

To achieve a more compact illustration of this trend, we averaged the 8 (sorted) matrices representing a particular experimental/stimulus condition. For the sake of visualization, we also apply a small degree of smoothing (2D Gaussian; $SD = 1.5$). Figure 4.7 provides a concise visual summary of what is observed in the unprocessed correlations. (It should be noted that, after

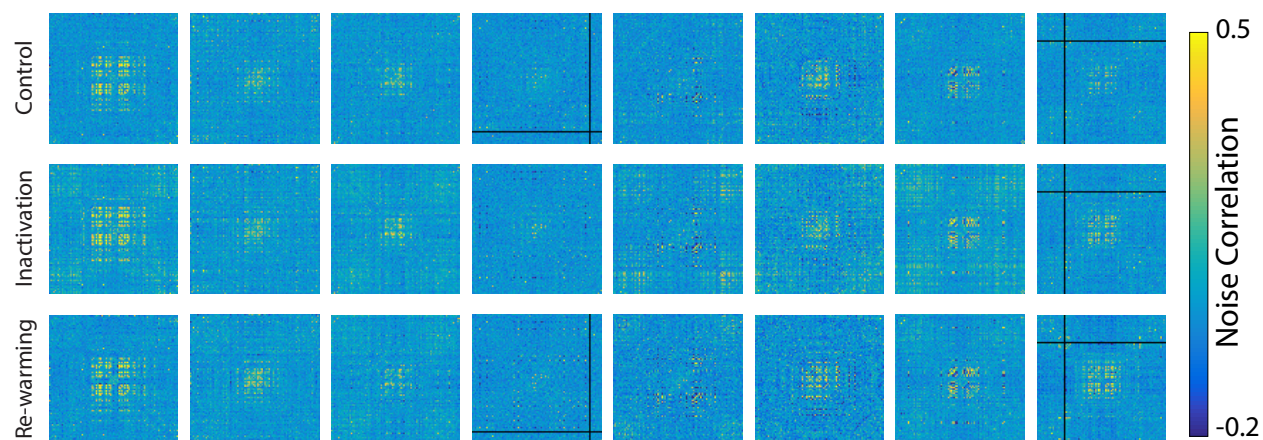


Figure 4.5 | Correlation matrices observed within the center-only data on each of the 8 recording days, visualized with the sorting determined by total distance.

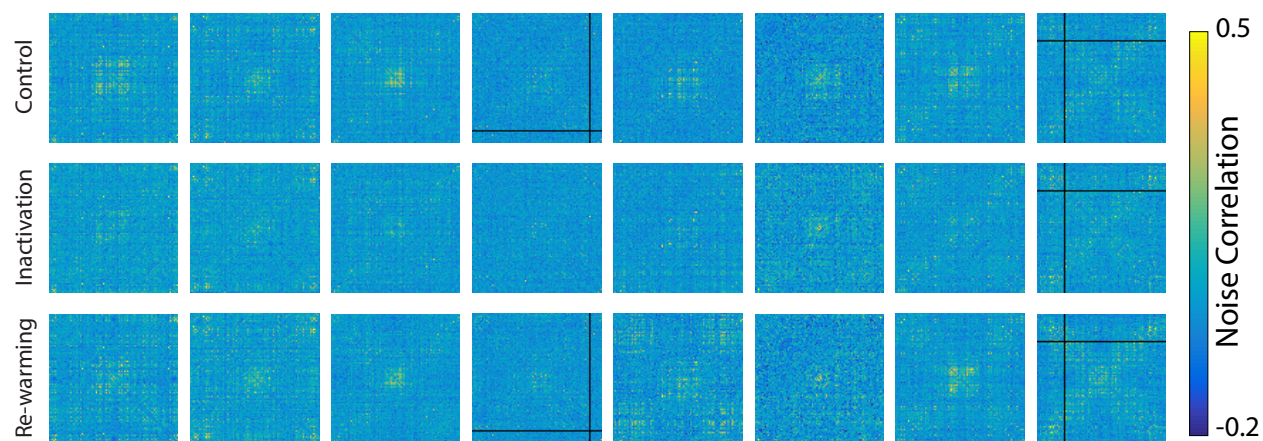


Figure 4.6 | Correlation matrices observed within the center+surround data on each of the 8 recording days, visualized with the sorting determined by total distance.

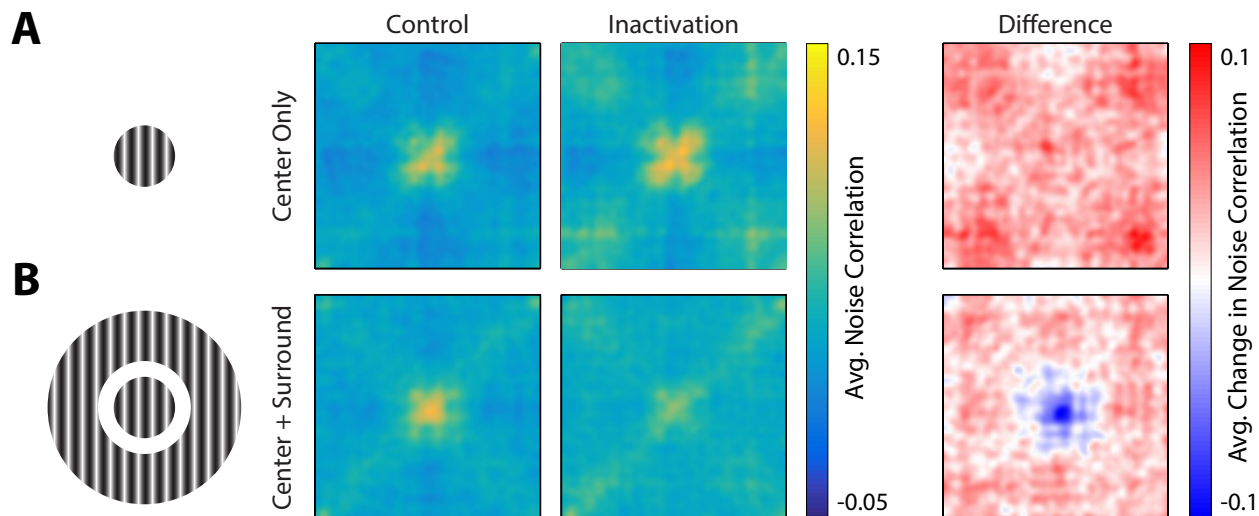


Figure 4.7 | Averages of the sorted correlation matrices across the 8 days. **A**, average center-only correlation structure during control (left) and inactivation (middle) and the difference between the inactivation and control correlation structures (right). **B**, same as A, for the center+surround data. After average across the 8 correlation matrices, the resulting matrix was smoothed with a 2D Gaussian ($SD = 1.5$) to make the visualization more salient.

averaging, these images no longer represent genuine correlation matrices but simply reflections of the common stimulus-aligned structure.) An examination of the correlation averages representing the center-only data support the observation that during inactivation the main effect is on pairs that are distant from the stimulus center but near one another, but one can also observe a previously unrecognized increase in correlations overall (Figure 4.7A, right). On the other hand, the effect of feedback inactivation on the correlation averages representing the center+surround data is best described as a loss of stimulus-aligned structure (Figure 4.7B). This is manifested as an increase in the correlations between most pairs of electrodes and a decrease in the correlations amongst the pairs representing the stimulus center.

Within this data, we are also able to separate the center+surround stimuli into iso-oriented and orthogonal sets and repeat the same analysis. The average of the total-distance-sorted correlation matrices representing the responses to these two stimulus sets are shown in Figure 4.8. The further subdivision of the data required for this analysis resulted in some loss of signal-to-noise,

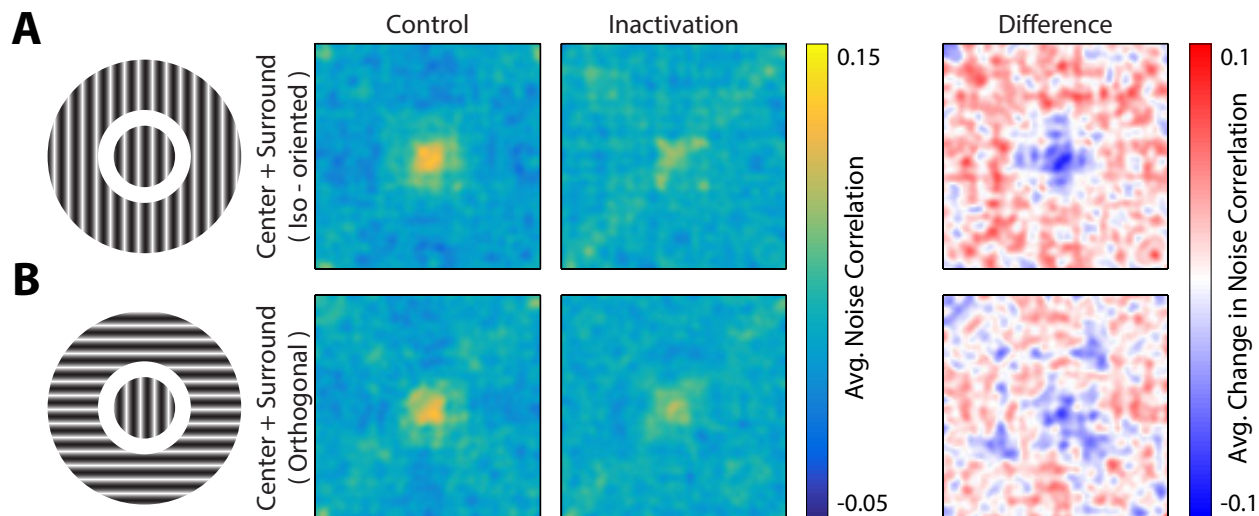


Figure 4.8 | Averages of the correlation matrices across the 8 days for iso-oriented and orthogonal center+surround stimuli. **A**, average correlation structure measured with an iso-oriented center+surround stimulus during control (left) and inactivation (middle) and the difference between the inactivation and control correlation structures (right). **B**, same as A, for the data measured with an orthogonal center+surround stimulus. After average across the 8 correlation matrices, the resulting matrix was smoothed with a 2D Gaussian ($SD = 1.5$) to make the visualization more salient.

but one can still observe the stimulus-dependent correlation structure and its feedback dependency in the iso-oriented data (Figure 4.8A). Interestingly, the correlation patterns look very nearly the same within the orthogonal data (Figure 4.8B). As reported previously (Section 3.5), we observed that the responses of the center-confined units were strongly influenced by feedback inactivation in a configuration-dependent manner within this time window. And, as is typically reported in the literature (but was not the case for sustained responses in monkey V), the influence of the surround is strongly determined by center/surround configuration (Section 3.3). As such, we expected that these structured correlations, given their apparent sensitivity to feedback inactivation, would similarly exhibit some dependency on center/surround configuration, either with regard to baseline correlations or a differential effect of inactivation. However, we did not observe any obvious indication of such. We do not rule out the possibility that correlation structure is shaped in some less obvious way by center/surround configuration, but the analyses we propose to search for such an effect (Section 4.5) are beyond the present scope.

4.4 Quantifying stimulus-dependent correlations and effects of inactivation

We wanted to provide a quantitative description of the correlation patterns illustrated in the previous section. In particular, we sought a means to quantify the effects of feedback inactivation so as to establish some confidence that were statistically meaningful. This required more than simply stating that the stimulus-localized correlations were affected by inactivation but also to be able to rule out potential confounds when reporting correlations.

Several known factors can cause correlations measured from spike counts to misrepresent the actual correlated variability within the underlying signals (Cohen & Kohn, 2011). Some of these issues are more of a concern when one is interested in overall levels of correlations, but our analyses focused on changes in correlations. As such, we wanted to rule out the possibility that these changes were due to changes in the confounding variables during inactivation. In particular, we wanted to disentangle changes in correlations from changes in overall firing rate and spike count variance. Due to the discrete nature of spike counts and noisy estimates of covariability caused by sampling with a limited number of trials, measured spike count correlations (even after z-scoring) will be biased towards smaller values as overall firing rate and variance decrease (Cohen & Kohn, 2011).

Our main finding with regard to correlations is that they depend not only on the distance between electrodes but also the proximity to the stimulus center, the combined effect of which can be captured with our total distance metric. As such, we sought to quantify this effect in a manner that would allow us to address possible confounds stemming from firing rate or variability. We chose as pairwise representations of these variables the geometric mean firing rate of the pair (the square root of the product of their mean firing rates) and the geometric mean Fano factor. To calculate

Fano factor for an individual multiunit, we calculated its Fano factor (firing rate variance divided by the firing rate mean) for each of the stimuli within the relevant subset of the data and then averaged these Fano factors together. The Fano factor of the pair was calculated as the geometric mean of the individual Fano factors. For brevity, we refer to the geometric mean firing rate as ‘geometric mean’ and the geometric mean Fano factor as ‘Fano factor.’ It should be noted that the distribution of Fano factors across all the pairs peaked very close to 1, but was heavily skewed towards larger values. To improve robustness to outliers, we excluded from this and the remainder of the analyses any electrode pairs with a Fano factor value greater than 3 (0.13% of pairs). Within the center+surround data, we observed that noise correlations were influenced by total distance, geometric mean, and Fano factor (Figure 4.9A-C). Furthermore, we observed that geometric mean and Fano factor varied as a function of total distance and, like noise correlations, were on average highest amongst the pairs with the smallest total distance values (Figure 4.9D-E).

This raised the possibility that the observed relationship between total distance and noise correlations was an artifact due to geometric mean and/or Fano factor. Such a concern identified the need to relate total distance to correlations in a manner that accounts for these other variables. This can be addressed by asking whether correlations depend on an interaction between total distance and geometric mean/Fano factor. We can gain an intuition for whether such interactions exist by fitting a surface to the set of points describing center+surround correlations, total distance and geometric mean (Figure 4.10A) or Fano factor (Figure 4.10B). (These surface fits were generated using the gridfit function available from the MathWorks File Exchange, the precise details of which are unimportant for the present analysis.) These illustrations suggest that geometric mean and/or Fano factor alone cannot explain the relationship between total distance

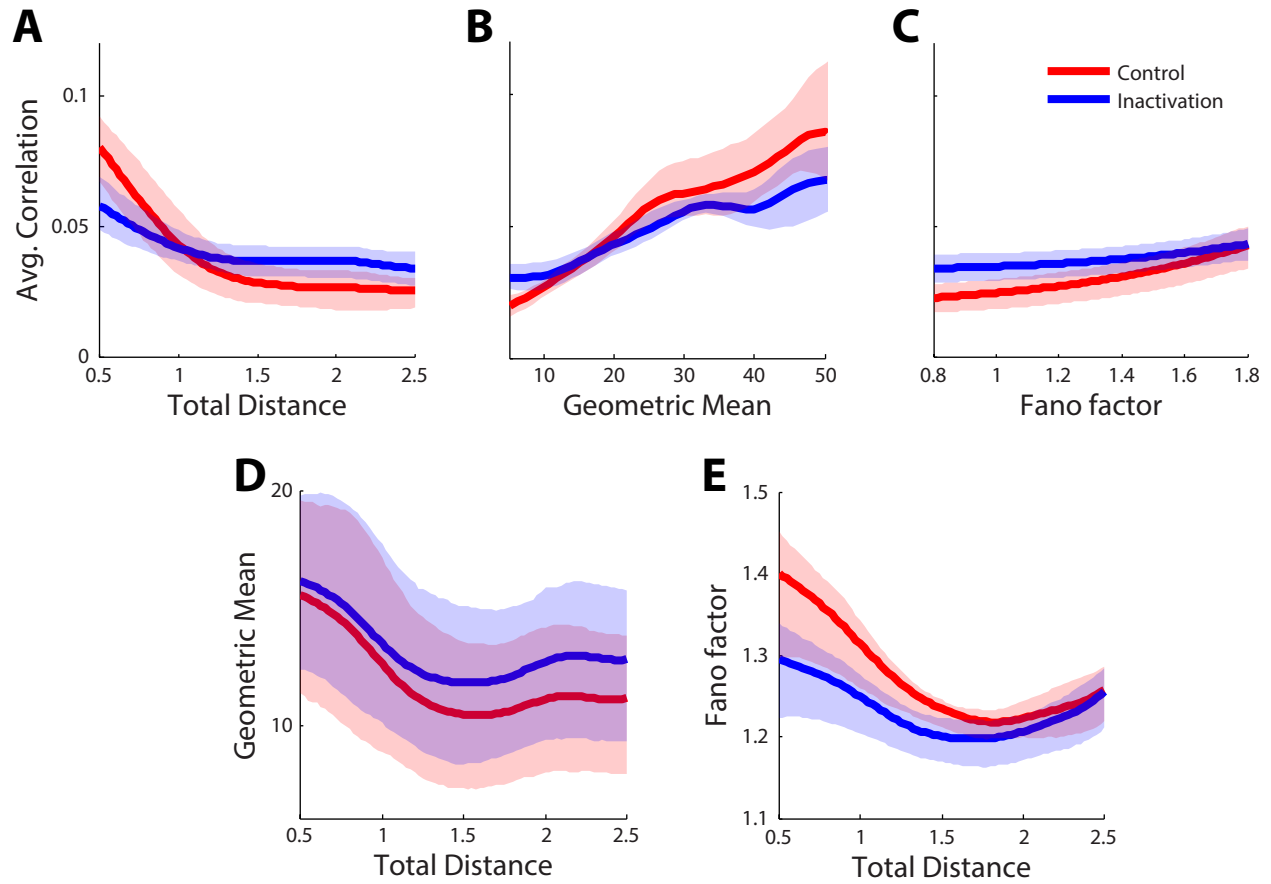


Figure 4.9 | The relationship between variables used to describe electrode pairs and correlations. A-C, average noise correlations as a function of total distance (A), geometric mean firing rate (B), and Fano factor (C). D-E, average geometric mean (D) and Fano factor (E) as a function of total distance. Solid lines indicate true averages, obtained with a sliding Gaussian window. Shaded regions indicate 95% confidence intervals about the mean obtained from bootstrapping.

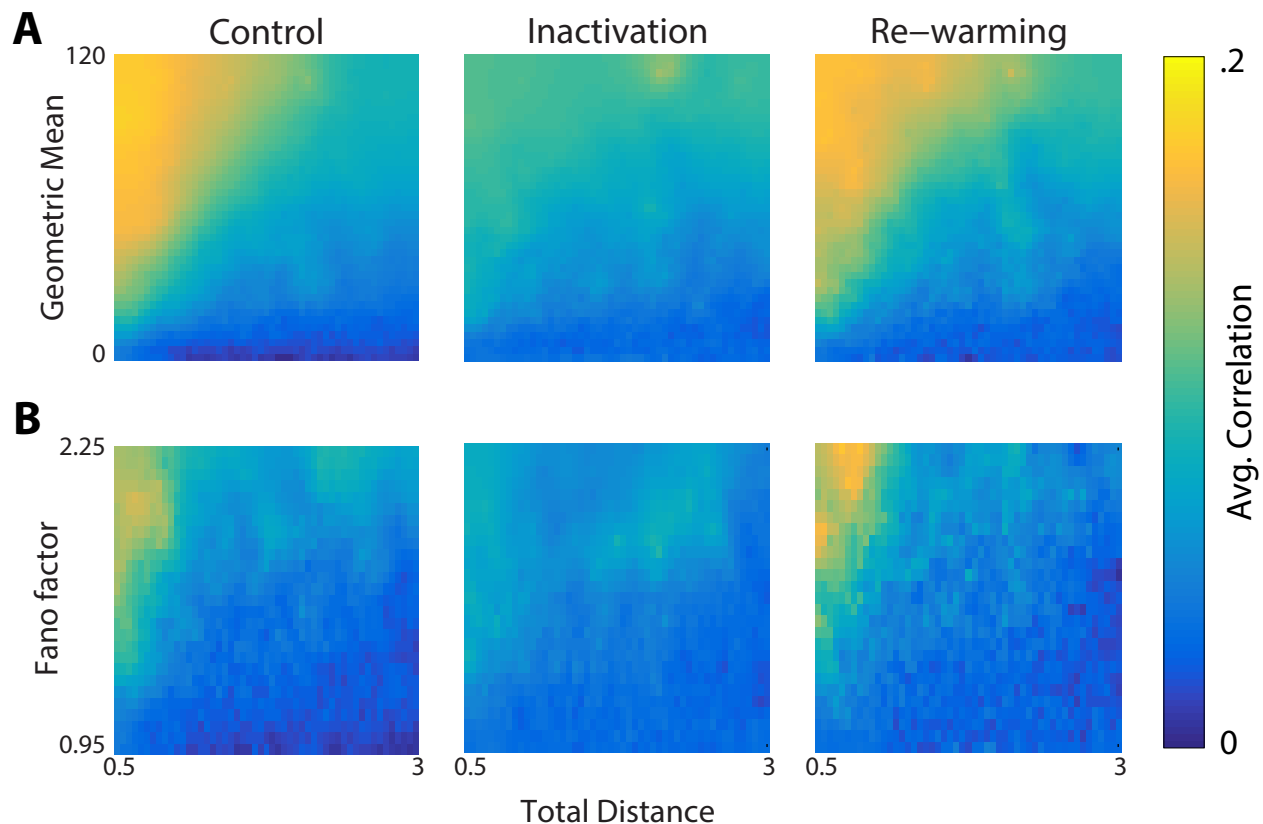


Figure 4.10 | Noise correlations reflected in the interactions between total distance, geometric mean, and Fano factor. **A**, surface fits (see main text) illustrating the average noise correlations at each combination of total distance and geometric mean, for control (left), inactivation (middle) and re-warming (right). **B**, same as A, for total distance and Fano factor.

and noise correlations since the relationships between noise correlations and geometric mean/Fano factor changed as a function of total distance. This explains why the largest correlations occurred when total distance was low and geometric mean/ Fano factor was high. Furthermore, these illustrations demonstrate that these interactions were weaker during inactivation and returned to normal during re-warming (Figure 4.10, middle and left panels).

To address this intuition quantitatively, we asked how, in the center+surround data, the slope relating geometric mean/Fano factor to noise correlations changed when focusing on groups of pairs with different total distance values. This slope can be measured with linear regression, or, in this case, weighted linear regression, where weights are assigned to each data point (i.e. each

electrode pair) and regression parameters are found by minimizing the weighted mean squared error of the fit. To estimate the effects of total distance on the slopes between geometric mean/Fano factor and noise correlations, we employed weighted linear regression, where weights were determined by total distance. We used a Gaussian weighting profile and “slid” it across total distance values, calculating a new weighted regression for each location. This method has the advantage of using identical weights at each regression step for control, inactivation, and re-warming data since total distance did not vary within an experimental session. To provide a formal description of this method, we defined the weighted linear model of the data as

$$w_{t,p} \cdot c_p = w_{t,p} \cdot [x_p \cdot S_t + \beta_t],$$

where the subscript p specifies the electrode pair, the subscript t specifies the current total distance value on which the regression window is “centered,” c_p is the correlation value of the pair, x_p is the regressor for the pair (geometric mean or Fano factor), S_t and β_t are the regression slope and offset term used to fit the data when the regression window is centered on total distance t , and $w_{t,p}$ is the weight assigned to the pair for the given regression window. This weight is calculated as

$$w_{t,p} = \exp\{-(0.5/\sigma^2) \cdot (TD_p - t)^2\},$$

where σ is the standard deviation of the Gaussian window, t is the total distance on which the window is centered, and TD_p is the total distance value of the pair. We used a σ of 0.2° for this analysis. We performed a “sliding” regression by fitting the regression slope and offsets across a range of t values. The regression slope is the parameter that characterizes the interaction described above.

This analysis confirmed our intuition that total distance interacted with geometric mean/Fano

factor to determine their relationship with noise correlations. The slopes relating geometric mean (Figure 4.11A) and Fano factor (Figure 4.11B) to noise correlations were highest when considering pairs with the lowest total distance values. The shaded regions in Figure 4.11 represent the 95% confidence intervals bootstrapped from the data. Importantly, we employed what is arguably the most conservative estimation of the variance about this measurement by bootstrapping the data by day rather than by something more fine-grained like electrode pair. Therefore, these confidence intervals provide a representation of the consistency of these effects across days, not simply across the population of electrode pairs. In the control data, the slope coefficients relating geometric mean/Fano factor to noise correlations decrease as total distance increases. And, even with this conservative bootstrapping approach, we can see that these decreases over the first degree of total distance are well outside the variability of the measurements themselves within the control data. Therefore, the effect in the control data illustrated by these plots is extremely unlikely to have arisen by chance.

Framing the effect of total distance with the regression slopes is useful not only because it accounts for potential confounds when establishing the control effect but also because it levels the comparison between control and inactivation data. The fact that the relationship between geometric mean/Fano factor becomes much more consistent across pairs with disparate total distance values provides the most compelling description of a meaningful change during inactivation (Figure 4.11). In the same style that we resampled the data by day to bootstrap confidence intervals, we permuted the data by day, thus breaking the relationship of interest, in order to establish a null distribution and thereby determine the statistical significance of the changes observed during inactivation. To do so, we iteratively permuted our data by randomly

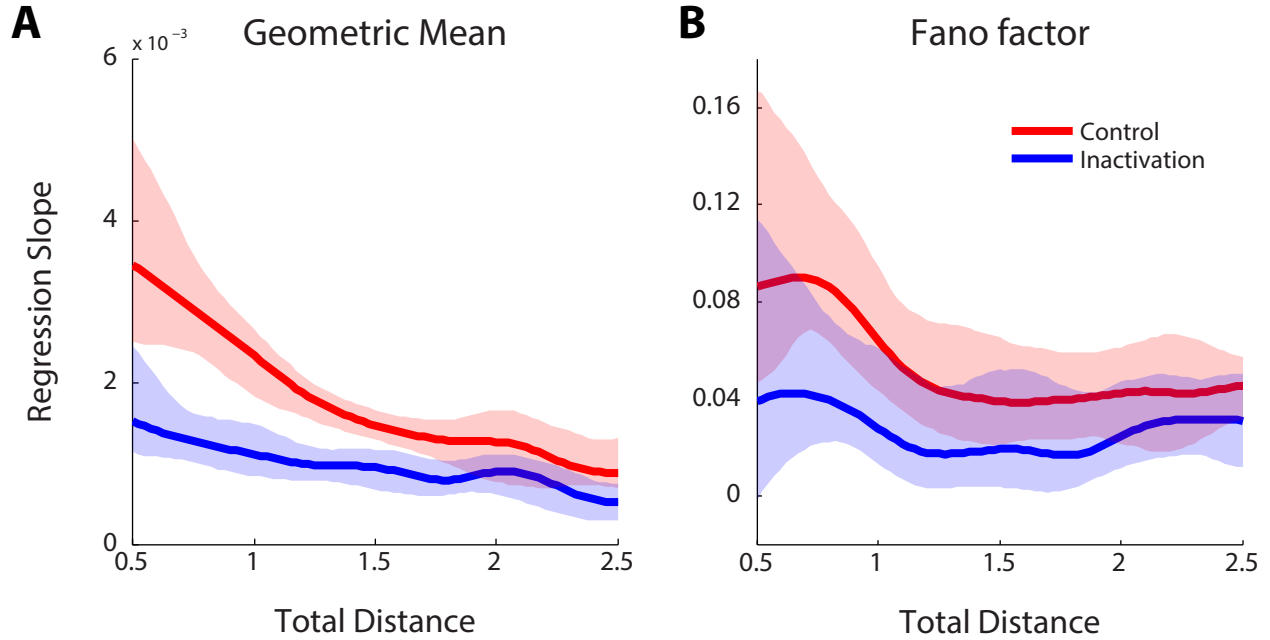


Figure 4.11 | Examining the relationship between geometric mean, Fano factor, and noise correlations along the range of total distances. **A**, coefficients from weighted linear regression (see main text) providing the slope between geometric mean and noise correlations as a function of total distance during control (red) and inactivation (blue). **B**, same as **A**, but for the slope between Fano factor and noise correlations. Solid lines indicate true averages, obtained with a sliding Gaussian window. Shaded regions indicate 95% confidence intervals about the mean obtained from bootstrapping.

switching the control and inactivation data within each of the eight sessions and repeated the sliding weighted regression analysis for each permutation. We observed that the differences between control and inactivation were significant at each total distance setting when considering geometric mean. The same was true for Fano factor within the regression windows centered on total distance values greater than 0.6° and less than 2° . Examining the p-value at each time point might be criticized for inviting the problem of multiple comparisons; however, we feel that the results are substantially unequivocal to warrant the use of this statistical characterization and point out that the typical solution of Bonferonni correction would yield a significance criterion that is less than the minimum p-value we could find with this particular permutation test (2^{-8}).

The observation that the regression coefficients were measurably larger for the control data even

for pairs that did not mutually overlap the stimulus center (total distance roughly larger than 1°) was somewhat surprising since correlations were larger for those pairs during inactivation (Figure 4.7B; 4.9A). However, our regression method also used an offset term and we observed this to be consistently higher during inactivation, suggesting that the increase in correlations was simply unrelated to our regressors (geometric mean and Fano factor), perhaps reflecting a general increase in baseline correlations.

4.5 Discussion

In this Chapter, we presented evidence that surround stimulation and feedback interact to structure noise correlations in V1. These interactions manifested as elevated correlations amongst multiunits whose RFs were both close to the stimulus center. Describing pairs in terms of total distance, a metric we used to quantify the overall distance between the RF centers and the stimulus center, revealed this pattern. Importantly, this pattern was visible in both center-only and center+surround responses and was highly unlikely to have arisen as an artifact produced by stimulus-dependent structure in overall firing rate or variability.

The most striking finding within this data was the effect of feedback inactivation on this correlation structure. We observed a substantial decrease in the stimulus-dependent correlation structure in exchange for a more spatially uniform pattern of correlations. We were able to identify this change in terms of total distance and in a manner that accounts for potential confounds produced by changes in overall firing rate or variability. Furthermore, the correlations returned to baseline in the re-warming data.

Our main goal when developing a rigorous quantification of this effect was to describe

correlations in terms of total distance while accounting for geometric mean/Fano factor, since these factors tended to covary (Figure 4.9). The analysis we settled on achieved this goal by focusing on interactions between total distance, noise correlations, and geometric mean/Fano factor. We are agnostic to how these interactions might be interpreted. For example, the tendency for correlations to be strongest between pairs with large geometric means could be interpreted as a consequence of sampling bias introduced by measuring noisy, discrete data (Cohen & Kohn, 2011). Alternatively, this tendency has been shown to be of computational use in the retina (Zylberberg et al., 2016; Franke et al., 2016), raising the possibility that it reflects a computational feature of the network that we are recording from. This would also suggest that individual pairs should be most correlated when responding to stimuli that produce their largest geometric mean. We observed a slight trend in this direction in the center+surround data but it did not appear to change when feedback was inactivated (data not shown). It is possible that a compelling interpretation of the interactions between total distance, geometric mean/Fano factor, and correlations, as well as their feedback-dependency, will result from different analyses. However, we leave that to future work.

Developing an interpretation of just the relationship between total distance and noise correlations is itself challenging. That is, we did not approach the analysis of noise correlations with any particular hypotheses concerning how they might be organized. The first possible explanation that came to mind was that the manner in which correlations tracked the stimulus center was a reflection of figure-ground segregation. V1 neurons have been shown to respond more strongly when their visual inputs are part of a textured figure, where the perception of a figure arises out of some form of feature-contrast with the background (Lamme, 1995; Zipser et al., 1996; Self et al., 2012; van

Kerkoerle et al., 2014). These response patterns are reminiscent of the different effects of iso-oriented and orthogonal center+surround stimuli (Sections 2.3.1 and 3.3) and we reasoned that the juxtaposition of the center grating and the annular surround in our stimuli may recruit the same mechanisms that allow neurons in V1 to distinguish figures from background, especially since this mechanism has been suggested to use feedback (Lamme & Roelfsema, 2000; Roelfsema, 2006). In a separate set of experiments, we recorded responses to Kanizsa squares, which offer an elegant way to manipulate the perception of illusory figures with low-level stimulus changes. We chose this stimulus approach because responses to illusory stimuli in V1 are rarely visible in the activity of single neurons (Lee, 2003), which gave us the best chance to separate first-order and second-order effects of illusory scene segmentation. Over multiple sessions of recording, we found no evidence that correlations were sensitive to the perceptual location of the illusory figure. This corroborates the finding that figure-ground segregation has been shown to have little effect on synchrony (Lamme & Spekreijse, 1998), which are noise correlations on very small timescales, and supports the observation that the link between perceptual grouping and rate covariance measured on a longer timescale (more like the one considered here) reflects attention (Roelfsema et al., 2004), which we did not manipulate. Therefore, we believe that an explanation of the correlation patterns we observed is unlikely to involve mechanisms devoted to scene segmentation. That is not to say that the correlation patterns we observed would not be visible with the class of figure-ground stimuli that rely on feature-contrast; we lack the data to make such a claim. We instead would argue that such feature-contrast (or general center/surround discontinuity) might be the requisite characteristic of the stimuli for observing this effect on correlations. Whether that should be regarded as figure-ground segregation is a matter of semantics.

Some insight into the nature of the correlations observed amongst low total distance pairs comes from the observation that their response variability was larger compared with pairs of larger total distance values and that this description was less applicable during feedback inactivation (Figure 4.9E). This might suggest that the elevated correlations amongst the low total distance pairs reflected a shared input related to the stimulus center. If this input was mediated through feedback projections, that would explain why the effects of their inactivation could be described as the loss of a source of shared variance. This description, however, doesn't readily explain the simultaneous increase in correlations observed for the majority of pairs, most of which did not align with the stimulus center.

How, then, might we interpret the influence of stimulus location on noise correlations? It is clear that feedback inactivation shapes this influence and this is only true in the presence of a surround stimulus. Considering the ways in which the surround has been shown to influence the representation of the stimulus center, it is rather surprising that we were unable to find any clear effect of center/surround configuration on noise correlations. This lack of effect is only briefly mentioned in the results portion of this Chapter, but it was consistently the result whenever we analyzed the data according to center/surround configuration.

However, it would seem premature to conclude that the analyses we used to explore the influence of stimulus location fully captured the richness of the correlation patterns within these data (let alone within V1). It is widely accepted that cortical neurons do not work in isolation and that even first-order response properties of these neurons reflect complex network interactions. And yet, while it is growing, the analytical toolkit for deciphering second-order measurements like noise correlations remains quite sparse. Furthermore, taking the analyses that yield insight into

the behavior of individual neurons and conceiving of how to adapt such analyses such that they might similarly help explain the behaviors of pairs of neurons is very difficult. This is a difficulty that we did not overcome, in the sense that the analyses of Chapters 3 and 4 are largely disconnected. The goal of future work should be to use a platform for thoroughly characterizing the structure of covariability in the data in such a way that we may be able to better connect the first-order descriptions of contextual modulation presented in Chapter 3 with the second-order descriptions presented here.

How might one approach such a goal? Rather than trying to develop and consider all the potentially relevant pairwise descriptions of the data with the hopes that some of them will offer explain correlations in an interpretable way, it is perhaps better to use analyses that start at the level of correlations but provide descriptions of the data that can be related to individual neurons. A popular example of this is principal component analysis, which provides a low-dimensional description of high-dimensional data using the eigenvectors of its covariance matrix. These eigenvectors provide a way to connect the structure of the correlations to individual units and can therefore make more parametric descriptions of second-order behavior tractable. However, the data we collected indicated flexibility within the correlations themselves and principal component analysis is best applied when fixed covariance is a reasonable assumption.

We are intrigued by the work of Karklin and Lewicki (2008). They developed a model for V1 responses based on the premise that V1 forms representations of the distributions (that is, distributions of pixel intensities) that describe patches of natural images. The goal of the system when presented with a given image patch is to represent the covariance matrix that the input image is mostly likely to have been drawn from. In this sense, an image is represented as a

generalization of all the images that stem from the same distribution and Karklin and Lewicki found that, over the course of learning, these representations took on properties of complex cells in V1. Importantly, this model encompasses the benefits of principal component analysis since it summarizes inputs according to covariance matrices—whose eigenvectors can be analyzed as described above. This point was excellently demonstrated by the authors when they showed how the covariance structure represented by one of their model neurons gave rise to the phase invariance associated with complex cells.

Their work started with a dataset composed of image patches; the system then learned the covariance “building blocks” such that any individual example from their dataset could be represented with a particular combination of these covariance structures. Each V1 neuron in their model encoded one such combination. Our application of this model would begin with the dataset analyzed above, which is composed of trials, where instead of pixels we have multiunits and instead of intensities we have firing rates. Conceivably, training a representation of our data using this model would provide access to the fundamental covariance structures that describe the data as a whole as well as the combinations of these structures that tend to exemplify a given trial. Furthermore, the representations used to describe a trial would be of immense interest. In the original work starting with natural image data, these representations mimicked V1. In our work, which would be starting with actual V1 data, these representations would, by analogy, serve as a model for V2/V3. Therefore, the most exciting possibility associated with this future computational project is that it will not only provide access to the flexible structure of the correlations that we observed but also that it will deepen our understanding of what properties of V1 have changed when V2/V3 are inactivated. In theory, this approach could allow us to quantify

each individual trial from the level of stimulus parameters to the level of high-order response structure, providing an unprecedented ability to tractably relate these descriptions of neural representation.

The combined insights from theory and experimentation point more and more to the importance of understanding the properties of natural images if we are to understand contextual modulation of early visual processing and how it is influenced by feedback. It is likely that understanding how these mechanisms interact to shape responses to natural stimuli will be essential if we are ever to meaningfully understand the role of either in cortical computation. To be certain, the continuation of the work presented in this dissertation will arrive at natural images and will require a method to understand how the responses of populations of neurons to these stimuli are affected by perturbations of feedback. By adapting flexible models of higher-order interaction to the study of neural populations, we hope to contribute not only a better understanding of our own data but, more generally, a method to understand the richness and intricacies of such complex data as one is sure to encounter when dealing with natural stimuli.

References

- Adesnik, H. & Scanziani, M. (2010). Lateral competition for cortical space by layer-specific horizontal circuits. *Nature*, 464(7292), 1155–1160.
- Albus, K. & Beckmann, R. (1980). Second and third visual areas of the cat: interindividual variability in retinotopic arrangement and cortical location. *The Journal of physiology*, 299, 247–276.
- Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14(2), 105–126.
- Alonso, J. M., Cudeiro, J., Pérez, R., Gonzalez, F., & Acuña, C. (1993a). Influence of layer V of area 18 of the cat visual cortex on responses of cells in layer V of area 17 to stimuli of high velocity. *Experimental brain research*, 93(2), 363–366.
- Alonso, J. M., Cudeiro, J., Pérez, R., Gonzalez, F., & Acuña, C. (1993b). Orientational influences of layer V of visual area 18 upon cells in layer V of area 17 in the cat cortex. *Experimental brain research*, 96(2), 212–220.
- Annegers, J. F., Hauser, W. A., Coan, S. P., & Rocca, W. A. (1998). A population-based study of seizures after traumatic brain injuries. *The New England journal of medicine*, 338(1), 20–24.
- Bair, W., Cavanaugh, J. R., & Movshon, J. A. (2003). Time course and time-distance relationships for surround suppression in macaque V1 neurons. *Journal of Neuroscience*, 23(20), 7690–7701.
- Bardy, C., Huang, J. Y., Wang, C., FitzGibbon, T., & Dreher, B. (2009). 'Top-down' influences of ipsilateral or contralateral postero-temporal visual cortices on the extra-classical receptive fields of neurons in cat's striate cortex. *NSC*, 158(2), 951–968.
- Barlow, H. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. *NPL Symposium on the Mechanization of Thought Process*, (pp. 535–539).
- Benucci, A., Saleem, A. B., & Carandini, M. (2013). Adaptation maintains population homeostasis

- in primary visual cortex. *Nature Publishing Group*, 16(6), 724–729.
- Blasdel, G. & Campbell, D. (2001). Functional retinotopy of monkey visual cortex. *Journal of Neuroscience*, 21(20), 8286–8301.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 12(12), 4745–4765.
- Bullier, J., Hupé, J. M., James, A., & Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of physiology, Paris*, 90(3-4), 217–220.
- Busse, L., Wade, A. R., & Carandini, M. (2009). Representation of Concurrent Stimuli by Population Activity in Visual Cortex. *Neuron*, 64(6), 931–942.
- Carandini, M. & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264(5163), 1333–1336.
- Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 17(21), 8621–8644.
- Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002a). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, 88(5), 2530–2546.
- Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002b). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *Journal of Neurophysiology*, 88(5), 2547–2556.
- Chance, F. S., Abbott, L. F., & Reyes, A. D. (2002). Gain modulation from background synaptic input. *Neuron*, 35(4), 773–782.
- Coen-Cagli, R., Dayan, P., & Schwartz, O. (2009). Statistical Models of Linear and Non-linear Contextual Interactions in Early Visual Processing. In *Advances in Neural Information Processing Systems*.

- Coen-Cagli, R., Dayan, P., & Schwartz, O. (2012). Cortical Surround Interactions and Perceptual Salience via Natural Scene Statistics. *PLoS Computational Biology*, 8(3), e1002405.
- Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience*, 18(11), 1648–1655.
- Cohen, M. R. & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7), 811–819.
- Cohen, M. R. & Newsome, W. T. (2008). Context-dependent changes in functional circuitry in visual area MT. *Neuron*, 60(1), 162–173.
- Cohen, M. R. & Newsome, W. T. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *Journal of Neuroscience*, 29(20), 6635–6648.
- Cumming, B. G. & Nienborg, H. (2016). Feedforward and feedback sources of choice probability in neural population responses. *Current opinion in neurobiology*, 37, 126–132.
- De Pasquale, R. & Sherman, S. M. (2013). A modulatory effect of the feedback from higher visual areas to V1 in the mouse. *Journal of Neurophysiology*, 109(10), 2618–2631.
- Destexhe, A., Rudolph, M., & Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nature reviews. Neuroscience*, 4(9), 739–751.
- Dodd, J. V., Krug, K., Cumming, B. G., & Parker, A. J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *Journal of Neuroscience*, 21(13), 4809–4821.
- Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 1(1), 1–47.
- Ferguson, P. L., Smith, G. M., Wannamaker, B. B., Thurman, D. J., Pickelsimer, E. E., & Selassie, A. W. (2010). A population-based study of risk of epilepsy after hospitalization for traumatic brain injury. *Epilepsia*, 51(5), 891–898.
- Franke, F., Fiscella, M., Sevelev, M., Roska, B., Hierlemann, A., & da Silveira, R. A. (2016).

- Structures of Neural Correlation and How They Favor Coding. *Neuron*, 89(2), 409–422.
- Freeman, J. & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14(9), 1195–1201.
- Freiwald, W. A. & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005), 845–851.
- Galuske, R. A. W., Schmidt, K. E., Goebel, R., Lomber, S. G., & Payne, B. R. (2002). The role of feedback in shaping neural representations in cat visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), 17083–17088.
- Gilbert, C. D., Li, W., & Piëch, V. (2009). Perceptual learning and adult cortical plasticity. *The Journal of physiology*, 587(Pt 12), 2743–2751.
- Haefner, R. M., Berkes, P., & Fiser, J. (2014). The implications of perception as probabilistic inference for correlated neural variability during behavior. *eprint arXiv:1409.0257*, (pp.--).
- Haider, B., Krause, M. R., Duque, A., Yu, Y., Touryan, J., Mazer, J. A., & McCormick, D. A. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron*, 65(1), 107–121.
- Hashemi-Nezhad, M. & Lyon, D. C. (2012). Orientation tuning of the suppressive extraclassical surround depends on intrinsic organization of V1. *Cerebral Cortex*, 22(2), 308–326.
- Herrero, J. L., Gieselmann, M. A., Sanayei, M., & Thiele, A. (2013). Attention-Induced Variance and Noise Correlation Reduction in Macaque V1 Is Mediated by NMDA Receptors. *Neuron*, 78(4), 729–739.
- Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148, 574–591.
- Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160, 106–154.
- Hubel, D. H. & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28, 229–289.

- Hupé, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology*, 85(1), 134–145.
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695), 784–787.
- Ito, M. & Gilbert, C. D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron*, 22(3), 593–604.
- Jeanne, J. M., Sharpee, T. O., & Gentner, T. Q. (2013). Associative Learning Enhances Population Coding by Inverting Interneuronal Correlation Patterns. *Neuron*, 78(2), 352–363.
- Jehee, J. F. M., Rothkopf, C., Beck, J. M., & Ballard, D. H. (2006). Learning receptive fields using predictive feedback. *Journal of physiology, Paris*, 100(1-3), 125–132.
- Jones, H. E., Wang, W., & Sillito, A. M. (2002). Spatial organization and magnitude of orientation contrast interactions in primate V1. *Journal of Neurophysiology*, 88(5), 2796–2808.
- Karklin, Y. & Lewicki, M. S. (2005). A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural computation*, 17(2), 397–423.
- Karklin, Y. & Lewicki, M. S. (2008). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, 457(7225), 83–87.
- Kohn, A. & Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, 25(14), 3661–3673.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural ...*, (pp. 1097–1105).
- Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 15(2), 1605–1615.
- Lamme, V. A. & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and

- recurrent processing. *Trends in neurosciences*, 23(11), 571–579.
- Lamme, V. A. & Spekreijse, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature*, 396(6709), 362–366.
- Lee, J. & Maunsell, J. H. R. (2009). A normalization model of attentional modulation of single unit responses. *PLoS ONE*, 4(2), e4651.
- Lee, J. & Maunsell, J. H. R. (2010). Attentional modulation of MT neurons with single or multiple stimuli in their receptive fields. *Journal of Neuroscience*, 30(8), 3058–3066.
- Lee, T. S. (2003). Computations in the early visual cortex. *Journal of physiology, Paris*, 97(2-3), 121–139.
- Li, N. & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, 321(5895), 1502–1507.
- Lochmann, T. & Denève, S. (2011). Neural processing as causal inference. *Current opinion in neurobiology*, 21(5), 774–781.
- Lochmann, T., Ernst, U. A., & Denève, S. (2012). Perceptual inference predicts contextual modulations of sensory responses. *Journal of Neuroscience*, 32(12), 4179–4195.
- MacEvoy, S. P., Tucker, T. R., & Fitzpatrick, D. (2009). A precise form of divisive suppression supports population coding in the primary visual cortex. *Nature Neuroscience*, 12(5), 637–645.
- Mariño, J., Schummers, J., Lyon, D. C., Schwabe, L., Beck, O., Wiesing, P., Obermayer, K., & Sur, M. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. *Nature Neuroscience*, 8(2), 194–201.
- Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huisoud, C., Clavagnier, S., Giroud, P., Sappey-Marini, D., Barone, P., Dehay, C., Toroczkai, Z., Knoblauch, K., Van Essen, D. C., & Kennedy, H. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral Cortex*, 24(1), 17–36.
- Martinez-Conde, S., Cudeiro, J., Grieve, K. L., Rodriguez, R., Rivadulla, C., & Acuña, C. (1999).

- Effects of feedback projections from area 18 layers 2/3 to area 17 layers 2/3 in the cat visual cortex. *Journal of Neurophysiology*, 82(5), 2667–2675.
- Martinez-Trujillo, J. C. & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current biology : CB*, 14(9), 744–751.
- Maunsell, J. H. R. & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in neurosciences*, 29(6), 317–322.
- McAdams, C. J. & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19(1), 431–441.
- McIlwain, J. T. (1964). Receptive fields of optic tract axons and lateral geniculate cells: peripheral extent and barbiturate sensitivity. *Journal of Neurophysiology*, 27, 1154–1173.
- Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological cybernetics*, 66(3), 241–251.
- Murphy, P. C., Duckett, S. G., & Sillito, A. M. (1999). Feedback connections to the lateral geniculate nucleus and cortical response properties. *Science*, 286(5444), 1552–1554.
- Nassi, J. J. & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nature Publishing Group*, 10(5), 360–372.
- Nassi, J. J., Gómez-Laberge, C., Kreiman, G., & Born, R. T. (2014). Corticocortical feedback increases the spatial extent of normalization. *Frontiers in systems neuroscience*, 8, 105.
- Nassi, J. J., Lomber, S. G., & Born, R. T. (2013). Corticocortical Feedback Contributes to Surround Suppression in V1 of the Alert Primate. *Journal of Neuroscience*, 33(19), 8504–8517.
- Ni, A. M., Ray, S., & Maunsell, J. H. R. (2012). Tuned Normalization Explains the Size of Attention Modulations. *Neuron*, 73(4), 803–813.
- Nienborg, H. & Cumming, B. G. (2006). Macaque V2 neurons, but not V1 neurons, show choice-related activity. *Journal of Neuroscience*, 26(37), 9567–9578.

- Nienborg, H. & Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, 459(7243), 89–92.
- Okun, M. & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, 11(5), 535–537.
- Olshausen, B. A. & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), 607–609.
- Ozeki, H., Finn, I. M., Schaffer, E. S., Miller, K. D., & Ferster, D. (2009). Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4), 578–592.
- Ponce, C. R., Hunter, J. N., Pack, C. C., Lomber, S. G., & Born, R. T. (2011). Contributions of indirect pathways to visual response properties in macaque middle temporal area MT. *Journal of Neuroscience*, 31(10), 3894–3903.
- Ponce, C. R., Lomber, S. G., & Born, R. T. (2008). Integrating motion and depth via parallel pathways. *Nature Neuroscience*, 11(2), 216–223.
- Rabinowitz, N. C., Goris, R. L., Cohen, M., & Simoncelli, E. P. (2015). Attention stabilizes the shared gain of V4 populations. *eLife*, 4.
- Ramalingam, N., McManus, J. N. J., Li, W., & Gilbert, C. D. (2013). Top-Down Modulation of Lateral Interactions in Visual Cortex. *Journal of Neuroscience*, 33(5), 1773–1789.
- Rao, R. P. & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19(5), 1736–1753.
- Reynolds, J. H. & Desimone, R. (2003). Interacting Roles of Attention and Visual Salience in V4. *Neuron*, 37(5), 853–863.
- Roberts, M., Delicato, L. S., Herrero, J., Gieselmann, M. A., & Thiele, A. (2007). Attention alters spatial integration in macaque V1 in an eccentricity-dependent manner. *Nature Neuroscience*,

10(11), 1483–1491.

Rockland, K. S. & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain research*, 179(1), 3–20.

Rockland, K. S. & Pandya, D. N. (1981). Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. *Brain research*, 212(2), 249–270.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual review of neuroscience*, 29, 203–227.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience*, 7(9), 982–991.

Ruff, D. A. & Cohen, M. R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. *Nature Neuroscience*, 17(11), 1591–1597.

Sandell, J. H. & Schiller, P. H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *Journal of Neurophysiology*, 48(1), 38–48.

Sceniak, M. P., Hawken, M. J., & Shapley, R. (2001). Visual spatial characterization of macaque V1 neurons. *Journal of Neurophysiology*, 85(5), 1873–1887.

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

Schmidt, K. E., Lomber, S. G., Payne, B. R., & Galuske, R. A. W. (2011). Pattern motion representation in primary visual cortex is mediated by transcortical feedback. *NeuroImage*, 54(1), 474–484.

Schuett, S., Bonhoeffer, T., & Hübener, M. (2002). Mapping retinotopic structure in mouse visual cortex with optical imaging. *Journal of Neuroscience*, 22(15), 6549–6559.

Schwabe, L., Obermayer, K., Angelucci, A., & Bressloff, P. C. (2006). The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. *Journal of Neuroscience*, 26(36), 9117–9129.

- Schwartz, O., Sejnowski, T. J., & Dayan, P. (2006). Soft mixer assignment in a hierarchical generative model of natural scene statistics. *Neural computation*, 18(11), 2680–2718.
- Schwartz, O. & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8), 819–825.
- Self, M. W., Kooijmans, R. N., Supèr, H., Lamme, V. A., & Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proceedings of the National Academy of Sciences of the United States of America*, 109(27), 11031–11036.
- Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(4), 1486–1510.
- Shen, Z.-M., Xu, W.-F., & Li, C.-Y. (2007). Cue-invariant detection of centre-surround discontinuity by V1 neurons in awake macaque monkey. *The Journal of physiology*, 583(Pt 2), 581–592.
- Shipp, S., Watson, J. D., Frackowiak, R. S., & Zeki, S. (1995). Retinotopic maps in human prestriate visual cortex: the demarcation of areas V2 and V3. *NeuroImage*, 2(2), 125–132.
- Shushruth, S., Ichida, J. M., Levitt, J. B., & Angelucci, A. (2009). Comparison of spatial summation properties of neurons in macaque V1 and V2. *Journal of Neurophysiology*, 102(4), 2069–2083.
- Shushruth, S., Mangapathy, P., Ichida, J. M., Bressloff, P. C., Schwabe, L., & Angelucci, A. (2012). Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. *Journal of Neuroscience*, 32(1), 308–321.
- Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378(6556), 492–496.
- Smolyanskaya, A., Haefner, R. M., Lomber, S. G., & Born, R. T. (2015). A Modality-Specific Feedforward Component of Choice-Related Activity in MT. *Neuron*, 87(1), 208–219.
- Snyder, A. C., Morais, M. J., Kohn, A., & Smith, M. A. (2014). Correlations in v1 are reduced by stimulation outside the receptive field. *Journal of Neuroscience*, 34(34), 11222–11227.

- Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision research*, 48(12), 1391–1408.
- Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *Journal of Neuroscience*, 30(9), 3531–3543.
- Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. *Vision research*, 51(6), 563–576.
- Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, 36(4), 739–750.
- Sun, Y. J., Wu, G. K., Liu, B.-h., Li, P., Zhou, M., Xiao, Z., Tao, H. W., & Zhang, L. I. (2010). Fine-tuning of pre-balanced excitation and inhibition during auditory cortical development. *Nature*, 465(7300), 927–931.
- Sundberg, K. A., Mitchell, J. F., & Reynolds, J. H. (2009). Spatial attention modulates center-surround interactions in macaque visual area v4. *Neuron*, 61(6), 952–963.
- Szegedy, C., Liu, W., Jia, Y., & Sermanet, P. (2015). Going deeper with convolutions. In *Proceedings of the ...*
- Szegedy, C., Toshev, A., & Erhan, D. (2013). Deep Neural Networks for Object Detection. *Advances in Neural Information ...*
- Treue, S. & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736), 575–579.
- Trott, A. R. & Born, R. T. (2015). Input-gain control produces feature-specific surround suppression. *Journal of Neuroscience*, 35(12), 4973–4982.
- Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., & McNaughton, B. L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 17(11), 4382–4388.
- van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward

- processing in monkey visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*.
- Vinje, W. E. & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456), 1273–1276.
- Vinje, W. E. & Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *Journal of Neuroscience*, 22(7), 2904–2915.
- Wang, C., Huang, J. Y., Bardy, C., FitzGibbon, T., & Dreher, B. (2010). Influence of 'feedback' signals on spatial integration in receptive fields of cat area 17 neurons. *Brain research*, 1328, 34–48.
- Wang, W., Jones, H. E., Andolina, I. M., Salt, T. E., & Sillito, A. M. (2006). Functional alignment of feedback effects from visual cortex to thalamus. *Nature Neuroscience*, 9(10), 1330–1336.
- Williams, Z. M., Elfar, J. C., Eskandar, E. N., Toth, L. J., & Assad, J. A. (2003). Parietal activity and the perceived direction of ambiguous apparent motion. *Nature Neuroscience*, 6(6), 616–623.
- Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., & de la Rocha, J. (2015). Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. *Nature Communications*, 6, 6177.
- Zagha, E., Casale, A. E., Sachdev, R. N. S., McGinley, M. J., & McCormick, D. A. (2013). Motor Cortex Feedback Influences Sensory Processing by Modulating Network State. *Neuron*, (pp. 1–12).
- Zipser, K., Lamme, V. A., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(22), 7376–7389.
- Zylberberg, J., Cafaro, J., Turner, M. H., Shea-Brown, E., & Rieke, F. (2016). Direction-Selective Circuits Shape Noise to Ensure a Precise Population Code. *Neuron*, 89(2), 369–383.