Inter-area communication in the brain: a population-level approach

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

João D. Semedo

B.S., Biomedical Engineering, Instituto Superior Técnico

M.S., Biomedical Engineering, Instituto Superior Técnico

Carnegie Mellon University

Pittsburgh, PA

April, 2018

Thesis committee:

Byron M. Yu, chair

Adam Kohn

Christian K. Machens

Marlene Cohen

Patrícia Figueiredo

Steve Chase

Acknowledgments

In the summer of 2012, I was given the now rare opportunity to initiate a Dual-PhD with no strings attached. I was free to seek out whatever subject or project I felt most passionate about. Having just finished a degree in Biomedical Engineering, I flew from Lisbon to Pittsburgh with one thing certain in my mind: I would definitely not work on anything remotely related to life sciences ever again. If you've read the title of dissertation, and I hope you did, you will know I spectacularly failed to achieve this. The reason was Byron.

After a week of interviewing with faculty all over CMU no one made me feel as excited about what I could achieve in my PhD. Talking with Byron was inspiring from day one, as his clarity of though covers statistical methods and neuroscience questions to the same (thorough) extent. Once I convinced Byron to take me in, we started thinking about who could co-advise me from the Portuguese side. Byron immediately suggested Christian, who had just joined the Champalimaud Center for the Unknown, just off of Lisbon. Returning from Pittsburgh to meet Christian completed what must have been the most inefficient trip ever taken from the center of Lisbon to its periphery. Meeting Christian again confirmed my (newly) formed conviction that I wanted to work in Neuroscience. Christian's background in physics shone through and provided, both then and throughout my PhD, a new take on subjects I though I knew. It is impossible to overstate how crucial this diversity of thought was for my development throughout the past ~~five~~ six years.

Both Byron and Christian work on computation neuroscience, and neither lab conducts experiments. As luck would have it, neuronal data is pretty important for neuroscience work. We turned to Adam, who had recently collected the amazing recordings analyzed in this dissertation. I had some ideas regarding the V1-V2 datasets. Adam politely clarified how wrong these ideas were. But fortunately he decided to stick around and work with me. Over the course of the next few years, he effectively became my third advisor. His input was invaluable in guiding the scientific questions we pursued, and he was instrumental in keeping our analyses and interpretations grounded. I cannot stress enough how deeply grateful I am to Adam, Byron and Christian for their support, their guidance, and for all the time they invested in me throughout this process (and it was a lot). I also wish to thank my thesis committee, Adam Kohn, Byron Yu, Christian Machens, Marlene Cohen, Patrícia Figueiredo and Steve Chase for the thoughtful questions and

guidance.

Nothing in science happens in a vacuum, and all ideas are undoubtedly shaped by the interactions that happen daily with colleagues and collaborators. The environment fostered by both Byron and Christian in their labs contributed to this in no small way, and made it an absolute pleasure to work alongside a number of talented students and postdocs, many of whom made critical contributions to the way I think about my work. I also want to express my gratitude towards the Scabby group, the weekly Chase/Batista/Yu lab journal club, for I could always count on their critical thinking to wrap my head around the latest work in neuroscience.

I have also received tremendous support from an entire team of people at the CMU-Portugal program, without whom navigating the bureaucratic process of simultaneously living in Portugal and the US would not have been possible. In particular, I must single out João Paulo Costeira and Ana Mateus. Were it not for João Paulo, I would not be in this program, and none of this work would have happened, pretty simple really. His endless support and optimism brought me in and saw me trough this process. Were it not for Ana, I would still be in Portugal trying to figure out how to apply for a VISA. She went above and beyond her duty in the way she took care of all PhD students in the CMU-Portugal program. I have also received generous financial support from Fundação para a Ciência e a Tecnologia and from Carnegie Mellon's Carnegie Institute of Technology through the John and Claire Bertucci Graduate Fellowship. Adam, Byron and Christian have also provided generous support through their Simon's Foundation joint grant.

Finally none of this would have been possible in the first place without the support of my partner and family. Asma, I will never be able to thank you enough for all you have done for me. You had to endure me at my worse and get me back on track. Thank you Mom, Dad, Raquel and Leonor. Not just for your encouragement, love and support. If it were not for your sacrifices, I would not have made it from Nisa, a small town of 3,000 people in the center of Portugal, to Lisbon and the United States.

Abstract

All brain functions, from seeing and moving to thinking, rely on the interaction of multiple, functionally distinct brain areas. We know, however, very little about how different areas interact at the level of networks of neurons, or what mechanisms are used to control the routing of information through the brain. Only very recently has technology evolved to the point where we can simultaneously monitor multiple neurons in various brain areas. While such experiments enable a host of new and exciting questions about inter-area interaction, they also pose significant analysis and interpretation challenges.

Here, we approach the problem of studying population-level interactions across brain areas using dimensionality reduction methods. In short, dimensionality reductions methods extract a small set of latent variables that summarize a given aspect of the data. Traditionally, these methods have been used to extract low-dimensional summaries of the population activity structure within a brain area. We propose to instead extract a set of latent variables that summarize the interaction between brain areas, i.e., instead of capturing the dominant features of the activity within an area, they capture the features that are relevant to its downstream targets.

We used this approach to characterize both the population-level structure and the dynamics of the interactions between populations of neurons in two cortical areas, visual areas V1 and V2. We found that V1-V2 interactions occur through a communication subspace: V2 fluctuations are related to a small subset of V1 population activity patterns, distinct from the largest fluctuations shared among neurons within V1. We propose that the communication subspace may be a general, population-level mechanism by which activity can be selectively routed across brain areas. Furthermore, we found these interactions to be dynamic and flexible, changing rapidly under different stimulus contexts. This work thus provides a foundation for studying how multiple populations of neurons interact and how this interaction supports brain function.

Contents

1 Introduction **1**

2 Previous work **5**

2.1 Studying large-scale neuronal activity 5

2.1.1 Dimensionality reduction 6

2.2 Studying inter-area communication 7

2.2.1 Communication through coherence (CTC) 8

2.2.2 Synchrony as a proxy for communication 8

3 V1-V2 neuronal recordings **11**

3.1 Physiology, visual stimulation and recordings 11

3.2 Data preprocessing 14

4 Inter-areal interactions occur through a communication subspace **17**

4.1 Strength of inter-area interactions 18

4.1.1 V1-V1 and V1-V2 interactions are similar in strength 18

4.1.2 Estimating predictable variability in the target populations 21

4.2 Structure of inter-area interactions 24

4.2.1 V1-V2 interactions use only a small number of dimensions 26

4.2.2 Low-dimensional V1-V2 interaction is not due to low-dimensional V2 activity 26

4.2.3 Using linear methods to study nonlinear computations 35

4.3 Discussion 40

4.4 Methods 44

4.4.1 Pairwise correlations . . . 44
4.4.2 Linear regression models . . . 44
4.4.3 Factor analysis . . . 46
4.4.4 Selective communication simulation . . . 46

5 Properties of the communication subspace **49**
5.1 Relationship to source activity structure . . . 50
5.1.1 V2 predictive dimensions are not aligned with target V1 predictive dimensions 50
5.1.2 The dominant dimensions of V1 are not the most predictive of V2 . . . 52
5.1.3 The communication subspace cannot be explained by a physical bottleneck . 54
5.1.4 Some dimensions of the source population activity remain private regardless of the stimulus . . . 58
5.2 Relationship to signal communication . . . 61
5.2.1 Predictive dimensions identified using residual activity can predict responses that include stimulus information . . . 62
5.3 Dependence on receptive field alignment, stimulus and anesthesia . . . 64
5.3.1 V1-V2 interaction structure is distinct for retinotopically-offset V2 populations 64
5.3.2 V1-V2 interactions driven by naturalistic stimuli also occur through a communication subspace . . . 66
5.3.3 Interactions between V1 and V4 in alert, behaving animals are also low-dimensional . . . 67
5.4 Discussion . . . 71
5.5 Methods . . . 73
5.5.1 Removing activity along the predictive dimensions . . . 73
5.5.2 Comparing dominant and predictive dimensions . . . 74

6 Dynamics of inter-areal interactions **75**
6.1 Strength of inter-area interactions across time . . . 76
6.1.1 V1-V2 interaction strength changes within and across trial epochs both in overall magnitude and temporal structure . . . 76
6.2 Structure of inter-area interactions across time . . . 78

6.2.1 V1-V2 interaction structure is distinct during evoked and spontaneous activity 78
6.3 Capturing intra- and inter- area dynamics: group latent auto-regressive analysis . . 83
6.3.1 Methods 84
6.3.2 Model validation 88
6.4 Discussion 93
6.5 Methods 96
6.5.1 Canonical correlation analysis 96
6.5.2 Ridge reduced rank regression 97

7 Summary and future directions **101**

List of Figures

3.1 A scaled representation of the cortical visual areas of the macaque 12

3.2 V1 and V2 recordings 13

4.1 V1-V1 and V1-V2 interactions are similar in strength 19

4.2 Similarity of predictive performance for target V1 and V2 is robust to bin size and number of trials 21

4.3 Absolute predictive performance can be explained by Poisson variability in the target population or subsampling from a large source population 23

4.4 Illustration of a low-dimensional interaction 25

4.5 V1-V2 interactions use only a small number of dimensions 27

4.6 Low-dimensional V1-V2 interaction is not due to low-dimensional V2 activity 29

4.7 The V1-V2 communication subspace was evident across a wide range of time bin widths, but was difficult to detect with limited numbers of trials 31

4.8 Predictive dimensions do not arise from capturing the activity of only a few target neurons 33

4.9 Difference in the number of V1 and V2 predictive dimensions is not due to differences in receptive field alignment 34

4.10 Predictive dimensions as local linear approximations of a globally nonlinear mapping 37

4.11 Reduced-rank regression recovers the correct number of predictive dimensions even when the mapping between populations is nonlinear 38

4.12 Communication subspaces enable selective communication with multiple downstream areas 43

5.1 V2 predictive dimensions are not aligned with target V1 predictive dimensions . . . 51
5.2 The dominant dimensions of V1 are not the most predictive of V2 54
5.3 The communication subspace cannot be explained by a physical bottleneck 56
5.4 Some dimensions of the source population activity remain private regardless of the stimulus 60
5.5 Predictive dimensions identified using residual activity can predict responses that include stimulus information 63
5.6 V1-V2 interaction structure is distinct for retinotopically-offset V2 populations . . . 65
5.7 V1-V2 interactions driven by naturalistic stimuli also occur through a communication subspace 67
5.8 Interactions between V1 and V4 in alert, behaving animals are also low-dimensional 70

6.1 V1-V2 interaction strength changes within and across trial epochs both in overall magnitude and temporal structure 77
6.2 V1-V2 interaction structure is distinct during evoked and spontaneous activity . . . 79
6.3 Differences in the number of predictive dimensions between evoked and spontaneous activity can be explained by differences in the source and target population statistics 81
6.4 Top predictive dimension is distinct across evoked and spontaneous epochs 82
6.5 Spontaneous activity predictive dimensions are well aligned with the dominant modes of spontaneous source V1 activity 83
6.6 Directed graphical models for multi-population activity 85
6.7 Comparing the optimal dimensionality for FA and pCCA 90
6.8 Model selection for AR-pCCA and gLARA 91
6.9 Leave-one-neuron-out prediction using gLARA 92
6.10 Temporal structure of coupling matrices for gLARA 93
6.11 Low firing rates during spontaneous activity can lead to overfitting 98

"In the beginning the Universe was created.

This has made a lot of people very angry, and has been widely regarded as a bad move."

- Douglas Adams, The Restaurant at the End of the Universe

Chapter 1

Introduction

Most brain functions involve interactions among multiple, distinct areas or nuclei. For instance, visual processing in primates requires the appropriate relaying of signals across dozens of distinct cortical areas. Yet our understanding of how populations of neurons in interconnected brain areas communicate is in its infancy.

Traditionally, recording technologies posed one of the biggest challenges to studying inter-area interaction at a population level. Indeed, a lot of past work on this subject relied on pairwise recordings (two electrodes), and focused on measures of synchrony between the spiking activity of individual neurons and population summary signals, such as local field potentials (LFPs). However, in the past decade, simultaneous recordings of large neuronal populations have become much more common, often including neurons spread across multiple brain areas. While these are exciting developments, enabling a host of new and exciting questions about inter-area interaction, they also pose significant analysis and interpretation challenges.

This dissertation has two main goals: (1) Develop and apply statistical tools, based on dimensionality reduction, to efficiently describe interactions between high-dimensional population recordings; (2) Take a population view of inter-area interaction, testing the plausibility of population-based mechanisms for the selective routing of information across brain areas. In other words, we aim to leverage recent developments in large-scale neuronal recording analysis to further the study of inter-area interactions. As such, we'll begin in Chapter 2 by providing a short background both on high-dimensional population analysis, namely dimensionality reduction approaches, and on past work on inter-area communication.

For all results presented in this dissertation, we used cutting edge electrophysiological recordings from our collaborators in Adam Kohn's group, one of a small number of groups to have recorded from large number of neurons in multiple brain areas. A description of all experimental procedures, as well as the pre-processing steps used in all analyses, are presented in Chapter 3.

The first contribution of this dissertation is the use of dimensionality reduction to provide a population-level characterization of inter-area interactions between visual areas V1 and V2. In Chapter 4, we show that while the strength of inter-area interactions is similar to that of interactions happening within V1, the structure is remarkably distinct. Specifically, we found that V1-V2 interactions occur through a communication subspace: V2 fluctuations are related to a small subset of V1 population activity patterns. Crucially, as we explore during our discussion of this finding, the existence of a communication subspace could subserve the selective routing of information between brain areas.

The finding that V1-V2 interactions occurs through a communication subspace implies that not all activity in V1 is effectively propagated to V2. In Chapter 5, we turn to understanding how the communication subspace relates to the structure of the activity within V1. We found that the communication subspace is not aligned with the largest shared fluctuations in V1 activity. In other words, the dominant patterns of activity in V1 are not the most effective at driving V2. Furthermore, while we determined the communication subspace by characterizing how activity fluctuations in V1 propagated to V2, we found that the communication subspace captured a significant portion of the stimulus induced variability. In an effort to determine whether the communication subspace might be a general property of inter-area interactions, we analyzed recordings under naturalistic stimuli, as well as recordings from V1 and V4 in behaving animals, and found these to be consistent with our finding for the V1-V2 interactions.

Having developed a framework for the study of population-level inter-area interactions, we proceeded in Chapter 6 to characterize how these interactions change across time and stimulus conditions. In other words, we turned to the study of the dynamics of inter-area interactions. We found that the population-level correlation between these areas was higher for spontaneous activity, while the temporal structure of the interaction changed from being feedforward dominated (V1 leading V2) early in the stimulus period to feedback dominated (V2 leading V1) late in the stimulus period and during spontaneous activity. We analyzed the population-level structure of

V1-V2 interactions and found that the dimensions of the V1 population activity that were involved during spontaneous activity were distinct from those involved for evoked activity. These results suggest that population-level interactions between V1 and V2 are dynamic and flexible – even for the simplest visual stimuli – and rich dynamics might mediate the way in which information is relayed between these areas. Motivated by this observation we developed a novel method, termed group Latent Auto-Regressive Analysis (gLARA), which enables us to directly capture intra- and inter-area dynamics simultaneously.

Finally, in Chapter 7, we summarize the main contributions of this dissertation and propose research directions that we believe will advance the goals of this project. The contributions of this dissertation have been published or submitted for publication as follows:

Chapter 4

J. D. Semedo, A. Zandvakili, C. K. Machens*, B. M. Yu*, A. Kohn* (2018). Cortical areas interact through a communication subspace. Under review at *Nature Neuroscience*.

B. R. Cowley, J. D. Semedo, A. Zandvakili, A. Kohn, M. A. Smith, B. M. Yu (2017). Distance covariance analysis. *Artificial Intelligence and Statistics (AISTATS)* (32% acceptance rate).

Chapter 5

J. D. Semedo, A. Zandvakili, C. K. Machens*, B. M. Yu*, A. Kohn* (2018). Cortical areas interact through a communication subspace. Under review at *Nature Neuroscience*.

Chapter 6

J. D. Semedo, A. Zandvakili, A. Kohn, C. K. Machens*, B. M. Yu* (2014). Extracting Latent Structure From Multiple Interacting Neural Populations. *Advances in Neural Information Processing Systems (NIPS) 27* (24% acceptance rate).

Chapter 2

Previous work

Most brain functions depend on the interaction of many neurons, spread across multiple brain areas. In recent years, recording technologies have evolved to the point where we can now record from dozens to hundreds of neurons simultaneously, and much work has focused on understanding how large populations of neurons work together to elicit behavior. How these large populations of neurons coordinate their activity across brain areas remains, however, elusive. Given a number of interconnected neuronal populations in different brain areas, how do these populations retain their functional specificity? What mechanisms control the transfer of information across brain areas?

This work sits at the intersection of two related questions: (1) How is neuronal activity structured across a large population of neurons? (2) How is inter-area communication structured? While we, as a field, are just beginning to attempt to jointly tackle these two questions, a lot of fruitful work as been put into approaching each of these questions separately. In this chapter, we provide a brief overview of this body of literature.

2.1 Studying large-scale neuronal activity

How can we leverage simultaneously recorded activity of a large population of neurons to generate and test hypothesis regarding network function? A substantial part of past work has focused on analyzing activity recorded from individual neurons, and studying how each neuron's responses related to the task/stimulus context under which the activity was recorded (Hubel and

Wiesel's work in the primary visual system is a fantastic example of this approach). Large-scale recordings potentially offer a straightforward improvement to this approach, offering a larger number of recorded neurons while being less susceptible to selection bias.

Given that computations in the brain are likely to occur at a population level (i.e., by many neurons working together), it is unclear that, at least for some systems, any one neuron will be particularly informative regarding the full-network function. At worse, studying single-neuron activity in isolation could even turn out to be misleading. How can we then simultaneously make sense of the activity of hundreds of neurons? One possible approach is to fit interaction models to the recorded activity. Briefly, these models seek to account for the activity in each and every neuron using some combination of all other neurons and possibly relevant task/stimulus parameters[1–3]. Studying these models can then reveal how the population activity as a whole is influenced by external parameters, or how to relate activity across any pair of neurons. However, while this may be a reasonable approach for small populations of neurons, the number of interactions grows with the square of the number of recorded neurons, which may make it difficult to summarize how larger populations of neurons interact[4].

2.1.1 Dimensionality reduction

An alternative to this approach comes through the use of dimensionality reduction methods. Briefly, this family of methods seeks to extract a small number of latent variables that summarize the recorded multivariate activity. Most dimensionality reduction methods can be categorized depending on how the latent variables are identified and how they relate to the recorded activity. As an example, consider one of the most commonly used dimensionality reduction methods, principal component analysis (PCA). In PCA, latent variables are identified by maximizing the total amount of variance explained (thereby summarizing the neuronal activity), and relate to the observed activity linearly.

Indeed, most dimensionality reduction methods used to characterize neuronal activity belong to the family of linear dimensionality reduction methods, where the latent variables are obtained through a linear combination of the multivariate activity. Because of this simple relationship, these methods can be seen as identifying the subspace of neuronal activity that is most relevant

for a given objective. PCA extracts the subspace that retains most variance. Factor analysis (FA), another linear dimensionality reduction method, extracts the subspace that retains most variance shared across neurons.

The application of dimensionality reduction methods has provided important insights in a variety of studies (see ref. 5 for a review), and has become an invaluable tool for exploratory analysis of large-scale recordings, providing good summaries of neuronal activity, and aiding in the formulation and testing of hypotheses. Interestingly, some recent studies have suggested that some of the subspaces identified through dimensionality reduction approaches might also have functional significance. Sadtler and colleagues[6] used a linear dimensionality reduction method, FA, to identify a small subspace of motor cortex activity that best summarized the shared variability across neurons. They then found, using a brain-machine interface paradigm, that monkeys could learn to control a cursor by modulating neuronal activity within this subspace, but failed to do so when required to modulate activity outside of the FA-defined subspace. Kaufman and colleagues[7] used a targeted dimensionality reduction approach to identify the subspace of motor cortex activity that was most predictive of muscle responses during a center-out reach task. They found that during the delay period of the task, where monkeys know what target they will be required to reach to, but must remain still, neuronal activity avoided this "potent subspace", remaining largely outside of it. This result suggested an effective mechanism for gating the propagation of motor cortex activity to the muscles.

In this dissertation, we aim to leverage dimensionality reduction approaches to study inter-area interaction. In other words, we will seek to identify sets of latent variables (or subspaces) that are specially adept at summarizing inter-area interactions between populations of neurons.

2.2 Studying inter-area communication

While the study of large-scale neuronal population activity has attracted a lot of interest over the past decade, we are only beginning to understand how populations of neurons interact across brain areas. Previous studies of inter-areal interactions have related the spiking activity of pairs of neurons in different areas[8–13], the spiking activity of a neuronal population in one area and a single neuron in another[14,15], the spiking activity of a neuron or group of neurons in one area and

the local field potential (LFP) in another[16–20], and the LFPs recorded in different areas[10,16,17,21,22]. Furthermore, be it directly or indirectly, most previous studies of inter-area communication have built on the idea that selective inter-area communication might be subserved, or achieved, by leveraging ongoing oscillations happening throughout the brain, particularly in the theta/gamma bands[16,23]. This has led to measures of interaction such as synchrony or phase locking/coherence[16–22].

2.2.1 Communication through coherence (CTC)

One of the most widespread theories of inter-area communication is that of communication through coherence (CTC)[24]. Briefly, the basis behind CTC is that ongoing oscillations cause periods of enhanced excitability within an area. This enhanced excitability makes it more likely that an incoming spike results in further spike generation. As a result, one can most effectively relay information from one area to another when both areas oscillate with the same frequency, and have a phase difference that is compatible with the spike propagation delay between the two areas. This type of organization would result in periods of spike generation in the upstream area being met with precisely timed periods of high excitability in the downstream area. Communication could then be"turned-off" either by changing the frequency of oscillation in one of the areas, or changing the phase relationship.

2.2.2 Synchrony as a proxy for communication

While the evidence for CTC is still a subject of debate, enhanced synchrony between brain areas is often linked to communication. Attending a specific visual location enhanced spike-LFP synchrony across neurons whose receptive field covered the attended location[16]. Visual working memory tasks elicited synchrony between prefrontal cortex (PFC) and posterior parietal cortex (PPC)[17]. Arce-McShane and colleagues[19] not only observed enhanced coherence between primary motor and sensory cortical areas during a tongue protrusion task, but found that this coherence emerged with task learning. Wong and colleages[20] found that neurons in PPC that were highly coherent with both local and long range LFPs (presumably involved in inter-area communication) were more informative of animals' decisions. These examples provide a compelling case for looking at synchrony as a hallmark of inter-area interaction. However, it is unclear whether

synchrony is promoting, and subserving, the interactions (as in CTC) or if it is a byproduct of effective communication, modulated through some other means.

Chapter 3

V1-V2 neuronal recordings

In all the work presented in this dissertation, we used cutting edge electrophysiological recordings from our collaborators in Adam Kohn's group, one of a small number of groups to have recorded from large number of neurons in multiple brain areas. Specifically, we analyzed recordings done simultaneously in primary visual cortex (V1) and visual area V2 in three sufentanil-anesthetized monkeys. Neurons consisted of both well-isolated single units and small multi-unit clusters.

Studying inter-area interactions, at a population-level, requires not only monitoring the activity of large neuronal populations across multiple brain areas, but also that these neuronal populations effectively interact. Primary visual cortex (V1) and visual area V2 are well suited for this. First, they are the two largest brain areas in the visual cortex, by cortical surface area[25,26] (Fig. 3.1). Second, they are densely, and reciprocally connected[26]. Furthermore, to maximize the probability that the recorded neuronal populations directly interact, the recordings analyzed here were performed in the output layers (2/3 and 4B) of V1 (88 to 159 neurons; mean: 112.8) and their primary downstream target, the middle layers of V2 (24 to 37 neurons; mean: 29.4)[27]. In addition, the recorded V1 and V2 populations had retinotopically-aligned receptive fields, again to maximize the probability of direct interactions[15].

3.1 Physiology, visual stimulation and recordings

Animal procedures and recording details have been described in previous work[15,29]. Briefly, animals (macaca fascicularis) were anesthetized with ketamine (10 mg/kg) and maintained on

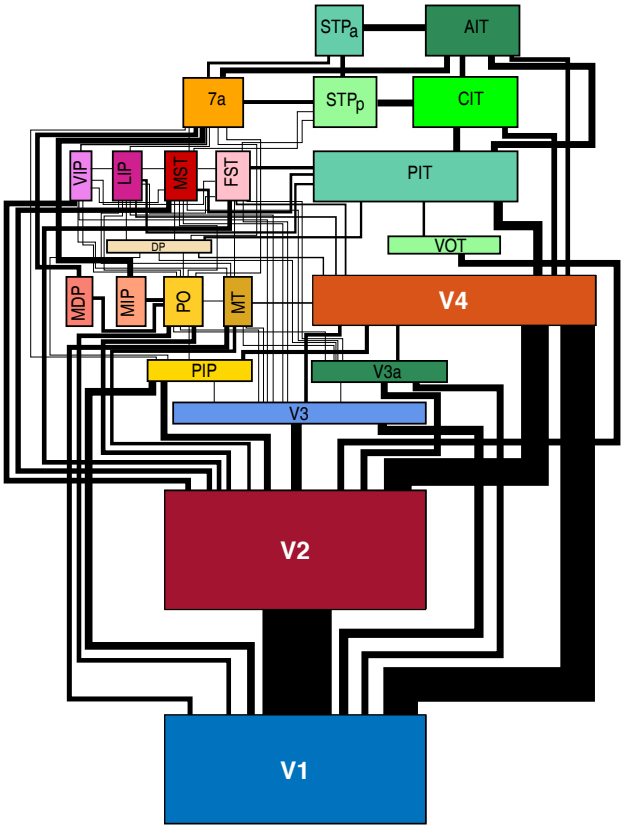

Figure 3.1: **A scaled representation of the cortical visual areas of the macaque.** (Adapted from Wallisch and Movshon (2008)[25], along with the reproduced caption.) Each colored rectangle represents a visual area, for the most part following the names and definitions used by Felleman and Van Essen (1991)[27]. The gray bands connecting the areas represent the connections between them. Areas right of the midline of the figure belong to the dorsal stream. Areas right of the midline of the figure belong to the ventral stream. Following Lennie (1998)[28], each area is drawn with a size proportional to its cortical surface area, and the lines connecting the areas each have a thickness proportional to the estimated number of fibers in the connection. The estimate is derived by assuming that each area has a number of output fibers proportional to its surface area and that these fibers are divided among the target areas in proportion to their surface areas. The connection strengths represented are therefore not derived from quantitative anatomy and furthermore represent only feedforward pathways, though most or all of the pathways shown are bidirectional. The original version of this figure was prepared in 1998 by John Maunsell.

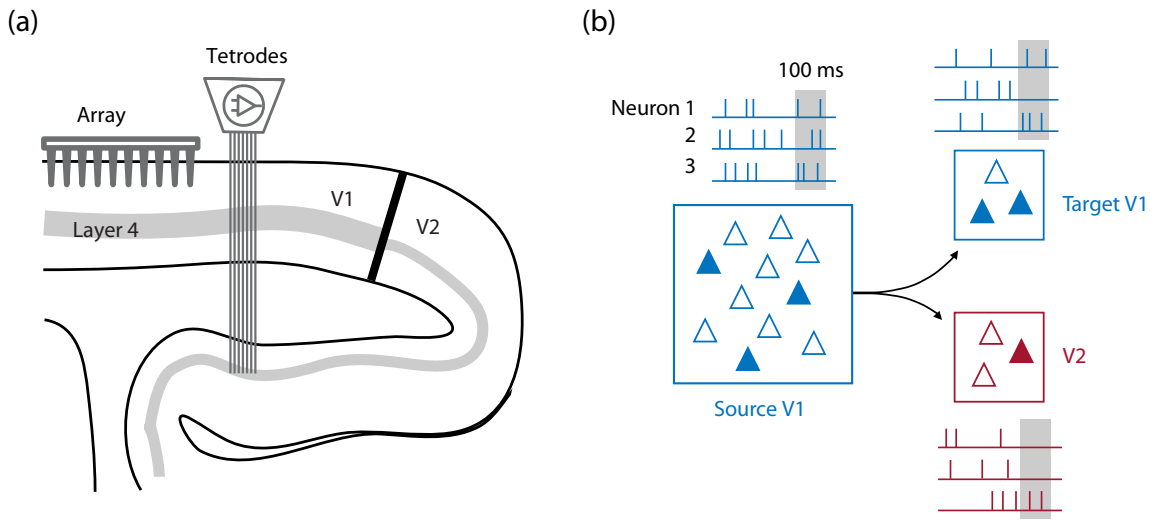

Figure 3.2: **V1 and V2 recordings. (a)** Schematic showing a sagittal section of occipital cortex and the arrangement of the recording apparatus. Recordings of V1 population activity were done using a 96-channel Utah array, while V2 population activity was recorded using a set of movable electrodes and tetrodes. **(b)** We related activity of the same V1 source population to a target V1 population and a V2 population. In this illustration, each triangle represents a neuron and the filled triangles indicate active neurons. Spike counts were taken in 100 ms bins.

isoflurane (1-2%) during surgery. Recordings were performed under sufentanil (typically 6-18 microgram/kg/hr) anesthesia. Vecuronium bromide (150 microgram/kg/hr) was used to prevent eye movements. All procedures were approved by the IACUC of the Albert Einstein College of Medicine.

The data analyzed here represent a subset of those reported in Zandvakili and Kohn, 2015[15], namely those that involved the largest and best retinotopically-aligned populations. V1 activity was recorded using a 96 channel Utah array (400 micron inter-electrode spacing, 1 mm length, inserted to a nominal depth of 600 microns; Blackrock, UT). V2 activity was recorded using a set of electrodes/tetrodes (interelectrode spacing 300 microns) whose depth could be controlled independently (Thomas Recording, Germany). These electrodes were lowered through V1, the underlying white matter, and then into V2. Within V2, we targeted neurons in the input layers. A schematic of the recording apparatus is shown in Fig. 3.2. Voltage snippets that exceeded a user-defined threshold were digitized and sorted offline. The sampled neurons had spatial receptive fields within $2-4^\circ$ of the fovea.

The measured responses were evoked by drifting sinusoidal gratings (1 cyc/°, drift rate of

$3 - 6.25$ Hz, $2.6 - 4.9°$ in diameter, full contrast) at 8 different orientations ($22.5°$ steps), on a calibrated CRT monitor placed 110 cm from the animal (1024 x 768 pixel resolution at 100 Hz refresh). Each stimulus was presented 400 times for 1.28 seconds. Each presentation was preceded by an interstimulus interval of 1.5 seconds.

Neuronal activity was recorded in a total of three animals. In two of the animals, activity was recorded in two different but nearby locations in V2, providing distinct middle-layer populations. We refer to each of these five recordings as a session. We treated responses to each of the 8 stimuli in each session separately, yielding a total of 40 data sets. All statistical tests treat the data sets as independent (with the exception of Fig. 4.1a, in which data are pooled across all stimuli, resulting in a single pairwise correlation value per pair per session). Repeating the same statistical tests across the five sessions (i.e., averaging the results across the 8 stimuli for each session) also returned significant results ($p < 0.05$) for all tests, with the exception of Fig. 4.1b, where we can no longer reject the null hypothesis that the average predictive performance is the same when predicting target V1 and V2.

3.2 Data preprocessing

Unless otherwise noted, we measured neuronal activity as spike counts in 100 ms bins, beginning 160 ms after stimulus onset and spanning a total of 1 second (10 bins per trial). To study how neuronal variability in the two areas is related, we subtracted the appropriate peri-stimulus time histogram from each single-trial response, and then analyzed the residuals for each orientation (henceforth referred to as data sets) separately. We confirmed that the temporal structure had little effect on our results by shuffling the data across trials while maintaining the temporal identity, as doing so reduced the predictive performance for the V2 population to 0. We obtained similar results after z-scoring both the source and target population responses, ruling out the possibility that our results were driven by a few high-firing neurons (see also Fig. 4.8). For all analyses, we excluded neurons which fired less than 0.5 spikes/s on average, across all trials.

We compared our analyses of V1-V2 interactions to the results of applying the same analyses to a held-out V1 population (V1-V1). The target population in the V1-V1 analyses was a held-out subset of the originally recorded population, which was matched in neuron count to the corre-

sponding V2 population (Fig. 3.2b). We also matched the firing rate distribution (mean-matched) to the V2 population separately for each stimulus condition (as in ref. 30). To do so, we binned the firing rate distribution of the V1 and V2 populations (for each neuron, the average firing rate was taken across time and trials for each data set), and determined the common firing rate distribution (i.e., for each firing rate interval, we took the minimum neuron count between the two populations). For each firing rate interval, we then randomly picked this minimum number of neurons from the corresponding bin in each population, without replacement. Because we had many more V1 than V2 neurons, the common distribution usually matched the V2 distribution and we selected an equal number of V1 neurons. The size of the matched populations ranged from 15 to 31 units across data sets (mean: 22.3). The V1 neurons that were not selected for the held-out population defined the source V1 population. V2 neurons that were not selected for the V2 mean-matched population were not used in the analysis. We repeated the mean-matching procedure 25 times, using different random, mean-matched subsets of neurons (and consequently producing a different source population). Results for each data set are based on averages across these repeats.

Chapter 4

Inter-areal interactions occur through a communication subspace

Interactions among brain areas are widely assumed to be essential to most brain functions. Yet we are only beginning to understand the interactions among neurons in distinct brain areas. Previous studies of inter-areal interactions have related the spiking activity of pairs of neurons in different areas[8–13], the spiking activity of a neuronal population in one area and a single neuron in another[14,15], the spiking activity of a neuron or group of neurons in one area and the local field potential (LFP) in another[16–20], the LFPs recorded in different areas[10,16,17,21,22], or the trial-averaged population activity in distinct areas[7]. These approaches have provided insight into how interaction strength changes with stimulus drive[8,10,22], attentional state[12,13,16,21], or task demands[11,17–20,7].

These previous approaches fall short, however, of elucidating how the spiking activity of neuronal populations – the signals thought to encode information in the brain – is related across areas on a trial-by-trial basis[31]. Pairwise correlations, by definition, ignore structure not evident in the interactions between two individual neurons. LFPs lump the activity of spiking populations into a single summary signal and thereby risk losing much of the richness of area-to-area interactions[32,33]. Trial-averaging allows one to study how mean signals (e.g., receptive field structure) are related, but not to understand how the moment-by moment changes in activity in one area relate to those in another area[17,34].

Here we leverage trial-to-trial co-fluctuations of V1 and V2 neuronal population responses,

recorded simultaneously in macaque monkeys, to understand the nature of population-level interactions between these areas. Within individual brain areas, these fluctuations involve multiple *dimensions* of activity shared among neurons[35,36,6,37–39], as identified using dimensionality reduction methods (see ref. 5 for a review). Each of these dimensions represents a characteristic way in which the activities of the recorded neurons covary (referred to as a population activity pattern). It is currently unknown whether all or only a subset of these dimensions are related *across* brain areas, and which dimensions are involved.

We find that, although interactions between V1 and V2 are similar in strength to those between subpopulations of neurons within V1, the structure of those interactions is strikingly distinct: V2 activity is related to a small subset of V1 population activity patterns. This selective routing of specific population activity patterns between V1 and V2 can be described by a low-dimensional communication subspace, which defines which activity patterns are effectively relayed between areas. We propose that the communication subspace can be a population-level mechanism by which activity can be selectively and dynamically routed between distinct populations.

4.1 Strength of inter-area interactions

We first characterized V1-V2 interactions by measuring the degree to which response fluctuations were shared between pairs of neurons (i.e., noise correlations), as in previous inter-areal studies[8,13], as well as between the two populations as a whole. To determine how V1-V2 interactions differ from interactions within V1, we divided the recorded V1 neurons into *source* and *target* populations (see Section 3.2). Briefly, for each data set, we matched the target V1 population to the neuron count and firing rate distribution of the measured V2 population. We then related the activity of the same source V1 population separately to the activity of the target V1 population (V1-V1 interaction) and that of the V2 population (V1-V2 interaction).

4.1.1 V1-V1 and V1-V2 interactions are similar in strength

The vast majority of V1-V2 pairs had correlations between 0 and 0.2 (Fig. 4.1a, red histogram; average correlation: 0.07 ± 0.06 S.D.). V1-V1 correlations were remarkably similar to those of V1-V2 pairs (Fig. 4.1a, blue histogram; average correlation: 0.07 ± 0.06 S.D.; two-sided Monte

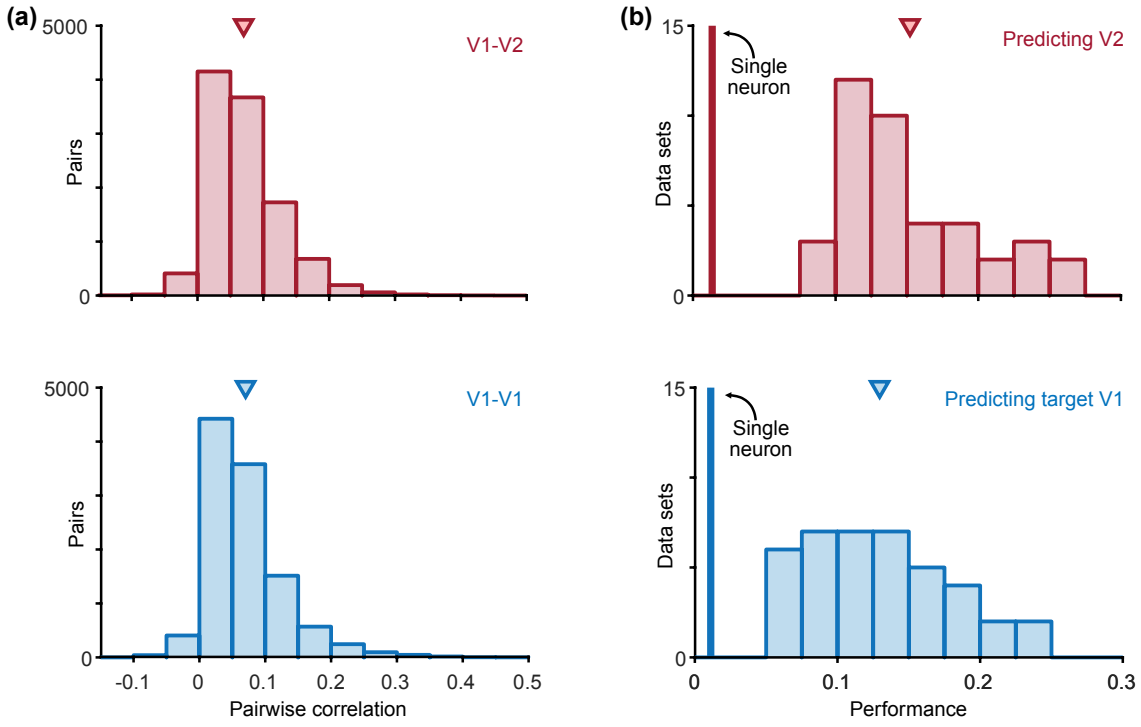

Figure 4.1: **V1-V1 and V1-V2 interactions are similar in strength. (a)** Pairwise correlation histograms for pairs of V1-V2 (red) and V1-V1 (blue) neurons. Triangles indicate average pairwise correlation. Total number of pairs in each histogram $n = 10,944$. **(b)** Prediction performance for V1-V2 (red) and V1-V1 (blue). Prediction was performed using a single V1 neuron at a time (solid lines) or using the entire source V1 population (histograms; triangles indicate mean). Prediction performance for each data set is defined as the average cross-validated r^2 across all selections of the target and source V1 populations.

Carlo paired permutation test, $p > 0.05$ for difference between V1-V1 and V1-V2). These weak correlations indicate that only a small fraction of a neuron's response variability can be explained by another individual neuron. Indeed, individual source V1 neurons could predict only $1.11 \pm 0.03\%$ and $1.35 \pm 0.03\%$ of the variability of the target V1 and V2 neurons, respectively (Fig. 4.1b, solid lines).

We next asked how well the variability of the target V1 and V2 neurons could be explained by the source V1 population using multivariate linear regression (see Section 4.4). On average, the source V1 population predicted $15.2 \pm 0.7\%$ of the V2 variability (Fig. 4.1b, red histogram), a substantial improvement over the performance afforded by individual V1 neurons. V1-V1 prediction quality (Fig. 4.1b, blue histogram; $12.9 \pm 0.8\%$; two-sided Monte Carlo paired permutation test, $p < 0.01$ for difference between V1-V1 and V1-V2) was similar to that of the V1-V2 prediction. V1-V2 and V1-V1 prediction performance were also similar to each other when activity was measured using a wide range of alternative time bin sizes (Section 4.1.1).

To assess whether the performance of the regression models is reasonable in absolute terms, we also implemented a basic model of population interactions using a linear feedforward network. Regression performance for these simulated data was similar to performance on the physiological data either when the target population had Poisson variability or when the observed source population was a subset of the full input population (Section 4.1.2).

In summary, both pairwise analysis and population-based regression models indicate that interactions between areas are similar in strength to those within a cortical area: fluctuations in the source V1 population can be used as effectively for predicting V2 activity as for predicting the fluctuations of other V1 neurons.

Similarity of predictive performance for target V1 and V2 is robust to bin size and number of trials

In all analyses, we used a 100 ms bin width for counting spikes and 400 trials per data set. Here we assess how predictive performance of target V1 and V2 varies with the bin width and number of trials. Using all 400 trials, we found that predictive performance increased with bin width (Fig. 4.2a). Crucially, for each bin width explored, the predictive performance for the target V1

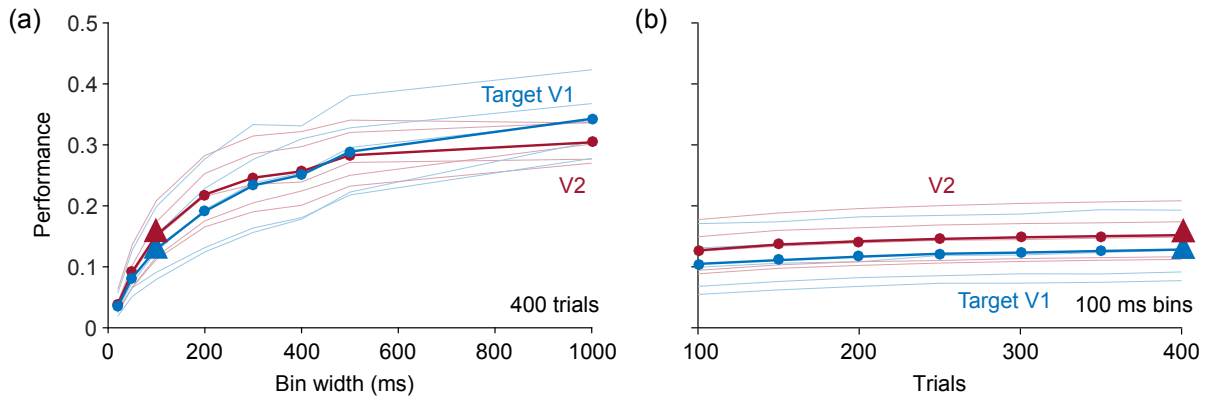

Figure 4.2: **Similarity of predictive performance for target V1 and V2 is robust to bin size and number of trials. (a)** The predictive performance for the target V1 and V2 populations remained similar across a wide range of bin widths. Solid line shows the average across all data sets. Faded lines show the average for each recording session. Triangle indicates the 100 ms bin width, used in all other analysis. **(b)** Predictive performance of both target populations depended modestly on the amount of data used. Triangle indicates the 400 trials used in all other analysis.

and V2 populations remained similar to each other. For a fixed bin width of 100 ms, we found that the predictive performance for the target V1 and V2 populations was largely independent of the amount of data used (Fig. 4.2b). Thus, the similarity in predictive performance for the target V1 and V2 populations was robust to the number of trials used to fit the regression model.

4.1.2 Estimating predictable variability in the target populations

Although we wish to focus on relative comparisons of predictive performance and interaction structure, we wondered whether the absolute performance of our regression models was reasonable. In particular, the predictive performance might be limited because of variability in the target population that is not related to the recorded source population. This could occur because of stochasticity in spike generation in the target population[40,41], or because the recorded source neurons are only a fraction of the relevant input neurons. In these cases, only part of the variability in the target population is predictable from the source population, regardless of the model used. Here we perform simulations to quantify the effect of stochasticity in spike generation in the target population, as well as of subsampling from a large source population. Our simulations indicate that the absolute predictive performance reported in Section 4.1 is consistent with either spike generation stochasticity in the target population or subsampling of the input population.

The predictive performance for the recorded activity was higher than expected if both of these effects contribute.

Contribution of Poisson-like variability

To assess the contribution of stochasticity, or noise, in the target population's spike generation process, we constructed surrogate target population responses (target V1 and V2) using linear combinations of the recorded source V1 activity and Poisson noise. We first applied a linear regression model (reduced-rank regression; see Chapter 4) to the recorded activity to identify a (number of source neurons by number of target neurons) matrix B, which relates the activity between the source and target populations. In these simulations, B represents the ground truth for generating the surrogate target population responses. While other choices of B are possible, we used a linear regression model to ensure the surrogate "residuals" qualitatively match the recorded residuals. The surrogate target residuals were obtained using $Y = XB$, where X is the recorded source residuals (a number of data points by number of source neurons matrix). We added the PSTH of the appropriate neuron and stimulus orientation to obtain a firing rate for each target neuron for each time bin and trial. We then drew spike counts from a Poisson distribution with the specified underlying rate. Note that we assumed Poisson spiking statistics and did not determine the level of noise from the data[42]. Using this approach, we generated surrogate target activity for both the V2 and target V1 populations for each of the recorded data sets, and then applied ridge regression to the residuals obtained from these surrogate data (to mimic the way in which we measured overall predictive performance in Section 4.1). In these simulations, the predictive performance can only be limited by the Poisson variability, as the interaction between populations is linear by design.

We found that the predictive performance of the linear regression model on the surrogate data was similar to the performance on the recorded activity (Fig. 4.3a; ratio of recorded activity performance to surrogate data performance: 0.7 ± 0.01 for V2 and 0.95 ± 0.01 for target V1). This outcome suggests that Poisson-like variability in the target population can largely account for the performance levels observed. That is, under the assumption that Poisson variability is "noise" and therefore not predictable, the linear model accounts for most of the predictable trial-to-trial

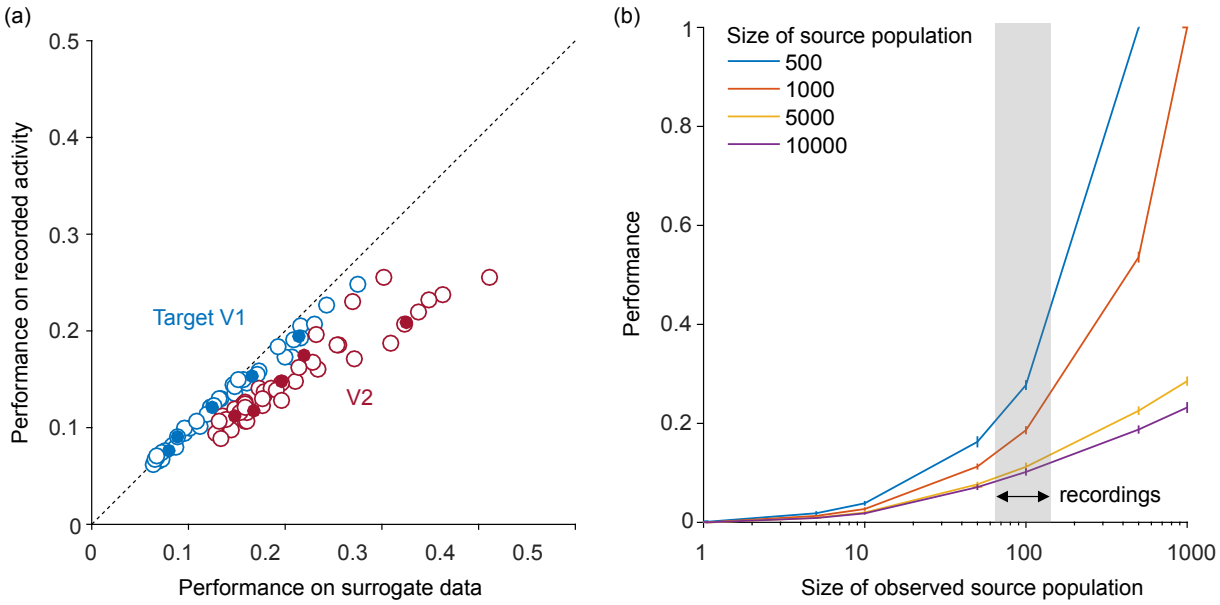

Figure 4.3: **Absolute predictive performance can be explained by Poisson variability in the target population or subsampling from a large source population. (a)** Absolute predictive performance on recorded activity matches performance performance on surrogate data with Poisson variability matched to the recorded target populations. Open circles show average predictive performance for each data set. Filled circles show average for each recording session **(b)** Observed predictive performance is also on par with the performance obtained under source population subsampling. Shaded area denotes recorded source population size range.

variability in the target population.

Contribution of source population sub-sampling

To mimic our recording from a fraction of the full neuronal source population, we conducted an additional simulation in which we subsampled from a large source population. Following the procedure of ref. 29, we generated the source activity (500 to 10,000 neurons) by drawing 4000 samples (matching the experimental data set size: 400 trials with 10 time bins each) from a multi-dimensional Gaussian distribution. The mean of the distribution was drawn from a uniform distribution between 0 and 100 for each neuron independently. To determine the covariance of the distribution, we first constructed a noise correlation matrix by assigning each neuron a "preferred orientation", which was drawn from a uniform distribution between $0°$ and $180°$. We then determined the correlation between each pair of neurons using the difference between preferred orientations. Correlations varied from 0 to 0.3; these values were chosen so that the eigenspectrum

of the resulting covariance matrix qualitatively matched that of the real data when matching the number of neurons. We then obtained the covariance matrix from the correlation matrix using the mean rates and assuming a Fano factor of 1. To generate activity in the target population (30 neurons), we defined the response of each target neuron using a weighted sum of the activity of the source population, with the weights chosen randomly from a standard Gaussian distribution. There is no added noise in this simulation. We then subsampled different proportions of the source population and used ridge regression to predict activity in the target population, as in Section 4.1. We repeated the process of subsampling from the large source population and drawing a set of weights 25 times. All presented results are the average across the 25 repeats, with the shading indicating the standard deviation.

As expected, predictive performance depended strongly on the size of the observed source population. When the source population is completely observed, predictive performance is perfect, since there is no noise added to the target population in this simulation. As the size of the observed population decreases, so does predictive performance, reaching a value close to 0 when a single source neuron is observed. Notably, when the size of the observed population roughly matched the size of the source V1 populations (64 to 135 neurons, indicated by the shaded area), the predictive performance was similar to that observed in the recorded data (average predictive performance across all recordings: 0.14; predictive performance on the surrogate data: 0.28 for 100 observed neurons out of a source population of 500, 0.19 for 100 out of 1000, 0.11 for 100 out of 5000, 0.10 for 100 out of 10000). These results show that recording from a subset of the V1 neurons projecting to V2 limits predictive performance, leaving a large proportion of V2 variability unexplainable by the recorded activity.

4.2 Structure of inter-area interactions

In Section 4.1 we found that within and across area interactions were similar in strength. Here, we asked whether the structure of these interactions is similar as well. Consider predicting the activity of a V2 neuron from a population of three V1 neurons using linear regression, as in Section 4.1:

$$V2^k = w_1 {V1_1}^k + w_2 {V1_2}^k + w_3 {V1_3}^k$$

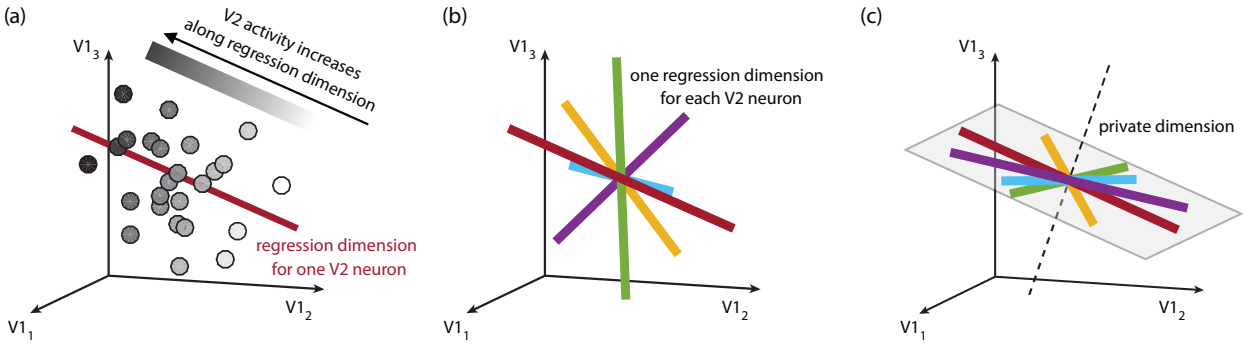

Figure 4.4: **Illustration of a low-dimensional interaction. (a)** Graphical depiction of linear regression between a population of V1 neurons and one V2 neuron. Each circle represents the activity recorded simultaneously in V1 (three neurons) and V2 (one neuron) during one timestep (100 ms). The position of the circle represents the V1 population activity and its shading represents the activity of the V2 neuron. The activity of the V2 neuron increases along the regression dimension (red line). **(b)** High-dimensional interaction. The regression dimensions for different V2 neurons (one regression dimension per V2 neuron) span the entire V1 population space. **(c)** Low-dimensional interaction. The regression dimensions for different V2 neurons span a subspace of the V1 population space. In this illustration, all regression dimensions lie in a 2-dimensional subspace (the grey plane). The basis vectors for this subspace are called predictive dimensions. Thus, two predictive dimensions are sufficient to capture the between area interaction. All dimensions that are not predictive of V2, and therefore lie outside of this subspace, are called private dimensions.

where $V2^k$ is the predicted activity of a V2 neuron on the kth trial, ${V1_1}^k$, ${V1_2}^k$ and ${V1_3}^k$ are the corresponding activities of the three V1 neurons on the same trial, and w_1, w_2 and w_3 are the regression weights. We can plot the activity of the V1 population on each trial as a point in a three-dimensional space, where each axis represents the activity of one of the V1 neurons (Fig. 4.4a). The weights can be represented as a *regression dimension*, which captures which aspects of the V1 population activity are predictive of the V2 neuron's activity. Specifically, the location of the V1 activity along the regression dimension is the predicted activity of the V2 neuron (Fig. 4.4a, shading).

In a basic multivariate regression model, each V2 neuron has its own regression dimension. These regression dimensions could, in principle, fully span the V1 activity space (Fig. 4.4b). If this were the case, any fluctuation in V1 population activity would be predictive of the fluctuations of one or more V2 neurons (i.e., changing the V1 population activity would change the location of the activity along at least one of the regression dimensions). Alternatively, if the regression dimensions span only a subspace of the V1 activity space (shown as a plane in Fig. 4.4c), certain V1 fluctuations (i.e., those orthogonal to the plane, Fig. 4.4c, dashed line) would not be predictive

of V2 fluctuations. We define *predictive dimensions* to be those which reside within the V1 subspace which is predictive of V2 fluctuations, and *private dimensions* as those which do not. The existence of private dimensions within the source population would allow for specific population activity fluctuations to be relayed downstream; any fluctuations along the private dimensions would be hidden from the target population.

4.2.1 V1-V2 interactions use only a small number of dimensions

To test whether our ability to predict V2 fluctuations involves only a subspace of V1 population activity, we used reduced-rank regression[43,44], a variant of linear regression in which the regression dimensions are constrained to lie in a low-dimensional subspace (see Section 4.4). If only a few dimensions of V1 activity are predictive of V2, then using a low-dimensional subspace should achieve the same prediction performance as the full regression model. For a representative data set (Fig. 4.5a), only two dimensions were needed to achieve a prediction performance that was indistinguishable from the full regression model (triangle). In contrast, when we applied the same analysis to the target V1 population, six dimensions of the source V1 population activity were needed to reach the performance of the full model (Fig. 4.5b). Across all data sets, consistently fewer dimensions were needed to predict fluctuations in the V2 population (2.2 ± 0.1) compared to the target V1 population (3.5 ± 0.1; one-sided Monte Carlo paired permutation test, $p < 10^{-8}$; Fig. 4.5c). We obtained similar results when activity was measured using a wide range of time bin widths (Fig. 4.7).

These results indicate that the V1 fluctuations that are predictive of V2 are confined to a small number of V1 dimensions. Notably, the number of dimensions needed to account for interactions between areas was smaller than the number of dimensions involved in interactions within an area.

4.2.2 Low-dimensional V1-V2 interaction is not due to low-dimensional V2 activity

A possible explanation for the lower-dimensional interaction between V1-V2 compared to within V1 is that the V2 population activity is itself less complex, or lower dimensional, than the target V1 activity. For example, if the measured V2 population consisted of neurons with identical responses, then predicting those responses would involve the same weighting of V1 activity (i.e.,

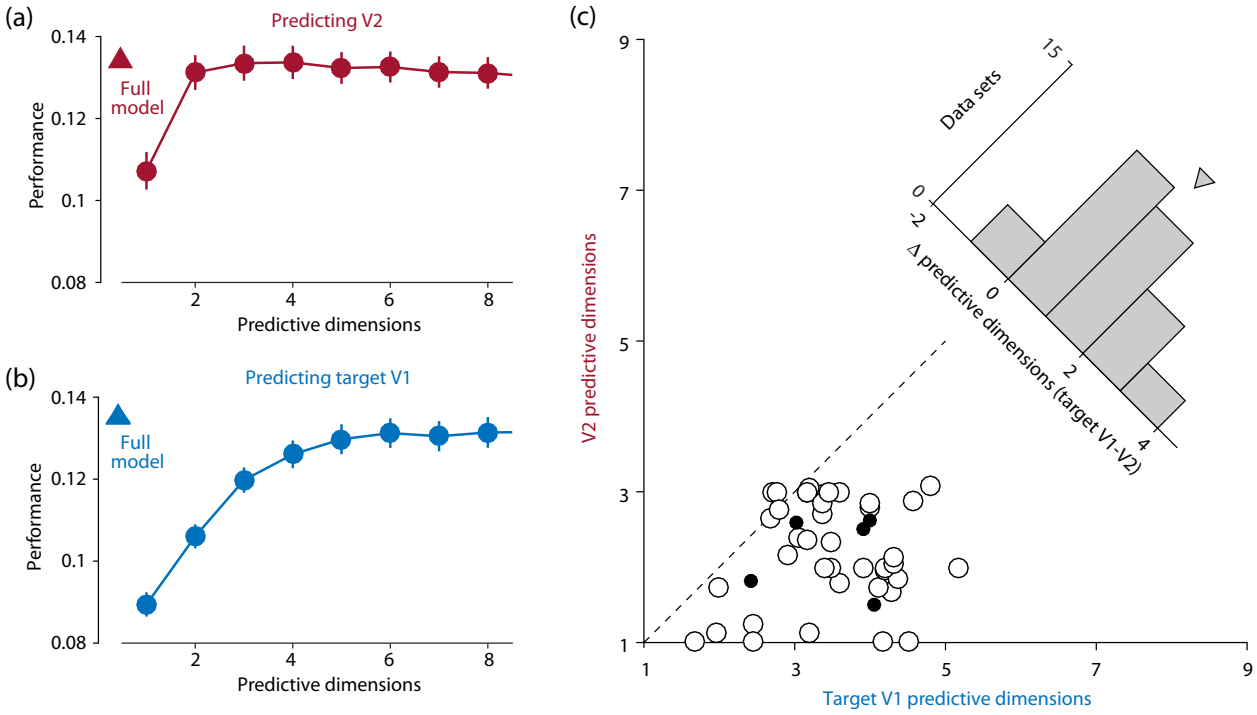

Figure 4.5: **V1-V2 interactions use only a small number of dimensions. (a)** Predicting V2 activity. The number of predictive dimensions (red circles; reduced-rank regression) needed to achieve full predictive performance (red triangle; ridge regression) is small (in this case, 2 dimensions). Across all data sets, reduced-rank regression achieved nearly the same performance as the full regression model (0.150 ± 0.007 for reduced-rank regression versus 0.152 ± 0.007 for the full regression model). The predictive performance slightly decreases with the number of predictive dimensions due to cross-validation. Error bars indicate S.E.M. across cross-validation folds. **(b)** Predicting target V1 activity. The number of predictive dimensions (blue circles; reduced-rank regression) needed to achieve full predictive performance (blue triangle; ridge regression) is large (in this case, 6 dimensions). Across all data sets, predictive performance was again similar for reduced-rank regression (0.123 ± 0.008) and the full regression model (0.129 ± 0.008). **(c)** The optimal number of predictive dimensions is smaller for predicting V2 than target V1. Each open circle corresponds to one data set. Filled circles indicate averages across data sets for each of the 5 sessions (see Section 4.4). Inset shows the difference between the optimal number of predictive dimensions needed when predicting the target V1 and V2 populations (target V1 minus V2).

one predictive dimension). More generally, the number of predictive dimensions will depend in part on the dimensionality of the target population activity. All else being equal, the lower the dimensionality of the target population activity, the smaller the number of predictive dimensions will be.

We used factor analysis to test whether the V2 population activity was lower-dimensional than the target V1 population activity. Factor analysis identifies factors (or dimensions) which capture shared activity fluctuations among neurons[31,35,36,6,37,39,45,30]. This analysis revealed that the dimensionality of the V2 activity was higher than that of the target V1 activity (Fig. 4.6a; 5.0 ± 0.2 for V2; 3.7 ± 0.1 for target V1; mean $\pm$ S.E.M.; one-sided Monte Carlo paired permutation test, $p < 10^{-8}$). Thus, the smaller number of V2 predictive dimensions cannot be explained by the V2 population response being less complex than the target V1 population response.

To assess how the complexity of the target population influenced the dimensionality of the interactions, we compared the number of predictive dimensions to the dimensionality of the target population activity. For V1-V1 interactions, the number of predictive dimensions closely matched the dimensionality of the target population activity in each data set (Fig. 4.6b, blue points). Although these two estimates of dimensionality are based on different analyses, their similarity suggests that the number of V1 predictive dimensions is as large as possible, given the complexity of the target population response. In contrast, for V1-V2 interactions, the number of predictive dimensions was consistently lower than the dimensionality of the target population (Fig. 4.6b, red points).

We conclude that the difference in the number of V1 and V2 predictive dimensions cannot be explained by the complexity of the respective target population responses, but rather reflects the nature of the interaction between these areas. Whereas the V1-V1 interaction uses as many predictive dimensions as possible, the V1-V2 interaction is more selective and is confined to a small subspace of source V1 population activity, which we term a *communication subspace*.

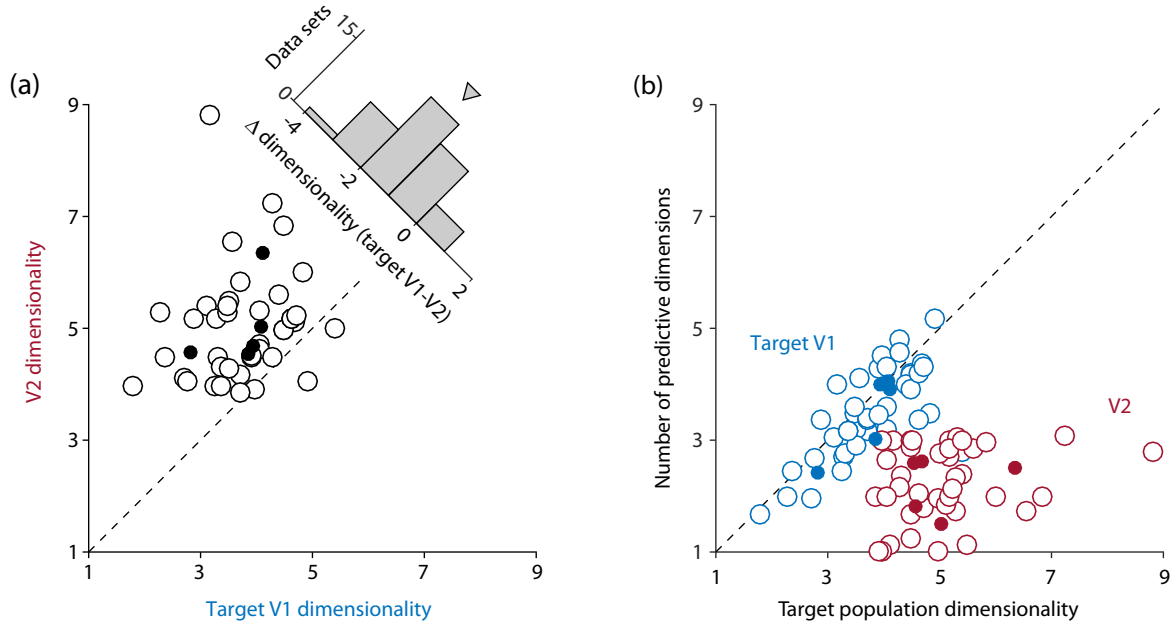

Figure 4.6: **Low-dimensional V1-V2 interaction is not due to low-dimensional V2 activity. (a)** Population activity is more complex in V2 than in target V1. Each open circle corresponds to one data set. Filled circles indicate averages across datasets for each of the 5 sessions. Inset shows the difference between the dimensionality (target V1 minus V2) of the population activity in target V1 and V2. **(b)** V1 and V2 interact through a communication subspace. The number of predictive dimensions identified for the V1-V2 interaction was always smaller than the dimensionality of the V2 population activity (red circles). The number of predictive dimensions required when predicting target V1 population activity was similar to the dimensionality of the target V1 population (blue circles). Each open circle corresponds to one data set. Filled circles indicate averages across data sets for each of the 5 sessions.

The V1-V2 communication subspace was evident across a wide range of time bin widths, but was difficult to detect with limited numbers of trials

To test whether our finding of a V1-V2 communication subspace depended on the choice of time bin width, we repeated the analyses presented in the previous sections for different bin widths and 400 trials per data set (Fig. 4.7, left column). We found that the V1-V2 communication subspace was present over a wide range of time bin sizes (Fig. 4.7a). The number of V2 predictive dimensions was smaller than the dimensionality of the V2 population responses for all bin widths. In contrast, for the target V1 populations, the number of predictive dimensions roughly matched the dimensionality of the target population. The difference in the number of estimated predictive dimensions for the target V1 and V2 populations was evident for all bin widths, but less evident for larger bins (Fig. 4.7b). Note that increasing the bin width reduces the amount of data available for model fitting; as we show below, it is more difficult to identify predictive dimensions with small amounts of data. Thus, results for larger bin widths should not be interpreted as indicating a timescale-dependence of the communication subspace. The estimated target population dimensionalities are consistent across a wide range of bin widths: V2 population activity was always higher dimensional than the target V1 activity (Fig. 4.7c).

To determine how the estimated dimensionalities depend on the amount of recorded data, we repeated the analyses after subsampling a fraction of the recorded trials (using a fixed bin width of 100 ms). For different numbers of trials, the number of V2 predictive dimensions was consistently smaller than the dimensionalities of the V2 populations (Fig. 4.7d). For the target V1 population, the number of estimated predictive dimensions roughly matched the dimensionality of the target population. Thus, the findings for the full data set were also evident when subsets of data were used in the analysis. However, we found that using less data for model fitting had a significant impact on the number of estimated predictive dimensions for both target populations (Fig. 4.7e). When we used 100 trials per data set we could only identify, on average, a single predictive dimension for each the target V1 and V2 populations. As we increased the amount of data, the difference between the number of estimated predictive dimensions for the target V1 and V2 populations increased (as do the number of estimated predictive dimensions for each target population). This trend suggests that had we been able to record more trials, we would have

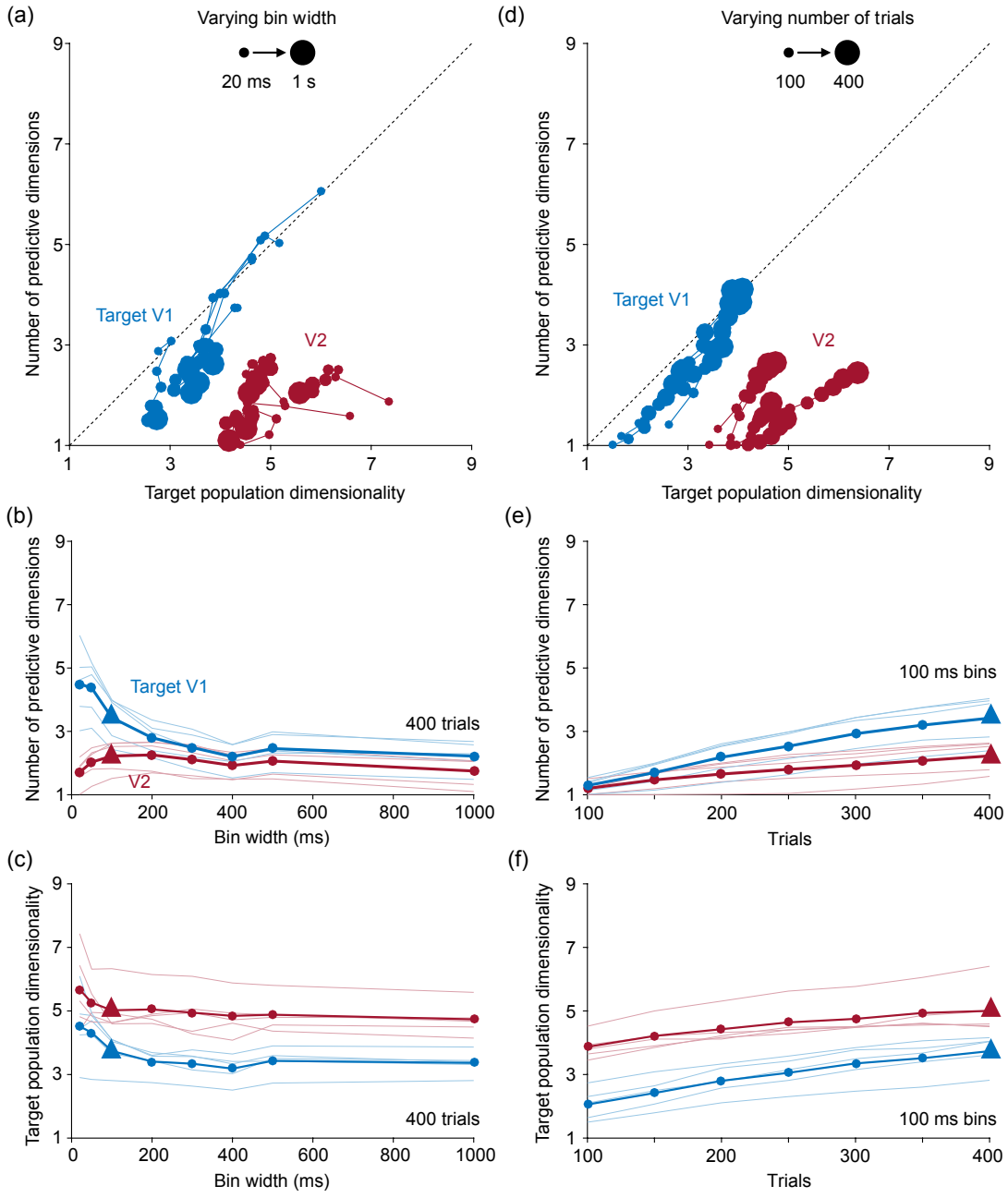

Figure 4.7: **The V1-V2 communication subspace was evident across a wide range of time bin widths, but was difficult to detect with limited numbers of trials. (a)** V1-V2 communication subspace remained visible for a wide range of bin widths. Each dot corresponds to the average dimensionality for one session with the dot size indicating the bin size. **(b)** Regardless of the bin width, less predictive dimensions were required to predict V2. Solid lines show the average across all data sets; faded lines show the average for each recording session. **(c)** Target population dimensionality was consistent for multiple choices of bin width. **(d)** Communication subspace becomes more evident as more trials are used for model fitting. Dot size indicates number of trials used for fitting. **(e)** Small number of trials made it difficult to detect differences between number of predictive dimensions for intra- and inter-area interactions. **(f)** Target population dimensionality was less affected by reducing the number of trials.

detected an even larger difference between the number of V1 and V2 predictive dimensions. The dimensionalities of the target populations increased with the number of trials, suggesting that the true dimensionality of the target populations in likely higher than what we could identify with 400 trials (Fig. 4.7f; see also ref. 39).

Predictive dimensions do not arise from capturing the activity of only a few target neurons

In principle, the small number of V2 predictive dimensions identified might be a consequence of predicting the activity of only a small number of V2 neurons. That is, rather than the small number of predictive dimensions predicting activity across the entire target population, those dimensions might only predict the activity of a handful of neurons, each using an independent predictive dimension. The regression model captured the activity fluctuations of some V2 neurons better than others (range of predictive performance from -0.09 to 0.52 across all the recorded V2 neurons), giving credence to this possibility. To directly test whether the small number of V2 predictive dimensions was due to explaining responses of only a few V2 neurons, we excluded (for each data set) the three target neurons for which the predictive performance was highest and then re-ran our analysis. We chose three target neurons because the highest number of V2 predictive dimensions, found across all data sets, was three (cf. Fig. 4.5c). If the identified predictive dimensions corresponded to these few target neurons for which the predictive performance was highest, then excluding these neurons should alter the number of identified predictive dimensions. We found that this was not the case (Fig. 4.8). The average number of V2 predictive dimensions was nearly unchanged after removing the three target neurons with highest predictability (Fig. 4.8, red circles; 2.20 ± 0.11 in the original analysis vs. 2.20 ± 0.09 after; two-sided Monte Carlo paired permutation test, $p > 0.05$; cf. Fig. 4.6b). The dimensionality of the V2 population also remained largely unchanged (5.05 ± 0.16 in the original analysis vs. 5.21 ± 0.18 after; two-sided Monte Carlo paired permutation test, $p > 0.05$). The same procedure was also applied to the Target V1 population (Fig. 4.8, blue circles), with similar effect. We conclude that the identified predictive dimensions are predictive of the population of recorded V2 neurons as a whole, rather than only a small handful of V2 neurons.

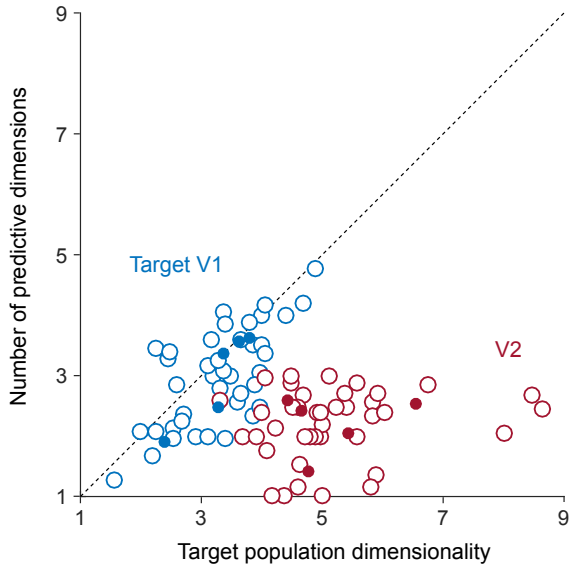

Figure 4.8: **Predictive dimensions do not arise from capturing the activity of only a few target neurons.** Each open circle corresponds to one data set. Filled circles indicate averages across data sets for each of the 5 sessions.

Difference in the number of V1 and V2 predictive dimensions is not due to differences in receptive field alignment

The V2 populations analyzed had spatial receptive fields that were closely aligned with the recorded source V1 populations. Because the target V1 population were selected from the same recording array as the source V1 population, these two populations also had well-aligned receptive fields. Nevertheless, one possible concern is that the difference in the number of V1 and V2 predictive dimensions was due to a subtle mismatch in receptive field alignment for V1-V1 compared to V1-V2.

On average, the V1-V1 receptive field distance (center-to-center population spatial receptive field distance) was smaller than for V1-V2 (Fig. 4.9a, top; average V1-V1 receptive field difference: $0.17 \pm 0.02°$; average V1-V2 receptive field difference: $0.58 \pm 0.06°$; one-sided Monte Carlo permutation test, $p < 10^{-3}$).

To test whether the differences we found between the V1-V1 and V1-V2 interaction structures (Fig. 4.9a, bottom, replicating Fig. 4.6b) could be explained by this difference in receptive field alignment, we repeated the analyses in the previous sections after matching the V1-V1 and V1-V2

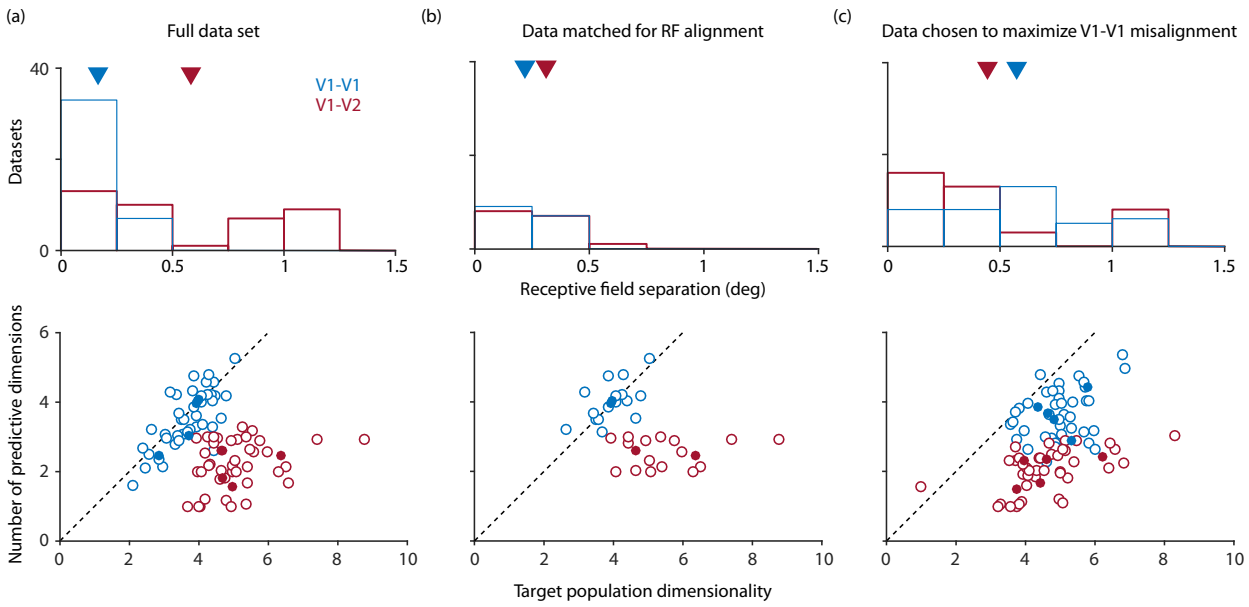

Figure 4.9: **Difference in the number of V1 and V2 predictive dimensions is not due to differences in receptive field alignment. (a)** (top) V1-V1 receptive field distance was on average smaller for V1-V1 than for V1-V2. (bottom) Replication of Fig. 4.6b. **(b)** Same as (a) after restricting the analyses to sessions for which V1-V2 alignment was as high as V1-V1 alignment. **(c)** Same as (a) after re-selecting the target V1 populations so as to minimize their alignment with the source V1 populations.

alignment. We did this matching in two ways: (1) analyzing only the sessions for which V1-V2 alignment was as high as V1-V1 alignment; (2) selecting the target V1 populations so as to minimize their alignment with the source V1 populations.

When we restricted our analysis to sessions for which V1-V2 alignment was as high as V1-V1 alignment (Fig. 4.9b, top; average V1-V1 receptive field difference: $0.21 \pm 0.04°$; average V1-V2 receptive field difference: $0.31 \pm 0.04°$; two-sided Monte Carlo permutation test, $p > 0.05$), we found V1-V1 and V1-V2 interactions were still strikingly distinct (Fig. 4.9b, bottom; compare with Fig. 4.9a, bottom). Specifically, fewer predictive dimensions were necessary to predict the V2 activity than V1 target population activity (2.6 ± 0.1 for V1-V2 vs 4.0 ± 0.2 for V1-V1; one-sided Monte Carlo paired permutation test, $p < 10^{-3}$), even though the V2 activity was higher dimensional than that of the target V1 population (5.5 ± 0.3 for V2 vs 4.0 ± 0.2 for target V1; one-sided Monte Carlo paired permutation test, $p < 10^{-3}$). Likewise, V2 predictive dimensions were also not aligned with the source V1 dominant dimensions in these sessions (not shown).

When we instead selected the target V1 population to minimize alignment with the source V1 populations (Fig. 4.9c, top; average V1-V1 receptive field difference: $0.57 \pm 0.05°$; average V1-V2 receptive field difference: $0.44 \pm 0.07°$; two-sided Monte Carlo permutation test, $p > 0.05$), we still found that fewer predictive dimensions were necessary to predict V2 than target V1 activity (Fig. 4.9c, bottom; 2.1 ± 0.1 for V1-V2 vs 3.7 ± 0.1 for V1-V1; one-sided Monte Carlo paired permutation test, $p < 10^{-3}$). In these data, both target populations had similar dimensionality (4.6 ± 0.2 for V2 vs 4.9 ± 0.1 for target V1; two-sided Monte Carlo paired permutation test, $p > 0.05$). This similarity arises because the firing rate distributions for the target V1 and V2 populations were not mean-matched, as in the the previous analyses; the requirement to select a specific subset of V1 neurons with offset receptive fields (i.e., those in one corner of the array) precluded rate-matching. It is still the case, however, that the smaller number of V2 than V1 predictive dimensions cannot be attributed to differences in target population dimensionality. In addition, it was still the case that V2 predictive dimensions were not aligned with the source V1 dominant dimensions (not shown).

Together these results show that the differences between V1-V1 and V1-V2 interactions cannot be explained by differences in retinotopic alignment with the two target populations.

4.2.3 Using linear methods to study nonlinear computations

The communication subspace as a local linear approximation to a non-linear computation

In most of the results presented thus far, we identified predictive dimensions for each stimulus separately. An important motivation for this approach is that if the interactions between areas are nonlinear[46], we can best apply linear methods to small perturbations in the population signals around a fixed operating point, which is precisely what the analysis of trial-to-trial fluctuations to a fixed stimulus accomplishes. However, a concern for this approach is that the predictive dimensions identified may change across stimulus conditions. Here, we investigate how the predictive dimensions identified from responses to different oriented gratings are related.

To assess the similarity of the communication subspaces identified for different gratings, we determined the predictive dimensions for each stimulus and then used these to predict responses evoked by the other stimuli. This was done by first projecting the source activity onto the identified communication subspace and then fitting a linear regression model between these projections

and the target activity. Predictive performance was then normalized by the performance achieved by a model fit directly to the responses to that stimulus (i.e., without first projecting the responses onto the subspace defined for responses to another stimulus). Performance was measured using 10-fold cross-validation (i.e., for each fold, the model was fit to a training set pertaining to one stimulus and then used for prediction in test sets of all stimuli). If the communication subspaces were entirely distinct for different stimuli, performance should plummet when the regression model is fit using the subspace derived from responses to other gratings. Alternatively, if the communication subspaces are similar for different gratings, performance should be similar regardless of which responses are used to define the subspace.

We found that the communication subspaces are similar for different gratings Fig. 4.10: the performance declined smoothly as the responses were evoked by stimuli of progressively different orientations, but the drop in performance was modest. When the predictive subspace identified for one orientation was used to predict responses to an orthogonal grating – the most challenging scenario – performance was roughly 75% of that achieved for the subspace identified for the orthogonal grating responses. Performance is averaged across all sessions.

For comparison, we applied the same analyses to data synthesized as in Section 4.1.2. Briefly, we first defined a fixed linear mapping, using the residual responses to activity pooled across all conditions. We then generated target population activity for each stimulus by passing the corresponding source population activity thought this fixed mapping, and added Poisson noise to each sample with mean given by the corresponding PSTH time bin. For these synthesized responses, we found that performance was uniformally high, when we identified the predictive dimensions using responses to one stimulus orientation and applied them to responses evoked by another (Fig. 4.10b). This analysis indicates that (1) the modest performance decrement in the physiological data (Fig. 4.10a) cannot be attributed to differences in the source population responses across orientations; (2) if the mapping between areas were strictly linear and fully identified by our analysis, we should not observe any decrease in performance. These results thus support the suggestion that the mapping between V1 and V2 is not strictly linear.

In summary, the communication subspaces identified for distinct stimuli were not identical. Furthermore, these differences could not be explain by changes in the statistics of the source and target populations. However, these subspaces changed smoothly and only moderately across all

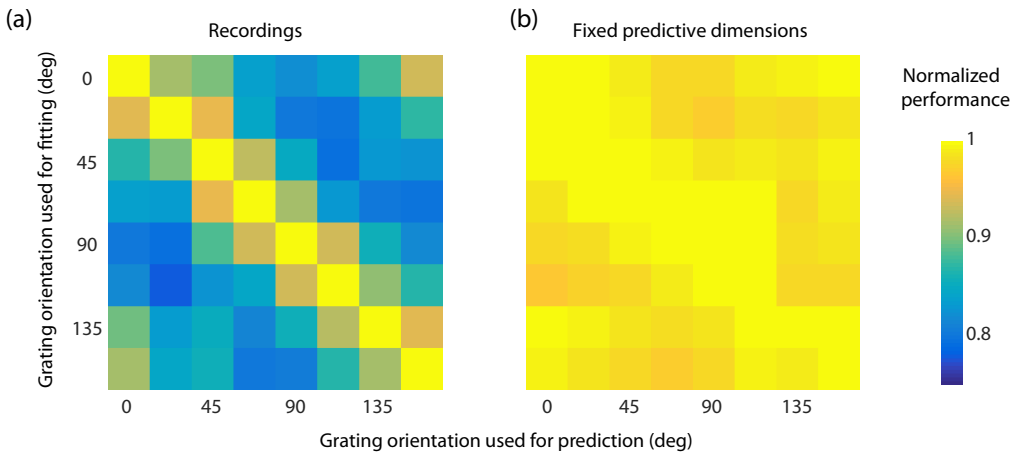

Figure 4.10: **Predictive dimensions as local linear approximations of a globally nonlinear mapping. (a)** Communication subspace changes smoothy and modestly across stimulus conditions. Each row corresponds to a different communication subspace and each column to the application of those subspaces to a different stimulus condition. The diagonal elements thus indicate the normalized performance of identifying and applying the communication subspace to responses evoked by the same stimuli. The off-diagonal elements have normalized predictive performance values less than 1, indicating that the communication subspaces are not identical across the 8 stimulus conditions. **(b)** The change in the identified predictive dimensions cannot be attributed to differences in the source population responses across orientations. Same conventions as in (a).

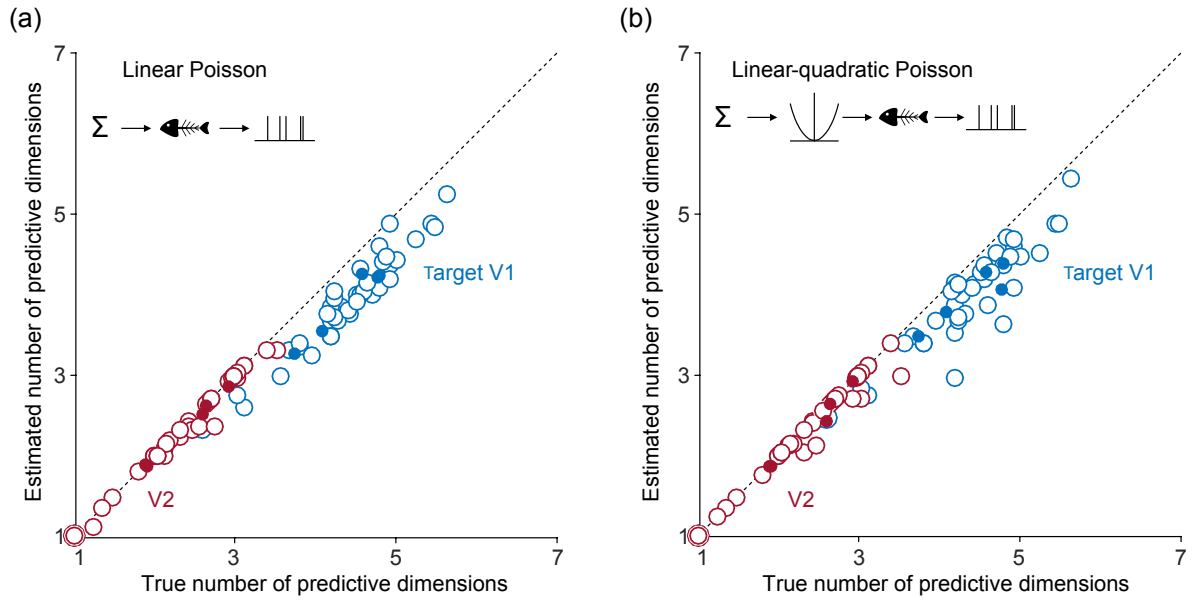

Figure 4.11: **Reduced-rank regression recovers the correct number of predictive dimensions even when the mapping between populations is nonlinear. (a)** Linear Poisson model. Open circles show average predictive performance for each data set Filled circles show average for each recording session. **(b)** Linear-quadratic Poisson model. Same conventions as in (a).

stimuli.

Reduced-rank regression recovers the correct number of predictive dimensions even when the mapping between populations is nonlinear

Reduced-rank regression is a linear dimensionality reduction method which assumes additive Gaussian noise. We were concerned that our estimates of the dimensionality of V1-V1 and V1-V2 interactions might be inaccurate, either because interactions between populations are likely nonlinear or because neuronal variability is Poisson-like (not additive Gaussian). To test these possibilities, we applied reduced-rank regression to surrogate data sets in which target population variability was determined either by (1) a linear mapping from the source to target population followed by Poisson noise (linear-Poisson, or LP, model); or (2) by a linear-nonlinear mapping followed by Poisson noise (linear-nonlinear-Poisson, or LNP, model)[47].

For the LP model, surrogate data were obtained using a similar procedure as in Section 4.1.2. Briefly, we defined the (number of source neurons by number of target neurons) mapping matrix B by applying reduced-rank regression, with the optimal number of dimensions, to the recorded

source and target populations (without subtracting the PSTHs). We then generated the surrogate target rates $Y = XB$, where X is the recorded source activity (a number of datapoints by number of source neurons matrix). We obtained the target activity by generating Poisson spike counts based on the rates Y. To estimate the underlying number of predictive dimensions in the surrogate data, we applied reduced-rank regression to the residual activity in the source and target populations. We found that the estimated number of predictive dimensions matched the underlying model closely, especially for interactions involving a small number of predictive dimensions .

The LNP model was identical to the LP models, but the linear combination of the spike counts was passed through a quadratic nonlinearity before generating Poisson spike counts. As with the LP model, we found the number of predictive dimensions estimated with reduced-rank regression closely matched the dimensionality of the mapping matrix (Fig. 4.11).

Nonlinear interactions are confined to the communication subspace

As observed above, RRR was able to identify the correct number of predictive dimensions, even when the underlying interaction was not linear. In general, and provided enough data, RRR should be able to identify the relevant subspaces for linear-nonlinear interactions, so long as the nonlinear interactions have some linear component (i.e., they do not induce 0 correlation between the source and target activity). Note that this does not mean that the predictions of the RRR model will be accurate (as they are still a linear function of the source activity), but rather that RRR should still identify the dimensions of source activity which influence downstream activity.

One might still worry, however, that for a fixed amount of data RRR might struggle more to identify dimensions for which the interaction with the target areas is highly nonlinear. In particular, if the V1-V2 interactions deviate more from linearity than the V1-V1 interaction, then this could provide an explanation for the difference in the number of predictive dimensions identified.

To test this, we proposed and employed a new dimensionality reduction method, termed distance covariance analysis (DCA)[46]. Briefly, DCA extracts two subspaces of activity, in the source and target populations, for which the corresponding projections are dependent. This means that any type of dependence, be it linear or even (purely) nonlinear, will be picked up by DCA. Conversely, all source and target activity outside of the subspaces extracted by DCA is guaranteed to

be independent.

We started by confirming that the V1-V2 interaction does indeed have nonlinear components (as hinted by the analysis above). To do this, we fit a linear model to the V1-V2 interaction (ridge regression), and then subtracted the linear predictions from the target activity. We confirmed that doing so made it such that we could no longer identify predictive dimensions for the V1-V2 interactions. Crucially, applying DCA to this same activity (after removing the linear components) still revealed a subspace of interaction between the two areas.

We then wondered whether these nonlinear interactions took place along the same dimensions previously identified by RRR, or if they happened outside of the communication subspace. If the latter turned out to be true, it would indicate that perhaps the V1-V2 interactions were not lower-dimensional after all, just more nonlinear. However, we found that this was not the case, as the subspaces identified by DCA were contained within the communication subspace identified by RRR.

4.3 Discussion

Nearly all previous studies of interactions between brain areas have used pairwise spike-spike or spike-LFP analyses. Here we investigated the structure of interactions between areas at the level of neuronal population spiking responses. We found a striking difference in the nature of V1-V1 and V1-V2 interactions: V2 activity was related to a small subset of population activity patterns in the source V1 population. In contrast, more activity patterns in the source V1 population were relevant for predicting the activity of other V1 neurons. Interactions between areas are thus defined by a communication subspace: V1 activity that lies within the communication subspace is communicated with V2, whereas V1 activity that lies outside this subspace is not.

Our analyses were designed to ensure a fair comparison of V1-V1 and V1-V2 interactions. First, we used the same V1 population to predict target V1 and V2 responses, ruling out any potential differences in the source population. Second, we matched the sizes of the target V1 and V2 populations as well as their firing rate distributions, ruling out differences in these basic target population properties. Third, we were able to predict fluctuations in the target V1 and V2 populations equally well (Section 4.1), so our results cannot be attributed to differences in the

strength of V1-V1 and V1-V2 interactions. Finally, the spatial receptive fields of both the target V1 and V2 populations overlapped those of the source population, and subtle differences in the retinotopic alignment of the respective populations could not account for our findings (Section 4.2.2).

It is important to note that the estimated number of predictive and dominant dimensions likely depends on the number of recorded neurons and trials[39]. Accordingly, our results do not define the dimensionality of V1-V2 interactions in absolute terms; rather, they indicate that those interactions are low-dimensional relative to V1-V1 interactions. In separate analyses, we found that as we considered more neurons or trials, the difference between V1-V1 and V1-V2 interactions became more prominent (see Section 4.2.2). Thus, the true difference between V1-V1 and V1-V2 interactions is likely larger than we were able to identify.

Dimensionality reduction analyses have provided important insights into neuronal population activity structure and its function (see ref. 5 for a review). However, such analyses have been applied almost uniquely to population responses recorded in a single brain area, rather than to the study of interactions between areas, as we have done. In a recent study, Kaufman and colleagues investigated the relationship between activity in motor cortex and muscles[7,48]. They found that preparatory motor activity avoids the potent (i.e., predictive) dimensions which relate cortical activity to muscles during movement, akin to our finding of private dimensions for V1-V2 interactions. However, our studies differ in that their analysis focused on trial-averaged responses (PSTHs) rather than relating trial-to-trial fluctuations in directly connected neuronal populations (i.e., those with functional alignment and in specific cortical laminae).

V2 is likely to perform non-linear operations on inputs received from V1[49,50]. Our approach to understanding V1-V2 interactions was to study local fluctuations around different set points (i.e., the trial-to-trial variability around the mean responses to a particular grating; see Section 4.2.3) – which function effectively as local linear perturbations in the non-linear transformation between V1 and V2. Our use of trial-to-trial fluctuations is consistent with most previous studies of inter-areal interactions[17,34,51,21,52], although these have used entirely distinct analyses such as spike-field coherence. To be sure that our estimates of V1-V2 interactions were not distorted by simple downstream non-linearities, we implemented several feedforward network models with standard non-linearities (e.g., squaring). In all cases, we found our analyses recovered interaction

dimensionality that closely matched the dimensionality of the linear weights (Section 4.2.3).

We propose that the communication subspace is an advantageous design principle of inter-area communication. The ability of a source area to communicate only certain activity patterns while keeping others "private" could be a means for the selective routing of signals between areas. To understand the computational benefit of structuring inter-areal communication in this way, we implemented a simulation which captures the common scenario of a source area projecting to two downstream areas, areas A and B (Fig. 4.12). If each downstream area reads from the source area using a different communication subspace, there will be dimensions of the source population activity that are relayed to area A, but not to area B (Fig. 4.12a), and vice versa (Fig. 4.12b). Crucially, if the interaction between these areas does not involve communication subspaces, then all fluctuations in the source population will be relayed to both downstream areas (Fig. 4.12c). The communication subspace is consequently a population-level mechanism whereby activity can be selectively routed between brain areas.

Allowing interactions between areas to be modulated by the alignment of population activity with a relevant communication subspace has several advantages over a well-known alternative: defining interaction strength by the phase-alignment of spikes to ongoing oscillations[24] (termed "communication through coherence", CTC). First, the communication subspace hypothesis does not require coordinated oscillations between the source and target areas, which can be difficult to achieve in practice[53]. Instead, the implementation of a communication subspace requires only that the target area takes a particular type of weighted combination of its inputs, namely a linear readout that is low-dimensional. This can be implemented in a neural circuit by setting the synaptic weights onto each downstream neuron from linear combinations of a set of basis "weights" (or predictive dimensions; see Fig. 4.4 and Fig. 4.12)[54,55]. Second, different target areas (or subpopulations within the same target area) can have different communication subspaces in the same source area (Fig. 4.12). CTC can also route distinct signals to downstream targets by using different oscillations within the source area, each of which is coherent with a different downstream target. However, the number of oscillations that can be distinguished by phase is limited by the temporal precision of neurons, and it is not clear whether the same source neurons can entrain to different oscillations at the same time[56].

The selective routing allowed by the communication subspace could be adjusted dynamically,

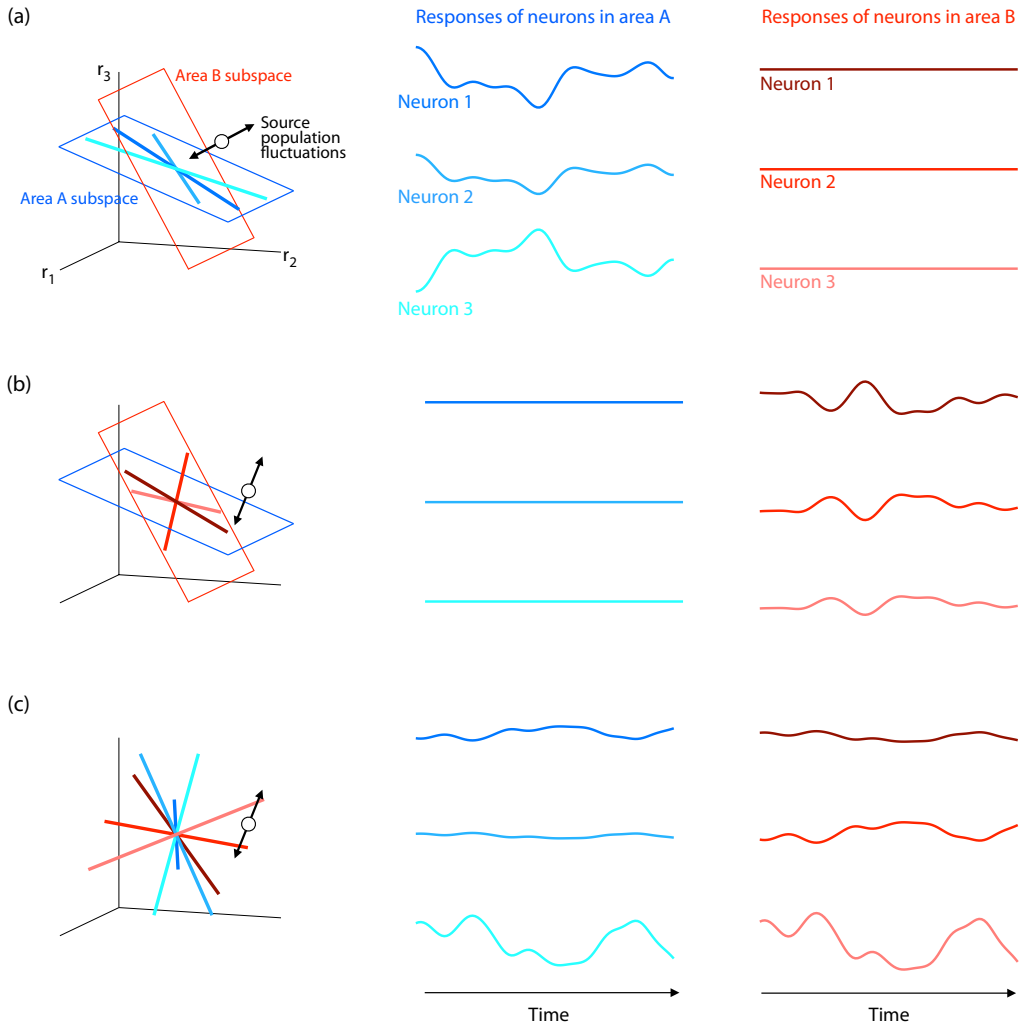

Figure 4.12: **Communication subspaces enable selective communication with multiple downstream areas.** **(a)** When two downstream areas interact with a source population using low-dimensional communication subspaces (2D planes in this illustration; left), some of the source activity is relayed to one area (area A; middle) while remaining private to the second area (area B; right). Left panel follows conventions of Fig. 4.4. The axes corresponds to the activity of each of three source neurons while the colored dimensions correspond to the predictive dimensions of three area A neurons. Blue plane corresponds to the communication subspace for area A. Red plane corresponds to the communication subspace for area B. Middle and right panels depict the rate of each of the three area A and area B neurons, respectively, when source activity varies along the dimension depicted by the black arrow in the left panel. **(b)** Conversely, source activity orthogonal to area A's communication subspace (blue plane; left) will not be relayed to this area (middle), but will drive activity in area B (right). **(c)** If no communication subspaces are present, i.e., both area A and area B read from the full source activity space, then all fluctuations in source activity are relayed to both downstream areas.

allowing moment-to-moment modulation of interactions between cortical areas. Dynamic routing could be accomplished by altering the structure of population activity in a source area; it need not involve changing the communication subspace itself. Much recent work has shown that the structure of population activity is highly and rapidly malleable, by stimulus drive[30,57], task demands[58,59], attention[60,61], and many other factors[62]. A critical implication of our work is that studying population activity changes within a given cortical area can be misleading. One must also understand how these altered population responses interact with the mapping to downstream areas.

4.4 Methods

4.4.1 Pairwise correlations

The pairwise correlation (r_{sc}) analysis in Figure 4.1a was based on a single mean-matching procedure which was done jointly for all stimulus conditions. Statistical evaluation for this analysis was performed after converting r_{sc} to Z-scores using the Fisher transformation[57]:

$$z = \frac{1}{2}\ln\left(\frac{1 + r_{sc}}{1 - r_{sc}}\right)$$

4.4.2 Linear regression models

We related trial-to-trial fluctuations in the source V1 population to those in the target populations using a linear model of the form:

$$Y = XB$$

where X is a $n \times p$ matrix containing the residual activity of the source V1 population and Y is a $n \times q$ matrix containing the residual activity of the target (V1 or V2) population (n represents the number of data points, p and q are the number of neurons in the source and target populations, respectively). The coefficient matrix B is of size $p \times q$. Each of the q columns of B linearly combines the activity of the p neurons in X to predict the activity of one neuron in Y. B can be found using

the ordinary least squares (OLS) solution which minimizes the squared prediction error:

$$B_{OLS} = (X^T X)^{-1} X^T Y$$

To reduce overfitting, we used ridge regression, a variant of classical linear regression, which gives the solution $B_{Ridge} = (X^T X + \lambda I)^{-1} X^T Y$, where I is a $p \times p$ identity matrix and λ is a constant that determines the strength of regularization. We chose the value of λ using 10-fold cross-validation. Specifically, we selected the largest λ for which mean performance (across folds) was within one S.E.M. of the best performance. To jointly select λ and quantify model performance, we employed 10-fold nested cross validation[63].

We sought to test whether the target population activity (V1 or V2) could be predicted using a subspace of the source V1 population activity. In other words, we asked if the linear model $Y = XB$ was still accurate when we impose B to be of a given rank, rank$(B) = m$. This constrained linear regression problem is known as reduced-rank regression (RRR)[43,44], and can be solved using the singular value decomposition:

$$B_{RRR} = B_{OLS} V V^T$$

where B_{OLS} is the ordinary least squares solution and the columns of the $q \times m$ matrix V contain the top m principal components of the optimal linear predictor $\hat{Y}_{OLS} = X B_{OLS}$. To predict target population activity using RRR, we computed:

$$\hat{Y}_{RRR} = X B_{RRR} = X B_{OLS} V V^T = X \bar{B} V^T$$

where $\bar{B} = B_{OLS} V$ is a matrix of size $p \times m$. The columns of $\bar{B}$ define which dimensions of the source population activity are used when generating predictions: they are the predictive dimensions. The sets of weights used to predict each target neuron (the columns of B_{RRR}) are themselves linear combinations of the columns of $\bar{B}$. Note also that the columns of $\bar{B}$ do not form an orthonormal basis. Rather, they are uncorrelated with respect to the source activity, i.e., $\bar{B}^T \Sigma \bar{B} = D$, where Σ is the covariance matrix of the source population activity and D is a diagonal matrix. Thus, the columns of $\bar{B}$ are linearly independent and rank$(\bar{B}) = m$.

To find the optimal dimensionality for the RRR model (the value of m), we used 10-fold cross-

validation and found the smallest number of dimensions for which predictive performance was within one S.E.M. of the peak performance.

4.4.3 Factor analysis

To quantify the dimensionality of the activity in the target populations we used Factor Analysis (FA)[35,39]. FA is defined by:

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$$

$$\mathbf{y}|\mathbf{z} \sim \mathcal{N}(L\mathbf{z} + \mu, \Psi)$$

where $\mathbf{y}$ is a q-dimensional vector containing the observed residuals at a given time point, L is the $q \times m$ loading matrix that defines the relationship between the m-dimensional ($m < q$) latent variable $\mathbf{z}$ and $\mathbf{y}$, μ is a q-dimensional vector and Ψ is a $q \times q$ diagonal matrix. We estimated the dimensionality of the latent variable $\mathbf{z}$ in two steps: (1) we found the number of dimensions m_{peak} that maximized the cross-validated log-likelihood of the observed residuals; (2) we fitted a FA model with m_{peak} dimensions and chose m, using the eigenvalue decomposition, as the smallest dimensionality that captured 95% of the variance in the shared covariance matrix LL^T. This procedure provides more robust estimates of the FA model dimensionality[39].

4.4.4 Selective communication simulation

In order to show how a communication subspace can subserve selective communication (Fig. 4.12), we simulated responses in a source population (3 neurons), as well as in two downstream populations (3 neurons each). The responses of each downstream neuron were generated as a linear combination of the activity of the neurons in the source population. In Fig. 4.12a-b, where both downstream areas interact with the source area via communication subspaces, the predictive dimensions for all neurons in each area were chosen to lie within a 2-dimensional subspace. Specifically, we generated these dimensions by creating randomly oriented unit vectors in the $r_1 r_2$ plane and then rotating these vectors $20°$ around the r_1 axis for the downstream area A neurons and $40°$ for the downstream area B neurons. In Fig. 4.12c, all predictive dimensions were generated by creating randomly oriented unit vectors in the 3-dimensional source activity space. To gener-

ate source activity we drew a sample from a Gaussian process with a squared exponential kernel (length scale $\ell^2 = 0.05$). This sample was then embedded into the 3-dimensional source activity space by projecting the activity along a chosen dimension, i.e., if the Gaussian process sample is represented as a $T \times 1$ vector $\mathbf{x}$ (where T represents the number of time points), and the chosen dimension is represented by the 3×1 vector $\mathbf{v}$, then the 3-dimensional source activity is given by $\mathbf{x}\mathbf{v}^T$. When a communication subspace was present, the source activity was chosen to lie along the private dimension of the relevant area. In Fig. 4.12c, any choice of $\mathbf{v}$ leads to qualitatively similar results, so we chose it to align with the dimension used in Fig. 4.12b. The response for each downstream neuron is given by the projection of the source activity onto the corresponding predictive dimension.

Chapter 5

Properties of the communication subspace

In the previous chapter, we found that the V1-V2 interaction involved only a subset of the activity patterns in V1, i.e., it occurred through a communication subspace. We now turn to characterizing these activity patterns. Is it the case that V2 "reads" from the largest shared fluctuations in V1? If this was the case, V2 might still see the bulk of V1 activity, only ignoring smaller, perhaps "noisy" fluctuations. How do these predictive dimensions relate to the signal components of the activity?

We found that the largest shared fluctuations in V1 (the dominant patterns of activity in V1) are not the most predictive of V2. In other words, the predictive dimensions identified are distinct from the largest shared fluctuations among V1 neurons. Furthermore, we found that there is high overlap between the dimensions that are most predictive of residual activity and the dimensions that are most predictive of the stimulus components of the activity.

We then wondered whether the communication subspace was specific to the V1-V2 recordings analyzed thus far, or whether it could be a more general principle for inter-area interactions. We found that the communication subspace depended on the retinotopic alignment between the source and target populations: for non-aligned populations, interactions tended to involve a single predictive dimensions, which coincided with the most dominant dimension in the source V1 population. We analyzed additional recordings where a naturalistic stimuli was used in place of the oriented gratings, and again found evidence of a communication subspace. Finally, we also

analyzed an additional set of recordings in V1 and V4 in behaving animals, and once again found the inter-area interaction structure to be consistent with a communication subspace.

5.1 Relationship to source activity structure

We next sought to understand the structure of the V1-V2 communication subspace. Specifically, we asked two related questions. First, we examined how the V1 and V2 predictive dimensions are related. Are the predictive dimensions for these target populations aligned or do they capture distinct activity fluctuations within the source V1 population? Second, we examined how the V1-V2 communication subspace relates to the structure of activity within the source V1 population. Is V2 activity predicted by the most dominant fluctuations within V1?

5.1.1 V2 predictive dimensions are not aligned with target V1 predictive dimensions

To characterize the relationship between V1 and V2 predictive dimensions, we made use of the fact that these dimensions are both defined within the source V1 activity space and capture the parts of the source population activity that are most relevant for predicting each target population. We thus removed the source V1 activity along the different predictive dimensions (see Section 5.5.1) and assessed whether the remaining source activity could still be used to predict activity in the target V1 and V2 populations.

We first confirmed that our method for removing activity along predictive dimensions was effective. As expected, our ability to predict V2 fluctuations quickly decreased as we removed the source V1 activity along the dimensions that were most predictive of V2 (Fig. 5.1a, filled circles). Across data sets, average predictive performance vanished when all source activity aligned with the V1-V2 communication subspace had been removed (Fig. 5.1b, filled bars; average normalized performance: -0.005 ± 0.001; value is negative due to cross-validation).

In contrast, after removing the source V1 activity that fell along the top V1 predictive dimensions, we were still able to predict V2 fluctuations (Fig. 5.1a, open circles). Across data sets, we retained a substantial fraction of our ability to predict fluctuations in V2 after removing the same number of V1 predictive dimensions as the number of predictive dimensions in the V1-V2 communication subspace (Fig. 5.1b, open bars; average normalized performance: 0.24 ± 0.01; one-sided

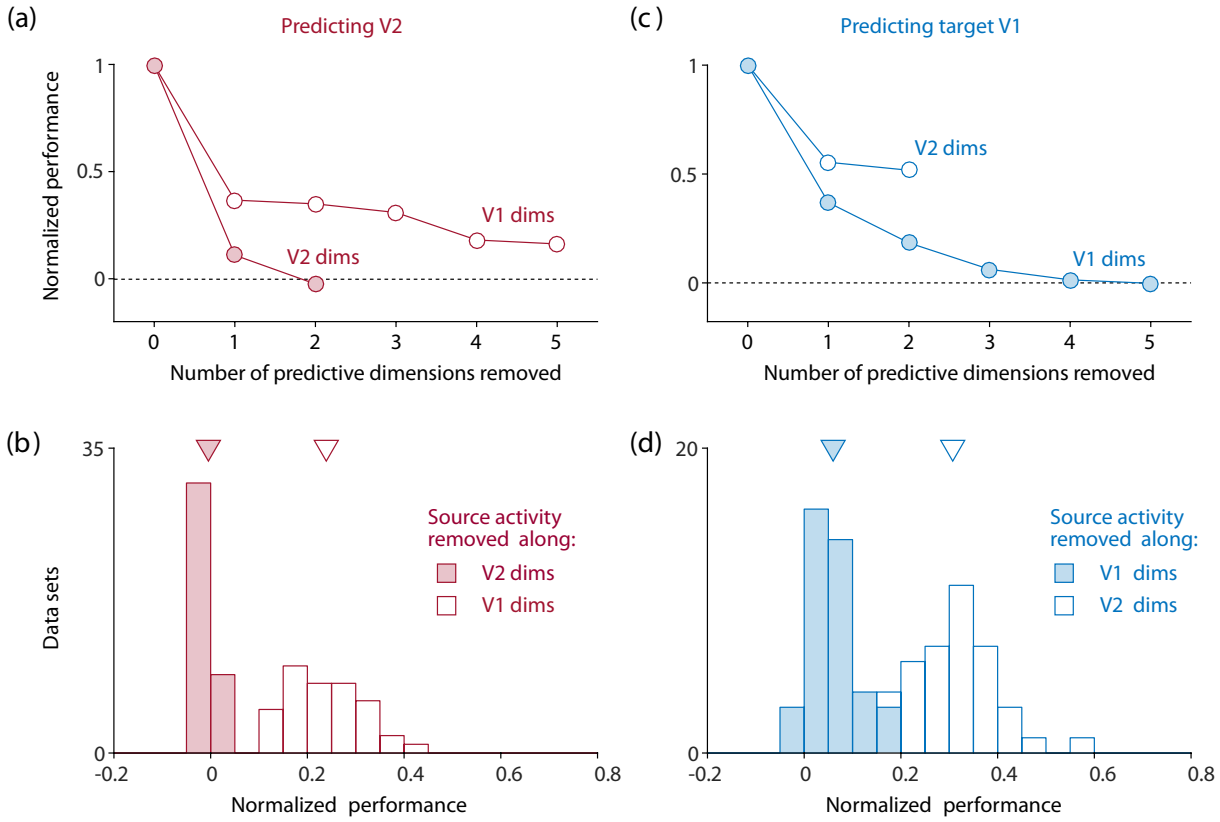

Figure 5.1: **V2 predictive dimensions are not aligned with target V1 predictive dimensions. (a)** Source V1 activity outside of the V1 predictive dimensions is still predictive of V2 activity. V2 predictive performance quickly decreased as we removed the source V1 activity along the V2 predictive dimensions (filled circles). Removing the source V1 activity along the target V1 predictive dimensions had a smaller impact on V2 predictive performance (open circles). Predictive performance was normalized by the performance of the reduced-rank regression model when no activity was removed. S.E.M. is smaller than plotted circles. **(b)** Across all data sets, removing all V2 predictive dimensions drove the V2 predictive performance to 0, as expected (red histogram). Removing the same number of V1 predictive dimensions, had a smaller impact on performance (white histogram), as we could still account for roughly one fifth of the predictable activity in V2. **(c)** Source V1 activity outside of the V1-V2 communication subspace still accounts for a substantial part of the explained activity in target V1. Target V1 predictive performance decreased faster when removing source V1 activity along the target V1 predictive dimensions (filled circles), when compared to removing source activity along the V2 predictive dimensions (open circles). S.E.M. is smaller than plotted circles. **(d)** Across all datasets, even after removing all source activity that was predictive of V2, we could still account for approximately a third of the predictable activity in target V1 (white histogram). Removing the same number of target V1 predictive dimensions had a much larger effect on target V1 predictive performance (blue histogram).

Monte Carlo paired permutation test, $p < 10^{-8}$). This indicates that the V2 predictive dimensions are not aligned with the leading V1 predictive dimensions.

We obtained similar results when predicting fluctuations in the target V1 population (Fig. 5.1c). Across data sets, predictive performance was significantly higher after removing source activity along V2 predictive dimensions (Fig. 5.1d, open bars; 0.31 ± 0.01) than after removing activity along the same number of V1 predictive dimensions (Fig. 5.1d, filled bars; 0.06 ± 0.01; one-sided Monte Carlo paired permutation test, $p < 10^{-8}$). Even after removing all source activity that fell within the V1-V2 communication subspace, we could still predict fluctuations in the target V1 population. Together, these analyses indicate the V1-V2 and V1-V1 interactions not only differ in the number of predictive dimensions, but also involve different patterns of source population activity.

5.1.2 The dominant dimensions of V1 are not the most predictive of V2

To understand how the V1-V2 communication subspace is related to the structure of the source V1 population activity, we used factor analysis to identify the dimensions of largest shared fluctuations within the source V1 population (termed *dominant dimensions*). We then predicted the activity of V2 neurons using linear regression based on the dominant dimensions only. This analysis is conceptually related to reduced-rank regression, which was used to identify the predictive dimensions. However, rather than identifying the subspace that is best for predicting fluctuations in the target population (as in reduced-rank regression), this analysis identifies a subspace that captures the largest shared fluctuations within the source population and then performs regression in that space.

If the dominant source V1 dimensions are able to predict V2 activity as well as the V2 predictive dimensions, for the same number of dimensions, this would indicate that the V1-V2 communication subspace preferentially involves the largest activity fluctuations of the V1 population. However, as shown for a representative data set, the dominant V1 dimensions (Fig. 5.2a, open circles) were not able to predict V2 to the same extent as the predictive dimensions (Fig. 5.2a, filled circles). In contrast, within V1, the predictive and dominant dimensions performed similarly (Fig. 5.2b). Across data sets, predicting V2 fluctuations almost always required more dominant V1

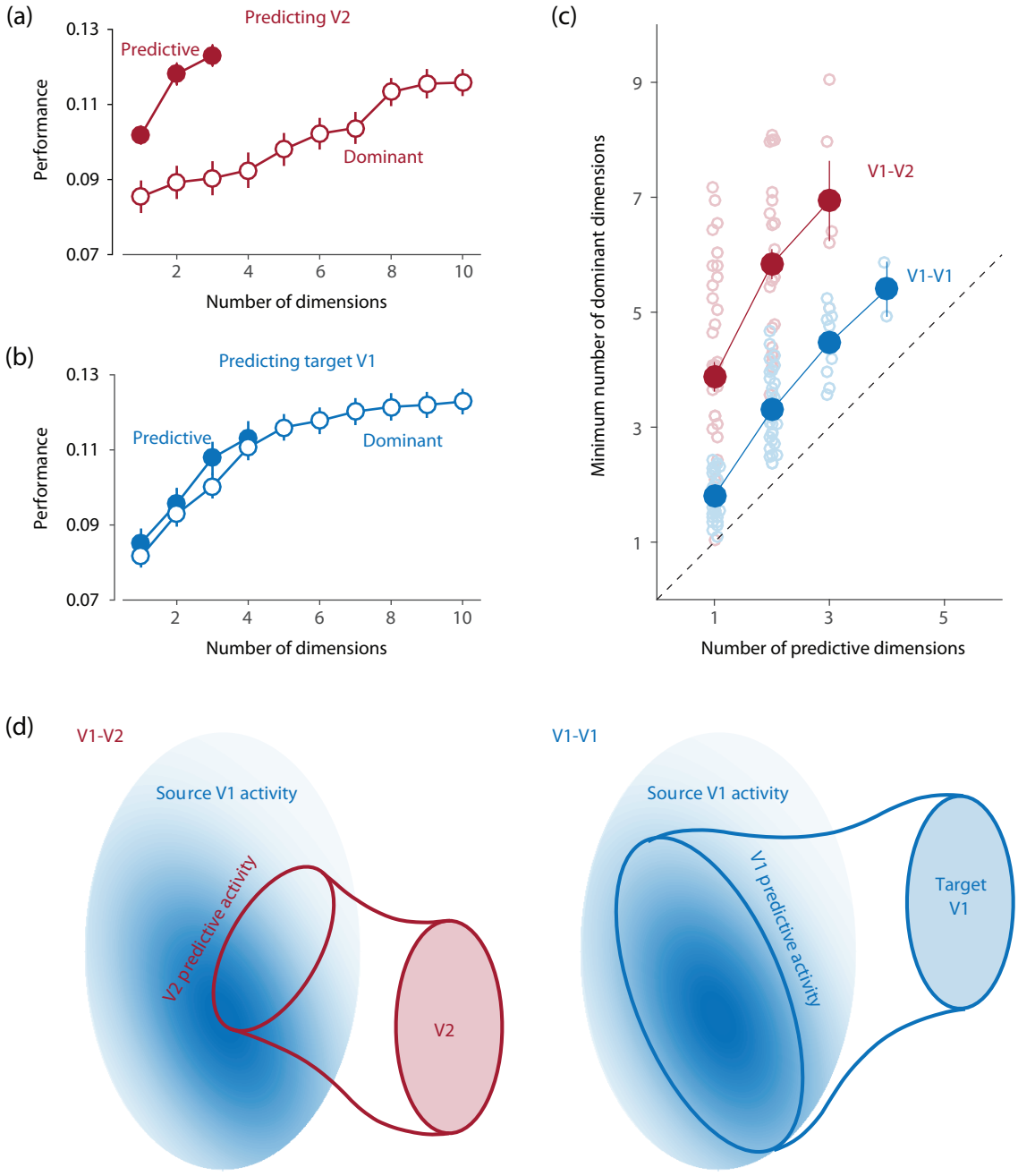
(a)
Predicting V2
Predictive
Dominant
Performance
0.13
0.11
0.09
0.07
2
4
6
8
10
Number of dimensions
(b)
Predicting target V1
Predictive
Dominant
Performance
0.13
0.11
0.09
0.07
2
4
6
8
10
Number of dimensions
(c)
V1-V2
V1-V1
Minimum number of dominant dimensions
9
7
5
3
1
1
3
5
Number of predictive dimensions
(d)
V1-V2
Source V1 activity
V2 predictive activity
V2
V1-V1
Source V1 activity
V1 predictive activity
Target V1

Figure 5.2: **The dominant dimensions of V1 are not the most predictive of V2. (a)** Predicting V2 activity using dominant and predictive dimensions. Dominant dimensions (open circles, factor-analysis regression) carried less predictive power than the same number of predictive dimensions (filled circles, reduced-rank regression). Error bars indicate S.E.M. across cross-validation folds. **(b)** Predicting target V1 activity using dominant and predictive dimensions. Predictive performance using dominant dimensions (open circles, factor-analysis regression) was similar to the predictive performance obtained for the same number of predictive dimensions (filled circles, reduced-rank regression). Error bars indicate S.E.M. across cross-validation folds. **(c)** For a given number of predictive dimensions, a larger number of dominant dimensions was required to reach (within a S.E.M., across folds) the same V2 predictive performance (red circles). When predicting target V1 activity, the number of dominant dimensions needed was only slightly greater than the number of predictive dimensions (blue circles). Error bars indicate S.E.M. across datasets. Faded circles show results for each data set, and were horizontally jittered for visual clarity. **(d)** Left: Schematic of V1-V2 results. Only a small number of activity patterns in the source V1 population was predictive of the V2 population. These predictive activity patterns did not correspond to the dominant patterns in the source V1 population. Large blue ellipse represents the set of all activity patterns observed in the source V1 population. Darker shading indicates more dominant activity patterns. Right: Schematic of V1-V1 results. A large number of activity patterns in the source V1 population was predictive of the target V1 population. These predictive activity patterns corresponded to the dominant patterns in the source V1 population.

dimensions than V2 predictive dimensions (Fig. 5.2c, red). However, for target V1 fluctuations, dominant dimensions of the source V1 population were nearly as informative as the predictive dimensions (Fig. 5.2c, blue; one-sided Monte Carlo permutation test for difference in the minimum number of dominant dimensions when predicting target V1 and V2, $p < 10^{-8}$ for 1 predictive dimension; $p < 10^{-8}$ for 2 predictive dimensions; $p < 0.01$ for 3 predictive dimensions).

These results indicate that the V1 predictive dimensions are well-aligned with the largest source V1 fluctuations. The V2 predictive dimensions, however, are distinct: not only are they less numerous, they are not aligned with the V1 predictive dimensions nor with the largest source V1 fluctuations (as portrayed in Fig. 5.2d).

5.1.3 The communication subspace cannot be explained by a physical bottleneck

Only a fraction of the neurons in the output layers of V1 directly project to V2[64,65]. Thus, it is likely that not all of the recorded V1 neurons project to the recorded V2 population. We wondered whether the small number of V2 predictive dimensions could be a consequence of the linear regression model relying on those few projecting V1 neurons. If this were case, the V1-V2 communication subspace would be a straightforward consequence of a "physical bottleneck" between these

areas. To test this possible explanation for our results, we performed two additional analyses.

First, we studied the structure of the predictive dimensions, after fitting the reduced-rank regression model to the (z-scored) recorded data. If the predictive dimensions rely on only a few V1 neurons, we would expect the magnitude of a few regression weights to be large and those for the remaining source neurons to be low. We found no evidence of this: the weights for the V1 neurons varied broadly and showed no signs of bimodality (Fig. 5.3a).

Second, we conducted a simulation in which a large source population influenced a target population via a physical bottleneck. In this simulation, not all source neurons project to the target population, and only a fraction of source and target neurons are observed, mimicking the situation in our recordings. We generated the source activity (10,000 neurons) in the same way as in Section 4.1.2. To generate the activity in the target population (30 neurons), we defined the response of each target neuron using a weighted sum of the activity of the source population, with the weights chosen from a standard Gaussian distribution. Importantly, the weights were chosen such that 95% of the source neurons did not project to the target population (i.e., 95% of the source neurons had their weights set to 0 for all target neurons) – a physical bottleneck in which only 5% of source neurons provide input to the target population. For comparison, we also conducted an additional simulation in which all neurons in the source population project downstream (i.e., all weights were chosen from a standard Gaussian distribution and none were set to zero so there is no physical bottleneck). The weights for this model were adjusted so that the covariance structure of the activity of the two target populations was matched. We did this using the singular value decomposition of the target population data matrix of the model without the physical bottleneck $Y_u = USV^T$, and then correcting it using:

$$\bar{Y}_u = Y_u V D V^T$$

where $\bar{Y}_u$ is the corrected target population data matrix of the model without the physical bottleneck. D is a diagonal matrix with entries σ_i^b/σ_i^u where σ_i^b and σ_i^u ($i = 1, ...30$) are the ordered singular values of the target populations with and without a physical bottleneck, respectively. This ensures both target populations have the same eigenvalues, and therefore any difference in the number of estimated predictive dimensions are not due to differences in the second order statis-

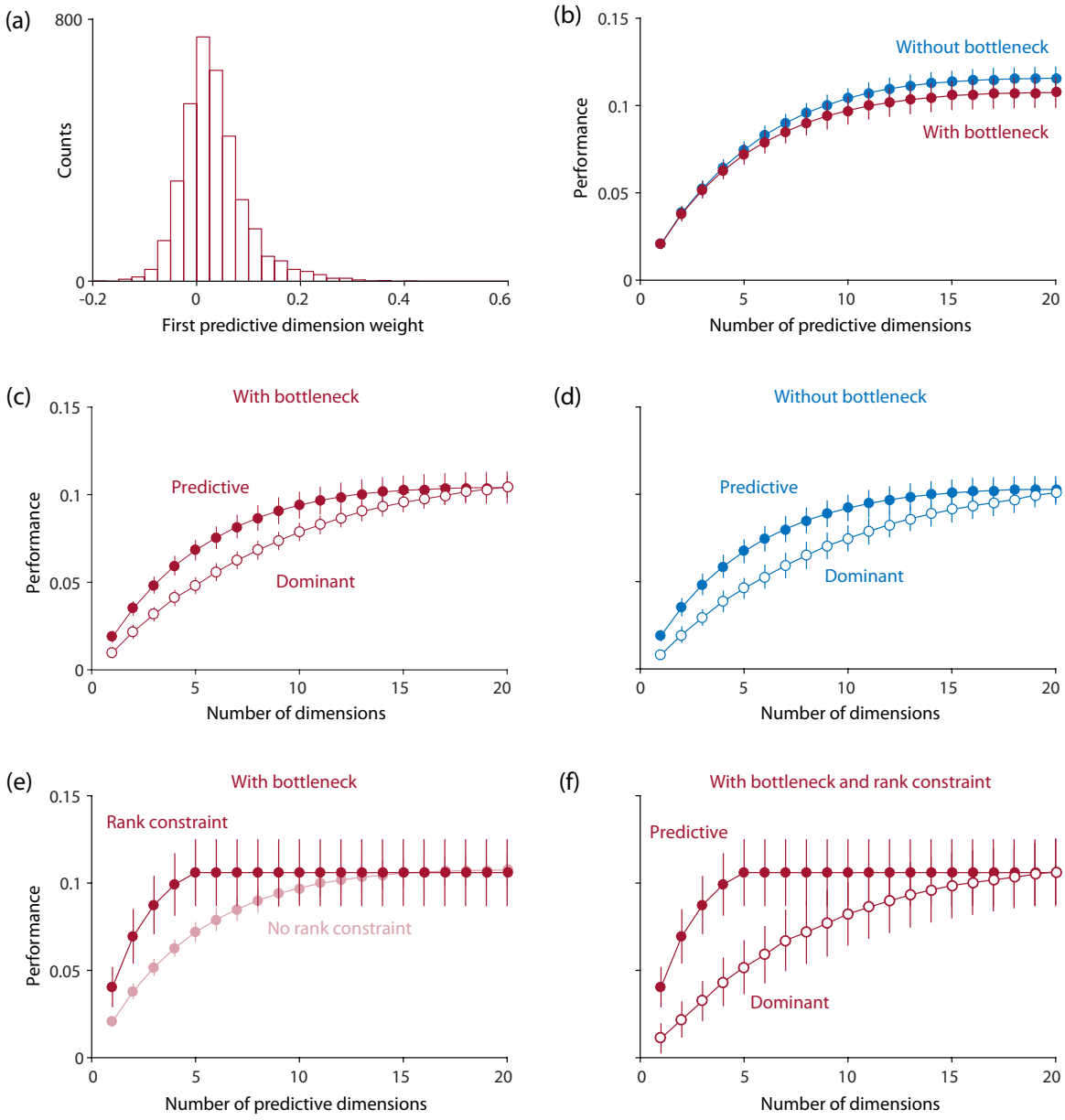

Figure 5.3: **The communication subspace cannot be explained by a physical bottleneck. (a)** Regression weights are broadly distributed, showing no signs of bimodality (across all data sets). Note that since the sign of the predictive dimension is arbitrary, we defined it for each data set as the sign for which most weights are positive. **(b)** Physical bottleneck does not account for low-dimensional V1-V2 interactions. **(c, d)** Physical bottleneck does not account for the misalignment between predictive and dominant dimensions found for the V1-V2 interaction. **(e)** Introducing a rank constraint led to a substantial decrease in the number of identified predictive dimensions. **(f)** Since the small number of predictive dimensions are randomly aligned with the source dominant dimensions, using the dominant dimensions for prediction led to worse performance. **(b-f)** Error bars indicate standard deviation across random source subsets and choices of the mapping matrices.

tics of the target populations (recall that we controlled for this in the recorded data by estimating the dimensionality of the target populations, Fig. 4.6). To estimate the mapping dimensionality, we applied reduced-rank regression to a random subset of 100 source neurons and all of the target neurons. Note that under this linear model generating 30 target responses is equivalent to randomly selecting 30 neurons from a larger target population. This entire process was repeated 25 times, using different random instantiations of the weighting matrices described above and different random selections of the observed source neurons.

We found that the physical bottleneck in the surrogate data had little influence on the number of estimated predictive dimensions. As shown in Fig. 5.3b, the optimal number of predictive dimensions was roughly the same whether or not a physical bottleneck was present (15.6 ± 0.3 with physical bottleneck and 15.2 ± 0.2 without physical bottleneck; two-sided Monte Carlo paired permutation test, $p > 0.05$; error bars indicate standard deviation across random source subsets and choices of the mapping matrices). It might seem counterintuitive that the estimated mapping dimensionality was the same whether or not all observed neurons projected to the target population. Furthermore, although only 5 observed neurons directly influence the target population on average in the simulation with a physical bottleneck, the estimated dimensionality of the interaction is three times that number. Both of these findings arise due to the covariance structure of the source population activity, which contains dimensions of covariability that are shared across many neurons. As a result, even if a neuron does not project to the target population, it can still be predictive of the target population activity if it covaries with neurons that do project.

In these simulations, the activity of the target populations was completely determined by the activity of the source population (i.e., no noise was added). Since we had 500 (or 10,000 when no physical bottleneck was present) source neurons projecting to 30 target neurons, the true dimensionality of the interaction was 30. However, due to the limited sampling of the source population and finite data, the estimated dimensionality is smaller than the true value.

We also tested whether the presence of a physical bottleneck could explain our finding that the dimensions that are most predictive of the V2 population are not well aligned with the dominant dimensions of the source V1 population (Fig. 5.2). We found that the difference in predictive performance between predictive and dominant dimensions was small in these simulations, whether or not a physical bottleneck was present (Fig. 5.3c: with physical bottleneck, Fig. 5.3d: without

physical bottleneck).

In order to qualitatively reproduce the findings for the recordings, we had to restrict the linear mapping between the source and target populations to be low-dimensional. That is, the sets of weights relating each V2 neuron to the source V1 population were chosen as a linear combination of a small set of basis weights (in this example, a 5 dimensional basis set). Imposing this structure reduced the estimated optimal number of predictive dimensions to 4.6 ± 0.1 (Fig. 5.3e).

Furthermore, due to the way in which the low-dimensional mapping was chosen, the predictive dimensions now lay in a subspace of the source population activity that was randomly oriented with respect to the dominant dimensions of this population Fig. 5.3f. As a result, using the source dominant dimensions to predict activity in the target population led to worse performance, when compared with the performance afforded by the identified predictive dimensions.

We conclude that that existence of a physical bottleneck cannot explain: (1) the communication subspace; or (2) the misalignment of the predictive and dominant dimensions observed in the recorded activity. Only by enforcing a low-dimensional mapping were we able to reproduce these findings.

5.1.4 Some dimensions of the source population activity remain private regardless of the stimulus

We showed that only a few dimensions of V1 activity are required to predict V2 responses. Consequently, a significant component of V1 population fluctuations remain private to V1 (from the perspective of V2). However, our analysis involved identifying predictive dimensions separately for each stimulus condition (i.e., data set). As a result, dimensions that are private for one stimulus condition may become predictive for another. Alternatively, some dimensions of the source V1 activity might remain private across all stimulus conditions. If so, these globally-private dimensions would constitute a subspace of V1 activity that is not related to V2 activity, perhaps representing processes internal to V1 which should not be relayed downstream.

To determine whether there are globally-private dimensions of V1 activity, we first identified a joint predictive subspace – the subspace of the source V1 activity that is predictive of V2 activity across all stimulus conditions. We did so by simultaneously fitting the reduced-rank regression

model to residual responses to all eight stimulus conditions in each recording session. This revealed that V1-V1 interactions involved as many predictive dimensions (9.7 ± 0.5) as the dimensionality of the target population (9.1 ± 0.5). V1-V2 interactions, on the other hand, involved fewer predictive dimensions (5.2 ± 0.3) than the target population dimensionality (9.6 ± 0.4). Thus, V1-V2 interactions occurred through a communication subspace. We then applied the analysis of Section 5.1.1. Namely, we assessed how well we could predict V2 activity as we removed V1 activity that fell along the identified joint communication subspace.

As shown in Fig. 5.4a for an example session, our ability to predict V2 activity quickly decreased as we removed the source V1 activity that fell along the V2 joint predictive subspace (filled circles). When we removed all activity in the joint communication subspace - in this case 6 dimensions - we were entirely unable to predict V2 responses. Removing the source V1 activity in the V1 joint predictive subspace had a smaller impact on V2 predictive performance (open circles).

Since the joint predictive subspace is defined across all stimulus conditions, the activity outside the joint predictive subspace should not be predictive of V2 under any condition. Indeed, across data sets, average predictive performance vanished when all source activity aligned with the V1-V2 communication subspace had been removed (Fig. 5.4b, filled bars; mean fraction of original predictive performance across data sets: -0.002 ± 0.001 ; t-test $p > 0.05$). Removing the same number of V1 predictive dimensions had a smaller impact on performance (Fig. 5.4b, open bars; mean fraction of original predictive performance: 0.12 ± 0.01; one-sided Monte Carlo paired permutation test for difference between removing V1 and V2 predictive dimensions, $p < 10^{-3}$).

We then performed a similar analysis, attempting to predict activity in the target V1 population. Specifically, we sought to determine whether the activity outside the V2 joint predictive subspace could be used to predict fluctuations in the target V1 population. The global private subspace would only be meaningful if there were substantial source activity that falls within it. As shown in Fig. 5.4c an example session, target V1 predictive performance decreased more quickly when we removed source V1 activity along the target V1 predictive dimensions (filled circles), compared to removing source activity along the V2 predictive dimensions (open circles). Importantly, we retained some ability to predict activity in the V1 target populations, after removing all V1 source activity along the V2 joint predictive subspace.

Across data sets, a substantial part of the target V1 activity could be predicted after remov-

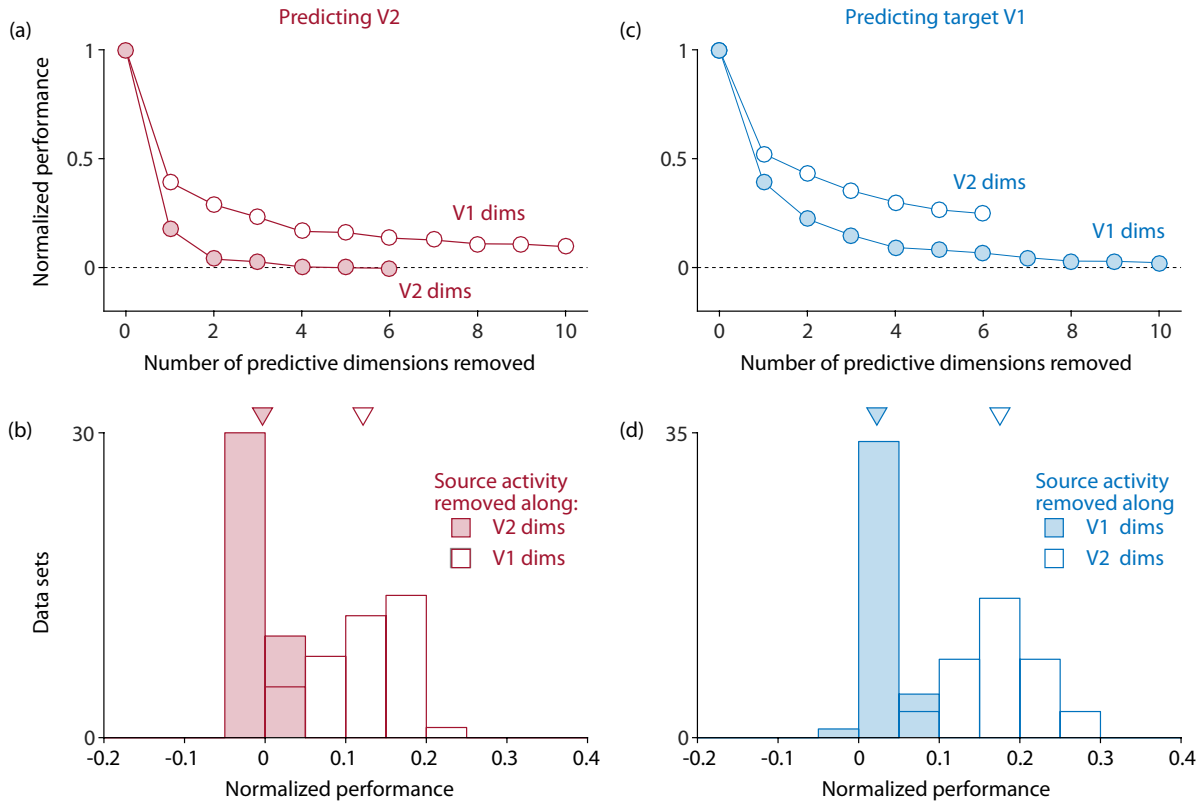

Figure 5.4: **Some dimensions of the source population activity remain private regardless of the stimulus. (a)** V2 predictive performance decreased more quickly when removing activity along the V2 joint predictive dimensions. Predictive performance was normalized by the performance when no activity was removed. S.E.M. (across folds) is smaller than plotted circles. **(b)** Across all datasets, removing source V1 activity along the V2 joint predictive dimensions resulted in no ability to predict activity in V2. **(c)** Even after removing all V2 joint predictive dimensions, we could still predict a substantial part of the target V1 activity. S.E.M. (across folds) is smaller than plotted circles. **(d)** This was true across all datasets.

ing the V2 joint predictive subspace (Fig. 5.4d, open bars; mean fraction of original predictive performance across conditions: 0.18 ± 0.01). Recall that when we identified the communication subspaces separately for each stimulus condition (Fig. 5.1d), we retained 0.31 ± 0.01 of the original predictive performance. Thus, the predictive performance for the target V1 activity is largely preserved when going from a condition-specific V1-V2 private subspace to an across-condition V1-V2 private subspace. Removing the same number of target V1 predictive dimensions had a much larger effect on target V1 predictive performance (Fig. 5.4d, filled bars; mean fraction of original predictive performance across conditions: 0.026 ± 0.003; one-sided Monte Carlo paired permutation test for difference between removing V1 and V2 predictive dimensions, $p < 10^{-3}$).

Together, these results indicate that a substantial component of V1 activity is globally-private (across the stimuli tested), and not predictive of the V2 population. This is consistent with the results reported in Section 4.2.3: while the predictive dimensions identified for the different stimuli are not identical, they differed only moderately, showing a high degree of overlap.

5.2 Relationship to signal communication

So far, we identified predictive dimensions using residual activity, i.e., after subtracting from each trial the corresponding PSTH. While there are advantages to this approach, it is important to understand how the predictive dimensions identified in this way relate to stimulus processing. Have we lost critical components of the V1-V2 interaction by removing the stimulus component (or PSTH) from the activity? Or do the predictive dimensions identified for the residual activity largely capture these stimulus components as well?

To answer these questions, we directly compared the communication subspaces identified using residual activity with those identified when we did not subtract the PSTHs (termed here the "full activity"). We did so by estimating the predictive dimensions either from the residual activity (pooled across conditions, as in Section 5.1.4) or from the full activity, and then used these subspaces to predict held-out sets of residual and full activity. For example, to predict full activity using the residual communication subspace, we first identified the residual predictive dimensions on a residual activity training set, then projected the full source V1 activity onto the subspace spanned by these dimensions and used these projections to predict the full V2 activity. The pre-

dictive performance was then normalized by the performance obtained from predicting the full activity using the subspace fit to the full activity training set. This process was repeated 10 times, using 10-fold cross-validation. There was no overlap between the data used to fit the models and the data used to quantify predictive performance.

5.2.1 Predictive dimensions identified using residual activity can predict responses that include stimulus information

When predicting the full activity, we found that the subspace identified using the residual activity retained 0.85 ± 0.04 of the predictive performance achieved by the predictive dimensions identified for the full activity itself (Fig. 5.5, red open circles). For comparison, we also quantified the performance retained when using a random subspace with the same dimensionality as the residual subspace. We selected these random predictive subspaces by fitting a reduced rank regression (RRR) model to the same training residual activity, but after shuffling the sample order of the two areas independently, thereby destroying V1-V2 covariability. These random communication subspaces retained a much smaller fraction of the performance (Fig. 5.5. black open circles; 0.16 ± 0.03).

Using the subspace identified for the full recordings to predict the residual activity resulted in an even higher fraction of performance retained (Fig. 5.5, red open triangles; 0.97 ± 0.01 vs. 0.06 ± 0.01 for random dimensions). This was expected, given that the full activity is composed of the stimulus component and trial-to-trial variability. Thus predicting the full activity involves also predicting the residual activity. Consistent with this statement, more predictive dimensions were needed to predict the full activity than to predict residual activity (6.75 ± 0.3 for full activity vs 5.1 ± 0.3 for residual activity).

One might worry that the ability of a subspace identified using residual activity to predict the full activity is not because this subspace captures the stimulus component, but rather because the stimulus component is small compared to the trial-to-trial variability (i.e., the full activity is dominated by trial-to-trial fluctuations). If so, the residual subspace would retain most of the full activity predictability even if it failed to capture the stimulus component. To test whether this was the case, we used the residual subspace to predict the stimulus component. We did so by creating

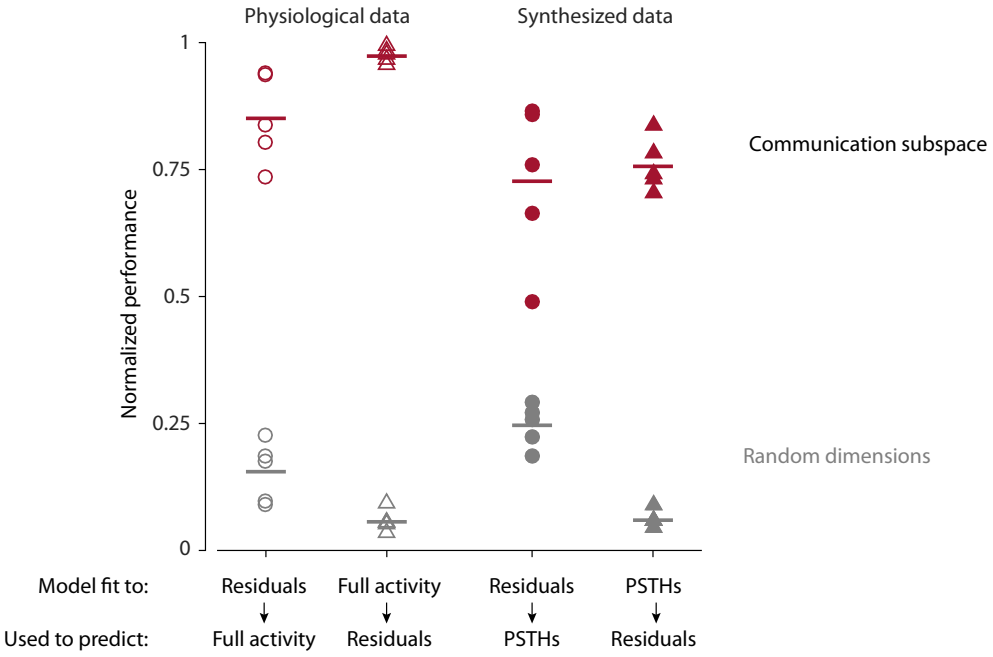

Figure 5.5: **Predictive dimensions identified using residual activity can predict responses that include stimulus information.** Each mark (circle of triangle) corresponds to one recording session. Horizontal bar indicate average across recording sessions.

surrogate data with realistic levels of trial-to-trial variability for which all covariability, within and across areas, was solely due to the PSTHs. In other words, the only interactions between the two areas were those induced by the PSTHs. Specifically, for each stimulus, we generated 400 trials (the same number as with the real data) of responses by drawing from a Poisson distribution where the mean, for each time bin, was given by each neuron's PSTH. We used this surrogate data because directly fitting a RRR model to the PSTHs leads to overfitting, due to the small number of samples. This approach (fitting RRR to surrogate data created in this way) can be thought of as a form of regularization, similar to that employed by ridge regression but using Poisson noise. For these surrogate data, the residual subspace retained most of the predictability achieved by identifying predictive dimensions using the surrogate data itself (Fig. 5.5, red filled circles; 0.73 ± 0.07 vs. 0.25 ± 0.02 for random dimensions, in black filled circles). The subspace identified using the surrogate data was also able to predict residual activity (Fig. 5.5, red filled triangles; 0.76 ± 0.02 vs. 0.06 ± 0.01 for random dimensions in black filled triangles).

These results show that there is high overlap between the dimensions that are most predictive of residual activity and the dimensions that are most predictive of the stimulus components of the activity. Thus, one can use covariations of trial-to-trial fluctuations to identify the relationship between mean activity of neurons in one area and those in another, at least for the limited stimulus ensemble used here.

5.3 Dependence on receptive field alignment, stimulus and anesthesia

5.3.1 V1-V2 interaction structure is distinct for retinotopically-offset V2 populations

In the data sets analyzed so far, the V1 and V2 populations had retinotopically-aligned spatial receptive fields (average receptive field distance: $0.58°$), maximizing the probability of direct feed-forward connections between the populations. In Section 4.2.2, we found that small changes in receptive field alignment could not account for the differences in the interaction structure for the V1-V1 and V1-V2 interactions. Here, we ask if this is still the case for V2 populations whose spatial receptive fields are more grossly misaligned with those sampled in V1. Specifically, we analyzed five additional sessions (each containing responses to gratings of 8 different orientations, for a total of 40 data sets) where the V1 and V2 receptive fields had a small degree of overlap (average

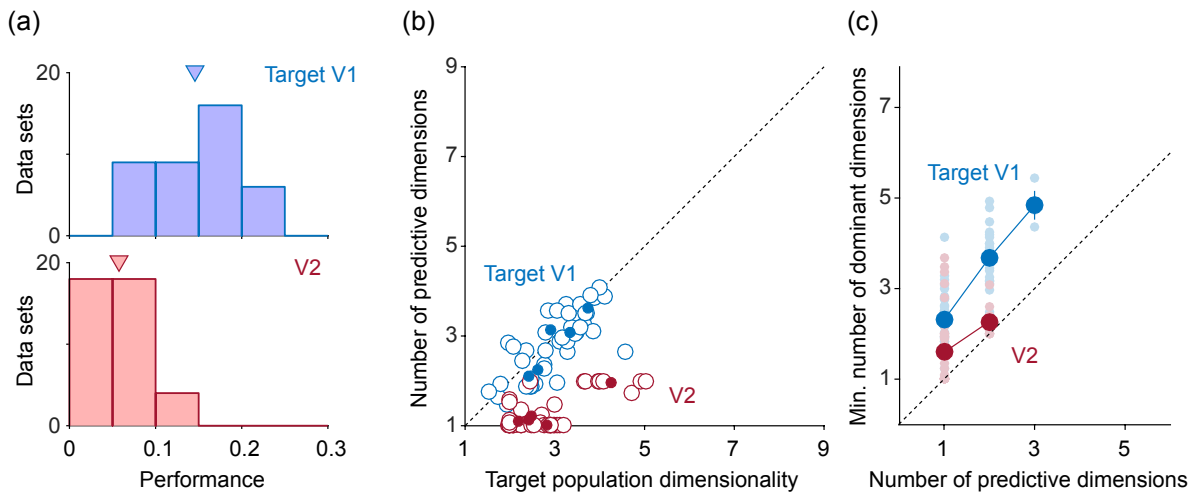

Figure 5.6: **V1-V2 interaction structure is distinct for retinotopically-offset V2 populations. (a)** Predictive performance was lower for retinotopically-offset V2 populations. Triangles indicate mean. **(b)** Interaction structure for retinotopically-offset V2 populations frequently hinged on a single predictive dimension. Each open circle corresponds to one data set. Filled circles indicate averages across data sets for each of the 5 sessions. **(c)** The predictive dimensions for the retinotopically-offset V2 populations were well aligned with the source dominant dimensions. Error bars indicate S.E.M. across datasets. Faded circles show results for each data set.

receptive field distance: 3.73°).

Our ability to predict responses in V2 was substantially lower when the source V1 population had retinotopically-offset receptive fields than when they are aligned (Fig. 5.6a; 0.06 ± 0.01 for offset, red histogram vs. 0.15 ± 0.01 for aligned, shown in Fig. 4.1b). The performance for the target V1 prediction, as expected, remained roughly the same (0.15 ± 0.01 for offset, blue histogram vs. 0.13 ± 0.01 for aligned, shown in Fig. 4.1b).

We then asked how many dimensions were necessary to predict the target populations in the data sets with retinotopically-offset receptive fields. When predicting target V1 activity, the number of predictive dimensions matched the dimensionality of the target V1 population (Fig. 5.6b, blue circles). When predicting the V2 population, on the other hand, the number of predictive dimensions was smaller than the dimensionality of the V2 population (Fig. 5.6b, red circles). Indeed, for most offset data sets only a single predictive dimension was needed to predict V2 (optimal dimensionality was one for 29 out of 40 data sets with retinotopically-offset receptive fields).

To understand which source V1 population fluctuations were captured by the V2 predictive

dimension(s), we assessed how similar the predictive dimensions were to the dominant dimensions in the source V1 population. Contrary to our findings for retinotopically-aligned recordings (cf. Fig. 5.2c), we found that for the retinotopically-offset data sets predictive and dominant dimensions achieved similar performance when predicting the V2 population (Fig. 5.6c).

In summary, we found that V1-V2 interactions depend on retinotopical alignment: predictive performance is lower for recordings from populations with retinotopically-offset receptive fields, and these predictions frequently required only a single predictive dimension. The predictive dimensions for the offset data were similar to the dominant dimensions of activity in the source population.

5.3.2 V1-V2 interactions driven by naturalistic stimuli also occur through a communication subspace

All results presented thus far are based on analyzing responses to oriented gratings. To understand the interaction between V1 and V2 for richer stimuli, we also analyzed V1 and V2 responses to repeated presentations of a natural movie.

These recordings were made in V1 (130 neurons) and V2 (18 units) in one anesthetized monkey, while presenting 300 repetitions of a 30s natural movie. Movies consisted of 750 unique frames, with each frame presented on 4 sequential monitor refreshes; given the refresh rate of 100 Hz, this yielded a video rate of 25 Hz. For our analysis, we divided the 30 s movie into shorter segments of either 1, 1.5, 3, 6, 10 or 30 s (the full movie), yielding 30, 20, 10, 5, 3 and 1 data sets, respectively. We then analyzed each data set independently. Note that different data sets (i.e., movie segments) correspond to activity evoked by distinct stimuli. As in the main text, we binned activity using a 100ms window, and subtracted the corresponding PSTH from each cell's response.

We found that the number of V1 predictive dimensions closely matched the dimensionality of the target V1 population (Fig. 5.7, blue symbols). V2 predictions, on the other hand, consistently required fewer predictive dimensions than the dimensionality of the V2 population (Fig. 5.7, red symbols).

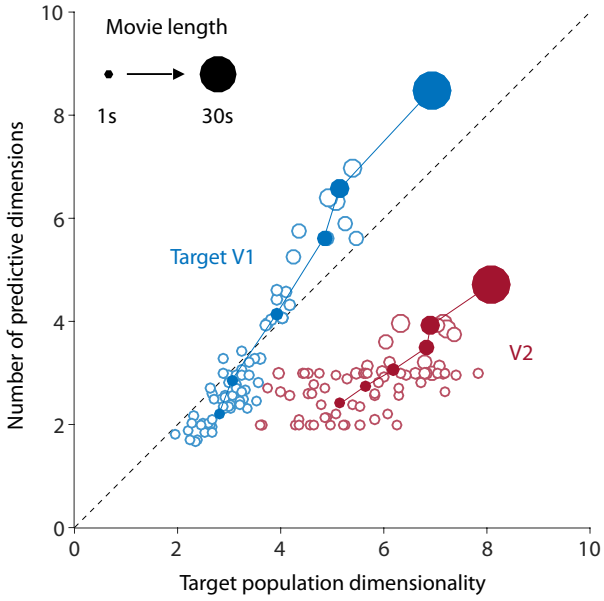

Figure 5.7: **V1-V2 interactions driven by naturalistic stimuli also occur through a communication subspace.** Filled circles show the average across all movies of a given length. Open circles show the estimates for each movie segment (averaged across 25 rate-matched samples of the V1 and V2 populations). Circle size indicates movie length.

5.3.3 Interactions between V1 and V4 in alert, behaving animals are also low-dimensional

The V1-V2 recordings were performed in sufentanil-anesthetized animals. To test whether interactions between cortical areas are also low dimensional in awake animals, we analyzed simultaneous recordings from small neuronal populations in V1 and V4, in two monkeys performing a grating orientation discrimination task. V4 is, after V1 and V2, one of the largest areas in the visual cortex. Furthermore, V1 and V4 are also directly and densely connected (see Fig. 3.1).

Animals were required to fixate on a small target ($0.1°$; fixation window diameter $1°$). After a delay of 200 ms, a drifting grating was presented for 200 ms (2-4 cpd, 6 Hz, 100% contrast). Animals were required to maintain fixation for an additional 200 ms, before reporting their decisions by making saccades to two choice targets. Saccades to the vertical choice target (i.e., above the fixation point) were rewarded if the orientation was less than $45°$; saccades to the horizontal choice target (i.e., on the horizontal meridian) were rewarded when stimulus orientation was larger than $45°$. Trials in which a $45°$ grating was presented were rewarded randomly.

After a training period of roughly 6 months, we implanted 48 channel Utah arrays in V1 and

V4. The spatial receptive fields of the two populations were overlapping, such that the V1 receptive fields lay entirely within the aggregate receptive fields of the V4 population. We then recorded from V1 and V4 simultaneously, while animals performed the discrimination task on gratings which covered the aggregate receptive fields of the recorded populations (stimulus diameters were $1 - 4°$, centered at eccentricities of $0.5 - 2.5°$). During recording sessions, we presented 7 orientations centered at $45°$ and chosen to straddle the slope of the psychometric function ($37.5°$ to $52.5°$), plus the two extremes ($0°$ and $90°$). The probability of presenting a $45°$ orientation in most sessions was twice that of the other orientations, whose presentation was equally likely. Additional details of the recording approach are provided in ref. 10, and ref. 66.

We analyzed 145 recording sessions as in the main text (see Online Methods). Due to the limited number of trials for each grating orientation, we jointly analyzed all stimulus conditions (after removing the corresponding stimulus PSTHs) and used a 200ms bin size. We treated each V1-V4 recording session as an independent data set. To form source and target V1 populations, we randomly split the V1 populations in half, and then randomly selected neurons from the V4 population to match the size of the target V1 population. For data sets for which the number of V4 neurons was smaller than the target V1 population, all V4 neurons were used and the extra target V1 neurons were re-assigned to the V1 source population. The process of randomly selecting the source and target V1 and V4 populations was repeated 50 times for each data set.

To ensure that our results could not be explained by differences in target population dimensionality or predictive performance, we matched the joint distributions for these quantities for both target populations. Briefly, we computed the joint histogram of target population dimensionality and predictive performance for both the target V1 and V4 populations (target population histogram edges: $[0, 0.5, ..., 4]$; predictive performance histogram edges: $[0, 0.025, ..., 1]$; excluding recording sessions for which the predictive performance for either target population was below 0.001).

Our goal was to select sessions for which the target V1 populations and the V4 populations both followed a minimum common distribution. This cannot be done exactly while retaining the pairings between each target V1 population and the corresponding V4 population. We thus used an approximate matching approach: we randomly chose a target population and bin, and then randomly selected a session for which the target population dimensionality and performance fell

within the selected bin. In parallel, we added the corresponding paired target population to the other histogram. We repeated this procedure until each bin, for each target population, had at least as many sessions as the common histogram. The approximate matching procedure was repeated 100 times. We then used the matching for which the sum of the total absolute errors between the resulting target population histograms and the common histogram was smallest. This procedure yielded 32 matched sessions, containing $5-38$ V1 neurons (average: 16 ± 7 s.d.) and $4-42$ V4 neurons (26 ± 10 s.d.), with $414-4500$ data points per session (1899 ± 1166 s.d.).

We found that fewer predictive dimensions were required to predict V4 activity, than target V1 activity (Fig. 5.8a; 0.66 ± 0.06 for V1-V4; 1.04 ± 0.09 for V1-V1; one-sided Monte Carlo permutation test, $p < 10^{-3}$), consistent with analyses of within and between area interactions in the V1-V2 recordings. As expected, due to the matching procedure, population dimensionality was similar for target V1 and V4 (Fig. 5.8b; 1.66 ± 0.13 for target V1; 1.48 ± 0.12 for V4; two-sided Monte Carlo permutation test, $p > 0.05$). For a given target population dimensionality, more predictive dimensions were necessary to predict activity in the target V1 population, compared with the V4 population (Fig. 5.8c). The higher the target population dimensionality, the clearer the difference we observed between V1-V1 and V1-V4 interactions. For example, considering only recording sessions for which the target population dimensionality was above 2 for both target populations, the average number of predictive dimensions was 1.04 ± 0.06 for V1-V4 and 1.56 ± 0.18 for V1-V1 interactions (one-sided Monte Carlo permutation test, $p < 10^{-2}$). Thus, recordings from larger populations would likely yield still larger differences between V1-V1 and V1-V4 interactions.

Although our analysis revealed a communication subspace between V1 and V4 populations recorded in awake animals, the results were less striking than for the V1 and V2 populations recorded in anesthetized animals. On the other hand, the awake recordings involved much smaller populations, which strongly reduces the dimensionality of the population activity[39] and thus any observable differences for between versus within area interactions.

For comparison, we thus re-analyzed the V1-V2 data sets used in the main text to match the data available from our awake recordings. To do so, we used a bin size of 200 ms, and randomly sub-selected the source and target V1 and V2 populations to match population size of the V1-V4 recording sessions. Specifically, for each of the 32 awake recording sessions shown above, we randomly selected one of the 40 V1-V2 data sets, and re-analyzed it after randomly sub-selecting

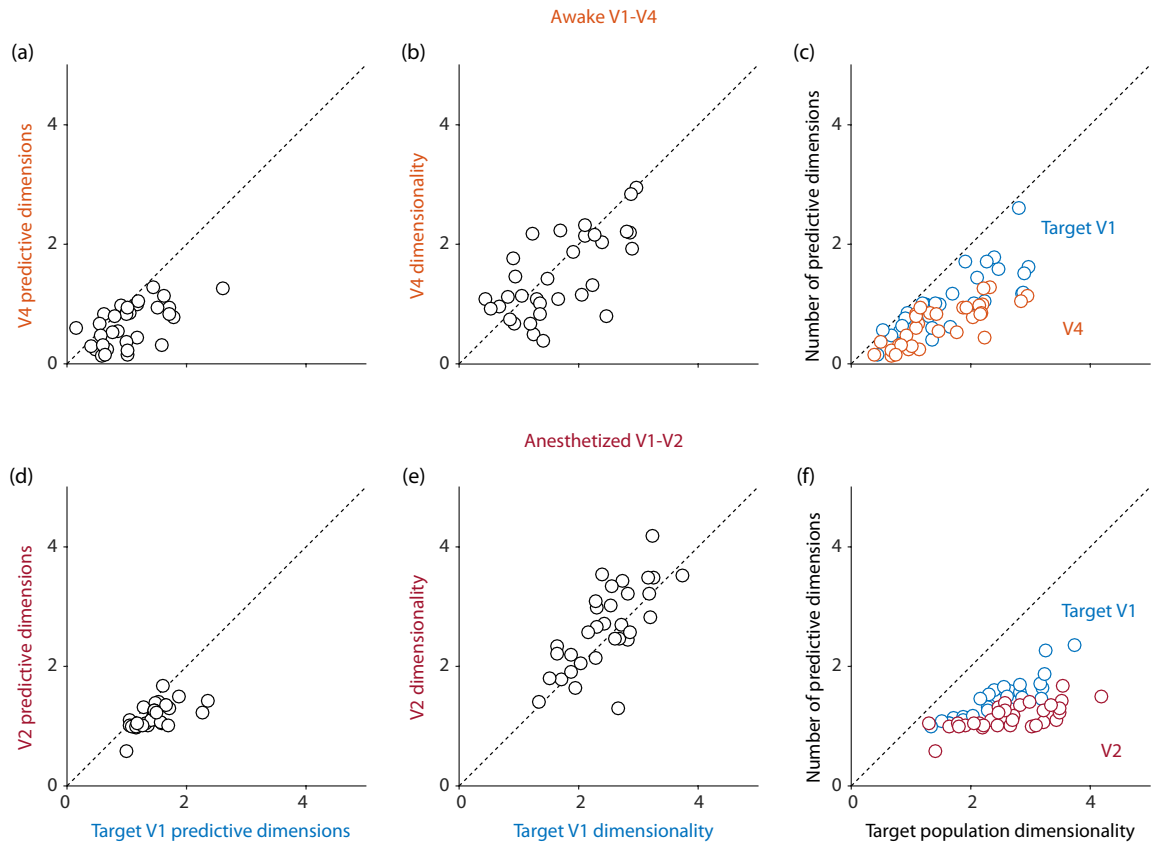

Figure 5.8: **Interactions between V1 and V4 in alert, behaving animals are also low-dimensional.** **(a)** V1-V4 interaction involved less predictive dimensions than the V1-V1 interaction. **(b)** As a result of the matching procedure, target population dimensionality was similar for target V1 and V4. **(c)** For a given target dimensionality, less predictive dimensions were required to predict V4, compared with target V1. **(a-c)** Each open circle corresponds to one of the 32 matched recording sessions. **(d)** Matching the source and target population sizes with those in the V1-V4 recordings, still revealed that predicting V2 activity required less predictive dimensions than for target V1 prediction. The difference is, however, smaller, and on par with that observed in the V1-V4 recordings. **(e)** Target V1 and V2 population dimensionalities were similar. **(f)** For a given target dimensionality, less predictive dimensions were required to predict V2, compared with target V1. The difference is, however, smaller, and on par with that observed in the V1-V4 recordings. **(d-f)** Each open circle corresponds to one of the 32 matched recording sessions.

source and target neurons to match the size of the populations in the corresponding awake recording session (this is repeated 50 times, for different sub-selections of the source and target neurons; each symbols shows the average across all 50 repetitions). Additionally, since we could not rate-match the target V1 and V4 populations, we also did not employ the rate-matching procedure to the sub-selected target V1 and V2 populations.

In these 'awake-matched' V1-V2 data, the number of V2 predictive dimensions was smaller than for target V1 (Fig. 5.8d; 1.16 ± 0.04 for V1-V2; 1.44 ± 0.06 for V1-V1; one-sided Monte Carlo permutation test, $p < 10^{-3}$), but the difference between the two cases was smaller than in the main text (Fig. 4.5) and more similar to the awake data (Fig. 5.8a). Target population dimensionality was higher for V2 than for target V1 (Fig. 5.8e; 2.64 ± 0.12 for V2; 2.45 ± 0.10 for target V1; two-sided Monte Carlo paired permutation test, $p = 0.02$), but substantially smaller than in the main text (Fig. 4.6) and again more similar to the awake data (Fig. 5.8b). Comparing the number of predictive dimensions with the dimensionality of the corresponding target populations revealed that, for a given target population dimensionality, more predictive dimensions were necessary to predict activity in the target V1 population, compared with the V2 population (Fig. 5.8f). The difference between V1-V2 and V1-V1 interactions were similar to those found for the awake recordings.

These results indicate: (1) a communication subspace also exists for interactions between V1 and V4 in awake primate cortex, and thus the results in the main text cannot be ascribed to peculiarities of V1-V2 interactions or to anesthesia; and (2) the quantitative differences between our results in awake and anesthetized animals are due primarily to differences in the size of the recorded populations and the number of recorded trials.

5.4 Discussion

In Chapter 4, we found that V1-V2 interactions occur through a communication subspace, implying that not all activity in V1 is effectively propagated to V2. In this chapter we turned to understanding the properties of this communication subspace, particularly how it relates to the structure of the activity within V1. We found that not only was V2 activity was related to a small subset of population activity patterns in the source V1 population, these patterns were distinct from the most dominant shared V1 fluctuations. In contrast, more activity patterns in the source

V1 population were relevant for predicting the activity of other V1 neurons, and the dominant fluctuations in the source population were the most predictive.

As discussed in Section 4.2.3, our approach of characterizing the V1-V2 interaction structure separately for each stimulus condition can be though of as performing local linear approximations to the underlying nonlinear computation. But given this logic, how can we be sure that the communication subspaces are not an oddity, perhaps defining private and communicated V1 fluctuations differently for each grating stimulus? First, we confirmed that a communication subspace was evident when we analyzed our grating data sets together (Section 5.1.4). Thus, it is not the case that V1 population fluctuations that are private during the presentation of one grating stimulus are relayed to V2 during the presentation of another. Consistent with the existence of a shared communication subspace, we also found that the communication subspace defined for responses to one grating could effectively predict responses to other gratings (Section 4.2.3). Second, we analyzed V1-V2 interactions during repeated presentations of brief naturalistic movies. These responses also revealed a communication subspace (Section 5.3.2), indicating that the low-dimensional V1-V2 interactions do not arise from the use of grating stimuli. Finally, we analyzed the relationship between the communication subspace and the mapping of stimulus-driven activity from V1 to V2 (i.e., the PSTHs) and found that the communication subspace was able to capture responses that included stimulus information (Section 5.2). These lines of evidence together indicate that the communication subspace is a fundamental aspect of V1-V2 interactions.

What is the basis of the communication subspace? One might think that our results reflect global population fluctuations, which involve all neurons increasing and decreasing their activity together[37,39,67] and may be more prevalent under anesthesia[37] (but see ref. 68, 69). However, since global fluctuations are one-dimensional, they cannot by themselves explain the V1-V2 interactions reported here, which typically involved more than a single dimension. In addition, the most predictive dimensions for the V1-V2 interaction were not aligned with the largest shared fluctuations in V1, nor with the dimensions that were most predictive of the target V1 activity. Finally, we observed a similar communication subspace in recordings from more limited neuronal populations recorded in V1 and V4 of alert, behaving monkeys (Section 5.3.3), ruling out any confounding influence of anesthesia.

A second possibility is that the low-dimensional communication subspace could simply reflect

that only a small subset of V1 neurons project to V2[64]. If only the activity of those neurons were related to V2, the communication subspace would arise trivially. However, we found that the V2 predictive dimensions involved a broadly-distributed weighting of all V1 neurons, rather than the activity of a small subset of neurons (Section 5.1.3). Using additional simulations, we also confirmed that an anatomical bottleneck between source and target areas, by itself, cannot explain our results (Section 5.1.3).

Our work did not test for a behavioral role for the communication subspace, focusing instead on developing a framework for understanding inter-areal population-population interactions. Our framework does, however, make clear predictions of how the communication subspace could contribute to behavior, which can be tested in future work. For instance, if attention involves altered inter-areal communication, this could be achieved by better alignment between population responses in a source area and the communication subspace relaying those responses to a relevant downstream area. Similarly, learning could involve achieving population activity patterns that are better aligned with an existing communication subspace[6], or perhaps altering the communication subspace itself. Finally, the degree to which the effects of perturbation experiments (e.g., patterned optogenetic stimulation) would propagate across areas would depend on their alignment with the relevant communication subspaces. As a result, studying how experimental manipulations alter population responses within a given cortical area in isolation can be misleading. It is crucial to understand how these altered population responses interact with the mapping to downstream areas.

5.5 Methods

5.5.1 Removing activity along the predictive dimensions

In order to remove the source population activity along the predictive dimensions, we projected the source activity onto the subspace that is uncorrelated with the predictive dimensions. Formally, we state that two dimensions defined by the vectors $\mathbf{u}$ and $\mathbf{v}$ are uncorrelated with respect to the source activity matrix X if:

$$\mathbf{u}^T \Sigma \mathbf{v} = 0$$

where Σ is the covariance matrix of the source activity. Let matrix $\bar{B}$ contain the predictive dimensions. The set of vectors in the uncorrelated subspace is:

$$\{\mathbf{v} : \bar{B}^T \Sigma \mathbf{v} = 0\}$$

In particular, it will be useful to find an orthonormal basis for this subspace:

$$\{Q : \bar{B}^T \Sigma Q = 0,\, Q^T Q = I\}$$

This can be accomplished using the singular value decomposition (SVD). Start by defining $M = \bar{B}^T \Sigma$ and consider its SVD $M = UDV^T$. Choosing Q as the last $p-m$ columns of V (corresponding to the 0 singular values) yields $MQ = 0$, $Q^T Q = I$, which makes Q an orthonormal basis for the uncorrelated subspace. We then projected the source population onto the uncorrelated subspace, $\hat{X} = XQ$, and predicted target activity using ridge regression between $\hat{X}$ and Y.

5.5.2 Comparing dominant and predictive dimensions

To identify the dominant dimensions in the source population, we fit a FA model, and determined the optimal dimensionality (as described above). Using this FA model, we estimated the latent variables $\hat{\mathbf{z}} = \mathbb{E}[\mathbf{z}|\mathbf{y}]$ for each $\mathbf{z}$, then performed an orthonormalization procedure to order the elements of $\hat{\mathbf{z}}$ by the amount of shared variance explained[39]. This allowed us to predict the target population activity using only the most dominant V1 dimension (first element of orthonormalized $\hat{\mathbf{z}}$), the top two most dominant V1 dimensions (first two elements of orthonormalized $\hat{\mathbf{z}}$), etc. We then compared the performance of the dominant and predictive dimensions for predicting activity of the target populations.

Chapter 6

Dynamics of inter-areal interactions

Complex tasks frequently require integrating information from multiple modalities. This information informs high-level decisions and eventually culminates in the generation of adequate motor commands. This process is not instantaneous, and information travels, in parallel or sequentially, through a number of brain areas as it is processed and transformed, until it ultimately impacts behavior. As a result, this processing pathway creates rich temporal dependencies in the activity of the areas involved. Furthermore, these interactions must also be flexible, changing rapidly under changing stimuli or task conditions. Attempts to study inter-area interaction therefore frequently focus on these temporal aspects, quantifying, for example, correlation[8–13,70] and/or coherence[10,16–22] between the activity in two areas during different task contexts.

To elucidate the interplay between interaction structure and dynamics, we studied the same recordings as in Chapters 4 and 5, both during evoked and spontaneous activity (the inter-trial period). Using related statistical methods, we characterized inter-area interactions under these two conditions, at multiple time scales. We found that the population-level correlation between these areas was higher for spontaneous activity, while the temporal structure of the interaction changed from being feedforward dominated (V1 leading V2) early in the stimulus period to feedback dominated (V2 leading V1) late in the stimulus period and during spontaneous activity. We analyzed the population-level structure of V1-V2 interactions and found that the dimensions of the V1 population activity that were involved for spontaneous activity were distinct from those involved for evoked activity.

These results suggest that population-level interactions between V1 and V2 are dynamic and

flexible – even for the simplest visual stimuli – and rich dynamics might mediate the way in which information is relayed between these areas. In order to directly capture inter-area interaction dynamics, we developed a novel method, termed Group Latent Auto-Regressive Analysis (gLARA). This method extracts a set of latent variables (where the number of latent variables, M, is smaller than the number of recorded neurons, N) for each neuronal population (it is also applicable to more than two populations) and approximates intra- and inter-area auto-regressive linear dynamics in these latent spaces.

6.1 Strength of inter-area interactions across time

A common way to characterize the temporal dependence of two signals is to compute a cross-correlogram. This pairwise correlation-based approach has seen widespread use in the systems neuroscience field, and has provided important insights. However, being a univariate method, it cannot be used in a straightforward manner to study population-level interactions. In other words, given the activity recorded from two populations of neurons, it is not obvious how to best leverage the cross-correlograms based on pairwise correlations (one for each pair of neurons in each area) to extract a single summary of the way in which the two populations of neurons are related. Here, we turned to Canonical Correlation Analysis (CCA), which extracts a pair of dimensions, one for each population, for which the correponding projections are maximally correlated. In its simplest interpretation, given the population activity in two neuronal populations, CCA returns the maximum correlation between any pair of projections, which we term the canonical correlation between the two populations. This essentially allows us to compute population-based, or multivariate, cross-correlograms.

6.1.1 V1-V2 interaction strength changes within and across trial epochs both in overall magnitude and temporal structure

We were interested in studying not just how the interaction between V1 and V2 is temporally structured (e.g., what is the lag between the two areas), but also how this temporal structure evolves through time, and across stimulus conditions. To achieve this, we applied CCA to fine timescale (1 ms resolution; 100 ms time windows) residual activity, obtained by subtracting the

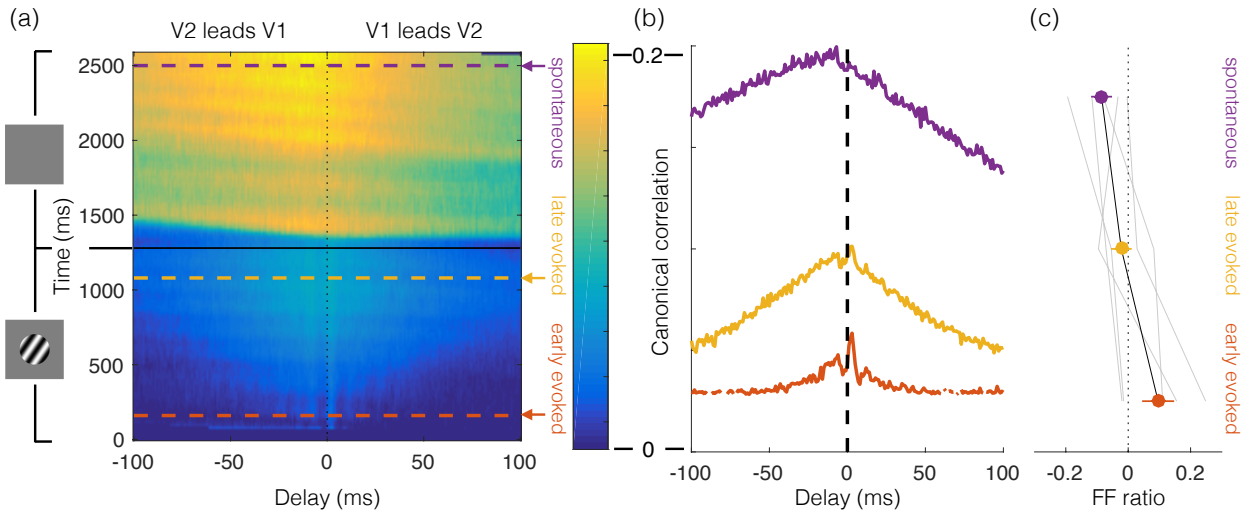

Figure 6.1: **V1-V2 interaction strength changes within and across trial epochs both in overall magnitude and temporal structure. (a)** V1-V2 across-area correlation at different time points and time lags, assessed using CCA, for an example recording session. **(b)** Three horizontal slices of the correlation map shown in (a), for early evoked activity (red), late evoked activity (yellow) and spontaneous activity (purple). **(c)** Summary across all five recording sessions. Feedforward ratio is given by the area of the difference between the positive and negative delay curves, normalized by the positive (feedforward) delay curve. Faded lines show the feedforward ratios for each of the 5 recording sessions separately. Error bars indicate SEM.

corresponding stimulus PSTHs from responses to each stimulus condition and then pooling across conditions. The resultant canonical correlation map (CCM; Fig. 6.1a) shows the maximal (canonical) correlation between the two areas, for each time point (relative to stimulus onset) and time lag. The CCM can be interpreted as a multivariate (population-level) extension of pairwise correlation. In particular, each horizontal slice through the CCM (Fig. 6.1b) is akin to a population cross-correlogram, positioned at a time point in the trial.

We found that V1-V2 canonical correlation is higher for spontaneous activity (Fig. 6.1a, intertrial period, from 1280 ms onwards) than for evoked activity (Fig. 6.1a, stimulus presentation period, 0-1280 ms). Furthermore, the temporal structure of the V1-V2 canonical correlation evolves from feedforward (V1 leading V2) early in the stimulus period (Fig. 6.1b, red trace; indicated by the red dashed line in Fig. 6.1a; 160 ms after stimulus onset), where a clear feedforward peak can be observed at 2 ms (2 ± 0.55 ms across all 5 recording sessions), to "broad" feedback (V2 leading V1) during the intertrial period (Fig. 6.1b, purple trace; indicated by the purple dashed line in Fig. 6.1a; 2500 ms after stimulus onset/1220 ms after stimulus offset). V1-V2 canonical correlation

also increases throughout the stimulus presentation period, while the initial feedforward peak becomes less pronounced (compare red and yellow cross-correlograms in Fig. 6.1b; mean peak canonical correlation across all sessions: 0.067 ± 0.008 for early evoked, 0.099 ± 0.010 for late evoked and 0.156 ± 0.015 for spontaneous). These effects were observed across all 5 recording sessions (Fig. 6.1c).

6.2 Structure of inter-area interactions across time

We then asked whether the dynamics of V1-V2 interactions (Fig. 6.1) also involve a change in the dimensions relating activity in the two areas (i.e., a change in interaction structure). We did this using the same approach employed to compare predictive dimensions across stimuli (Section 4.2.3).

6.2.1 V1-V2 interaction structure is distinct during evoked and spontaneous activity

We took 1 s of evoked and spontaneous activity, starting 160 ms after stimulus onset and offset, respectively, and binned spike counts in 100ms windows (as in previous Chapters). We then fit reduced rank regression models (specifically, ridge reduced rank regression; see Section 6.5) to 200 ms time windows throughout the evoke and spontaneous activity periods. We opted to study the structure of inter-area interactions using RRR because this method makes it easier to estimate the number of predictive dimensions necessary, and is easier to regularize in order to prevent overfitting (see Section 6.5).

The communication subspaces identified by these models are then used to predict activity for all other time windows, both in the evoked and the spontaneous epochs (see Section 4.2.3 for more details). If the communication subspaces were entirely distinct for different time windows, performance should plummet when using the communication subspace of one time window to predict another. Alternatively, if the communication subspaces are similar across time windows, performance should be similar regardless of which responses are used to define the subspace.

We found that while the communication subspaces generalized well within the evoked and spontaneous activity epochs, they did not generalize as well between the two epochs (Fig. 6.2a). In particular, communication subspaces identified for spontaneous activity generalized less well to

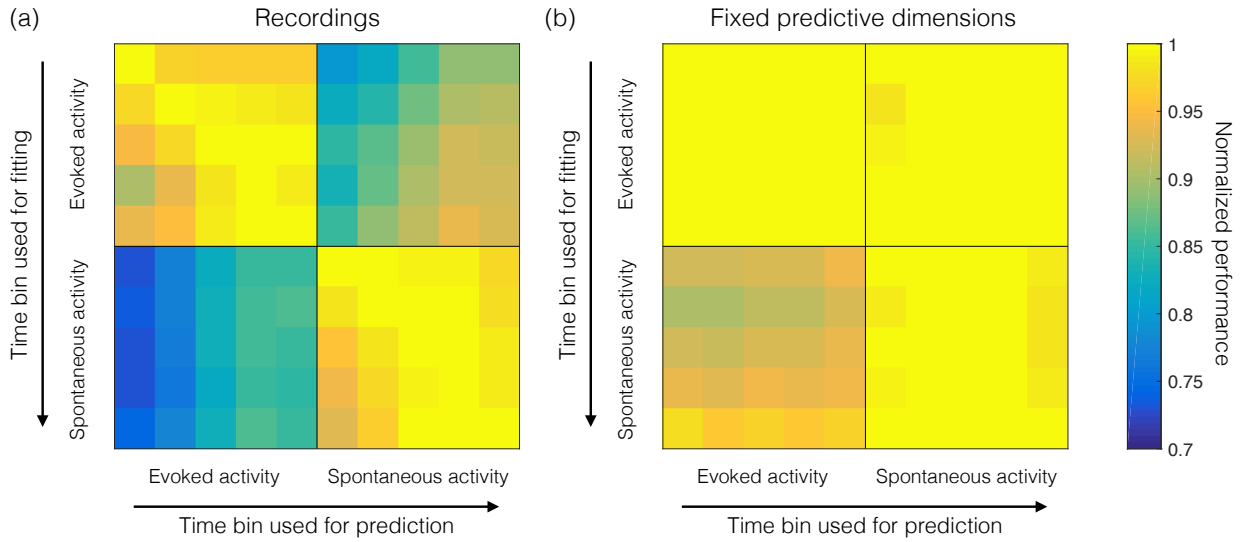

Figure 6.2: **V1-V2 interaction structure is distinct during evoked and spontaneous activity.** **(a)** Communication subspace changes for different task epochs. Each row corresponds to a different communication subspace and each column to the application of those subspaces to a different time window. The diagonal elements thus indicate the normalized performance of identifying and applying the communication subspace to responses evoked by the same stimuli. Off-diagonal elements can have normalized predictive performance values less than 1, indicating that the communication subspaces are not identical across time windows. Performance was measured using 10-fold cross-validation (i.e., for each fold, the model was fit to a training set pertaining to one time window and then used for prediction in test sets of all time windows). Each entry shows the average normalized performance across all recording sessions. **(b)** The change in the identified predictive dimensions cannot be attributed to differences in the source population responses across time. Same conventions as in (a).

early evoked activity (Fig. 6.2a, bottom-left block, first column; average normalized performance, across all recording sessions and spontaneous activity communication subspaces: 0.73 ± 0.03) than to late evoked activity (Fig. 6.2a, bottom-left block, fifth column; 0.85 ± 0.01). Conversely, communication subspaces fit to late evoked activity (top-right block, last row; average normalized performance, across all recording sessions and spontaneous activity time windows: 0.90 ± 0.01) capture spontaneous interactions better than those fit to early evoked activity (top-right block, first row; average normalized performance, across all recording sessions and spontaneous activity time windows: 0.85 ± 0.02).

This is consistent with the differences in temporal structure found in the previous section: early evoked activity is feedforward dominated, while late evoked activity is less so, and seems to display signatures of both early evoked activity and spontaneous activity (Fig. 6.1b-c).

For comparison, we applied the same analyses to data synthesized as in Section 4.1.2. Briefly, we first defined a fixed linear mapping, using the residual responses across all time windows. We then generated target population activity for each time window by passing the corresponding source population activity thought this fixed mapping, and added Poisson noise to each sample with mean given by the corresponding PSTH time bin. For these synthesized responses, we found that performance was substantially higher, compared to the recorded activity, when we identified the predictive dimensions using responses from one time window and applied them to responses from another (Fig. 6.2b). This analysis indicates that the performance decrement in the physiological data (Fig. 6.2a) cannot be solely attributed to differences in the source population responses across time.

One interesting feature on Fig. 6.2a is the asymmetry between how well evoked models generalize to spontaneous activity windows and spontaneous models generalize to evoked activity windows. One possible explanation is that evoked communication subspaces might involve a higher number of predictive dimensions than the spontaneous communication subspaces. The larger the communication subspace, the less restrictive it is (in the extreme, a communication subspace that involves as many predictive dimensions as the number of V1 neurons will always perfectly generalize, since no activity is lost by projecting onto it).

Indeed, we found that evoked communication subspaces involved more predictive dimensions than the communication subspaces found for spontaneous activity. (Fig. 6.3a, red traces; average number of predictive dimensions across all evoked time windows and recording sessions: 3.64 ± 0.19; 2.68 ± 0.27 for spontaneous time windows; Monte Carlo permutation test: $p < 10^{-2}$). However, the same was true for the surrogate data (Fig. 6.3a, black traces; number of predictive dimensions: 3.64 ± 0.19 for evoked; 2.68 ± 0.27 for spontaneous; Monte Carlo permutation test: $p < 10^{-2}$), where the linear mapping is the same of all time windows (and therefore the true number of predictive dimensions is also the same). This can be observed in Fig. 6.2b: the communication subspaces identified for the evoked epoch fully generalize to the spontaneous epoch, but the the communication subspaces identified for the spontaneous epoch generalize less well (compare top-right and bottom-left blocks), even though the linear mapping is the same for all time windows. Given that we showed in Section 4.2.3 that we could recover the correct number of predictive dimensions in the evoked period, these results suggest that we might be underestimat-

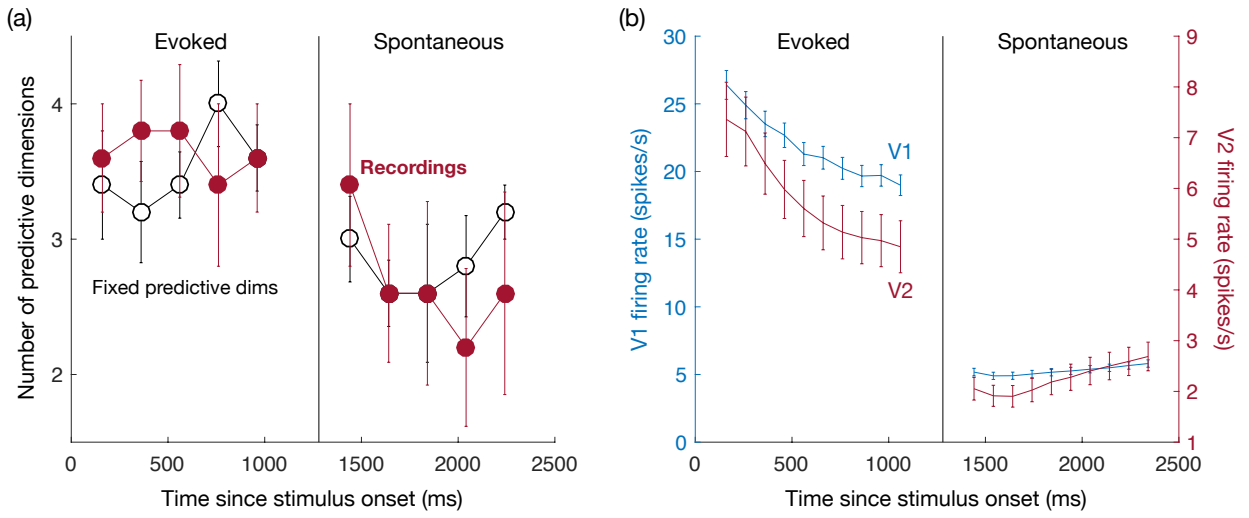

Figure 6.3: **Differences in the number of predictive dimensions between evoked and spontaneous activity can be explained by differences in the source and target population statistics.** **(a)** Number of predictive dimensions identified for evoked activity was higher than for spontaneous activity. This effect was also observed for the surrogate data, where the linear mapping was the same for both epochs, suggesting this difference is caused by differences in the V1/V2 population activity structures, and not actual differences in the size of the communication subspaces. Traces indicate average number of predictive dimensions across the five recording session. Error bars indicate SEM. **(b)** The smaller number of predictive dimensions identified for spontaneous activity could be a reflection of much lower firing rates during this period. Traces indicate average firing rate across all neurons in the five recording sessions. Error bars indicate SEM.

ing the number of predictive dimensions involved in spontaneous interactions. One factor that could account for these differences in the number of estimated predictive dimensions is the overall firing rates in the V1 and V2 populations. When using linear models to predict spike counts, it is generally the case that higher firing rates lead to better fits. Indeed, both V1 and V2 firing rates are much lower during the spontaneous activity period (Fig. 6.3b).

Note that while firing rate differences across the two epochs might offer an explanation for differences in the estimated number of predictive dimensions, and the asymmetry in the generalization between the two periods, they cannot account for the overall lack of generalization across epochs: Fig. 6.2b shows that under a fixed mapping the evoked communication subspaces generalized perfectly to the spontaneous period, which is not true for the recordings.

One simple way to account for the difference in the number of predictive dimensions when quantifying the across-epoch generalization is to fix all communication subspaces to have the same dimensionality. Given that all identified communication subspaces contain at least one predictive

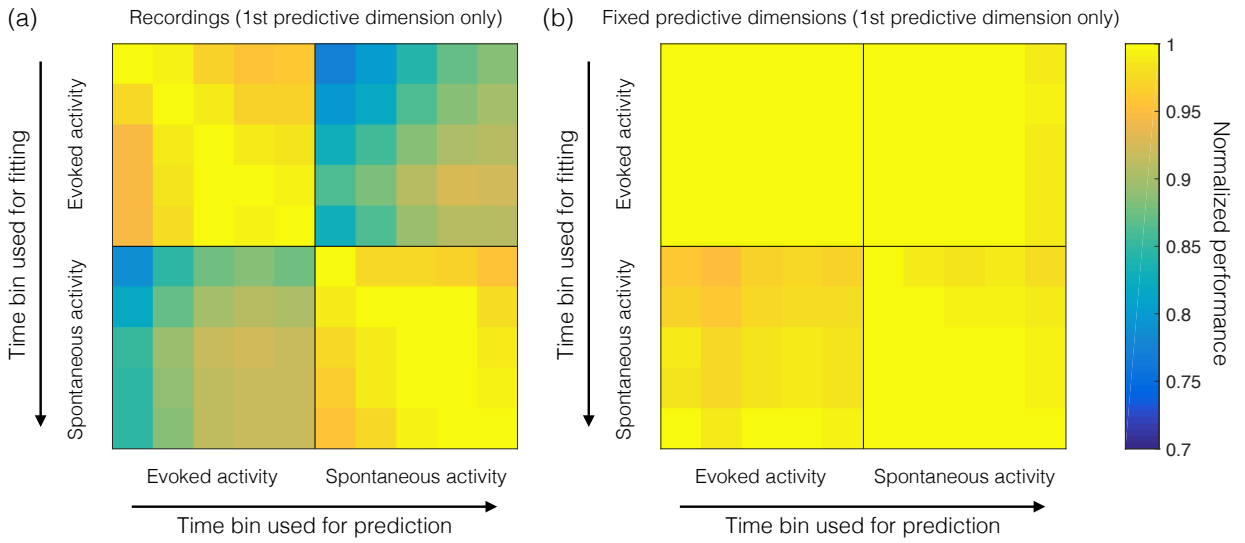

Figure 6.4: **Top predictive dimension is distinct across evoked and spontaneous epochs.** **(a)** First predictive dimenision changes for different task epochs. **(b)** The change in the first predictive dimension cannot be attributed to differences in the source population responses across time. **(a-b)** Same conventions as in Fig. 6.2.

dimension, we can repeat the analysis shown in Fig. 6.2 using only the top predictive dimension (Fig. 6.4). Doing so revealed much the same structure that can be observed in Fig. 6.2, the main difference being that the overall asymmetry in across-epoch generalization is now largely absent. Importantly, it is still the case that the communication subspaces identified for spontaneous activity generalize better for late evoked activity than for early evoked activity.

Having observed a difference between the predictive dimensions for evoked and spontaneous activity, and that during evoked activity V1-V2 predictive dimensions were distinct from source V1 dominant dimensions (see Section 5.1.2), we wondered if the same was true for spontaneous activity. We found that this was not the case: predictive dimensions had a similar predictive performance over V2 as source V1 dominant dimensions (Fig. 6.5; compare with Fig. 5.2). This suggests that not only are predictive dimensions distinct between evoked and spontaneous activity, their relationship to the source population activity is also altered.

In summary, we found that the top V1-V2 predictive dimensions involved during evoked activity are distinct from those involved during spontaneous activity. Furthermore, predictive dimensions identified for spontaneous activity were more similar to those identified for late evoked activity than to the predictive dimensions identified earlier in the evoked epoch. Conversely, pre-

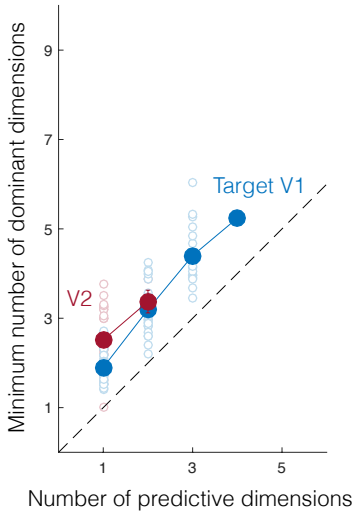

Figure 6.5: **Spontaneous activity predictive dimensions are well aligned with the dominant modes of spontaneous source V1 activity.** For a given number of predictive dimensions, a similar number of dominant dimensions was required to reach (within a S.E.M., across folds) the same V2 predictive performance (red circles). The same was true for the target V1 population (blue circles). Error bars indicate S.E.M. across datasets. Faded circles show results for each data set. This analysis was performed in the same way as in Fig. 5.2, with the exception that it was applied to the spontaneous epoch.

dictive dimensions fit to late evoked activity were better at capturing spontaneous interactions than those fit to early evoked activity. In the previous section we had found that the temporal signature of the V1-V2 interaction changes across time and trial epochs. Here we find that these changes are accompanied by changes in the very structure of the interaction, such that the patterns of activity shared across these areas are modified as well.

6.3 Capturing intra- and inter- area dynamics: group latent auto-regressive analysis

The results in the previous sections suggest that both the strength and the structure of the V1-V2 interaction exhibit interesting temporal dynamics. However, the approaches taken thus far relied on fitting interaction models (be it CCA or RRR) separately to different temporal epochs in the trial. Here, we propose a novel latent variable model, termed group latent auto-regressive analysis (gLARA), that can explicitly, and simultaneously, capture intra- and inter- area dynamics

at various time delays.

We begin by considering canonical correlations analysis (CCA), specifically its probabilistic formulation (pCCA)[71], which identifies a single set of latent variables that explicitly captures the between-population correlation structure. To understand how the different neural populations interact on different timescales, we propose extensions of pCCA that introduce a separate set of latent variables for each neuronal population, as well as dynamics on the latent variables to describe their interaction over time. We then apply the proposed methods to populations of neurons recorded simultaneously in visual areas V1 and V2 to demonstrate their utility.

6.3.1 Methods

We consider the setting where many neurons are recorded simultaneously, and the neurons belong to distinct populations (either by brain area or by neuron type). Let $\mathbf{y}_t^i \in \mathbb{R}^{q_i}$ represent the observed activity vector of population $i \in \{1, ..., M\}$ at time $t \in \{1, ..., T\}$, where q_i denotes the number of neurons in population i. Below, we consider three different ways to study the interaction between the neural populations. To keep the notation simple, we'll only consider two populations ($M = 2$); the extension to more than two populations is straightforward.

Factor analysis and probabilistic canonical correlation analysis

Consider the following latent variable model, that defines a linear-Gaussian relationship between the observed variables, $\mathbf{y}_t^1$ and $\mathbf{y}_t^2$, and the latent state, $\mathbf{x}_t \in \mathbb{R}^p$:

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{0}, I) \tag{6.1}$$

$$\begin{bmatrix} \mathbf{y}_t^1 \\ \mathbf{y}_t^2 \end{bmatrix} \mid \mathbf{x}_t \sim \mathcal{N}\left(\begin{bmatrix} C^1 \\ C^2 \end{bmatrix} \mathbf{x}_t + \begin{bmatrix} \mathbf{d}^1 \\ \mathbf{d}^2 \end{bmatrix}, \begin{bmatrix} R^{11} & R^{12} \\ R^{12^T} & R^{22} \end{bmatrix} \right) \tag{6.2}$$

where $C^i \in \mathbb{R}^{q_i \times p}$, $\mathbf{d}^i \in \mathbb{R}^{q_i}$ and:

$$\begin{bmatrix} R^{11} & R^{12} \\ R^{12^T} & R^{22} \end{bmatrix} \in \mathbb{S}_{++}^q$$

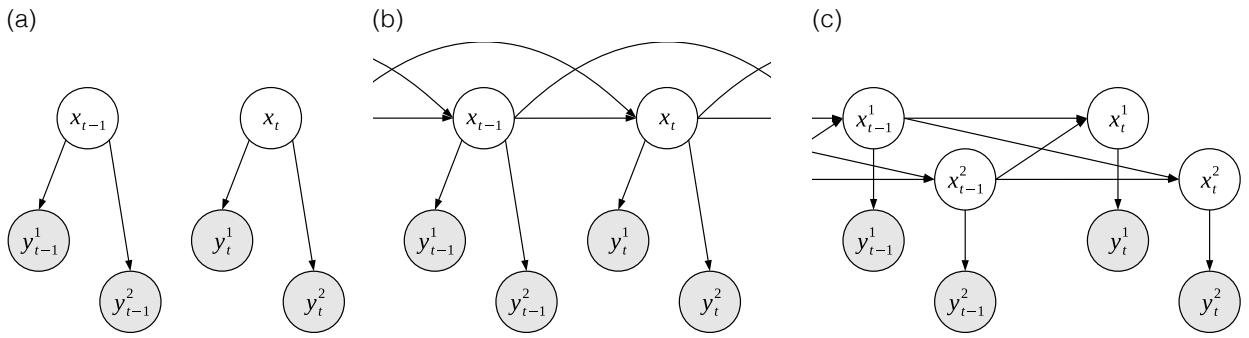

Figure 6.6: **Directed graphical models for multi-population activity. (a)** Probabilistic canonical correlation analysis (pCCA). **(b)** pCCA with auto-regressive latent dynamics (AR-pCCA). **(c)** Group latent auto-regressive analysis (gLARA). For clarity, we show only two populations in each panel and auto-regressive dynamics of order 1 in panel (c).

with $q = q_1 + q_2$. According to this model, the covariance of the observed variables is given by:

$$\text{cov}\left(\begin{bmatrix} \mathbf{y}_t^1 \\ \mathbf{y}_t^2 \end{bmatrix}\right) = \begin{bmatrix} C^1 \\ C^2 \end{bmatrix} \begin{bmatrix} C^{1T} & C^{2T} \end{bmatrix} + \begin{bmatrix} R^{11} & R^{12} \\ R^{12T} & R^{22} \end{bmatrix} \tag{6.3}$$

Factor analysis (FA) and probabilistic canonical correlation analysis (pCCA) can be seen as two special cases of the general model presented above. FA assumes the noise covariance to be diagonal, i.e., $R^{11} = \text{diag}(r_1^1, ..., r_{q_1}^1)$, $R^{22} = \text{diag}(r_1^2, ..., r_{q_2}^2)$ and $R^{12} = \mathbf{0}$. This noise covariance captures only the independent variance of each neuron, and not the covariance between neurons. As a result, the covariance between neurons is explained by the latent state through the observation matrices C^1 and C^2. pCCA, on the other hand, considers a block diagonal noise covariance, i.e., $R^{12} = \mathbf{0}$. This noise covariance accounts for the covariance observed between neurons in the same population. The latent state is therefore only used to explain the covariance between neurons in different populations. The directed graphical model for pCCA is shown in Fig. 6.6a.

Auto-regressive probabilistic canonical correlation analysis (AR-pCCA)

While pCCA offers a succinct picture of the covariance structure between populations of neurons, it does not capture any temporal structure. There are two main reasons as to why this time structure may be interesting. First, pCCA is modelling the covariance structure at zero time lag, which may not capture all of the interactions of interest. If the two populations of neurons correspond to

two different brain areas, there may be important interactions at non-zero time lags due to physical delays in information transmission. Second, the two populations of neurons may interact at more than one time delay, for example if multiple pathways exist between the neurons in these populations. To take the temporal structure into account we will first extend pCCA by defining an auto-regressive linear-Gaussian model on the latent state:

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{0}, I), \quad \text{if } 1 \leq t \leq \tau \tag{6.4}$$

$$\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, ..., \mathbf{x}_{t-\tau} \sim \mathcal{N}\left(\sum_{k=1}^{\tau} A_k \mathbf{x}_{t-k}, Q\right), \quad \text{if } t > \tau \tag{6.5}$$

where $A_k \in \mathbb{R}^{p \times p}$, $\forall k$, $Q \in \mathbb{S}^p_{++}$ and τ denotes the order of the autoregressive model. We term this model AR-pCCA, which is defined by the state model in Eq.(6.4)-(6.5) and the observation model in Eq.(6.2) with $R^{12} = \mathbf{0}$. Although the observation model is the same as that for pCCA, the latent state here accounts for temporal dynamics, as well as the covariation structure between the populations. The corresponding directed graphical model is shown in Fig. 6.6b.

Group latent auto-regressive analysis (gLARA)

According to AR-pCCA, a single latent state drives the observed activity in both areas. As a result, it's not possible to distinguish the within-population dynamics from the between-population interactions. To allow for this, we propose using two separate latent states, or one per population, that interact over time. We refer to the proposed model as group latent auto-regressive analysis (gLARA):

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{0}, I), \quad \text{if } 1 \leq t \leq \tau \tag{6.6}$$

$$\mathbf{x}^i_t \mid \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, ..., \mathbf{x}_{t-\tau} \sim \mathcal{N}\left(\sum_{j=1}^{2} \sum_{k=1}^{\tau} A^{ij}_k \mathbf{x}^j_{t-k}, Q^i\right), \quad \text{if } t > \tau \tag{6.7}$$

$$\begin{bmatrix} \mathbf{y}^1_t \\ \mathbf{y}^2_t \end{bmatrix} \mid \mathbf{x}_t \sim \mathcal{N}\left(\begin{bmatrix} C^1 & \mathbf{0} \\ \mathbf{0} & C^2 \end{bmatrix} \begin{bmatrix} \mathbf{x}^1_t \\ \mathbf{x}^2_t \end{bmatrix} + \begin{bmatrix} \mathbf{d}^1 \\ \mathbf{d}^2 \end{bmatrix}, \begin{bmatrix} R^1 & \mathbf{0} \\ \mathbf{0} & R^2 \end{bmatrix}\right) \tag{6.8}$$

where $\mathbf{x}_t$ is obtained by stacking $\mathbf{x}^1_t \in \mathbb{R}^{p_1}$ and $\mathbf{x}^2_t \in \mathbb{R}^{p_2}$, the latent states for each population, $C^i \in \mathbb{R}^{q_i \times p_i}$, $A^{ij}_k \in \mathbb{R}^{p_i \times p_j}$ and $Q^i \in \mathbb{S}^{p_i}_{++}$, $\forall k$ and $i \in \{1, 2\}$. Note that the covariance structure

observed on a population level now has to be completely reflected by the latent states (there are no shared latent variables in this model) and is therefore defined by the dynamics matrices A_k^{ij}, allowing for the separation of the within-population dynamics (A_k^{11} and A_k^{22}) and the between-population interactions (A_k^{12} and A_k^{21}). Furthermore, the interaction between the populations is asymmetrically defined by A_k^{12} and A_k^{21}, allowing for a more in depth study of the way in each the two areas interact by comparing these across the various time delays considered. Note that gLARA represents a special case of the AR-pCCA model. The corresponding directed graphical model is shown in Fig. 6.6c.

Parameter estimation for gLARA

The parameters of gLARA can be fit to the training data using the expectation-maximization (EM) algorithm. To do so, we start by defining the augmented latent state $\bar{\mathbf{x}}_t \in \mathbb{R}^{p\tau}$, with $p = p_1 + p_2$:

$$\bar{\mathbf{x}}_t = \begin{bmatrix} \bar{\mathbf{x}}_t^1 \\ \bar{\mathbf{x}}_t^2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_t^{1T} \dots & \mathbf{x}_{t-\tau}^{1\ T} & \mathbf{x}_t^{2T} \dots & \mathbf{x}_{t-\tau}^{2\ T} \end{bmatrix}^T \tag{6.9}$$

and the augmented observation vector $\bar{\mathbf{y}}_t \in \mathbb{R}^q$, with $q = q_1 + q_2$:

$$\bar{\mathbf{y}}_t = \begin{bmatrix} \mathbf{y}_t^{1T} & \mathbf{y}_t^{2T} \end{bmatrix}^T \tag{6.10}$$

for $t \in \{\tau, ..., T\}$. Using the augmented latent state $\bar{\mathbf{x}}$, the dynamics equation (Eq.(6.6) and (6.7)) can be rewritten as:

$$\bar{\mathbf{x}}_t \sim \mathcal{N}(\mathbf{0}, I), \quad \text{if } t = \tau \tag{6.11}$$

$$\bar{\mathbf{x}}_t \mid \bar{\mathbf{x}}_{t-1} \sim \mathcal{N}\left(\bar{A}\bar{\mathbf{x}}_{t-1}, \bar{Q}\right), \quad \text{if } t > \tau \tag{6.12}$$

for appropriately structured $\bar{A} \in \mathbb{R}^{p\tau \times p\tau}$ and $\bar{Q} \in \mathbb{S}_{++}^{p\tau}$. The observation model (Eq.(6.8)) can be rewritten as:

$$\bar{\mathbf{y}}_t \mid \bar{\mathbf{x}}_t \sim \mathcal{N}\left(\bar{C}\begin{bmatrix} \bar{\mathbf{x}}_t \\ 1 \end{bmatrix}, \bar{R}\right) \tag{6.13}$$

for appropriately structured $\bar{C} \in \mathbb{R}^{q \times (p\tau+1)}$ and $\bar{R} \in \mathbb{S}_{++}^q$.

We fit the model parameters using the EM algorithm. In the E-step, because the latent and

observed variables are jointly Gaussian, $P(\bar{\mathbf{x}}_t \mid \bar{\mathbf{y}}_1, ..., \bar{\mathbf{y}}_T)$ is also Gaussian and can be computed exactly by applying the forward-backward recursion of the Kalman smoother[72] on the augmented vectors. In the M-step, we directly estimate the original parameters $\theta = \{C^i, \mathbf{d}^i, R^i, A_k^{ij}\}$, as opposed to estimating the structured form of the augmented parameters $\bar{\theta} = \{\bar{C}, \bar{R}, \bar{A}\}$ (without loss of generality, we set $Q^i = I$):

$$\begin{bmatrix} C^i & \mathbf{d}^i \end{bmatrix} = \left(\sum_{t=1}^{T} \mathbf{y}_t^i \begin{bmatrix} \mathbb{E}(\mathbf{x}_t^{iT}) & 1 \end{bmatrix} \right) \left(\sum_{t=1}^{T} \begin{bmatrix} \mathbb{E}(\mathbf{x}_t^i \mathbf{x}_t^{iT}) & \mathbb{E}(\mathbf{x}_t^i) \\ \mathbb{E}(\mathbf{x}_t^{iT}) & 1 \end{bmatrix} \right)^{-1} \tag{6.14}$$

$$\begin{aligned} R^i = \frac{1}{T} \sum_{t=1}^{T} \{ & (\mathbf{y}_t^i - \mathbf{d}^i)(\mathbf{y}_t^{iT} - \mathbf{d}^i) - C^i \mathbb{E}(\mathbf{x}_t^i)(\mathbf{y}_t^i - \mathbf{d}^i)^T \\ & - (\mathbf{y}_t^i - \mathbf{d}^i) \mathbb{E}(\mathbf{x}_t^{iT}) C^{iT} + C^i \mathbb{E}(\mathbf{x}_t^i \mathbf{x}_t^{iT}) C^{iT} \} \end{aligned} \tag{6.15}$$

$$\begin{bmatrix} A_1^{11} & \dots & A_k^{11} & A_1^{12} & \dots & A_k^{12} \\ A_1^{21} & \dots & A_k^{21} & A_1^{22} & \dots & A_k^{22} \end{bmatrix} = \left(\sum_{t=2}^{T} \mathbb{E}\left(\bar{\mathbf{x}}_t \bar{\mathbf{x}}_{t-1}^T \right) \right) \left(\sum_{t=2}^{T} \mathbb{E}\left(\bar{\mathbf{x}}_{t-1} \bar{\mathbf{x}}_{t-1}^T \right) \right)^{-1} \tag{6.16}$$

To initialize the EM algorithm, we start by applying FA to each population individually, and use the estimated observation matrices C^1 and C^2, as well as the mean vectors $\mathbf{d}^1$ and $\mathbf{d}^2$ and the observation covariance matrices R^{11} and R^{22}. The A_k^{ij} matrices are initialized at $\mathbf{0}$.

6.3.2 Model validation

To validate the methods described above, we applied them to one of the V1-V2 recording sessions (see Chapter 3 for further details). In particular, we used 1.23s of data in each trial, from 50ms after stimulus onset until the end of the evoked response, and proceeded to bin the observed spikes with a 5ms window. This recording session included a total of 97 units in V1 and 31 units in V2 (single- and multi-units). We chose to analyze a subset of the trials for rapid iteration of the analyses, as the cross-validation procedure is computationally expensive for the full dataset. Given that 1000 trials provides a total of 246,000 timepoints (at 5 ms resolution), this provides a reasonable amount of data to fit any of the models with the 128 observed neurons. For model comparison, we performed 4-fold cross-validation, splitting the data into four non-overlapping

test folds with 250 trials each.

We again sought to investigate how trial-to-trial population variability in V1 relates to the trial-to-trial population variability in V2. For these gratings stimuli (which are relatively simple compared to naturalistic stimuli[49]), there is likely richer structure in the V1-V2 interaction for the trial-to-trial variability than for the stimulus drive. To this end, we subtracted the appropriate PSTH from the binned spike counts on each trial to obtain a single-trial "residual". The residuals across all neurons and conditions were considered together in the analyses shown in Section 6.3.2. Note that the methods considered in this study could also be applied to the PSTHs of sequentially recorded neurons in multiple areas.

We started by asking how many dimensions are needed to describe the between-population covariance, relative to the number of dimensions needed to describe the within-population covariance. This was assessed by applying pCCA to the labeled V1 and V2 populations, as well as FA to the two populations together (which ignores the V1 and V2 labels). In this analysis, pCCA captures only the between-population covariance, whereas FA captures both the between-population and within-population covariance. By comparing cross-validated data likelihoods for different dimensionalities, we found that pCCA required three latent dimensions, whereas FA required 40 latent dimensions (Fig. 6.7). This indicates that the zero time lag interaction between V1 and V2 is confined to a small number of dimensions (three) relative to the number of dimensions (40) needed to describe all covariance among the neurons. The difference of these two dimensionalities (37) describes covariance that is 'private' to each population (i.e., within-population covariance). The FA and pCCA curves peak at similar cross-validated likelihoods in Fig. 6.7 because the observation model for pCCA Eq.(6.2) accounts for the within-population covariance (which is not captured by the pCCA latents).

The distinction between within-population covariance and between-population covariance is further supported by re-applying pCCA, but now randomly shuffling the population labels. The cross-validated log-likelihood curve for these mixed populations now peaks at a larger dimensionality than three. The reason is that the shuffling procedure removes the distinction between the two types of covariance, such that the pCCA latents now capture both types of covariance (of the original unmixed populations). The peak for mixed pCCA occurs at a lower dimensionality than for FA for two reasons: i) because the mixed populations have the same number of neurons as the

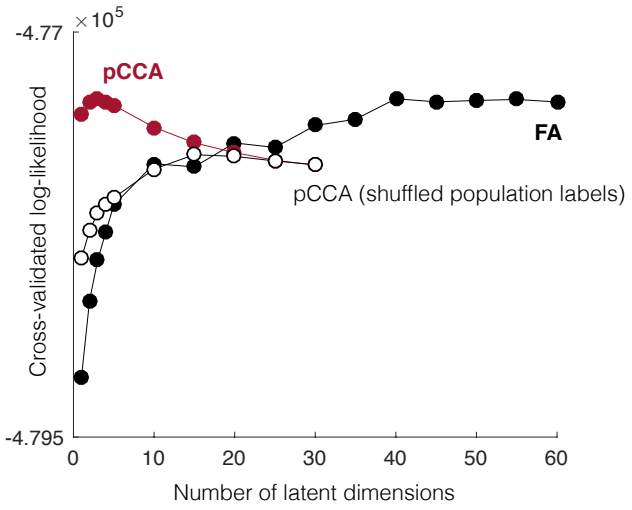

Figure 6.7: **Comparing the optimal dimensionality for FA and pCCA.** Cross-validated log-likelihood plotted as a function of the dimensionality of the latent state for FA (black) and pCCA (red). pCCA was also applied to the same data after randomly shuffling the population labels (white). Note that the maximum possible dimensionality for pCCA is 31, which is the size of the smaller of the two populations (in this case, V2).

original populations (97 and 31), the maximum number of dimensions that can be identified by pCCA is 31, and ii) for the same latent dimensionality, pCCA has a larger number of parameters than FA, which makes pCCA more prone to overfitting.

Together, the analyses in Fig. 6.7 demonstrate two key points. First, if the focus of the analysis lies in the interaction between populations, then pCCA provides a more parsimonious description, as it focuses exclusively on the covariance between populations. In contrast, FA is unable to distinguish within-population covariance from between-population covariance. Second, the neuron groupings for V1 and V2 are meaningful, as the number of dimensions needed to describe the covariance between V1 and V2 is small relative to that within each population.

We then analyzed the performance of the models with latent dynamics (AR-pCCA and gLARA). The cross-validated log-likelihood for these models depends jointly on the dimensionality of the latent state, p, and the order of the auto-regressive model, τ. For gLARA, p is the sum of the dimensionalities of each population's latent state, $p_1 + p_2$, and we therefore want to jointly maximize the cross-validated log-likelihood with respect to both p_1 and p_2. AR-pCCA required a latent dimensionality of $p = 70$, while gLARA peaked for a joint latent dimensionality of 65 ($p_1 = 50$ and $p_2 = 15$) (Fig. 6.8a). When computing the performance of AR-pCCA we considered models with

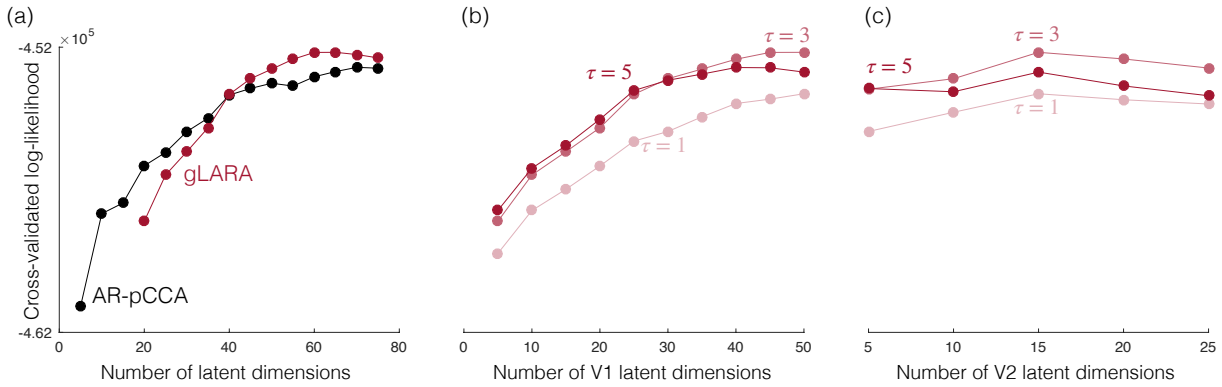

Figure 6.8: **Model selection for AR-pCCA and gLARA. (a)** Comparing AR-pCCA and gLARA as a function of the latent dimensionality (defined as $p_1 + p_2$ for gLARA, where p_2 was fixed at 15), for $\tau = 3$. **(b)** gLARA's cross-validated log-likelihood plotted as a function of the dimensionality of V1's latent state, p_1 (for $p_2 = 15$), for different choices of τ. **(c)** gLARA's cross-validated log-likelihood plotted as a function of the dimensionality of V2's latent state, p_2 (for $p_1 = 50$), for different choices of τ.

$p \in \{5, 10, ..., 75\}$ and $\tau \in \{1, 3, ..., 7\}$ (Fig. 6.8a shows the $\tau = 3$ case). To access how gLARA's cross-validated log-likelihood varied with the latent dimensionalities and the model order, we plotted it in Fig. 6.8b, for $p_2 = 15$ and $p_1 \in \{5, 10, ..., 50\}$, for different choices of τ. This showed that the performance is greater for an order 3 model, and that it saturates by the time p_1 reaches 50. In Fig. 6.8c, we did a similar analysis for the dimensionality of V2's latent state, where p_1 was held constant at 50 and $p_2 \in \{5, 10, ..., 25\}$. The cross-validated log-likelihood shows a clear peak at $p_2 = 15$ regardless of τ. We found that, for both models, the cross-validated log-likelihood peaks for $\tau = 3$ (see Fig. 6.8b and Fig. 6.8c for gLARA, results not shown for AR-pCCA).

Finally, we asked which model, AR-pCCA or gLARA, better describes the data. Note that gLARA is a special case of AR-pCCA, where the observation matrix in Eq.(6.8) is constrained to have a block diagonal structure (with blocks C^1 and C^2). The key difference between the two models is that gLARA assigns a non-overlapping set of latent variables to each population. We found that gLARA outperforms AR-pCCA (Fig. 6.8a). This suggests that the extra flexibility of the AR-pCCA model leads to overfitting and that the data are better explained by considering two separate sets of latent variables that interact.

The optimal latent dimensionalities found for AR-pCCA and gLARA are substantially higher than those found for pCCA, as the latent states now also capture non-zero time lag interactions between the populations, and the dynamics within each population. For gLARA, the between-

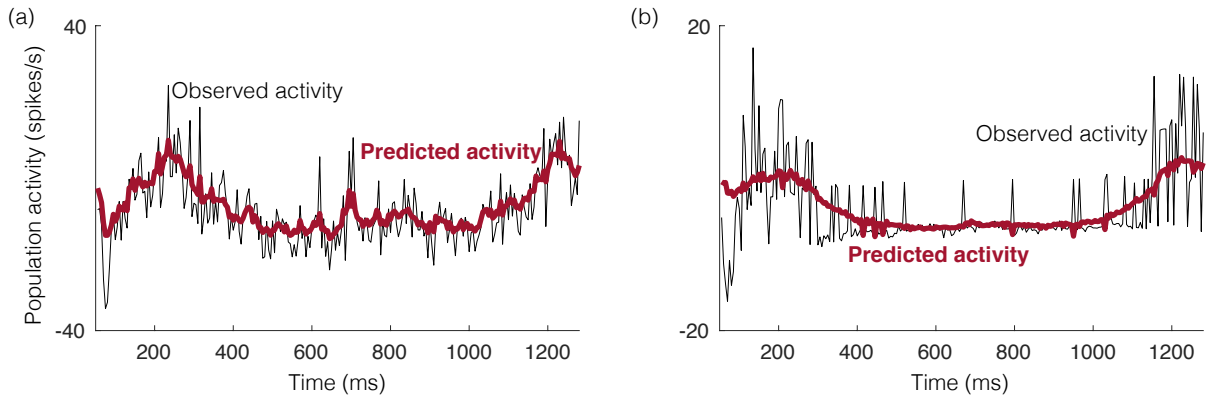

Figure 6.9: **Leave-one-neuron-out prediction using gLARA.** Observed activity (black) and the leave-one-neuron-out prediction of gLARA (blue) for a representative held-out trial, averaged over **(a)** the V1 population and **(b)** the V2 population. Note that the activity can be negative because we are analyzing the single-trial residuals.

population covariance must be accounted for by the interaction between the population-specific latents, $\mathbf{x}_t^1$ and $\mathbf{x}_t^2$, because there are no shared latents in this model. Thus, the interaction between V1 and V2 is summarized by the A_k^{12} and A_k^{21} matrices. Also, both AR-pCCA and gLARA outperform FA and pCCA (comparing vertical axes in Fig. 6.7 and Fig. 6.8), showing that there is meaningful temporal structure in how V1 and V2 interact that can be captured by these models.

Having performed a systematic, relative comparison between AR-pCCA and gLARA models of different complexities, we asked how well the best gLARA model fit the data in an absolute sense. To do so, we used $3/4$ of the data to fit the model parameters and performed leave-one-neuron-out prediction[35] on the remaining $1/4$. This is done by estimating the latent states $\mathbb{E}\left(\mathbf{x}_{1,\ldots,T}^1 \mid \mathbf{y}_{1,\ldots,T}^1\right)$ and $\mathbb{E}\left(\mathbf{x}_{1,\ldots,T}^2 \mid \mathbf{y}_{1,\ldots,T}^2\right)$ using all but one neuron. This estimate of the latent state is then used to predict the activity of the neuron that was left out (the same procedure was repeated for each neuron). For visualization purposes, we averaged the predicted activity across neurons for a given trial and compared it to the recorded activity averaged across neurons for the same trial. We found that they indeed tracked each other, as shown in Fig. 6.9 for a representative trial.

Finally, we asked whether gLARA reveals differences in the time structures of the within-population dynamics and the between-population interactions. We computed the Frobenius norm of both the within-population dynamics matrices A_k^{11} and A_k^{22} (Fig. 6.10a) and the between-population interaction matrices A_k^{12} and A_k^{21} (Fig. 6.10b), for $p_1 = 50$, $p_2 = 15$ and $\tau = 3$ ($k \in \{1, 2, 3\}$), which is

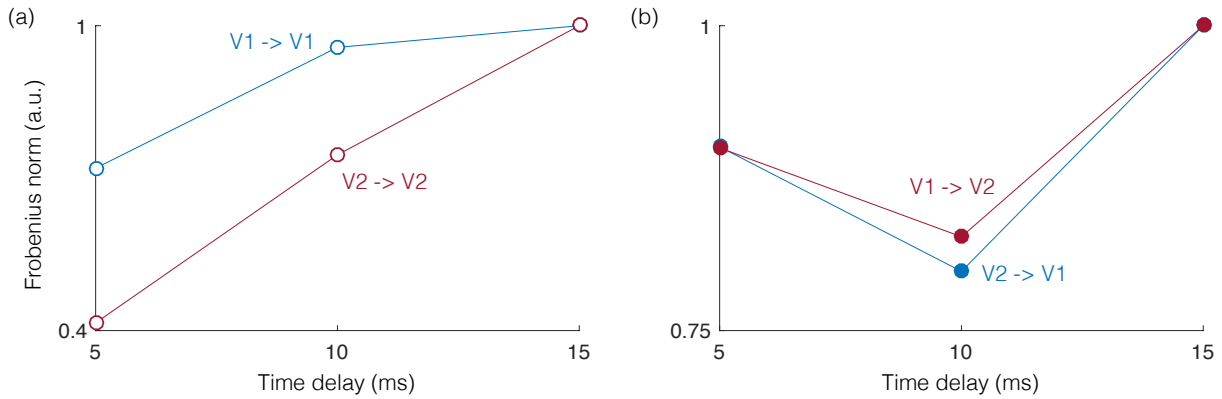

Figure 6.10: **Temporal structure of coupling matrices for gLARA. (a)** Frobenius norm of the within-population dynamics matrices A_k^{11} and A_k^{22}, for $k \in \{1, 2, 3\}$. Each curve was divided by its maximum value. **(b)** Same as (a) for the between-population interaction matrices A_k^{12} and A_k^{21}.

the model for which the cross-validated log-likelihood was the highest. The time structure of the within-population dynamics appears to differ from that of the between-population interaction. In particular, the latents for each area depend more strongly on its own previous latents as the time delay increases up to 15 ms (Fig. 6.10a). In contrast, the dependence between areas is stronger at time lags of 5 and 15 ms, compared to 10 ms (Fig. 6.10b). Note that the peak of the cross-validated log-likelihood for $\tau = 3$ (Fig. 6.8) shows that delays longer than 15ms do not contribute to an increase in the accuracy of the model and, therefore, the most significant interactions between these areas may occur within this time window. The structure seen in Fig. 6.10 is not present if the same analysis is performed on data that are shuffled across time (results not shown). Because the latent states may have different scales, it is not informative to compare the magnitude of A_k^{12} and A_k^{21} or A_k^{11} and A_k^{22} (A_k^{11} and A_k^{22} also have different dimensions). Thus, we divided the norms for each A_k^{ij} matrix by the respective maximum across k.

6.4 Discussion

In this chapter, we focused on characterizing the dynamics of the V1-V2 interactions. We found that the strength of these interactions depended on stimulus context, time since stimulus onset/offset, and temporal lag between the two areas. Furthermore, the structure of interaction also depended on stimulus context and elapsed time, suggesting that not only do these interactions

exhibit different temporal signatures, the patterns of activity shared across the two areas are also distinct. In an effort to build a parsimonious model for these observations, we proposed a novel statistical method, termed group latent auto-regressive analysis (gLARA). This method simultaneously extracts two (or more) sets of latent variables, one for each neuronal population, while explicitly, and simultaneously, capturing intra- and inter-area dynamics.

Our finding that inter-area interactions are stronger during spontaneous activity mimics what has been reported within neuronal populations for pairwise correlations[73]. However, this need not be the case, and recent studies have shown that changes in intra- and inter-area pairwise correlation do not always follow the same trend[13]. Interestingly, while V1 provides large input to V2 and the recordings analyzed here targeted the output layers of V1 and the input layers of V2, we only found a clear signature of feedforward interaction early in the evoked activity period. Indeed, during spontaneous activity, when the two areas were most correlated, peak correlation happened for negative lags, where V2 activity leads V1 activity. This shift from a feedforward dominated interaction early in the evoked epoch to a feedback dominated interaction in the spontaneous epoch was accompanied by a change in the communication subspace and its relationship to the source population activity.

It is important to note that there are significant differences between our strength and structure characterizations, chief among them the difference in time scale: while the canonical correlation maps (CCM) were computed using 1 ms bins, the reduced rank regression models were fit to spike counts in 100 ms epochs. The larger time bins make it so that the predictive dimensions identified could either reflect feedforward or feedback interactions. However, given that we find significantly different temporal signatures for the same time windows for which we identify distinct predictive dimensions, it is at least plausible that interaction directionality is responsible for the differences in predictive dimensions. This is not to say, for example, that predictive dimensions identified early in the evoked epoch exclusively reflect "feedforward" communication, but rather that the feedforward interaction might have a larger impact on the identified predictive dimensions during this period.

When identifying predictive dimensions during the spontaneous activity period, we found that the statistics of the neuronal populations, particularly their firing rates, lead to a potential underestimation of the number of predictive dimensions involved. One might wonder whether

the differences we found between the evoked and spontaneous activity communication subspaces might simply reflect the fact that it is harder to identify predictive dimensions during spontaneous activity, leading to bad approximations. Two reasons suggest this is not the case: (1) Evoked communication subspaces did not generalize well to spontaneous activity, even though they contained a larger number of predictive dimensions. Furthermore, we showed in Section 4.2.3 that we could recover the correct number of predictive dimensions in the evoked epoch; (2) When we did match the predictive dimensions between the evoked and spontaneous epochs, we found that the predictive dimensions identified generalized much better across epochs. In particular, for these surrogate data, dimensions identified for evoked activity generalized almost perfectly to spontaneous activity.

Motivated by the rich temporal structure of the V1-V2 interaction, we proposed a new statistical method, gLARA, that aims to capture the intra- and inter-area dynamics using a dimensionality reduction approach. We showed that gLARA provided a better approximation to the recorded activity than probabilistic canonical correlation analysis and an extension we proposed to capture activity dynamics, auto-regressive pCCA (AR-pCCA). Applying gLARA to the V1-V2 recordings suggested that the bulk of these interactions occur within a 15ms windows, and that the across-area dynamics were stronger at 5 and 15 ms delays, but less so at 10 ms delay.

In the context of studying the interaction between populations of neurons, capturing the information flow is key to understanding how information is processed in the brain[23,74,16,34,17,75]. To do so, one must be able to characterize the directionality of these between-population interactions. Previous studies have sought to identify the directionality of interactions directly between neurons, using measures such as Granger causality[2] (and related extensions, such as directed transfer function (DTF)[76]), and directed information[3]. Here, we proposed to study between-population interaction on the level of latent variables, rather than of the neurons themselves. The advantage is that this approach scales better with the number of recorded neurons and provides a more succinct picture of the structure of these interactions. To detect fine timescale interactions, it may be necessary to replace the linear-Gaussian model with a point process model on the spike trains[77].

6.5 Methods

6.5.1 Canonical correlation analysis

Canonical correlation analysis (CCA) seeks for pairs of dimensions, one in each area, for which the projections are maximally correlated. Let X be a $n \times p$ matrix containing the residual activity of the source V1 population and Y a $n \times q$ matrix containing the residual activity of the target (V1 or V2) population (n represents the number of data points, p and q are the number of neurons in the source and target populations, respectively).

We start by defining the extended sample covariance matrix:

$$S = \frac{1}{N} \begin{bmatrix} XX^T & XY^T \\ YX^T & YY^T \end{bmatrix} = \begin{bmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{bmatrix} \tag{6.17}$$

Consider the projections of each area's activity onto a pair of dimensions: $\mathbf{a}_1{}^T X$, $\mathbf{a}_1 \in \mathbb{R}^q$ and $\mathbf{b}_1{}^T Y$, $\mathbf{b}_1 \in \mathbb{R}^p$. The correlation between the two sets of projections is given by:

$$\rho = \frac{\mathbf{a}_1{}^T S_{XY} \mathbf{b}_1}{(\mathbf{a}_1{}^T S_{XX} \mathbf{a}_1 \mathbf{b}_1{}^T S_{YY} \mathbf{b}_1)^{1/2}} \tag{6.18}$$

CCA searches for the vectors $\mathbf{a}_1$ and $\mathbf{b}_1$ that maximize $\mathbf{a}_1{}^T S_{XY} \mathbf{b}_1$ subject to $\mathbf{a}_1{}^T S_{XX} \mathbf{a}_1 = \mathbf{b}_1{}^T S_{YY} \mathbf{b}_1 = 1$ (note that ρ does not depend on the norm of $\mathbf{a}_1$ and $\mathbf{b}_1$). This is attained by setting:

$$\mathbf{a}_1 = S_{XX}^{-1/2} \alpha_1, \quad \mathbf{b}_1 = S_{YY}^{-1/2} \beta_1 \tag{6.19}$$

where α_1 and β_1 are the eigenvectors associated with the largest eigenvalue of the matrices $N_1 = KK^T$ and $N_2 = K^T K$, respectively, with $K = S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2}$. In general, the m first canonical pairs are obtained by picking the eigenvectors of N_1 and N_2 associated with their m largest eigenvalues.

Note that $N_1 \in \mathbb{R}^{q \times q}$ and $N_2 \in \mathbb{R}^{p \times p}$ and consequently $m \leq \min(p, q)$. Also, while we have ${\alpha_i}^T \alpha_j = {\beta_i}^T \beta_j = \delta_{ij}$ (where $\delta_{ij} = 1$ if $i = j$, 0 otherwise) the $\mathbf{a}_i$ and $\mathbf{b}_i$ are not, in general, orthogonal (that only happens in the special case where the within-set covariance matrices are isotropic). They are, like the predictive dimensions returned by reduced rank regression (RRR),

uncorrelated, and guaranteed to be linearly independent.

CCA can be seem as a symmetric version of RRR, that takes into account the within-area covariance in both areas, i.e. it implicitly whitens the population activity in both areas (whereas RRR only whitens the source population activity X). A direct consequence of this observation is that if $S_{YY} = \sigma I$, for any σ, the dimensions $\mathbf{a}_i$ correspond to the predictive dimensions of RRR. Another interesting scenario is when the target population activity Y is univariate. In this case, CCA can be shown to reduced to linear regression. This observation forms the basis of the illustration in Fig. 4.4, where we present linear regression as seeking the dimension of the source population activity for which the corresponding projections are most correlated with the activity of the observed V2 neuron.

6.5.2 Ridge reduced rank regression

One of the challenges of characterizing the interaction structure during the inter-trial period was that firing rates were low, much lower than for the evoked activity period. This lead RRR to overfit the activity during this period. One way to see the effect of overfitting in our estimates of the communication subspace is to apply RRR to data where the underlying interaction is linear and fixed, but the statistics of the source and target populations are matched to those observed for the recordings, for both the evoked and spontaneous epochs (Fig. 6.11a; this is also the approach taken in Fig. 6.2b). We find that when we identify the communication subspace on one of the evoked activity windows it generalizes across all other time windows (rows 1 through 5, corresponding to models fit to evoked activity windows, have full generalization performance across all other time windows). However, communication subspaces fit to spontaneous activity do not generalize as well to evoked activity windows, even though the underlying mapping is the same.

To see why this might happen, suppose there is a number of V1 neurons that are active during the evoked epoch, but silent during the spontaneous activity period. Fitting the RRR model to evoked activity will correctly assign weights to these neurons, thereby leveraging their activity to predict activity in V2. During the spontaneous activity period these same neurons are silent, and their activity cannot therefore be used to predict activity in V2, despite the assigned weights (Fig. 6.11a, top-right block). When we fit the RRR model to spontaneous activity, on the other

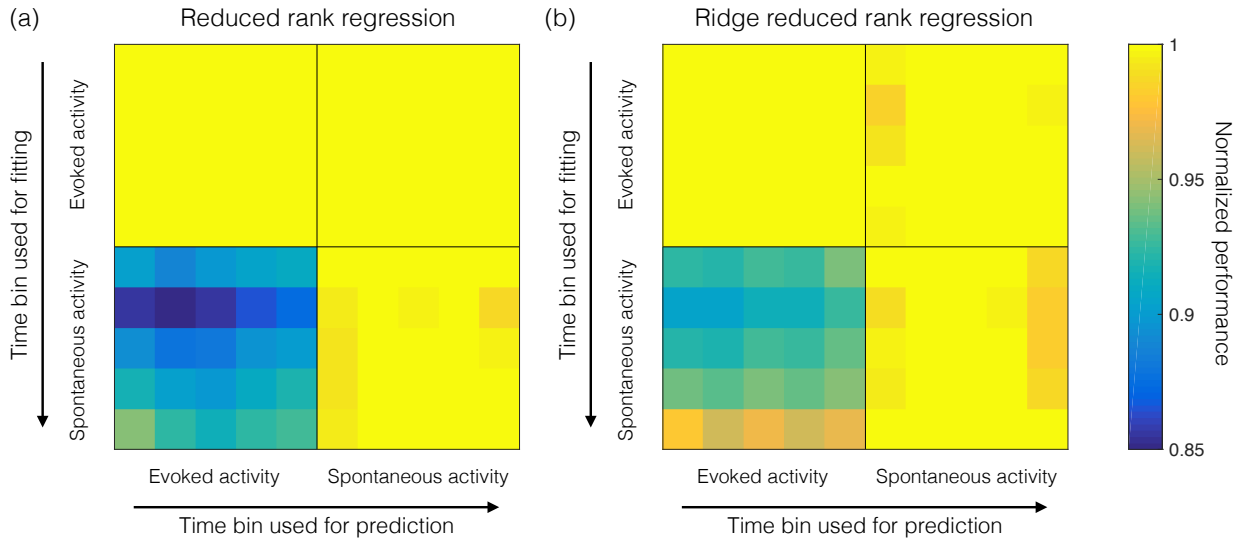

Figure 6.11: **Low firing rates during spontaneous activity can lead to overfitting. (a)** Reduced rank regression models fit to spontaneous activity windows did not generalize well to the evoked activity period, even though the underlying mapping was the same for both periods. Each row corresponds to a different communication subspace and each column to the application of those subspaces to a different time window. The diagonal elements thus indicate the normalized performance of identifying and applying the communication subspace to responses evoked by the same stimuli. Off-diagonal elements can have normalized predictive performance values less than 1, indicating that the communication subspaces did not generalize well across time windows. Performance was measured using 10-fold cross-validation (i.e., for each fold, the model was fit to a training set pertaining to one time window and then used for prediction in test sets of all time windows). Each entry shows the average normalized performance across all recording sessions. **(b)** Using ridge reduced rank regression substantially improved generalization. Same conventions as in (a).

hand, any set of weights can be assigned to the silent neurons without hurting predictive performance during this period (any linear combination of silent neurons will result in no predictions). However, when we use these communication subspaces to predict evoked activity, the neurons are no longer silent, and their activity gets combined using an overfitted set of weights, decreasing predictive performance during this period (Fig. 6.11a, bottom-left block).

To try to overcome this issue, we used an extension of RRR, which we termed ridge reduced rank regression (λRRR). The motivation behind this approach is to impose a ridge-like penalty on the weights of the RRR model. Intuitively, the weights assigned to the neurons that are silent during the spontaneous activity period should be decreased to 0, since they are not useful for prediction, thereby eliminating the erroneous predictions these weights produce during the evoked period. Note that this approach cannot fully solve the generalization problem: since the neurons are silent during the spontaneous activity period it will never be possible to correctly identify the weights associated with these neurons. It does however noticeably improve generalization performance across the two epochs (Fig. 6.11b).

To see how the λRRR is defined, recall the solution to the RRR problem:

$$B_{RRR} = B_{OLS} V V^T$$

where B_{OLS} is the ordinary least squares solution and the columns of the $q \times m$ matrix V contain the top m principal components of the optimal linear predictor $\hat{Y}_{OLS} = X B_{OLS}$. For the λRRR model, we simply replace the ordinary least squares solution, B_{OLS}, with the ridge regression solution, $B_{Ridge} = (X^T X + \lambda I)^{-1} X^T Y$:

$$B_{\lambda RRR} = B_{Ridge} V V^T$$

where the columns of V now contain the top m principal components of the ridge linear predictor $\hat{Y}_{Ridge} = X B_{Ridge}$. We chose the value of λ using 10-fold cross-validation.

Chapter 7

Summary and future directions

In this dissertation, we leveraged a dimensionality reduction approach to study inter-area interactions in the brain. This enabled us to provide a description of inter-area interaction, and propose a mechanism for selective communication based on the structure found. With this framework in place, we then ventured into the study of the dynamics of inter-area interaction, which led to the proposal of two new tools: the canonical correlation map (CCM) and group latent auto-regressive analysis (gLARA).

Future directions

The work presented in this dissertation has led to a number of interesting questions, which we feel will form the basis for future work.

How do the predicted dimensions in V2 relate to the dominant dimensions in this area? We have characterized the way in which V1 activity relates to the communication subspace, but we haven't yet explored how the V2 activity is impacted by the input from V1. The results in Section 4.2.2 suggest that the population activity in V2 spans a larger subspace than that related to the V1 activity. We would like to quantify how much of the shared variability in V2 is predicted by V1 and how the predicted dimensions relate to the dominant dimensions in the target areas.

Do current computational models of V2 induce a communication subspace? One possibility is that the low-dimensional communication subspace we identified for the V1-V2 interaction emerged due to the need to selectively route information between these areas. Alternatively, the computation implemented by the V2 population on the inputs it receives from V1 might be such that, effectively, it only relies on a subspace of the V1 activity. We want to test the latter hypothesis by applying the same approach used on the V1-V2 recordings to analyze data generated from a spiking model of the V1-V2 computation created in Christian Machens' group. We are also collaborating with Alex Pouget's group, who are developing a general spiking model with multiple neuronal populations.

How does the communication subspace relate to stimulus processing? In all the work developed so far we focused on how activity fluctuations in V1 relate to activity fluctuations in V2, under a fixed stimulus, which in these experiments were sets of oriented gratings. We plan to investigate how the communication subspace identified in this way relates to the stimulus representation in these areas. For example, how does the communication subspace compare with the subspace in the V1 population activity where visual information is encoded (the stimulus subspace)? Do they overlap? In other words, is it the case that all the activity relayed between these areas is stimulus related, while non-stimulus related information stays private? We have made some progress in this direction (see Section 5.2), but fully identifying the stimulus subspace will likely require a richer set of visual stimulus. Adam Kohn's group is currently performing new V1-V2 recordings using both finer grating orientation changes, as well as more naturalistic stimuli, which will provide an exciting avenue to pursue this question.

Bibliography

[1] Pillow, J. W. *et al.* Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995–999 (2008). URL http://www.nature.com/nature/journal/v454/n7207/abs/nature07140.html.

[2] Kim, S., Putrino, D., Ghosh, S. & Brown, E. N. A granger causality measure for point process models of ensemble neural spiking activity. *PLoS Comput Biol* **7**, e1001110 (2011). URL http://dx.doi.org/10.1371/journal.pcbi.1001110.

[3] Quinn, C. J., Coleman, T. P., Kiyavash, N. & Hatsopoulos, N. G. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *Journal of Computational Neuroscience* **30**, 17–44 (2011). URL http://link.springer.com/article/10.1007/s10827-010-0247-2.

[4] Macke, J. H. *et al.* Empirical models of spiking in neural populations. In *NIPS*, 1350–1358 (2011). URL https://papers.nips.cc/paper/4289-empirical-models-of-spiking-in-neural-populat

[5] Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience* **17**, 1500–1509 (2014). URL http://www.nature.com/neuro/journal/v17/n11/abs/nn.3776.html.

[6] Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* **512**, 423–426 (2014). URL http://www.nature.com/nature/journal/v512/n7515/abs/nature13665.html.

[7] Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null

space: permitting preparation without movement. *Nature Neuroscience* **17**, 440–448 (2014). URL http://www.nature.com/neuro/journal/v17/n3/abs/nn.3643.html.

[8] Nowak, L. G., Munk, M. H. J., James, A. C., Girard, P. & Bullier, J. Cross-Correlation Study of the Temporal Interactions Between Areas V1 and V2 of the Macaque Monkey. *Journal of Neurophysiology* **81**, 1057–1074 (1999). URL http://jn.physiology.org/content/81/3/1057.

[9] Roe, A. W. & Ts'o, D. Y. Specificity of Color Connectivity Between Primate V1 and V2. *Journal of Neurophysiology* **82**, 2719–2730 (1999). URL http://jn.physiology.org/content/82/5/2719.

[10] Jia, X., Tanabe, S. & Kohn, A. Gamma and the Coordination of Spiking Activity in Early Visual Cortex. *Neuron* **77**, 762–774 (2013). URL http://www.sciencedirect.com/science/article/pii/S0896627313000445.

[11] Pooresmaeili, A., Poort, J. & Roelfsema, P. R. Simultaneous selection by object-based attention in visual and frontal cortex. *Proceedings of the National Academy of Sciences* **111**, 6467–6472 (2014). URL http://www.pnas.org/content/111/17/6467.

[12] Oemisch, M., Westendorff, S., Everling, S. & Womelsdorf, T. Interareal Spike-Train Correlations of Anterior Cingulate and Dorsal Prefrontal Cortex during Attention Shifts. *Journal of Neuroscience* **35**, 13076–13089 (2015). URL http://www.jneurosci.org/content/35/38/13076.

[13] Ruff, D. A. & Cohen, M. R. Attention Increases Spike Count Correlations between Visual Cortical Areas. *The Journal of Neuroscience* **36**, 7523–7534 (2016). URL http://www.jneurosci.org/content/36/28/7523.

[14] Truccolo, W., Hochberg, L. R. & Donoghue, J. P. Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nature Neuroscience* **13**, 105–111 (2010). URL http://www.nature.com/neuro/journal/v13/n1/abs/nn.2455.html.

[15] Zandvakili, A. & Kohn, A. Coordinated Neuronal Activity Enhances

Corticocortical Communication. *Neuron* **87**, 827–839 (2015). URL http://www.sciencedirect.com/science/article/pii/S0896627315006480.

[16] Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-Frequency, Long-Range Coupling Between Prefrontal and Visual Cortex During Attention. *Science* **324**, 1207–1210 (2009). URL http://science.sciencemag.org/content/324/5931/1207.

[17] Salazar, R. F., Dotson, N. M., Bressler, S. L. & Gray, C. M. Content-Specific Fronto-Parietal Synchronization During Visual Working Memory. *Science* **338**, 1097–1100 (2012). URL http://science.sciencemag.org/content/338/6110/1097.

[18] Menzer, D. L., Rao, N. G., Bondy, A., Truccolo, W. & Donoghue, J. P. Population interactions between parietal and primary motor cortices during reach. *Journal of Neurophysiology* **112**, 2959–2984 (2014). URL http://jn.physiology.org/content/112/11/2959.

[19] Arce-McShane, F. I., Ross, C. F., Takahashi, K., Sessle, B. J. & Hatsopoulos, N. G. Primary motor and sensory cortical areas communicate via spatiotemporally coordinated networks at multiple frequencies. *Proceedings of the National Academy of Sciences* **113**, 5083–5088 (2016). URL http://www.pnas.org/content/113/18/5083.

[20] Wong, Y. T., Fabiszak, M. M., Novikov, Y., Daw, N. D. & Pesaran, B. Coherent neuronal ensembles are rapidly recruited when making a look-reach decision. *Nature Neuroscience* **19**, 327–334 (2016). URL http://www.nature.com/neuro/journal/v19/n2/full/nn.4210.html.

[21] Bosman, C. *et al.* Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas. *Neuron* **75**, 875–888 (2012). URL http://www.sciencedirect.com/science/article/pii/S089662731200623X.

[22] Roberts, M. *et al.* Robust Gamma Coherence between Macaque V1 and V2 by Dynamic Frequency Matching. *Neuron* **78**, 523–536 (2013). URL http://www.sciencedirect.com/science/article/pii/S0896627313002274.

[23] Fries, P. A mechanism for cognitive dynamics: neuronal communication

through neuronal coherence. *Trends in Cognitive Sciences* **9**, 474–480 (2005). URL http://www.sciencedirect.com/science/article/pii/S1364661305002421.

[24] Fries, P. Rhythms for Cognition: Communication through Coherence. *Neuron* **88**, 220–235 (2015). URL http://www.sciencedirect.com/science/article/pii/S0896627315008235.

[25] Wallisch, P. & Movshon, J. A. Structure and Function Come Unglued in the Visual Cortex. *Neuron* **60**, 195–197 (2008). URL http://www.sciencedirect.com/science/article/pii/S0896627308008519.

[26] Sincich, L. C. & Horton, J. C. THE CIRCUITRY OF V1 AND V2: Integration of Color, Form, and Motion. *Annual Review of Neuroscience* **28**, 303–326 (2005). URL https://doi.org/10.1146/annurev.neuro.28.061604.135731.

[27] Felleman, D. J. & Essen, D. C. V. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex* **1**, 1–47 (1991). URL http://cercor.oxfordjournals.org/content/1/1/1.1.

[28] Lennie, P. Single Units and Visual Cortical Organization. *Perception* **27**, 889–935 (1998). URL https://doi.org/10.1068/p270889.

[29] Smith, M. A. & Kohn, A. Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex. *The Journal of Neuroscience* **28**, 12591–12603 (2008). URL http://www.jneurosci.org/content/28/48/12591.

[30] Churchland, M. M. *et al.* Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience* **13**, 369–378 (2010). URL http://www.nature.com/neuro/journal/v13/n3/abs/nn.2501.html.

[31] Semedo, J., Zandvakili, A., Kohn, A., Machens, C. K. & Yu, B. M. Extracting Latent Structure From Multiple Interacting Neural Populations. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q. (eds.) *Advances in Neural Information Processing Systems 27*, 2942–2950 (Curran Associates, Inc., 2014). URL http://papers.nips.cc/paper/5625-extracting-latent-structure-from-multiple-inte

[32] Jia, X., Smith, M. A. & Kohn, A. Stimulus Selectivity and Spatial Coherence of Gamma Components of the Local Field Potential. *Journal of Neuroscience* **31**, 9390–9403 (2011). URL http://www.jneurosci.org/content/31/25/9390.

[33] Ray, S. & Maunsell, J. H. R. Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biol* **9**, e1000610 (2011). URL http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000610.

[34] Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X. & Kastner, S. The Pulvinar Regulates Information Transmission Between Cortical Areas Based on Attention Demands. *Science* **337**, 753–756 (2012). URL http://science.sciencemag.org/content/337/6095/753.

[35] Yu, B. M. *et al.* Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. *Journal of Neurophysiology* **102**, 614–635 (2009). URL http://jn.physiology.org/content/102/1/614.

[36] Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012). URL http://www.nature.com/nature/journal/v484/n7392/abs/nature10918.html.

[37] Ecker, A. *et al.* State Dependence of Noise Correlations in Macaque Primary Visual Cortex. *Neuron* **82**, 235–248 (2014). URL http://www.sciencedirect.com/science/article/pii/S0896627314001044.

[38] Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. *eLife* **4**, e04677 (2015). URL https://elifesciences.org/content/4/e04677v1.

[39] Williamson, R. C. *et al.* Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. *PLOS Computational Biology* **12**, e1005141 (2016). URL http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005141.

[40] Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. Noise in the nervous system. *Nature Reviews Neuroscience* **9**, 292–303 (2008). URL http://www.nature.com/nrn/journal/v9/n4/abs/nrn2258.html.

[41] Goris, R. L. T., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nature Neuroscience* **17**, 858–865 (2014). URL http://www.nature.com/neuro/journal/v17/n6/abs/nn.3711.html.

[42] Tolhurst, D. J., Movshon, J. A. & Dean, A. F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research* **23**, 775–785 (1983). URL http://www.sciencedirect.com/science/article/pii/0042698983902006.

[43] Izenman, A. J. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis* **5**, 248–264 (1975). URL http://www.sciencedirect.com/science/article/pii/0047259X75900421.

[44] Kobak, D. *et al.* Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016). URL https://elifesciences.org/content/5/e10989v2.

[45] Santhanam, G. *et al.* Factor-Analysis Methods for Higher-Performance Neural Prostheses. *Journal of Neurophysiology* **102**, 1315–1330 (2009). URL http://jn.physiology.org/content/102/2/1315.

[46] Cowley, B. *et al.* Distance Covariance Analysis. In *PMLR*, 242–251 (2017). URL http://proceedings.mlr.press/v54/cowley17a.html.

[47] Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural characterization. *Journal of Vision* **6**, 13–13 (2006). URL http://jov.arvojournals.org/article.aspx?articleid=2192881.

[48] Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M. & Cunningham, J. P. Reorganization between preparatory and movement population responses in motor cortex. *Nature Communications* **7**, 13239 (2016). URL http://www.nature.com/ncomms/2016/161027/ncomms13239/full/ncomms13239.html.

[49] Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience* **16**, 974–981 (2013). URL http://www.nature.com/neuro/journal/v16/n7/abs/nn.3402.html.

[50] Yu, Y., Schmid, A. M. & Victor, J. D. Visual processing of informative multipoint correlations arises primarily in V2. *eLife* **4**, e06604 (2015). URL https://elifesciences.org/content/4/e06604v2.

[51] Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. *Science* **291**, 1560–1563 (2001). URL http://science.sciencemag.org/content/291/5508/1560.

[52] Pesaran, B., Nelson, M. J. & Andersen, R. A. Free choice activates a decision circuit between frontal and parietal cortex. *Nature* **453**, 406–409 (2008). URL https://www.nature.com/articles/nature06849.

[53] Ray, S. & Maunsell, J. H. R. Do gamma oscillations play a role in cerebral cortex? *Trends in Cognitive Sciences* **19**, 78–85 (2015). URL http://www.sciencedirect.com/science/article/pii/S136466131400254X.

[54] Salinas, E. & Abbott, L. F. Vector reconstruction from firing rates. *Journal of Computational Neuroscience* **1**, 89–107 (1994). URL https://link.springer.com/article/10.1007/BF00962720.

[55] Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nature Neuroscience* **9**, 690–696 (2006). URL http://www.nature.com/neuro/journal/v9/n5/abs/nn1691.html.

[56] Remme, M. W. H., Lengyel, M. & Gutkin, B. S. Democracy-Independence Trade-Off in Oscillating Dendrites and Its Implications for Grid Cells. *Neuron* **66**, 429–437 (2010). URL http://www.cell.com/neuron/abstract/S0896-6273(10)00298-9.

[57] Kohn, A. & Smith, M. A. Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. *The Journal of Neuroscience* **25**, 3661–3673 (2005). URL http://www.jneurosci.org/content/25/14/3661.

[58] Cohen, M. R. & Newsome, W. T. Context-Dependent Changes in Functional Circuitry in Visual Area MT. *Neuron* **60**, 162–173 (2008). URL http://www.sciencedirect.com/science/article/pii/S0896627308006752.

[59] Bondy, A. G., Haefner, R. M. & Cumming, B. G. Feedback determines the structure of correlated variability in primary visual cortex. *Nature Neuroscience* 1 (2018). URL https://www.nature.com/articles/s41593-018-0089-1.

[60] Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience* **12**, 1594–1600 (2009). URL https://www.nature.com/articles/nn.2439.

[61] Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron* **63**, 879–888 (2009). URL http://www.sciencedirect.com/science/article/pii/S0896627309006953.

[62] Kohn, A., Coen-Cagli, R., Kanitscheider, I. & Pouget, A. Correlations and Neuronal Population Information. *Annual Review of Neuroscience* **39**, 237–256 (2016). URL https://doi.org/10.1146/annurev-neuro-070815-013851.

[63] Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning*, vol. 1 (Springer series in statistics Springer, Berlin, 2001). URL http://statweb.stanford.edu/ tibs/book/preface.ps.

[64] Sincich, L. C., Jocson, C. M. & Horton, J. C. V1 Interpatch Projections to V2 Thick Stripes and Pale Stripes. *Journal of Neuroscience* **30**, 6963–6974 (2010). URL http://www.jneurosci.org/content/30/20/6963.

[65] Sincich, L. C. & Horton, J. C. Input to V2 Thin Stripes Arises from V1 Cytochrome Oxidase Patches. *Journal of Neuroscience* **25**, 10087–10093 (2005). URL http://www.jneurosci.org/content/25/44/10087.

[66] Arandia-Romero, I., Tanabe, S., Drugowitsch, J., Kohn, A. & Moreno-Bote, R. Multiplicative and Additive Modulation of Neuronal Tuning with Population Activity Affects Encoded Information. *Neuron* **89**, 1305–1316 (2016). URL http://www.sciencedirect.com/science/article/pii/S089662731600091X.

[67] Schlvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D. & Carandini, M. Cortical State

Determines Global Variability and Correlations in Visual Cortex. *The Journal of Neuroscience* **35**, 170–178 (2015). URL http://www.jneurosci.org/content/35/1/170.

[68] Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses. *Science* **273**, 1868–1871 (1996). URL http://science.sciencemag.org/content/273/5283/1868.

[69] Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *eLife* **4**, e08998 (2015). URL https://elifesciences.org/content/4/e08998v3.

[70] Ruff, D. A. & Cohen, M. R. Stimulus Dependence of Correlated Variability across Cortical Areas. *The Journal of Neuroscience* **36**, 7546–7556 (2016). URL http://www.jneurosci.org/content/36/28/7546.

[71] Bach, F. R. & Jordan, M. I. A probabilistic interpretation of canonical correlation analysis (2005). URL http://www.stat.berkeley.edu/ jordan/688.pdf.

[72] Anderson, B. D. & Moore, J. B. *Optimal filtering* (Courier Dover Publications, 2012).

[73] Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012). URL http://www.nature.com/nature/journal/v487/n7405/abs/nature11129.html.

[74] Crowe, D. A. *et al.* Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nature Neuroscience* **16**, 1484–1491 (2013). URL http://www.nature.com/neuro/journal/v16/n10/abs/nn.3509.html.

[75] Vzquez, Y., Salinas, E. & Romo, R. Transformation of the neural code for tactile detection from thalamus to cortex. *Proceedings of the National Academy of Sciences* **110**, E2635–E2644 (2013). URL http://www.pnas.org/content/110/28/E2635.

[76] Kaminski, M. J. & Blinowska, K. J. A new method of the description of the information flow in the brain structures. *Biological Cybernetics* **65**, 203–210 (1991). URL http://link.springer.com/article/10.1007/BF00198091.

[77] Smith, A. C. & Brown, E. N. Estimating a state-space model from point process observations. *Neural Computation* **15**, 965–991 (2003). URL http://dx.doi.org/10.1162/089976603765202622.