

**PATTERNS OF POPULATION ACTIVITY
ASSOCIATED WITH THE INTENTIONAL
CONTROL OF SINGLE NEURON FIRING RATES
IN PRIMATE MOTOR CORTEX**

by

Andrew S. Whitford

BS in Biomedical Engineering, Tulane University, 2004

MS in Electrical Engineering, University of Rhode Island, 2006

Submitted to the Graduate Faculty of
the Swanson School of Engineering in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH
SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Andrew S. Whitford

It was defended on

July 13, 2016

and approved by

Andrew B. Schwartz, Ph.D., Distinguished Professor, Department of Neurobiology

Peter L. Strick, Ph.D., Distinguished Professor, Department of Neurobiology

Zhi-Hong Mao, Ph.D., Associate Professor, Department of Bioengineering

Steven M. Chase, Ph.D., Associate Professor, Department of Biomedical Engineering,

Carnegie Mellon University

Dissertation Director: Andrew B. Schwartz, Ph.D., Distinguished Professor, Department of

Neurobiology

**PATTERNS OF POPULATION ACTIVITY ASSOCIATED WITH THE
INTENTIONAL CONTROL OF SINGLE NEURON FIRING RATES IN
PRIMATE MOTOR CORTEX**

Andrew S. Whitford, PhD

University of Pittsburgh, 2016

Although it is firmly established that the primate motor cortex contributes to intentional behavior, the exact relationship between patterns of cortical activity and effected action is not well-understood. A particular matter of controversy is the degree to which such patterns subserve a singular function in the control of behavior, and are constrained by this specific role. The reported capability of nearby motor cortical neurons to vary independently of one another – and to be dissociated from overt movement – has often been cited as evidence that patterns of cortical activity are rather labile; that they are readily adjusted to suit varied behavioral demands. Such claims are rooted in evidence from neurophysiology experiments that entail direct reinforcement of variation in the firing rates of single cortical neurons – that is, single unit operant conditioning experiments. However, such investigations have largely been limited to the consideration of relationships among neurons recorded on the same electrode, or among small populations ($N < 5$) of nearby neurons. Here, we explore cortical flexibility by characterizing patterns of population ($N > 20$) activity as single neurons are targeted for conditioning. We find that a substantial percentage of the populations tend to co-vary, even though behavioral goals (i.e., reward) depend only on the activity of single neurons. Further, stable patterns of covariation can change with those goals. We conclude that cortical activity can be reshaped to suit behavioral demands, but that intrinsic structure – perhaps due to network connectivity patterns – likely constrains the flexibility of population activity.

TABLE OF CONTENTS

PREFACE	viii
1.0 INTRODUCTION	1
1.1 Early developments in behavioral electrophysiology	1
1.2 The relationship between cortex and muscles	3
1.3 An exploratory cortical conditioning paradigm	6
1.4 Characterizing interactions among cortical neurons	8
1.5 The significance of correlated activity	11
1.6 A question of mechanism and scope	12
1.7 Natural movement as a bridge to cortical control	15
1.8 Local and global patterns of population activity	18
1.9 Two kinds of constraints on cortical activity patterns	21
1.10 Simple tasks for profound flexibility	22
1.11 Revisiting "natural" constraints	24
1.12 Relaxing assumptions about "natural" activity patterns	26
1.13 Re-assessing the significance of learning	29
1.14 Explaining a diversity of results	30
2.0 METHODS	33
2.1 Experimental setup	33
2.2 Center-out task	34
2.3 Rings task	39
2.4 Data analysis	40
2.4.1 Rate estimation	41

2.4.2 Modulation test	43
3.0 RESULTS	45
3.1 Learning to control single unit activity	45
3.2 Non-contingent covariation: single unit modulation	47
3.3 Non-contingent covariation: pairwise correlation	51
3.4 Comparison with a population-based behavioral contingency	53
3.5 Population covariance structure	61
4.0 DISCUSSION	69
4.1 Covariation of non-contingent cortical activity	70
4.2 Relevance to "natural" behavior	72
4.3 Dissociation and differential conditioning	75
4.4 Conclusions	76
BIBLIOGRAPHY	77

LIST OF FIGURES

1	Array placement for NHP-C.	35
2	Experimental setup.	36
3	Illustration of the trial structure for the center-out and rings tasks.	37
4	Illustration of rate estimation and the modulation test in the 2D center-out task.	42
5	Illustration of firing rate estimation and the modulation test in the rings task.	44
6	Illustration of the data.	46
7	Examples of task acquisition.	48
8	Example population raster plots for NHP-A, during single unit conditioning.	49
9	Example population raster plots for NHP-C, during single unit conditioning.	50
10	Engagement of the population, across single unit conditioning sessions.	52
11	Scatter plot of firing rate correlations in pairs of rings task conditions, for NHP-A.	54
12	Scatter plot of firing rate correlations in pairs of rings task conditions, for NHP-C.	55
13	Engagement of the population, across center-out conditions.	57
14	Summary of population engagement in the two tasks.	58
15	Scatter plot comparison of maximum modulation depths, in the two tasks, for NHP-A.	59
16	Scatter plot comparison of maximum modulation depths, in the two tasks, for NHP-C.	60
17	Scatter plot of 2D principle component scores, for both tasks, for NHP-A.	62
18	Scatter plot of 2D principle component scores, for both tasks, for NHP-C.	63

19	Pie chart comparison of PCA decompositions of both sets of task data, for NHP-A.	65
20	Pie chart comparison of PCA decompositions of both sets of task data, for NHP-C.	66
21	Pareto plot comparison of rings and center-out data, for NHP-A.	67
22	Pareto plot comparison of rings and center-out data, for NHP-A.	68

PREFACE

It's not an easy task to acknowledge everyone that has helped me to reach this point. I have been pursuing this degree for nearly a decade, and my preparation for this position extended for years prior to that. It hasn't been a smooth road, and I appreciate the patience and understanding of those that stuck with me. To have a hope of individually and adequately thanking everyone that contributed, I realistically should've started writing this preface years ago. As it stands, this text can only capture an urgent and fumbling approximation of the gratitude that I feel. Hopefully, I have personally expressed this gratitude to each of you mentioned here, as well as to those that I've left out. I sincerely consider myself extremely fortunate to have shared my years with such people.

As impossible as it seems to adequately acknowledge all of those that have contributed to my academic accomplishments, to describe the importance and extent of my family's influence seems even more daunting. Without question, I owe the majority of my success to the support of my father, mother, and brother. They shaped who I am today, and planted the seeds that I needed to compete and to thrive. The capability and persistence that saw me through this degree is a reflection of the good in them. Completing the degree is as much their success as it is mine. I can't say it enough: thank you. I hope that I can repay you someday.

I'd also like to thank my extended family – aunts, uncles, cousins and grandparents. You shaped my development, and I've been lucky to have your support. I especially want to mention Nan, Leah, and Mike. I regret that they could not be here to see me finish, as they were exceptionally encouraging and engaging. Although encouragement was not exactly his forté, I also regret that Lawrence could not be a part of this. I miss them all.

BIG thanks to all of the members of the Schwartz lab. You've all come to be like an extended family. Thanks to Andy for this opportunity, for taking a chance on me, and especially for being understanding during these latter years, when personal factors really impacted my work. The original crew, during my first several years in Pittsburgh, were especially influential: Andy, Meel, Chance, Michael, Sam, Steve, Marco, George, Sagi, Jeff, Ingrid, Laurel, and later Angus, Rob, and Morgan. That time truly was exhilarating, and will always be remembered as a remarkable period of my life. During more recent years, Jeong-Woo, Jordan, Big Steve Suway, Sharlene, Kathy, Rex, Scott, Xiao, Emrah, Josue, and Heather have been amazing. If I were to call you each out individually, for everything that you've done for me, this preface would extend for several pages. You all helped me to grow a *lot*. You've been a big part of my life, and I won't forget that. I learned volumes from you all, and it truly was a life-changing experience. I want to emphasize that, as it isn't something that I say nonchalantly: the people that I've interacted with during my time in Pittsburgh – past or present – have been among the most intelligent and capable I've ever met. In that regard, it was an outstanding experience – and one that I am much better for. Thank you.

Thank you to my committee: Andy, Peter, Mao, and Steve. I appreciate your patience, and willingness to accommodate my somewhat chaotic and rushed schedule.¹ Your insight has been valuable. And... well... thanks for just being nice about it. Steve and Andy, especially: Thank you for your patience. It hasn't been easy for any of us. I'd also like to thank Rob Kass and Byron Yu, for being part of my original committee and my dissertation proposal.

Thank you to all of the people at Pitt that contributed, in some way or another. Nick Mance and Prof. Schroff helped a lot in these final stages. All of the McGowan people that aren't part of the Schwartz lab: it's been great working with you. Especially Gene: thanks for the encouragement and good-natured ribbing during the last few years.

¹And let's be honest: my habit of cutting things close.

Thanks to the people of the Center for the Neural Basis of Cognition, and the Bioengineering department at the University of Pittsburgh. I am thankful for all of the support. I've truly been part of an exceptional academic community. Thanks also to the medical students of the Federal University of Piau (especially Anatalia), and to Wolfgang Maass and his group in Graz – for being superb hosts, and for facilitating my first international experiences.

Thanks to the people of the clinical project – and to Jan, especially – for making my foray into patient-based experiments a positive experience.

Thanks to the people back home, and my friends / acquaintances during my formative years. Npt has always been a place that I knew I could go back to and feel at home. Thank you for that sanctuary.

Thanks to Arthur, Willy, and all of the others; but especially to Chase. Thanks for making "Chase quality" a synonym for great data.²

As exhilarating as the early years of my PhD experience were, these latter years have been nearly as trying. Never before have I needed people so much... or at least I've never before been so aware of that need. I truly learned the importance of supportive friends. I'm convinced that I wouldn't have made it through, had it not been for the kindness of an army of good people; far too many to properly address. I truly, sincerely apologize to all of those that dealt with my shit during the worst times. Thanks to Michael, Nancy, Emily, and many others, who went well beyond what should typically be expected of good friends. Thanks to people in the lab – especially Scott and Steve Suway, and my office mates (Sam, Morgan, Michael, Sharlene) – who listened to countless hours of me venting. Thanks again, to my family: mom, dad, Bob... I couldn't have done it without you. Knowing that I always had you guys there meant a lot to me. Finally, thanks to all of the people of Pittsburgh and the Southside who really made me feel at home during the last 2 years. In particular, thanks to the people at the various coffee shops – especially the Delanies crew – that became my second homes, as I worked on this dissertation. Thanks to my friends in New Orleans, that really helped to lift me up during some of the bleakest times. Thank you all.³

²lolz

³And sorry.

Whether or not my doctoral research went as expected, it was a phenomenal experience – one that I was exceedingly lucky to have. I’ve been privileged and fortunate to have the opportunity to participate in research at this level. I am grateful for those that made all of this possible. Hopefully, I’ve learned from this experience, and will go on to produce things that we are all proud of. *Thank you.*

Here, I was going to quote Borges’ *On Exactitude in Science* – since it seems appropriate – but I don’t have time to deal with copywriting permission... So I’ll just communicate the sentiment by quoting myself.

– Andrew Whitford

1.0 INTRODUCTION

1.1 EARLY DEVELOPMENTS IN BEHAVIORAL ELECTROPHYSIOLOGY

During the 1960s and 1970s, Dr. Eberhard Fetz and colleagues pioneered the practice of operantly conditioning the electrophysiological activity of cortical neurons in awake primates. Recording from electrodes implanted in the motor cortex of rhesus macaques, Fetz [17] developed a paradigm for behaviorally reinforcing changes in the rate of action potential discharge from single cortical neurons. Using a custom "activity integrator" circuit, animals were provided with auditory or visual feedback of the real-time firing rate of a target neuron. This feedback acted as a discriminatory stimulus, which the subject was trained to control. When the firing rate passed a threshold, specified by the experimenter, a reward was administered. Fetz showed that non-human primates were capable of intentionally and systematically varying the activity of single cortical neurons. Such variation most commonly manifested as phasic increases in firing rate, which were referred to as "operant bursts" of activity. After initial training sessions, average firing rates were reported to change, under reinforcement, by as much as 500% of their baseline levels. Once trained, "monkeys consistently and rapidly increased the activity of newly isolated cells", indicating that such control was not overly challenging for the animals. These results set the foundation for decades of subsequent research in cortical conditioning, and brain interface, research.

Fetz had come to develop the cortical conditioning paradigm as a response to the innovative work of Evarts, which had established behavioral electrophysiology as a means for studying the cortical control of movement. In the years preceding Fetz's initial operant conditioning experiments, Evarts had become the first to successfully record from cortical

neurons in awake, behaving primates. He had convincingly demonstrated that the firing rates of pyramidal tract neurons (PTNs) correlated with gross motor behavior [16], and had suggested that this correlation was indicative of a relatively direct, and causal, relationship between PTN activity and volitional movement [15]. Moreover, he reported evidence suggesting that the activity of PTNs was more closely associated with the force required to realize a particular movement, rather than the movement itself [14].

During the course of these seminal studies, an increasingly refined characterization of the relationship between cortical activity and volitional movement was made possible by an increasingly constrained behavioral paradigm. Initially, Evarts had studied "natural" and spontaneous movements, because they "allowed PT neuron discharge to be observed under a wide range of physiological conditions". However, he considered such passive observation to be "unsuitable for precise quantitative analysis" [16]. He noted that "such movements allowed detection of grossly apparent relations... but were entirely too uncontrolled to permit analysis of certain other relations" [15] – such as the manner in which cortical activity varied with the type and phase of movement. In order to achieve such experimental control, Evarts turned to operant conditioning.

Developing a constrained behavioral paradigm enabled Evarts to more reliably identify a relationship between cortical activity and particular movement-related variables, but it simultaneously limited the scope of the conclusions that he could make. The highly stereotyped movements encouraged by operant conditioning served to stabilize an inconsistent relationship with cortical activity, but they also induced behavioral correlations that confounded broader causal analysis. This fundamentally restricted interpretations to relationships among only those variables included in the behavioral design.

As a particular consequence of this confound, Evarts cautioned that the strong relationship between PTNs and force could not be taken to imply a direct relationship between PTNs and specific skeletal muscles – the effectors of force. In his own words [14]:

Even a seemingly simple movement involves the activity of many muscles and a slow, controlled movement of the type which has been investigated in the present study involves cocontraction of agonist and antagonist muscle groups. In addition, wrist movement depends not only on the forearm musculature but also on the maintenance of arm posture by the musculature of the entire upper extremity. Thus, when a unit is found to be related to the wrist movement, it cannot be inferred that this unit is specifically related to one of the

prime movers of the wrist. Ideally, the monkeys employed in this study should have been trained to carry out a wide repertoire of movements which involved differential activity of each of the numerous muscles of the upper extremity and the trunk, but the difficulty of achieving this degree of training seemed so great that it was not attempted.

It was this challenge that motivated Fetz to develop the cortical conditioning paradigm. Later recapitulating this earlier assessment, Fetz and Finocchio noted that Evarts had reinforced response patterns that "involved coordinated activity of many muscles of the responding limb, and therefore were not designed to resolve the question of which specific muscles a given cortical cell may influence" [18]. However, instead of pursuing experiments involving a "wide repertoire of movements", as Evarts proposed, they instead suggested that an "animal could be trained to activate specific cells or muscles directly" [17] to overcome this obstacle. They proposed that biofeedback could serve as a means for isolating and manipulating each element of the system, in turn, and thereby testing "functional relations between neurons and muscles".

1.2 THE RELATIONSHIP BETWEEN CORTEX AND MUSCLES

In their initial attempts to act on this proposal, Fetz and Finocchio [18, 19] aimed to "determine the degree to which precentral cell activity may be correlated with specific limb muscles and to test the stability of such correlations during different behaviors". Toward this end, they utilized the operant conditioning paradigm to reinforce distinct patterns of "motor activity" involving movement of the arm, isometric contraction of arm muscles, and intentional bursts of activity in the arm/hand area of motor cortex. When the limbs were manipulated by the experimenter, as the monkey sat passively, Fetz and Finocchio found that the firing rates of most cells varied with the movement. Similarly, the activity of most cells varied during active movements, in the manner reported by Evarts. Fetz and Finocchio's most significant results, however, were observed during isolated control of individual muscles or neurons.

In addition to feedback derived from cortical activity, the "activity integrator" device developed by Fetz [17] was capable of delivering muscle-related biofeedback. This apparatus enabled Fetz and Finocchio to train subjects to repeatedly contract individual muscles of the limb, in relative isolation, as they recorded from cortical neurons. Fetz and Finocchio [18, 20] provide an illuminating example use scenario for the activity integrator:

The use of the activity integrator can be illustrated by the procedure for differentially conditioning isolated activity in a specific muscle. If isolated biceps activity was desired, the weighting factor for that channel was made positive so that biceps activity drove the integrator toward the reinforcement level. To condition simultaneous suppression of the other three muscles, the weight factors on these channels were made negative so their activity drove the integrator voltage away from reinforcement level and prevented reinforcement. To eliminate unit activity from the reinforcement contingency, the weighting factor for that channel was made zero. At the beginning of a reinforcement period the monkey typically emitted simultaneous bursts of EMG activity in several arm muscles every 2-3 sec. The gains were then set to reinforce approximately half of these burst responses. As the monkey emitted a greater proportion of reinforced bursts the gains were continually adjusted to differentially reinforce only the closest approximations to the required pattern. Terminal performance typically consisted of repeated bursts of EMG activity in the reinforced muscle with negligible coactivation of the other three. After recording 50-100 responses with a given muscle, the procedure was repeated with each of the other three muscles. After approximately 8 weeks of training one monkey could reliably contract each of the four muscles in isolation in a given session; the time required to shape isolated activity in a given muscle decreased to a few minutes.

Fetz and Finocchio observed that the majority of observed neurons ($N=18$) tended to co-activate, to varying degrees, with several muscles [20]. Some neurons, however, did not correlate with activation of any of the observed muscles. Still others correlated with all observed muscles, in a uniform manner. They tentatively interpreted these patterns as suggestive of a "higher order relation than a one-to-one correlation with specific muscles", and noted that neighboring neurons in cortex could "exhibit quite different sets of correlations with the same set of muscles". They contrasted these characteristics with "motoneurons" of the spinal cord, which tend to exhibit fairly uniform connectivity patterns, such that motor units are restricted to single muscles.

Fetz and Finocchio questioned whether the reverse might be true: would reinforcement of cortical neuron activity also elicit co-activation of multiple muscles? They defined the "motor field" of a cortical neuron to be "the set of muscles coactivated with operant bursts of the cell", and conditioned several cortical neurons to emit bursts of activity on command. They

did, indeed, observe co-activation of muscles for 4 of the 6 neurons tested.¹ In agreement with the patterns observed during isometric contraction, they found that these neurons seemed to have divergent motor fields. They noted some overlap, but significant differences, in the motor fields of nearby neurons. Interestingly, the relative amplitude of average EMG activity in different muscles tended to remain consistent across single unit reinforcement sessions; that is, the amplitude of muscle activity correlated with a particular cell might vary, but the pattern of relative activity – across muscles – was usually stable, for a given cell. As Fetz and Baker [21] put it, the rank order of correlated muscle activity was a stable property for a given cortical cell. This was not necessarily true for different units; the pattern of muscle activity sometimes changed when the reinforcement contingency changed. This supports the concept of stable "motor fields" associated with individual motor cortical neurons [20].

What do these correlations mean? Do they imply a functional or anatomical relationship? Fetz and Finnochio suggest that a "powerful test of the consistency of an observed correlation is to directly reinforce its dissociation". They regarded consistency across behavioral contexts to be an important feature of "strong" correlations: if an observed correlation is not consistent, then it suggests that the functional relationship between neuron and muscle is a consequence of the behavioral context (i.e., whatever behavior is operantly conditioned) to a greater extent than it is of intrinsic, or anatomical, connectivity. With this in mind, they attempted to dissociate, via differential conditioning, the activity of neurons and muscles that had previously been found to correlate. Attempts to reinforce elevated cortical activity, while suppressing muscle activity, were "invariably successful": the monkey was able to control the activity of cortical neurons (N=4) without significantly correlated activity in any of the observed muscles. The inverse dissociation – reinforcement of muscle activity with simultaneous suppression of cortical activity – was twice attempted. Only one case was reported, and it was considered unsuccessful by the authors.

Despite small sample sizes, almost exclusively qualitative results, and a failure to reverse the dissociation protocol, the authors interpreted these results as evidence for a fairly flexible motor system. To summarize their results, Fetz once again contrasted cortical motor neurons

¹Both units that were not associated with co-activation of arm muscles responded to passive shoulder movement, and the authors speculate that they might have been more closely associated to (unmeasured) proximal muscles.

with those of the spinal cord. He noted that, unlike spinal motor units, "whose activity cannot be readily dissociated – if at all – from activity of homonymous motor units, motor cortex cells appear to have more flexible relations to muscle activity and force" [23]. During subsequent decades, Fetz and Finnochio's work would be cited as evidence for a profoundly labile cortico-motor system [38, 25].

1.3 AN EXPLORATORY CORTICAL CONDITIONING PARADIGM

Although the biofeedback paradigm was developed for the purpose of directly testing relationships between specific cortical neurons and specific skeletal muscles, Fetz and colleagues discovered that it could serve a more exploratory function. Contrasting such a "reverse experimental strategy" with the constrained behavioral paradigm of Evarts, they note that "motor activity was allowed to occur under relatively unrestrained conditions" during cortical reinforcement [21]. They hypothesized that "if a cell in the precentral motor cortex were involved in a specific movement, this movement might be emitted with conditioned bursts of activity in the cell". In constrained movement experiments, it is often the case that recorded cortical neurons will show little correlation with the specific behavior that was a priori chosen for reinforcement. For example, Evarts [15] recorded from 182 neurons while a monkey performed wrist flexion and extension, but only 57 of these showed a relationship to the movement suitable for his analysis. Fetz and Baker reasoned that they might more efficiently focus their inquiry by observing overt behavior associated with operantly conditioned cortical activity.

Fetz and Baker reported overt motor activity during reinforcement of firing rate changes for most, but not all, cortical neurons. Monkeys were seated in a primate chair, with head restrained, but with arms and legs relatively free to move. Emitted movement patterns varied from cell to cell, suggesting that the subjects "did not simply adopt the strategy of producing a single generalized motor pattern which was effective in activating all recorded precentral units". The authors note substantial diversity in responses, especially for such a localized cortical area. While reinforcing high firing rates, Fetz and Baker grouped overt

movement into 3 classes: specific, variable, and none. An example of a specific movement that might be emitted with each operant burst would be flexion or extension of a single joint. Variable movements were characterized by gross motor behavior that was not consistent from operant burst to operant burst. In some sessions, no correlated movements were observed, and the subjects might sit "quite still".

Once a motor field had been roughly identified, through observation, it could be more carefully characterized by adjusting behavioral constraints. Fetz and Baker [21] provided a concrete example of how this might happen. They noted a particular cortical conditioning session, wherein "it appeared that operant bursts were associated with dorsiflexion of the ankle". To quantify this covariate, Fetz and Baker fitted the foot with a cast that was hinged at the ankle. A potentiometer recorded the ankle joint motion in the sagittal plane, and confirmed that the most common motor correlate was dorsiflexion. However, plantar flexion was found to accompany operant bursts in approximately one third of trials. As a further refinement, electromyographic activity was also recorded from an ankle flexor and extensor, and isometric contraction was reinforced, in the manner of Fetz and Finocchio. In this way, Fetz and Baker iteratively refined their understanding of the relationship between the cortical neuron and movement – starting from a relatively coarse description of the non-contingent motor response – by successively adding more specific behavioral constraints, and realizing successively more specific behavioral contingencies. Largely owing to the technical burden of such an experimental paradigm, however, such iterative refinement were limited. Fetz and Baker note that a "severe limitation" of their approach was "the degree of precision with which any concomitant motor response could be documented".

Non-contingent covariates were not limited to muscle activation and motor behavior. Fetz and Baker [21] extended the concept of a cell's "motor field" to include other cortical neurons:

Although delivery of reinforcement was made contingent on the activity of a single cortical unit, that unit was obviously not the only one whose activity changed. With operant bursts, one can argue a priori that there should be some correlated activity in other cells and probably in muscles. At the least one would expect those cells involved in controlling the reinforced unit to undergo correlated changes in firing rate. In addition one might expect to see concomitant activity in other neurons and muscles involved in a wider correlated response pattern, of which the operant unit burst may be a part. Since any influence of

the motor cortex on movements would ultimately involve the correlated output of many thousands of neurons, the relationship between activity in different precentral cells is of considerable interest.

To assess interactions among nearby units, Fetz and Baker made an effort to isolate and simultaneously record pairs of units on the same electrode. Initially, they adopted the exploratory approach: firing rates for a primary unit were reinforced, and the behavior of the secondary unit was observed. Typically, the units were positively correlated, such that the average rate of the secondary unit also increased during reinforcement periods. In 1 of 9 sessions, the average rate of the non-contingent unit was slightly suppressed. In 4 sessions, the secondary unit actually showed a larger percentage change in the average rate than the primary, conditioned unit – as much as 729% of the baseline rate. Clearly, the unit upon which reward was made contingent was not the only unit to systematically vary during reinforcement periods.

In much the same way that Fetz and Finocchio had tested dissociation of the activity of specific neurons and muscles, Fetz and Baker refined their characterization of "cortical fields" by testing dissociation of the activity of pairs of neurons. They set up a differential reinforcement contingency, in which reward was dispensed for increasing the firing rate of one cell, while suppressing the other. They tested this contingency for 3 unit pairs. In 2 cases, the target activity pattern was successfully conditioned. Fetz and Baker interpret their results as convincing evidence that the firing rates of pairs adjacent cortical units – separated by no more than a few hundred microns – can be controlled independently. As with Fetz and Finnochio, the results of Fetz and Baker [21] came to be cited as evidence for profound cortical plasticity [25], despite the largely anecdotal character of their report.

1.4 CHARACTERIZING INTERACTIONS AMONG CORTICAL NEURONS

Fetz and Baker were neither the first, nor the last, to consider the importance of correlation among the firing rates of motor cortical neurons, and its relevance to behavior.

Early-on, Evarts [16] had remarked on the apparent significance of interactions among nearby cortical neurons and movements. In 13 cases, Evarts was able to compare the activity of adjacent units, measured simultaneously, on the same electrode. He found that the "correlation between discharge frequencies of adjacent PT neurons varied depending upon the particular movement which the monkey was making". At times, the pairs could be positively correlated, and at times negatively. Evarts notes that the correlation pattern could be quite "striking". Phase relationships were also observed. The relationship between firing patterns appeared to depend on the type and/or stage of movement. Patterns of correlation were also observed between cortical discharge and electromyographic (EMG) activity of the contralateral arm.

Around the same time that Fetz was developing the cortical conditioning paradigm, Humphrey, Schmidt, and Thompson were taking a different approach to refining Evarts' results [44]. They were similarly motivated by the question of what motor variables are "controlled" by motor cortex. Adopting Evarts' behavioral paradigm, Humphrey et al. extended his electrophysiological technique to record simultaneously from multiple, chronically-implanted electrodes. Significantly, they showed that it was possible to predict the temporal evolution of movement-related variables, with relatively high resolution. They suggested that "information about a given movement is carried not simply in the discharge patterns or spike trains of individual cortical neurons, but to a significant extent by the temporal relations between them". Humphrey et al. noted that "neuronal spike trains can be used for quantitative prediction of simple motor responses", which supports the feasibility of using "such signals for control of prosthetic devices". This observation largely anticipated Schmidt's subsequent research trajectory.

Several years later, Schmidt et al. [40] sought to extend the work of Fetz et al., by recording from the same neurons for periods of hours to days. Fetz et al. had endeavoured to characterize cortical activity across as many motor behavioral conditions as possible, but were limited by the 6-10 hours span within which they could reliably record from an isolated neuron [19]. Consequently, their reported results are a patchwork of permutations of neurons and conditions. As previously noted, this restricted them to largely qualitative observations, and prevented strong statistical conclusions from being drawn. Schmidt et al.

successfully demonstrated a chronic recording technique that allowed them to condition the same neurons, across multiple experimental sessions. Over the course of 99 days, Schmidt et al. monitored the activity of what they estimated to be 28 distinct neurons, from 6 chronic electrodes, which they subjected to a reinforcement contingency derived from the work of Fetz et al. They report observing what appeared to be a stable neuron for as long as 28 days. Thus, Schmidt et al. demonstrated the feasibility of characterizing individual cortical neurons, across a wide array of behavioral conditions, and an extended interval. They did not, however, attempt to reproduce the muscle and movement experiments of Fetz et al. [21, 20].

However, Schmidt et al. did extend the results of Fetz and Baker [21], by carefully quantifying correlations between the firing rates of nearby neurons [40]. During each experimental session, the activity of a "control neuron" was monitored, but not included in the operant conditioning reward contingency. Generally, this was chosen to be the most discriminable unit, on any electrode except the one on which the conditioned unit was found. We will here refer to the unit targeted for reinforcement as the "contingent neuron", and the control neuron as the "non-contingent neuron" of the pair. A correlation coefficient was computed for each of 29 unit pairs, by dividing successful trials into 50 bins of 5 ms width. The electrodes with the most highly-correlated pairs of units were 3.8 mm apart. Of the 29 unit pairs – on different electrodes – the average correlation magnitude was greater than 0.2 for 8 (28%) pairs, and greater than 0.3 for 5 (17%) pairs. Thus, the activity of non-contingent neurons co-varied with that of contingent neurons, during reinforcement, to at least some extent.

Although Schmidt et al. did not report statistical significance, or controls, their correlation results furnish an interesting extension to the pairwise comparisons of Fetz and Baker. They provided slightly more than a qualitative characterization, and showed that correlated activity was not limited to units recorded on the same electrode. These preliminary data support the notion that volitional control of single unit activity is accompanied by correlated patterns of variation in local population activity – or across a distributed "cortical field".

1.5 THE SIGNIFICANCE OF CORRELATED ACTIVITY

Ultimately, Fetz and colleagues cautioned that the correlations observed in the operant conditioning paradigm were not adequate for reliably assessing connectivity in the motor system. Fetz and Finocchio note that temporal correlations are not sufficient to establish that a functional connection exists, since "a precentral cell and muscle could be co-activated during a given response without being connected" [20]. Inversely, they point out that a neuron and muscle could, in fact, be varied independently, even if there is a direct synaptic connection between the two. This is because of the convergence of neurons onto muscles, and the need for coordinated excitation from a large number of neurons, in order to activate a muscle. Thus, strong temporal correlation is not even a necessary consequence of anatomic connection.² Fetz and Finocchio note that "more direct physiological tests" are necessary to test for synaptic connections, with confidence. However, they consider strong correlations between a neuron and muscle to "operationally define a 'functional relationship'" – where the correlation "strength" is "proportional to both the intensity and consistency of ... coactivation under different behavioral conditions". A functional relationship cannot prove the existence of anatomic or functional connectivity, but it strongly suggests that the two elements are relevant to each other, and provides guidance for further exploration and refinement, A similar logic applies to correlations between the activity of cortical neurons.

This perspective – that the combination of the operant conditioning paradigm and coarse correlation analysis was not sufficient for generating reliable connectivity estimates – might have prompted Fetz to concentrate on what he considered to be more powerful alternatives. Operant conditioning was a means to an end for Fetz: it represented an experimental tool for illuminating the functional relationship between cortical neurons and muscles. By 1975, Fetz et al. expressed the opinion that more sensitive methodologies – such as cortical stimulation and spike-triggered averages (STA) – were better suited to this problem. Such alternatives became the focus of Fetz's research for the next several decades. Fetz's subsequent com-

²This seems to contradict the prior statement, by Fetz and Finocchio [18], that a "consistent temporal correlation between two events, such as precentral cell activity and some component of the motor response (force, position, or activity of a specific muscle) is necessary but never sufficient evidence for a causal relation between the correlated events".

mentary on operant conditioning theory and methods [23, 24] was largely limited to the recapitulation of previous results, and reflections on their interpretation, until a resurgence of interest, around the turn of the millennium.

1.6 A QUESTION OF MECHANISM AND SCOPE

Although Fetz turned his attention away from cortical conditioning, an early clinical collaboration evolved into a parallel research program that would complement and extend his foundational work. As he was developing the motor cortical conditioning paradigm, Fetz established a relationship with Dr. Allan Wyler, a neurosurgeon. Together, they sought to explore the use of operant conditioning as a technique for investigating and treating epilepsy [22, 30]. They hypothesized that epilepsy might be alleviated by training patients, using biofeedback, to control pathological patterns of cortical activity. The cortical conditioning paradigm continued to be a primary focus of Wyler’s research program well into the 1980s. Of particular relevance to the present discussion are a series of studies in which Wyler and colleagues characterized the selectivity of single unit operant conditioning, with the aim of illuminating the mechanism underlying the volitional control of cortical activity.

Wyler et al. were interested in the question of how cortical ”operant control is mediated” [31]. They considered two strategies for such control – primarily distinguished by their degree of specificity. First, they suggested that changes in cortical activity patterns could be due ”merely to the monkey alerting and increasing muscle tone”. Second, they suggested that the monkey might search for ”subtle specific peripheral movements that coactivate with the firing of the central unit”. They reasoned that former strategy would activate a relatively disperse population of cortical neurons – that is, it would have a global effect – whereas the influence of the latter would be restricted to a more localized population.³

³It’s worth noting that Fetz had previously dismissed the former possibility, to some extent, citing his ”yoked control” results [17] and the diversity of movement patterns emitted during reinforcement of different cells [21].

To test these alternatives, Wyler et al. recorded unit activity bilaterally, from homologous regions of precentral cortex. While simultaneously observing the activity of a single unit in each cortical hemisphere, they applied a cortical conditioning paradigm to each unit in turn, and found that conditioning was successful for the majority of the 38 unit pairs considered. Across all cases, they noticed a change in activity patterns, from baseline, for both the contingent and non-contingent neurons. However, the activity of contingent neurons continued to change – becoming more task appropriate – with practice in the operant task, whereas the non-contingent neurons did not. Moreover, the effect was reversed when the contingency was transferred from one neuron to the other. In no case was any significant correlation between changes in the bilateral units’ firing patterns found. Since the activity of one unit was largely independent of the unit in the opposite hemisphere, Wyler et al. concluded that control of single unit firing rates is accomplished via specific, rather than generalized, behavioral processes.

Wyler next sought to further “investigate the specificity of single-unit operant conditioning” [32]. He interpreted existing evidence to indicate that “the operant response was a general one rather than specific to the neuron”, but it remained to determine just how general it was. Having been convinced of the absence of global correlation, he chose to explore the opposite extreme, by assessing correlation between cortical neurons in close proximity. Adopting the approach of Evarts, Fetz, and Baker before him, Wyler isolated pairs of neurons recorded on the same electrode. He attached special importance to the idea that recorded units were situated within the same “cortical column”, and could therefore be expected to share similar functional properties. He expressed difficulty in reconciling this with the apparent independence of such pairs, as reported by Evarts [15] and Fetz and Baker [21].

While reinforcing changes in activity patterns for one unit, Wyler computed firing rate statistics for each pair. For 78% of 133 unit pairs, “the contingent unit was brought under operant control without a significant or consistent change in the firing pattern of the second unit”. Firing rates were estimated using 15 second bins, and correlation coefficients were computed across 5 minute blocks. It is worth emphasizing that these were rather coarse

firing rate estimates; 3000 times the length of those considered by Schmidt et al. [40].⁴ In 90% of cases, the correlation coefficient was not significant for any block. For 98% of unit pairs, Wyler was especially surprised to report that correlations during reinforcement blocks were decreased, relative to time-out blocks. There were approximately equal numbers of positive and negative correlations, across all blocks. Moreover, "the magnitude and sign of the correlation coefficients varied randomly, from one operant period to the next, throughout individual experiments".

Altogether, Wyler conceded that these results did not paint a coherent picture. A lack of significant correlation among cortical neurons suggests that operant conditioning was, indeed, highly specific to single units. Wyler considered this to be an unexpected and confusing outcome. In particular, it did not explain "natural" correlations, reported previously [21, 40]. In this regard, Wyler suggested that "it may be that the dissociation that Fetz and Baker found between units was not entirely dependent on the operant paradigm, but may have been also secondary to the normal variability in firing rates that two adjacent units may have".

Not satisfied with this result, Wyler revisited the pairwise correlation analysis [33]. Using the same experimental paradigm and technique, he confirmed his previous observation: in 80% of cases, there was no significant correlation between 15 second firing rate fluctuations. He then took the analysis a step further, by computing fine-timescale, spike-triggered correlation histograms. For each spike of the unit upon which reinforcement was contingent, he counted the number of spikes of the non-contingent unit in 1 ms bins, across an interval spanning from 13 ms before to 13 ms after the primary spike event. This millisecond timescale is more germane to questions of local neural network interactions, whereas broader timescale correlations are more likely to reflect global, behavioral influences on single unit firing rates. Wyler might have reasoned that the dearth of broad timescale correlations reflected the relative specificity of cortical control to the local population.

⁴Fetz and Baker [21] reported many results in terms of firing rates averaged over 1 minute blocks, but also considered peri-event time histograms with finer resolution.

The revised analysis prompted a revised conclusion. In 67% of pairs, Wyler identified what was deemed a significant coincidence of firing, according to an ad hoc significance test. Significance did not seem to depend on correlated fluctuation of the 15 second firing rate estimates. When significant, coincidence of firing was relatively consistent between behavioral epochs. Thus, fine timescale correlations were deemed more common and consistent than broad timescale firing rate fluctuations. Wyler surmised that the initial analysis had not been sensitive enough. He concluded that single unit operant conditioning did, in fact, induce patterns of correlated variability in a population of nearby neurons – perhaps as a consequence of common synaptic input.

In the pursuit of an explanation for the mechanism via which cortical neuron activity might be intentionally varied, Wyler et al. had come to explore the related issue of how specific such variation is. They had asked how distributed patterns of co-variation were – with respect to the local cortical population – during single unit conditioning. Like Schmidt et al. [40], they concluded that the activity of surrounding, non-contingent neurons did co-vary with the activity upon which behavioral goals were contingent, but that patterns of covariation were restricted to the local population, and were not a generalized cortical phenomenon.

What is the nature and significance of such patterns? Humphrey et al. had previously demonstrated that information about simple arm movements could be found in correlations among firing rates of cortical neurons within a local population [44]. Is similar information evident during reinforcement of single unit activity? To what extent are patterns of non-contingent co-variation equivalent to the patterns observed during natural movement? Possible clues would come from a more naturalistic approach to cortical conditioning.

1.7 NATURAL MOVEMENT AS A BRIDGE TO CORTICAL CONTROL

Whereas derivatives of the work of Fetz et al. were often characterized as efforts to directly condition the activity of cortical neurons – without reference to the ”normal” functional role of those neurons – parallel efforts took a more ”natural” approach. Such investigations used

patterns of cortical activity induced by movement as a bridge for transitioning between operant responses involving gross motor behavior (like those reinforced by Evarts) and operant responses involving only modulation of cortical activity (like those reinforced by Fetz). In order to accomplish this, however, they first had to reliably characterize the relationship between natural movement and patterns of cortical activity.

A particularly influential series of experiments came from Georgopoulos et al. (for review, see [8, 11]). Inspired by the observation of movement-related correlations in parietal cortex, Georgopoulos trained monkeys to perform instructed reaching movements between planar targets in extrapersonal space. Like Evarts' early button-press task, subjects were conditioned to make relatively natural movements, under a behavioral contingency that depended only on the motion of the hand, and not on the evolving state of the underlying musculoskeletal system. However, unlike Evarts' task, Georgopoulos et al. systematically distributed targets throughout the workspace, in order to enforce large-amplitude movements, involving multiple muscles and joints. As the subjects performed this task, they recorded the activity of neurons in primary motor cortex.

Georgopoulos et al. [6, 4] found that the discharge rate of M1 neurons appeared to vary, systematically, with the reach target. They showed that this relationship was well-fitted by a cosine function, and formalized this observation in what came to be known as the 'cosine tuning model'. This model suggests that the firing rate of an M1 neuron (dependent variable) is 'tuned' to the angle of movement (independent variable), and that the two quantities are related by a cosine function. The peak of the cosine – the point of highest discharge rate – is situated at what is referred to as the 'preferred direction' of the cell. The firing rate of a cell declines as the angular difference between reach direction and the cell's preferred direction increases.

Georgopoulos et al. [6, 5, 3] further showed that the cosine tuning model could be used to infer the intended direction of hand movement, from populations of M1 cells. According to the model, knowledge of the target movement direction would be sufficient for predicting the expected firing rate of a cortical neuron. However, the redundancy of the cosine transformation – along with the imperfect / noisy nature of the model – made it necessary to observe two or more cells in order to compute the inverse transformation. Among measured cells,

preferred directions tended to be uniformly distributed in space, so that the representation of movement improved with the size of the observed population. By scaling the preferred direction of each cell by its firing rate, and finding the weighted average across neurons, they showed that it was possible to reliably estimate the direction in which the monkey intended to reach, prior to the initiation of movement. They referred to this resultant vector method as the 'population vector algorithm',

Georgopoulos et al. [3] demonstrated that time-varying estimates of movement intention could be synthesized into faithful movement trajectories. Such 'neural trajectory' estimates of movement intention were formed by computing population vectors at 20 ms intervals and integrating across time, to obtain a time series of hand position estimates. Schwartz [2] extended this analysis to more complex movements, by training monkeys to trace various shapes (e.g., spirals) in space. He showed that the population vector method reliably predicted movement trajectories, even when the direction of intended movement changed frequently. Moran and Schwartz [1] verified that the length of the population vector scaled with the speed of the movement, further validating the integrated population vector algorithm. Altogether, these results suggested that hand movement trajectories could be accurately predicted, from the activity of populations of M1 neurons, in real-time.

An important implication of the cosine tuning paradigm is that the activity of M1 hand/arm area neurons must be substantially correlated during hand motion. The model suggests that M1 neurons are driven by some common, low-dimensional, underlying process, so that movement-related variation must be shared across the entire population. In other words, every neuron of the population responds to the same 3 variables (hand velocity), which constrains the set of population patterns to 3 dimensions. The model does not, however, specify to what extent other factors influence population activity, or how stable such patterns of shared variation are across behaviors. Does correlated, movement-related activity reflect a labile structure – some sort of cell assembly [39] – dynamically organized to suit the prevailing behavioral context (i.e., reaching to targets)? Or, is the pattern of shared variation a relatively fixed property – perhaps reflecting the intrinsic connectivity of the network – which does not vary with behavioral goals?

To explore these questions, Taylor et al. [7] manipulated the relationship between population activity and behavioral goals. Beginning with the standard center-out reaching paradigm, developed by Georgopoulos et al., they transferred the behavioral contingency from the actual motion of the hand to a population-derived estimate of that motion, as obtained via the population vector algorithm. This closed the loop between cortical activity and effected action, making behavioral goals dependent on the observed population exclusively, and causing motion of the arm to become functionally irrelevant. They report that arm movement gradually subsided, following the change in contingency, and that measured EMG became attenuated. Although center-out task performance initially declined – presumably reflecting mismatch between intended and ”decoded” motion – it gradually improved, with continued practice.

Taylor et al. [7] reported that patterns of shared variation in cortical activity changed with the behavioral contingency. They quantified such changes in terms of the cosine tuning model – in particular, they considered changes in preferred directions. They reported consistent shifts in the preferred directions of individual units, from day to day, but a lack of consistency in the direction of the shifts across the population. Preferred directions continued to shift with practice, following the change in the behavioral contingency. Significantly, such gradual shifts correlated with improving task performance, indicating that the changes were non-random, and were driven by a behaviorally-relevant process.

1.8 LOCAL AND GLOBAL PATTERNS OF POPULATION ACTIVITY

Altogether, these results indicate that the cosine tuning model does not perfectly generalize across behaviors, and that patterns of shared variation implied by the model are, in some sense, labile. The results of Taylor et al. suggested that changes were driven by a process of behavioral optimization, but the nature of this optimization was not entirely clear. How did behavioral goals drive changes in patterns of shared variation in the observed population? How were such changes organized? One obstacle to answering these questions was the

fact that the change in the behavioral contingency could not be precisely quantified, since the mapping between the observed population and arm movement was only approximately known.

Jarosiewicz et al. [9] sought to more systematically explore behaviorally-driven changes in patterns of population activity, by more carefully controlling changes in the behavioral contingency. In a manner similar to the brain-computer interface experiment (BCI) of Taylor et al. they first established a contingency based on the population vector algorithm. This provided a baseline condition, so that the mapping between population activity and behavior was precisely known. Once performance of the center-out task had stabilized, in this condition, Jarosiewicz et al. systematically perturbed that mapping. Specifically, for a subset of units – either 25% or 50% – the preferred directions used to obtain the population vector estimate were rotated by 90 degrees. Initially, this caused a substantial performance deficit, but the subjects corrected for the perturbation with practice. As reported by Taylor et al., the shift in performance was correlated with a shift in patterns of population activity.

Jarosiewicz et al. proposed three possible ways in which population activity might be adjusted to compensate for the perturbation. They referred to these mechanisms as reaiming, reweighting, and remapping. Each of these adaptive strategies modified the local tuning properties of the observed population to varying degrees. At one extreme, reaiming entails no change in local tuning properties, and affects all neurons of the population equally. At the opposite extreme, remapping entails specific corrections of the tuning of individual units. Reweighting also falls at this more local extreme.

Jarosiewicz et al. reported that a re-aiming mechanism seemed to best account for the compensatory changes to patterns of population activity [9]. Re-aiming occurs when the subject aims at an imaginary target in the task space, such that effected cursor motion counteracts the net rotation caused by the perturbation. This is effectively a global mental rotation, and manifests as a single, common rotation of the preferred direction of all units. Arguably, this is the most "intuitive" possibility, since only high level behavior, and not the relationships between neurons, needs to be modified. Re-aiming was deemed to be a sub-optimal strategy, since it will result in a shortened overall length of the population vector, given the same unit modulation.

Jarosiewicz et al. [9] also observed some local changes, at the level of single units. On average, they found that the units subjected to perturbation showed a significant change in the modulation depth (i.e., amplitude of the cosine tuning curve) following the perturbation, whereas the remaining units did not. They interpreted this as evidence for the re-weighting mechanism. Moreover, they reported significant shifts in the direction of the imposed perturbation, at the level of individual units. This is consistent with the re-mapping strategy. Although local tuning properties were modified, the effects were more subtle than for the re-aiming corrections. Jarosiewicz et al. concluded that the "dominant response was a global change in the activity level of all of the neurons".

All of the units that Jarosiewicz et al. considered were part of the behavioral contingency, but previous cortical conditioning studies had reported correlations between contingent and non-contingent neurons. Could the global effect depend on the size and composition of the population used to define the behavioral contingency? If nearby neurons are not directly relevant to behavioral goals, then are behaviorally-induced changes still evident in their activity?

Ganguly et al. sought to answer this question by excluding a subset of units from the behavioral contingency [35]. Monkeys were required to perform a center-out cursor task, where motion of the cursor was a linear function of the firing rates of a small population of 10-15 M1 or PMd neurons. This behavioral contingency was similar, but not equivalent, to the population vector algorithm of Taylor et al. During each experimental session, the monkey also performed a center-out arm-movement task. Ganguly et al. reported changes in the preferred directions of all recorded neurons – between cortical conditioning and reaching blocks – whether or not they were part of the behavioral contingency. This suggests a relatively global effect, like re-aiming. These changes were stable over time, for both groups. However, once proficient control had been established – after several days of practice – they reported that modulation depths differed less, between hand movement and cortical conditioning, for the non-contingent neurons than they did for behaviorally-contingent neurons. This recalls the results of Wyler et al. [31]. They compared this observation to the re-weighting

mechanism proposed by Jarosiewicz et al., and interpreted it as evidence for "differential modification" of behaviorally-relevant neurons. This is a non-global effect. Altogether, the results of Ganguly et al. seemed to corroborate the observations of Jarosiewicz et al.

1.9 TWO KINDS OF CONSTRAINTS ON CORTICAL ACTIVITY PATTERNS

What determines the extent to which activity patterns can be shaped to behavioral goals? The early results of Fetz and Baker [21] suggest that the activity of individual motor cortical neurons can be conditioned to satisfy new behavioral contingencies. The results of Georgopoulos et al. suggest that there are limits to this flexibility, perhaps imposed by the role of motor cortex in the control of movement. The results of Taylor et al. and Jarosiewicz et al. seem to confirm this suspicion, while emphasizing that aggregate behavior is likely a mixture of local and global processes. Addressing the "degree of volitional control" of cortical activity, Fetz proposed two limitations on the extent to which individual elements of a cortical population can vary independently [25].

First, Fetz suggests that independent variation of cortical neurons might be limited by incomplete knowledge concerning the relationship between cortical activity and effected action. In that case, a subject only imperfectly understands the environmental consequence of emitting particular activity patterns. He maintains that this could be especially prohibitive in the case of "complex transforms", like that of the population vector algorithm of Georgopoulos et al., where actions are determined by the collective activity of a large number of units. The implication is that each neuron involved in the contingency adds to the dimensionality of the set of behaviors through which a monkey must search, to control cortical activity. He suggests that "in contrast to the relatively simple task of driving one or two cells in bursts while allowing free performance of any correlated responses, the requirement to modulate activity of a population to accurately control a transformed function may be more difficult because the effect of any particular cell is largely submerged in the population function". This is fundamentally a learning problem.

Second, Fetz proposes "ensemble interactions" as a less mutable limitation on the independence of cortical neurons. Noting that "internal representations depend on relationships between the activities of neurons in an ensemble", he suggests that "the processing of these representations involves corresponding constraints on the independence of those activities". In other words: to represent or manipulate any form of information, a cortical population must necessarily establish some manner of functional relationship between neurons. Fetz suggests that the functional relationships induced by the "natural" role of a given population might restrict the degree to which that population can be re-purposed for operant behavior. He offers an analogy with Henneman's size principle, which constrains the order in which motor units and muscle fibers can be recruited: high threshold motor units cannot be activated independently of low threshold motor units. Thus, operant conditioning of high-threshold motor units fundamentally cannot dissociate those units from the low-threshold group, as a consequence of physiological connectivity. A similar constraint might apply to cortical neurons.

1.10 SIMPLE TASKS FOR PROFOUND FLEXIBILITY

To explore these limits on cortical activity patterns, Fetz et al. compared motor cortical activity during natural movements with that observed during cortical conditioning. Employing a single unit operant conditioning procedure, derived from that of Fetz and Baker [21], Moritz et al. [26] reported fast learning of the task contingency. Each monkey learned to control nearly every tested unit (44 of 45), within 10 minutes of exposure to a new contingency. Moritz et al. then compared performance in the operant task with the strength of directional tuning of the target unit, during a wrist-torque task (8-target center-out). They found that tuning was only marginally correlated with task performance, and interpreted this result as an indication that subjects could easily control arbitrarily-selected motor cortex neurons, "regardless of any previous association to movement". This, they suggest, is evidence for profound flexibility in the patterns of activity that can be generated by cortical populations,

which "considerably expands the source of control signals for brain-machine interfaces". It was interpreted as evidence against the notion that ensemble interactions substantially limit the flexibility of a cortical population.

Moritz and Fetz [27] provided a much more thorough account of their single unit operant conditioning results. They report results for 240 cells recorded from the pre-central gyrus, and 16 cells from the post-central gyrus, with no significant difference in single unit conditioning performance between the two locations. All cells considered were successfully controlled. Monkeys rapidly improved performance at least 2-fold within each conditioning session, and could improve substantially across days, with stable unit recordings. As reported by Moritz et al. [26], performance during single unit conditioning was not correlated with strength of tuning during the wrist torque task.

Moritz and Fetz again reiterate the conclusion that "arbitrary single cortical neurons, regardless of the strength of directional tuning, are capable of controlling cursor movements in a one-dimensional" task. The authors note that they find it surprising that brain-computer interface (BCI) devices based on population decoding schemes – wherein tens or hundreds of cortical cells are operantly conditioned to vary in a particular pattern – do not perform as well as one might expect, given the demonstration of "rapid and robust operant control" for single units. They further suggest that the problem of learning to control the activity of single units might be "inherently simpler than controlling large numbers of neurons simultaneously", thus avoiding learning-related limitations on cortical flexibility. Altogether, Moritz and Fetz seem to conclude that neither learning barriers nor ensemble interactions appear to substantially limit control, or the flexibility of cortical population patterns.

Taking their investigation a step further, Milovanovic et al. [29] tested whether or not monkeys could "independently modulate cortical units and contralateral wrist torque, regardless of the strength of directional tuning". Specifically, they sought to test the hypothesis that units only weakly correlated with wrist torque would be most robust to interference, when the monkey was simultaneously required to modulate cortical activity and make wrist movements; that is, that they would be more flexible. The authors frame their investigation as a question of practical importance to the development of clinically-viable brain-computer

interfaces (BCI). Similar to their previous work, Milovanovic et al. report that all tested pairs of cortical units were successfully used to control cursor motion, and they found no relation between tuning strength and cortical conditioning task performance.

Milovanovic et al. designed a dual control task to test dissociation of hand motion and concurrent cortical activity. This task combined the wrist torque paradigm with the cortical conditioning paradigm, to move a cursor in a 2 dimensional tracking task. As a consequence of the dual control task design, movement to targets along a diagonal axis in the task space required dissociation of pairs of tuned units from the correlation pattern that they had "naturally" exhibited in the wrist tracking task. That diagonal axis was labeled the "hard" axis. Since the orthogonal axis corresponded to wrist motion and unit activity that were naturally in agreement, and therefore had a facilitatory effect, that axis was labeled "easy". A significant difference in performance was reported, between 'easy' and 'hard' targets, for tuned units. The difference for untuned units was not significant. The authors ultimately conclude that "a monkey can learn to control the activity of single units in motor cortex independently of ongoing contralateral limb movements". As stated, the implication is one of perfect dissociation from the natural role of cortical neurons. In other words, constraints deriving from ensemble interactions were effectively deemed inconsequential.

1.11 REVISITING "NATURAL" CONSTRAINTS

Independently of Fetz's group, Law et al. [34] reported similar conclusions. Like Milovanovic et al., they conditioned cortical activity by mapping the firing rates of a small ensemble of neurons – recorded via microelectrode arrays implanted in the motor cortex – to the motion of a cursor. Instead of carefully choosing units with related tuning to movement, however, Law et al. mapped groups of 1-4 randomly-selected units to cursor motion. The mapping was a simple average of firing rates, and the task was again reduced to a single dimension. As for Fetz et al., learning did not seem to be an issue: they report that subjects typically achieved control of such ensembles rather quickly, and that performance improved throughout each experimental session. Performance also improved with larger ensembles.

Since the mapping to cursor movement was selected arbitrarily, Law et al. reasoned that established patterns of correlation among ensemble units during "natural" movement might restrict performance of the cortical conditioning task. This is equivalent to Fetz's suggestion that "ensemble interactions" limit control. In particular, they suggested that ensembles of similarly-tuned neurons might be easier to control than neurons with dissimilar tuning. To quantify the relative similarity of tuning among ensemble neurons, Law et al. computed the average angle between the preferred directions of every possible pairwise combination of active ensemble neurons – where the preferred directions were estimated from data obtained during a center-out joystick task. They compared this value with a performance metric in the cortical control task, and found no significant predictive value. The authors interpret this to suggest that "the pre-existing synaptic architecture responsible for the natural correlations in firing exhibited during normal behavior of the native limb does not limit the rapid adaptation of small numbers of M1 neurons" to an arbitrary behavioral contingency, defined in terms of the activity of those neurons. They suggest that motor cortical activity can be varied "with little constraint based on the normal relationships of the individual neurons to native limb movement"

Law et al. also characterized patterns of activity among neurons that were not included in the behavioral contingency. Such non-contingent neurons were found to co-vary with the behaviorally-relevant ensemble, but the average depth of modulation was consistently less. Moreover, the average modulation depths of contingent neurons increased as the monkeys practiced the task, within an experimental session, whereas the modulation depths of non-contingent neurons did not. The authors interpret this as convincing evidence that control of cortical activity was somewhat specific to the contingent ensemble, rather than a general population phenomenon. This supports the conclusion that small populations of motor cortical neurons are profoundly flexible.

Law et al. contrast their results with that of Ganguly et al. [35]. Noting that Ganguly et al. reported that non-contingent neurons were initially (for the first 2-3 days) modulated as much during cortical conditioning as during a hand movement task, Law et al. suggest that the apparent disagreement with their conclusions might be explained by the number of neurons involved in the contingency: including more neurons in the behavioral contingency

could recruit more non-contingent neurons to co-modulate. They also suggest that neurons recorded from different regions of cortex (PMd, for Ganguly et al.) might behave differently – that is, they might participate in a more distributed, or generalized, control mechanism than neurons of primary motor cortex.

Law et al. also consider similar work from Hwang et al. [41]. Although they experimented with neurons located in the parietal cortex, and adopted a conditioning paradigm more like that of Fetz et al., Hwang et al. interpreted their results in terms of the investigation of Jarosiewicz et al. Citing the suggestion from Fetz et al. [25, 26] that cortical neurons are profoundly plastic, Hwang et al. sought to test "the assumption that individual neurons can be conditioned independently, regardless of their natural movement association". They reported that "monkeys could learn to elicit arbitrarily assigned movement patterns" – agreeing with the observations of Fetz et al. and Law et al. – but that "the seemingly arbitrary patterns always belong to the response set for natural reaching movements". In agreement with Jarosiewicz et al., they conclude that cortical activity is controlled via a relatively global mechanism, like re-aiming, by "preferentially exploring and exploiting" the "natural movement repertoire". They strongly contrast this with the results of Fetz et al. and Law et al. Both Hwang et al. and Law et al. suggest that this difference might reflect a fundamental difference in the properties of precentral and parietal cortex.

1.12 RELAXING ASSUMPTIONS ABOUT "NATURAL" ACTIVITY PATTERNS

It is difficult to fully explore – using only constrained movement experiments – the idea that the flexibility of cortical population activity is constrained by the "natural" role of the population. This is because the relationship between cortical neurons and "natural" behavior is not very well-understood. The activity of motor cortical neurons has been shown to co-vary with a wide array of behaviorally-relevant variables [11]. Consequently, one cannot ensure that an experimental design will sufficiently exercise exactly those natural behavioral variables that are relevant to the population, without a priori knowledge regarding what

those variables are. It is necessary, therefore, to choose – based on prior experience – an experimental task that can be expected to adequately engage the population. However, it is difficult to assess whether such an approach exercises population activity to its full extent. For example, recall that the activity of only 57 of 182 neurons that Evarts recorded from were found to significantly co-vary with the wrist movements that he reinforced [15]. The remainder could have been related to any number of movement-related variables, or combinations of those variables.

Is it possible to characterize "natural" interactions among neurons, without making any assumptions about how such interactions relate to overt movement behavior? Chase and Schwartz [10] proposed that direct cortical conditioning could serve as a means for "identifying the volitional control signals that affect single neurons in motor cortex, without resorting to any kind of explicit motor task". They emphasized that a permissive approach avoids a priori assumptions "about what these driving signals might actually encode". They adopted a technique similar to that of Schmidt et al. [40] – basing their paradigm on that of Fetz, while leveraging modern chronic array recording technology. They trained a monkey to modulate the firing rates of single neurons, while simultaneously monitoring the activity of a surrounding neuron population.

Chase and Schwartz report that "invariably there are typically large correlations in the firing rates of many of the other simultaneously recorded cells". They interpret this to indicate that "the volitional control signal used to drive the target neuron also influences the firing rates of other neurons". In other words, the single unit activity that is relevant to behavioral goals is part of a wider pattern of activity, in the manner of the "motor field" / "cortical field" concept proposed by Fetz et al. They reported that patterns of correlation among cortical neurons were stable within experimental blocks, but varied with the reinforcement contingency. This is reminiscent of the finding, by Fetz and Baker [21], that patterns of muscle activation tended to similarly vary with the single unit reinforcement contingency. The idea that common drivers influence the activity of neurons of a local population matches the conclusions of Wyler et al. and Schmidt et al.

Seemingly well-defined patterns of cortical variability exist, during cortical conditioning, but do these patterns reflect constraints on population activity? Do patterns of interaction limit the flexibility of cortical control? Such questions cannot be rigorously addressed without some sort of intervention, like the dissociation experiments of Fetz et al. To make such an intervention possible, Chase and Schwartz used principal components analysis (PCA) to summarize patterns of population activity observed during reinforcement of several cortical units. They point out that the signed score of a population activity pattern – with respect to a particular principal component eigenvector – is a statistical measure of the similarity between that pattern, and the pattern represented by the principal component. Thus, by treating this statistic as a discriminatory stimulus, it is possible to reinforce patterns of population activity, and to ”condition on the overlap of the population activity with a particular target pattern”. Moreover, because principal component eigenvectors form an orthogonal basis for the entire space of population activity patterns, this method can also be used to reinforce patterns of activity that are specifically not similar to those that have been previously observed. Effectively, this furnishes a method for guiding a subject to explore the set of feasible activity patterns. Chase and Schwartz indicated preliminary success with this paradigm, and an intention to pursue further variations, but they did not report any specific results.

Sadtler et al. [43] did, however, deliver on that promise. Asking ”if some neural activity patterns are easier to generate than others”, they sought to verify limits on the flexibility of cortical activity. As Chase and Schwartz suggested, they constructed a model to summarize patterns of population activity during natural, or intuitive, behavioral conditions. They built this model using dimensionality reduction – accomplished via factor analysis. They referred to the linear vector space represented by this model as the ”intrinsic manifold” of the population, which ”presumably reflects constraints imposed by the underlying neural circuitry”. The concept of the intrinsic manifold is equivalent to the ensemble interactions proposed by Fetz [25]. Sadtler et al. report that performance was substantially better for behavioral contingencies defined in terms of this manifold, than for arbitrary contingencies. They concluded that, ”on a timescale of hours, it seems difficult to learn to generate neural

activity patterns that are not consistent with the existing network structure”. Their ”results suggest that the existing structure of a network can shape” behavior, much as Fetz [25] suggested. This contrasts with the conclusions of Moritz et al. and Law et al.

1.13 RE-ASSESSING THE SIGNIFICANCE OF LEARNING

In addition to the issue of ”natural” interaction constraints, Chase and Schwartz [10] also addressed the other limitation to cortical flexibility proposed by Fetz [25]: an incomplete understanding of the behavioral contingency. In this regard, they seemed to come to a similar conclusion as Fetz: learning-related constraints can be mitigated by adopting a simple, iterative behavioral paradigm. Specifically, they propose ”two sequential processes” for introducing subjects to a behavioral contingency based on cortical population activity: first identify network-based constraints on population patterns – via a permissive cortical conditioning and dimensionality reduction paradigm, as described above – and then allow the subject to gradually learn to act within those constraints. In effect, this amounts to a separation of the processes of discovery and coordination of behavioral degrees-of-freedom, which they suggest will ”facilitate learning”.

How significant are such learning-related constraints? Both Fetz and Chase / Schwartz suggested that mismatch between a subject’s internal beliefs about a behavioral contingency, and the actual contingency, could substantially limit the flexibility of a cortical population. However, whereas Sadtler et al. [43] had effectively quantified the effect of the second constraint proposed by Fetz [25], evidence for the first constraint remained largely anecdotal.

Golub et al. [42] sought to supply this missing evidence. They developed a methodology ”for interpreting neural population activity in the context of how prior beliefs guide the transformation of sensory input to motor output”. Within the context of a cortical conditioning they sought to test the hypothesis that behavioral ”errors arise from a mismatch between the subject’s internal model” of the behavioral contingency, and the actual contingency. Toward this end, they adopted an experimental paradigm – like that of Taylor et al. and Jarosiewicz et al. – wherein patterns of population activity were mapped to the motion

of a cursor, in a center-out task. In a post-hoc analysis, they fitted the joint probability distribution of the behavioral target, the observed cortical activity, and the cursor motion that it effected. They parameterized this model such that they could extract an estimate of the monkeys' beliefs about the relationship between cortical activity and cursor motion. In this way, they were able to predict the behavioral output that the monkeys intended. They found that roughly 65% of movement errors could be accounted for by mismatch between this internal model and the true relationship. Importantly, they verified the absence of significant biases in behavior. This, they argue, is evidence that the subject did not face "systematic difficulties producing the neural activity patterns required to drive the cursor in all directions in the 2D workspace". In other words, stable structure – due to "ensemble interactions" or an "intrinsic manifold" – did not functionally limit task performance. This supports Fetz's suggestion that learning-related constraints could limit cortical flexibility – at least outside of the context of single unit conditioning experiments.

1.14 EXPLAINING A DIVERSITY OF RESULTS

The literature seems to paint a confusing picture of cortical activity patterns.

The classic experiments of Fetz and colleagues [17, 21, 20] are often cited as evidence for the idea that such patterns are highly flexible, and that cortical neurons are highly independent. This is interpreted to imply the absence of obligatory linkages between the activity of behaviorally contingent and non-contingent neurons – as well as non-contingent motor activity – during cortical conditioning. In recent years, Fetz et al. [25, 26, 27, 29] have reiterated the perspective that patterns of motor cortical activity are profoundly flexible, with only mild constraints due to intrinsic ensemble interactions and "natural associations" with movement behavior. Law et al. support this view. Altogether, these authors paint a picture of a cortical population that is easily adapted to the demands of any arbitrary behavioral contingency.

On the other hand, the work of Georgopoulos, Schwartz, and colleagues [6, 5, 3, 4] suggests that the activity of motor cortical neurons covaries in a rather systematic fashion, and that such patterns of population activity are strongly-linked to natural behavioral considerations – such as the direction of intended movement. They showed that changes in these patterns – prompted by the need to adjust to a changing behavioral contingency – are dominated by relatively ”global” changes, which affect the activity of all neurons of the population [9]. Ganguly et al. [35] showed that such global changes were also reflected in the activity of non-contingent neurons. Hwang et al. conclude that volitional control of cortical activity is achieved by leveraging the ”natural movement repertoire” – at least in parietal cortex. Combining chronic array recording with single unit conditioning experiments inspired by those of Fetz et al., Chase and Schwartz presented preliminary evidence that patterns of M1 population activity consistently co-vary with reinforcement of single unit activity. Sadtler et al. show that such patterns represent constraints on learning and control of cortical activity, in the context of behavioral contingencies based on population activity.

How might we resolve this controversy? Early-on, Fetz and Finocchio suggested that ”strong” correlations ”operationally define a functional relationship”, between two elements of the motor system, warranting further consideration. They adopted criteria for assessing the strength of correlation involving ”examination of the intensity and consistency of the covariation under different behavioral conditions”, or contingencies [20]. Stable functional relationships between cortical neurons would constrain the flexibility with which cortical population activity can be tailored to suit an arbitrary behavioral contingency. Thus, the intensity and consistency of pairwise correlations among cortical neurons – across contingencies – can be considered to be an indicator of the flexibility of the population.

Systematic, quantitative investigations of patterns of correlation among a population of cortical neurons during single unit conditioning has yet to be studied. Due to the technical limitations of the time, Fetz and colleagues were limited to cursory consideration of a small number of neurons, where pairwise correlations were assessed only for units recorded on the same electrode. Later, Schmidt et al. and Wyler et al. were able to more systematically consider interactions among cortical neurons, but these analyses were similarly limited to

single unit pairs. Even more recently, Moritz et al. evaluated the same scenario, but did not evaluate the strength of interactions, except within the context of movement-related tuning. Law et al. and Ganguly et al. did evaluate correlations among non-contingent neurons, but only as a secondary point, and only within the context of a single change in behavioral contingency. Chase and Schwartz considered patterns of population activity, during single unit conditioning, but report only preliminary results.

We performed a series of experiments in which we operantly conditioned the activity of single motor cortical neurons, while using chronic multi-electrode arrays to simultaneously monitor the activity of a surrounding population. Of particular interest is the strength of correlation patterns, as defined by Fetzi; that is, the intensity and consistency of such patterns, across behavioral contingencies. We compare patterns observed during single unit conditioning to those observed during a task in which the entire population is part of a behavioral contingency based on the model of the natural relationship between motor cortex neurons and behavior, as proposed by Georgopoulos et al. [5, 3, 2]. Altogether, we expect that this analysis will illuminate the flexibility of cortical populations, and the degree to which cortical activity is constrained by ensemble interactions, associations to natural movement, and/or intrinsic structure.

2.0 METHODS

The results reported here were obtained from a series of behavioral electrophysiology experiments. In short, two rhesus macaques – referred to here as non-human primates NHP-A and NHP-C – performed a behavioral task while the activity of a small population of primary motor cortex (M1) neurons was recorded. Both subjects had previously been trained to perform reaching tasks, and both had previously participated in cortical conditioning, or brain-computer interface, experiments.

2.1 EXPERIMENTAL SETUP

Arrays of microelectrodes (Blackrock Microsystems, Salt Lake City, UT) were implanted into the primary motor cortex (M1) of each subject. Prior to surgery, stereotaxic coordinates were prepared to roughly estimate the position of M1. After performing a craniotomy, array placement was decided via visual inspection of anatomical landmarks. The approximate placement of the array for NHP-C is pictured in Figure 1. Arrays for both subjects were implanted at least several months prior to these experiments, and were considered stable at the time of these experiments. The array implanted in NHP C continued to deliver consistent signals – which included well-isolated, single units – for several years following the experiments reported here.

Signals from 96 implanted electrodes were acquired and pre-processed using a Plexon (Dallas, TX) MAP system. Action potentials were isolated online using Plexon’s RASPUTIN software suite. Spike waveforms were discriminated using time-amplitude windows or prin-

principal component features. Custom software obtained spike events from the MAP system in real-time, and delivered binned spike counts to the software that drove the behavioral task. Spike events were counted in approximately 30ms time bins.

Operant conditioning of cortical activity was accomplished by providing the subjects with visual feedback of some transformation of cortical firing rates, and by delivering a liquid reward when certain criteria on that feedback were met. The experimental setup is pictured in Figure 2. Each subject was seated in a primate chair facing a 3D video display (Dimension Technologies Inc., Rochester, NY, USA), which provided task feedback. A linear decoding algorithm translated cortical activity into the motion of a cursor that was presented on the display. Subjects were required to drive the cursor to indicated target locations. For all tasks, the decoding algorithm had the form

$$\mathbf{k} = \mathbf{W}^T (\mathbf{r} - \boldsymbol{\mu}_r) \quad (2.1)$$

where \mathbf{k} are the kinematics of the cursor, \mathbf{r} are the population rate estimates, $\boldsymbol{\mu}_r$ is a vector of mean rates, and \mathbf{W} is a matrix of decoding weights. The tasks were organized into blocks of trials, where each trial was structured as diagrammed in Figure 3. In short, each trial required a subject to move the cursor to some starting position, at which point a new target was presented. If the subject caused the cursor to move from the starting position to the new target location, then the trial was considered a success, and a liquid reward was dispensed. Each trial phase was time-limited, such that a trial would fail if the subject were too slow to respond. Delays in receipt of reward motivated subjects to avoid failure. In the parlance of operant conditioning techniques, the task contingency was defined by the decoding parameters, the target, and the trial structure.

2.2 CENTER-OUT TASK

The center-out task is a canonical tool in movement electrophysiology, and is well-documented in the literature [6, 3]. A cursor and targets were rendered as spherical objects in a 3D workspace. For the cortically-controlled center-out task, neural activity determined cursor

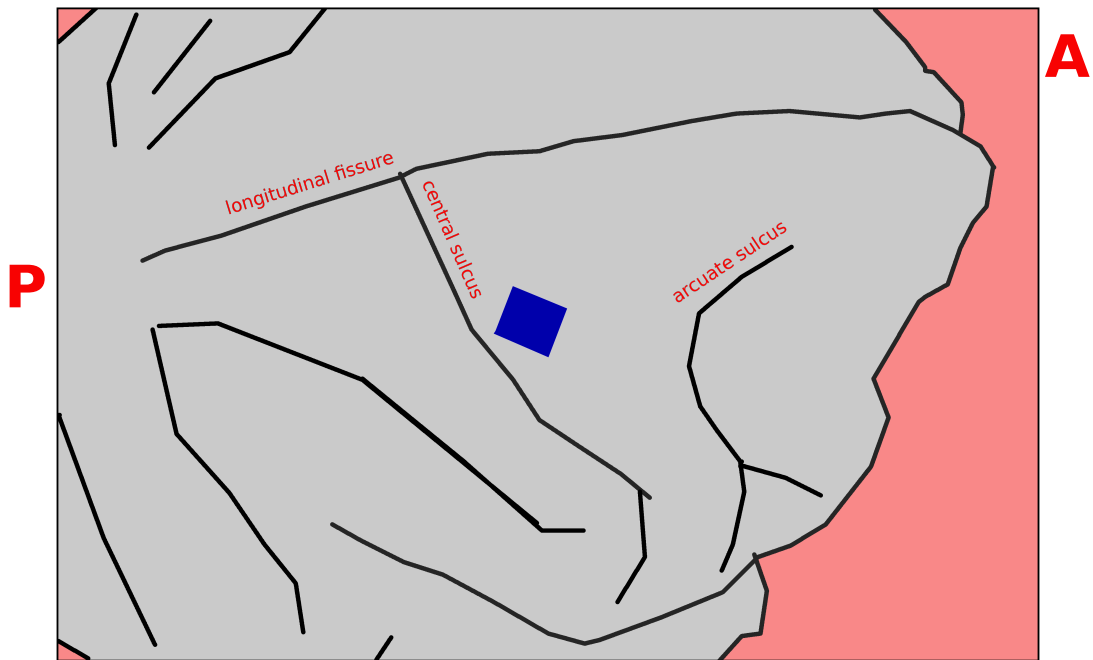


Figure 1: Array placement for NHP-C. The blue rectangle indicates the approximate location of the array. This diagram is derived from a post-mortem photograph.

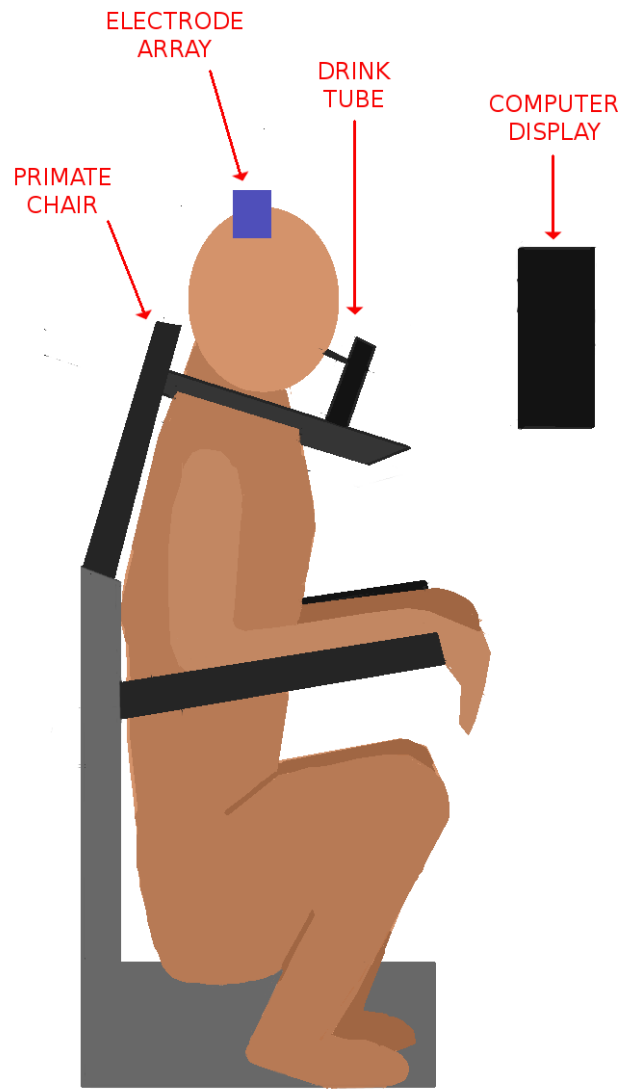


Figure 2: Experimental setup. A non-human primate performed a task presented on a computer display, as activity was recorded from a multi-electrode array implanted in the primary motor cortex. See text for details.

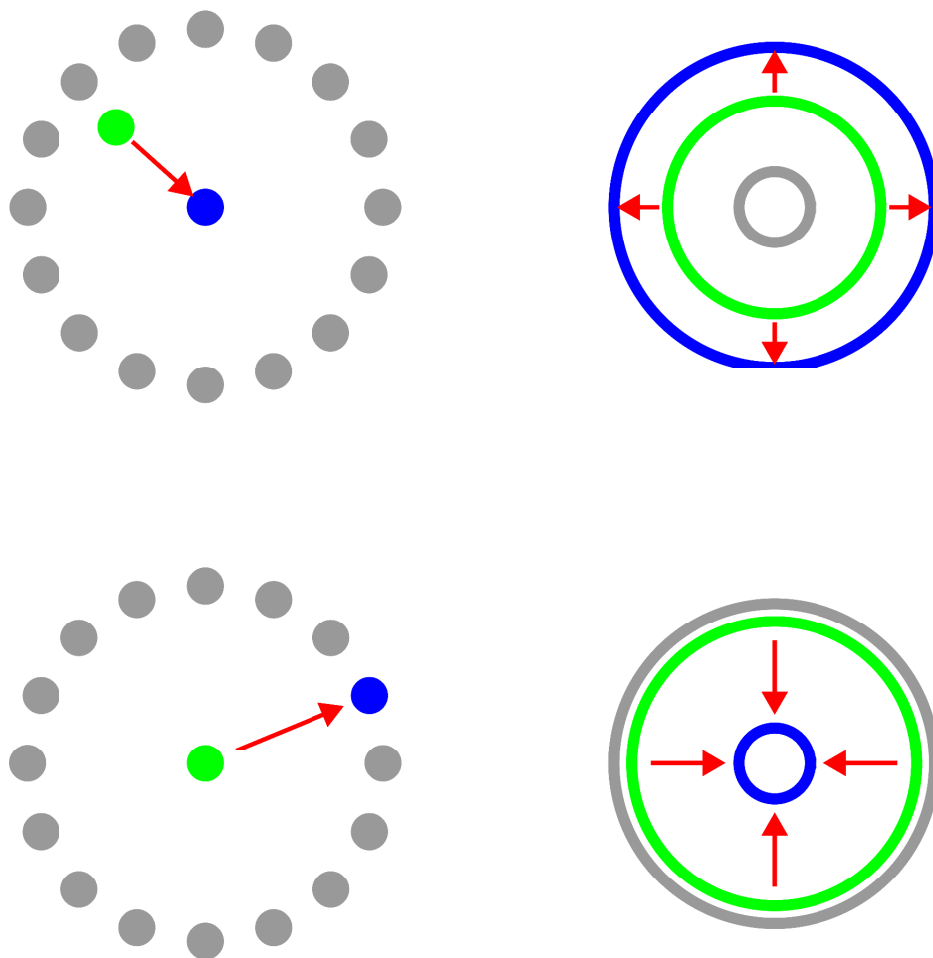


Figure 3: Illustration of the trial structure for the center-out (left column) and rings (right column) tasks. The cursor is shown in green, and the target in blue. Red arrows indicate the direction of motion of the cursor. For a trial to be successful, the subject must move the cursor from some initial position (Target A) to some final position (Target B). In the center-out task (shown here in 2D), this movement was from a central target to one of a set of peripheral targets. For the rings task, the subject was required to move from either a large-radius / high-rate target to a small-radius / low-rate target.

velocity – so that the vector \mathbf{k} of Equation 2.1 was a 2D or 3D vector of velocities. The firing rates \mathbf{r} were estimated online, every 30ms, by averaging spike counts recorded during the previous 150 ms. The decoding matrix \mathbf{W} had dimensions $N_u \times 2$ (2D task) or $N_u \times 3$ (3D task), where N_u is the number of sorted units that were applied to control. Targets were configured to lie in a 8 or 16 target circular configuration for the 2D task, and the cursor was locked to the plane of the targets. For the 3D task, targets were organized in an 8 target cubical configuration or a 26 target spherical configuration, and the cursor was permitted to vary in 3 dimensions.

The weight matrix \mathbf{W} of the decoding Equation 2.1 was calibrated at the start of every experimental session. This was accomplished using a gradual transition from an assisted task to a full cursor control task. At the start of each session, the subject was presented with several blocks of 8-targets. Within these blocks, targets had a circular configuration for the 2D task, or a cubical configuration for the 3D task. During any given trial, the subject was required to move the cursor from a central target to one of the peripheral targets. Initially, the cursor was constrained to lie on the line that connected the target pair, and the radial distance along this line was computed using some initial set of decoding parameters. After an initial block of 8 targets, a weight matrix \mathbf{W} was estimated by regressing firing rates against movement direction. In subsequent blocks of trials, the subject was given partial control of the cursor, via a weighted average of decoded cortical activity and automated computer control. During this phase of calibration, deviation from a straight-line path was attenuated by scaling the component of the commanded motion that was orthogonal to the straight line path. Initially, decoded cortical activity contributed less than 25% to cursor motion, and the attenuation coefficient was near one. After the second block of trials, cortical control of the cursor was increased, and attenuation was reduced. This process typically continued for 3-5 blocks of trials, until full control was achieved. The weight matrix \mathbf{W} was re-estimated after each block. After no more than 6 calibration blocks – typically lasting no more than 5 minutes – the decoding parameters were fixed, and the subject proceeded with full control.

In the center-out task, trials were fairly rigidly structured, as illustrated in Figure 3. Unless otherwise noted, trials began with both the cursor and the target placed in the center of the workspace, in order to avoid inter-trial cursor drift due to imperfect velocity decoding.

Typically, the subject was not required to hold the cursor in this position, or was only required to hold for a very short time. Subjects were prompted to move by the appearance of a peripheral target (*Target B*), as the central target (*Target A*) was extinguished. The subject was required to move the cursor to the target before the expiration of a timeout clock – typically between 1 and 3 seconds. Once reaching the peripheral target, the subject was required to maintain the cursor within the target for a short interval. If the subject successfully moved the cursor to the target in the allotted time, and held for the required interval, then the trial was considered a success. If not, then the trial failed, and the task reset. Trials were organized into blocks of 8, 16, or 26. The subject was required to successfully move to each target before progressing to the next block of trials.

2.3 RINGS TASK

The structure of the rings task shares many similarities with the center-out task, but differs primarily in the form of feedback delivered, in the dimensionality of the task, and in the specifics of the decoding Equation 2.1. The rings task was rendered in 2D, and the cursor and targets had the form of concentric rings, rather than spheres. Modulation of cortical activity dictated the radius of the cursor ring, and the subject was required to increase or decrease this radius in order to match that of the target ring. As such, the task was 1-dimensional and “directionless” (i.e., it did not favor movement in any particular direction in the workspace).

The decoding Equation 2.1 became a scalar inner-product between a weighted indexing vector and filtered firing rates. The kinematics vector \mathbf{k} became the scalar radius (position) of the cursor ring, whereas the weights \mathbf{W} became a vector with only one non-zero element. This element corresponded to a single cortical input channel, such that the cursor radius was driven entirely by a single unit. Unit selection was at the discretion of the experimenter. Typically, well-isolated units, or units that had been observed to modulate reliably, were selected. The parameters \mathbf{W} and μ_r were largely obtained through trial-and-error, or from estimates of a unit’s dynamic range that were derived from previous recording sessions (e.g.,

the center-out session). Typically, the experimenter would conduct short blocks of trials to adjust these parameters in such a manner as to balance low chance rates of success with moderate task difficulty. Once adequate parameters were identified, these parameters were fixed for blocks of tens of trials. The firing rates of vector \mathbf{r} were filtered using a 5-10 tap digital (FIR) filter, with either a flat (i.e., moving average) or exponential shape.

The trial structure of the rings task – illustrated in Figure 3 – was similar to that of center-out. In early versions of the task, trials initiated automatically. In a later version of the task, a trial was initiated by moving the ring cursor to the middle of the workspace, where a purple ring target was displayed. No cursor hold epoch was effectively enforced. Once the trial initiated, a blue ring target was placed at one of the two extremes of the workspace – that is, a target with either a very large radius or a very small radius appeared. In the modified task, the choice of which target to present first was random. This trial initiation and randomization protocol was adopted to reduce the possibility of random successes. The subject was expected to move to the presented target before the expiration of a movement timeout clock, but this timeout tended to be long (i.e., $> 1\text{sec}$). When the subject passed through the first target (*Target A*), the target ring was moved to the opposite extreme of the workspace. The subject was expected to reverse cursor direction, and to move to the opposite target (*Target B*). Only then was a trial considered successful, and reward delivered. No holds were required at any of the targets. Tens or hundreds of trials were obtained for each controlling unit. The lengths of trial block were chosen at the discretion of the experimenter. Typically, the control parameters would be changed once the subject had demonstrated proficient control for more than 20, but less than 100, trials.

2.4 DATA ANALYSIS

Sorted units were excluded from analysis if they did not fire during any trial within either task. For the center-out task, blocks of trials were excluded if fewer than 10 successful trials were recorded. For the rings task, the minimum number of successes per block was set at 15. All other data were considered for analysis.

2.4.1 Rate estimation

For the center-out task, the procedure for estimating firing rates is illustrated in the left column of Figure 4. In polar or spherical coordinates, cursor velocity consists of two components that can contribute to modulation: speed and angular direction. Cursor speed is only loosely constrained by the go-cue and trial timeouts, and varies somewhat from trial to trial. However, center-out movements are primarily ballistic, and the relative time at which the cursor transitioned from rest to maximum speed tended to be consistent. Thus, in an effort to represent the maximum range of modulation, we chose to estimate rates by anchoring an averaging interval to this peak ballistic speed. For each experimental session, we estimated the average time at which speed peaked. This estimate is indicated by a blue circle in the top panel of the left column in Figure 4. The horizontal error bars – obtained via resampling – are too narrow to be seen on the plot. Rates were estimated by counting spike events, for each unit, from this peak speed time to 150 ms prior. These averaging windows are indicated by colored boxes in the rate plots of Figure 4, where the horizontal limits indicate the temporal bounds of the window, and the vertical limits indicate the 1-standard deviation bounds about the estimated mean rate. The colors of the left column are matched to the corresponding, diametrically-opposed center-out targets in the right column. The like-colored line intersecting each box, in the plots of the left column, is the time-varying firing rate of the relevant unit, averaged across trials. The time axis in each plot is aligned to the end of the Hold A period (i.e., the "go cue" event). A unit is considered *modulated* if the mean rates of the two targets are significantly different (i.e., if the vertical space between the colored box centers is sufficiently large). In this example, the unit is considered modulated in the first and third plots, but not the middle plot. See text for details.

For the rings task, the estimation of firing rates is more straightforward, since the relevant behavioral variable is more directly linked to population activity than in the center-out task. More specifically, the behavioral variable of interest is the radius of the ring cursor, which is directly proportional to the firing rate of a single unit. Thus, averaging windows are

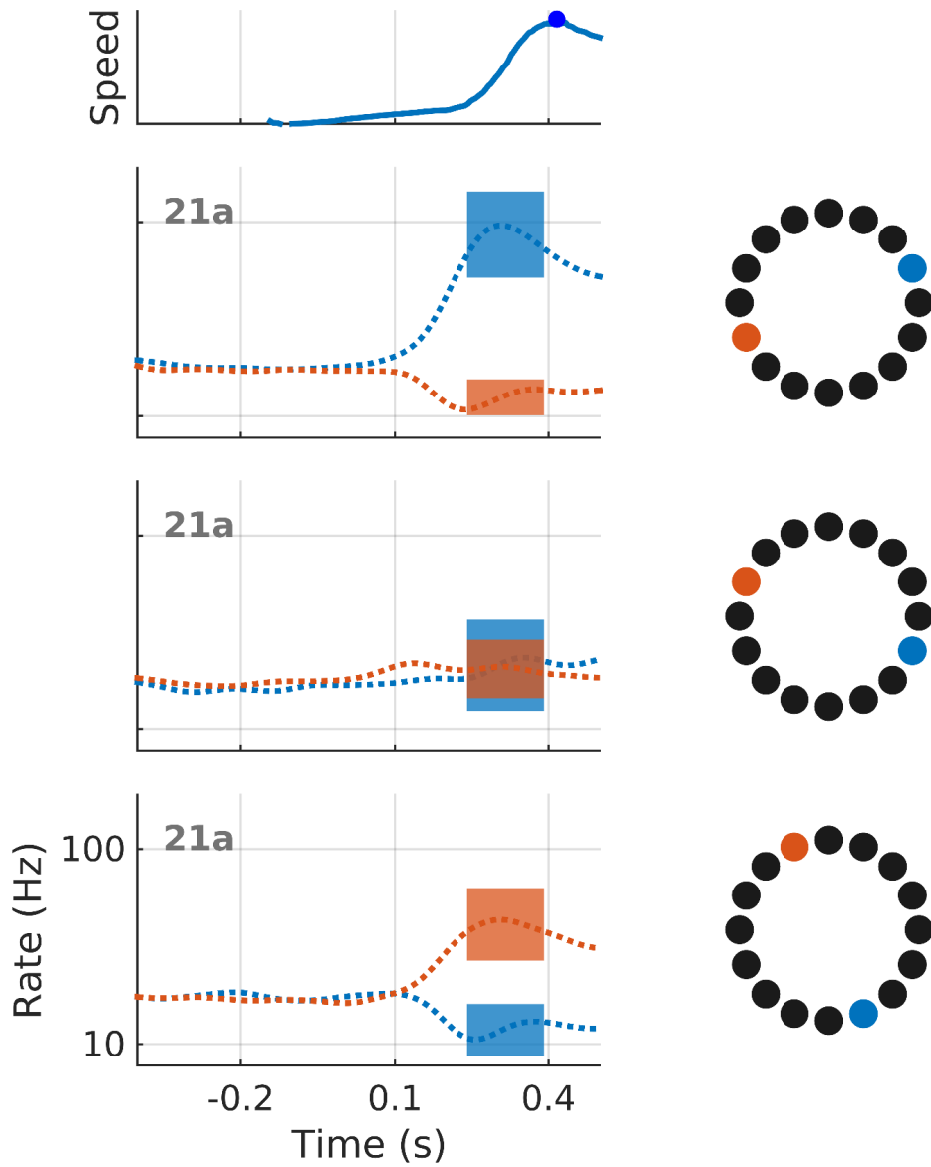


Figure 4: Illustration of rate estimation and the modulation test in the 2D center-out task, for one unit and three target pairs. Each panel in the right column shows the pair of diametrically-opposed targets for which color-matched rates are plotted in the left column. A unit is considered *modulated* if the mean rates of the two targets are significantly different (i.e., if the vertical space between the colored box centers is sufficiently large). In this example, the unit is considered modulated in the first and third plots, but not the middle plot. See text for details.

anchored to target impact events; i.e., the instant at which the ring cursor intersects with the ring target. We use the same 150 ms window, and compute one rate estimate for the high rate target and one of the low rate target. This procedure is illustrated in Figure 5.

2.4.2 Modulation test

A recorded unit is said to be *modulated* if the firing rate of that unit co-varies with some measured behavioral variable. In the present context, the relevant variables are the kinematics of cursor motion, as defined by Equation 2.1. In these terms, a unit u is considered to be modulated between kinematic condition A and condition B if the estimated rates – r_u^A and r_u^B – associated with each condition are distinct. Since firing rates estimates vary not only from target to target, but also from trial to trial, a statistical test is used to evaluate modulation. Here, we employed a two-sample, unequal-variance t-test to assess the significance of modulation. Specifically, we computed the mean and standard deviation of firing rates, across all trials to a given kinematic target. If the mean rates were significantly different, then the unit was considered modulated between that target pair.

For the center-out task, the second component of cursor velocity – angular direction – is the independent variable with respect to which modulation is evaluated. Ultimately, modulation might be observed between any two kinematics targets. However, we make the simplifying assumption that modulation will be greatest between targets that are most distant (in the kinematics space), to avoid having to consider every possible pair of kinematics targets (325 pairs, for 26 target 3D). Thus, the modulation test is considered only for diametrically-opposed targets, as diagrammed in the right column of Figure 4.

The analysis for the rings task, once again, more straightforward. As illustrated in Figure 5, the t-test is applied to assess the significance of the difference in unit rates at the high and low targets.

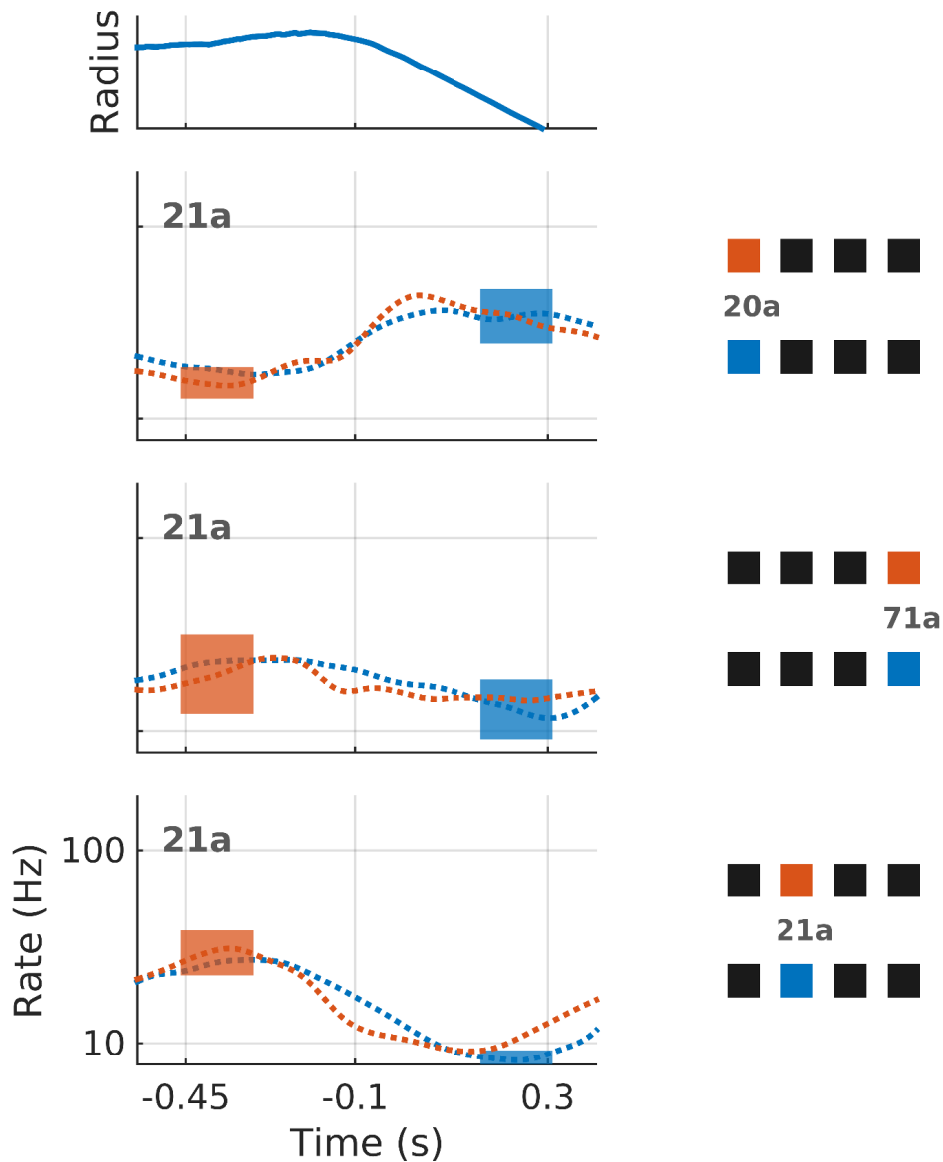


Figure 5: Illustration of firing rate estimation and the modulation test in the rings task, for one observed unit and three unit conditioning configurations. Compare with Figure 4. The same unit is illustrated in the left columns as for the center-out figure, and the interpretation is similar. Here, rate averaging windows are anchored to trial events (i.e., high and low target impact) within a given trial. The diagrams in the right column indicate which unit – of 4 total, for this particular experimental session – was active for conditioning.

3.0 RESULTS

Figure 6 shows the structure of the data reported here. A total of 7 sessions were collected for NHP-A, and 6 sessions for NHP-C. Each row of Figure 6 represents a separate experimental session, with NHP-A toward the bottom of the figure and NHP-C toward the top. For the center-out task, there were always 8 conditions (16 targets), except for the last 3 sessions for NHP-A, within which the task was switched to 26-target 3D center-out. In each task, the number of trials always exceeded 200. The number of successful trials per target in center-out ranged from 10 to 23 for NHP-A, and from 12 to 65 for NHP-C. The number of rings trials per condition for NHP-A ranged from 23 to 99. For NHP-C, the range was from 16 to 105.

3.1 LEARNING TO CONTROL SINGLE UNIT ACTIVITY

Both subjects (NHP-A and NHP-C) eventually achieved control of single units in a majority of tested cases, but the process of adapting to a new conditioning configuration varied by unit. Typically, but not always, this entailed a period of adjustment, during which task performance was diminished, as a subject learned how to modulate the target unit. Figure 7 plots the cursor radius versus time for 3 epochs that immediately followed the start of a new conditioning session. In the first case (top trace), successes were sporadic for an interval of approximately 2 minutes, during which the success rate gradually increases. Presumably, this reflects a search, on the part of the subject, for a control strategy via which to control the target unit. In contrast, the success rate in the middle trace very abruptly increases at around 35-40 seconds, and remains high. We interpret this to reflect a sudden, and

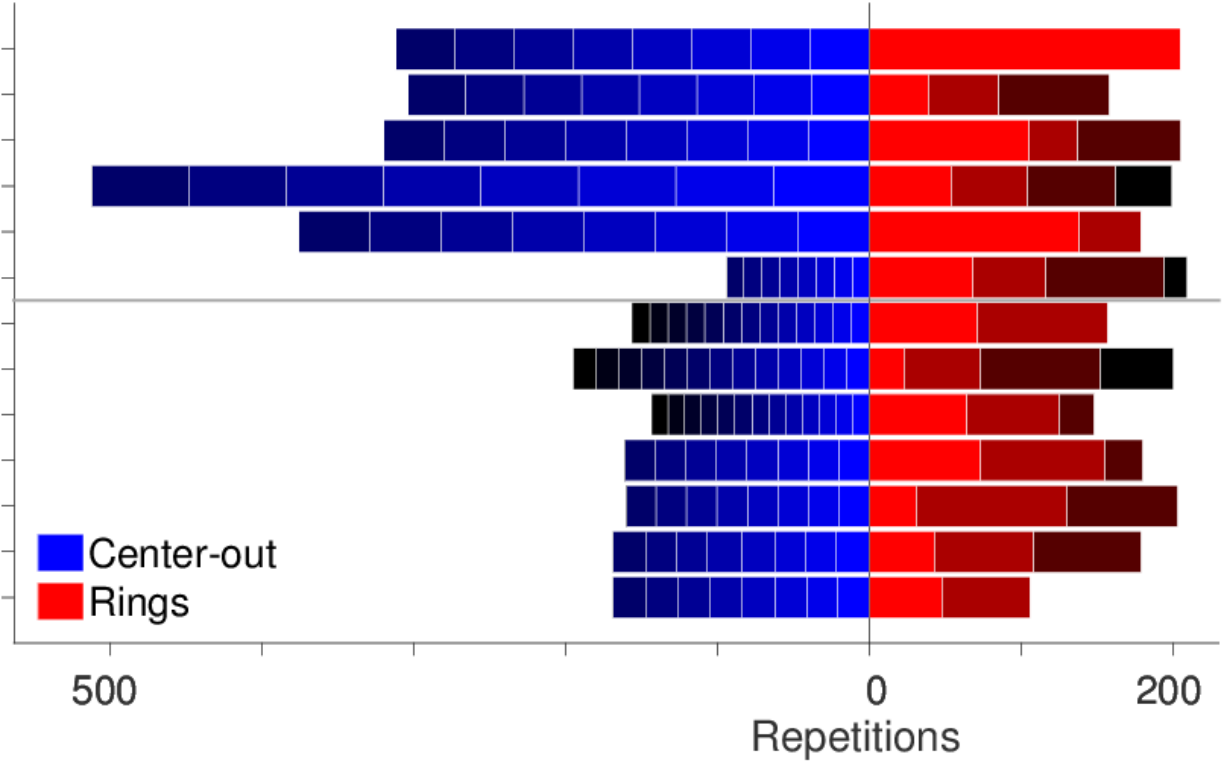


Figure 6: Illustration of the data. Number of repetitions per experimental session, task, and condition. Each row of bars represents a different experimental session. The left side of the figure (blue) displays the number of trials successfully collected in the brain-control center-out task, and the right side (red) displays the trial counts for the rings task. White lines divide the trial count bars by condition. The number of repetitions is equal to the number of trials in the rings task, and half the number of trials in the center-out task (since diametrically-opposed targets are considered in pairs). Data for NHP-A are plotted below the horizontal gray line, and data for NHP-C above.

perhaps accidental, discovery of a control strategy. Finally, a high success rate is immediately apparent in the bottom trace. Importantly, the conditioned unit co-modulated with the unit conditioned in the previous experimental block, so that only a small shift in strategy was apparently required to maintain task performance.

3.2 NON-CONTINGENT COVARIATION: SINGLE UNIT MODULATION

Robust patterns of population activity were typically observed during single unit conditioning sessions. Figure 8 and Figure 9 display raster plots for 3 different conditioned units from each subject. In each plot, the units are ordered by the correlation structure for the top-most panel. Shaded (green) areas on the raster and radius plots indicate the duration of each successful trial. For the radius trace, the upper and lower limits of the shaded trial boxes also indicate the upper and lower target radii. Finally, the outcome of a modulation t-test is indicated at the left-most margin of the raster plot: a red circle indicates that the unit is significantly modulated ($\alpha = 0.01$). The modulation test is computed using all trials within the plotted condition; even those that are not pictured.

There are a few things to notice in these raster plots. Foremost among these observations is that the conditioned unit does not modulate in isolation. There are clear patterns of co-modulation across the population, in every case. It is also clear that co-modulation of non-contingent units can be correlated (e.g., the first 4 units in the top panel of Figure 8) or anti-correlated (e.g., the first and last unit in the top panel of Figure 9), and that interactions can be temporally disperse (e.g., the first 4 units in the bottom panel of Figure 8) or fine. Regarding this latter point, it was often true that the responsiveness of conditioned control varied substantially by unit: some units could quickly modulate from one rate target to another, whereas others followed a meandering course, resulting in a range of average success rates.

The number of co-modulated units in the population varied by rings task condition. Figure 10 shows the number of modulated units (here, including the conditioned unit) in each of the 37 task conditions that we considered. Clearly, the number of modulated units

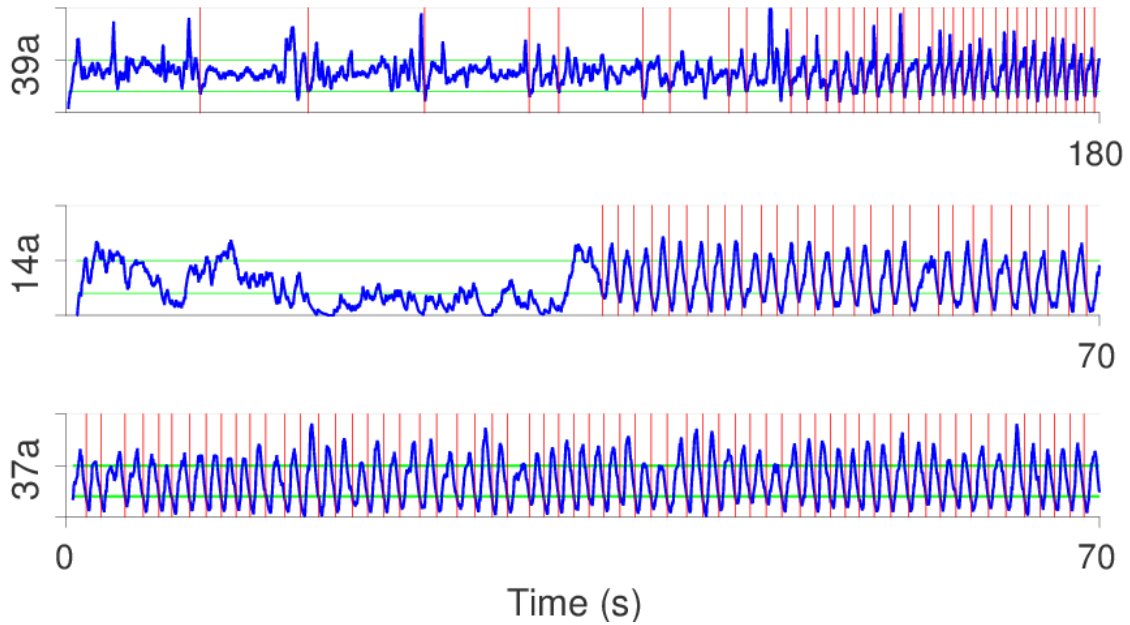


Figure 7: Examples of task acquisition. Cursor radius plotted during the epoch immediately following the onset of a new block of single-unit conditioning trials, for 3 different target units. All plots represent data from the same experimental session, and are derived from consecutive blocks (ordered top-to-bottom). The labels on the ordinates specify the units that are linked to ring cursor control. Vertical red lines indicate successful trial completion. Green lines indicate the target radii. The time scale (abscissa) of the top-most plot is longer than that of the two bottom plots to emphasize the gradual convergence to a high success rate. The scales of the vertical axes are identical.

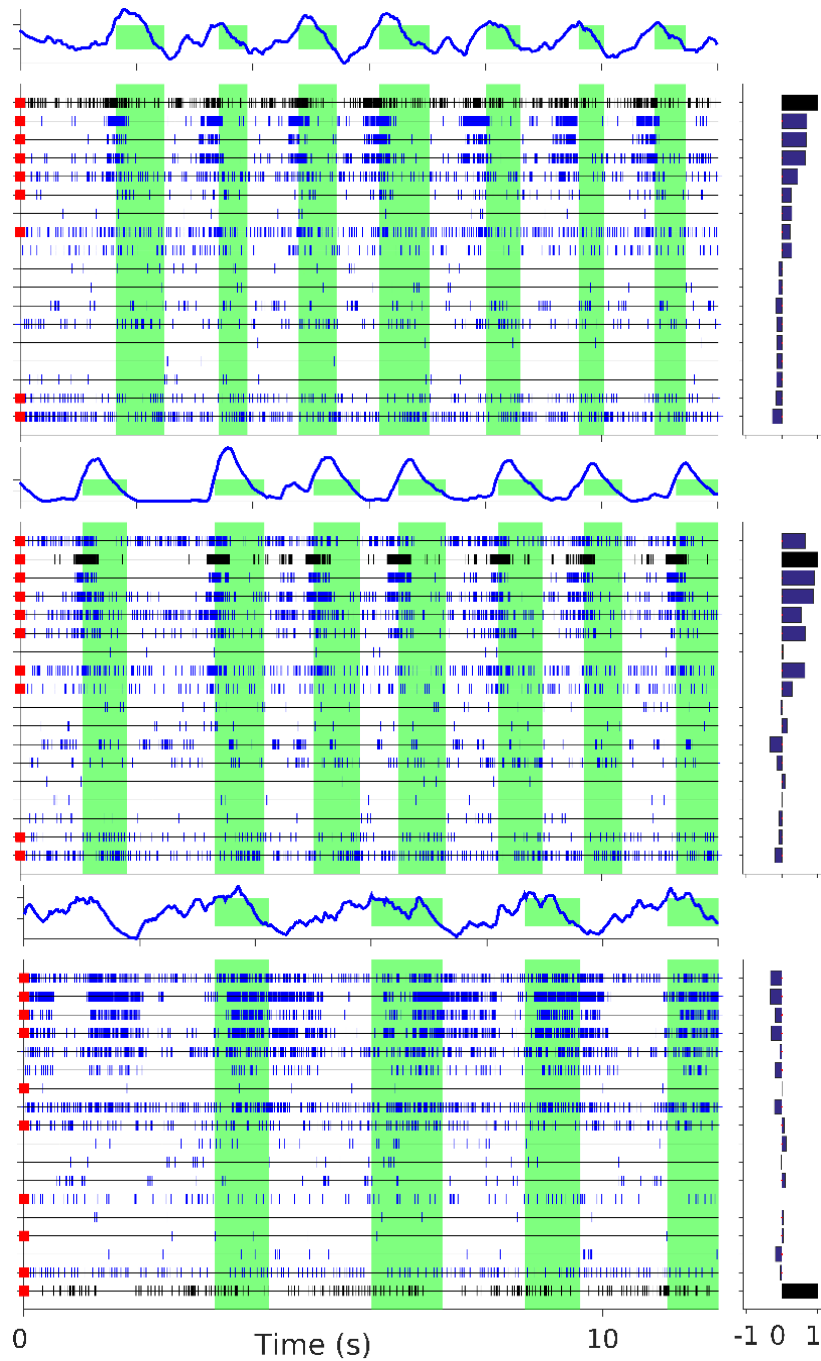


Figure 8: Raster plots for 3 different single unit conditioning configurations for NHP-A. Each panel is a snapshot from a different block of the rings task. Within each panel, the raster is plotted below the cursor radius trace (as pictured in Figure 7) and to the left of a bar plot indicating the correlation between each unit and the conditioned unit. In each raster and correlation plot, the conditioned unit is plotted in black. See text for further details.

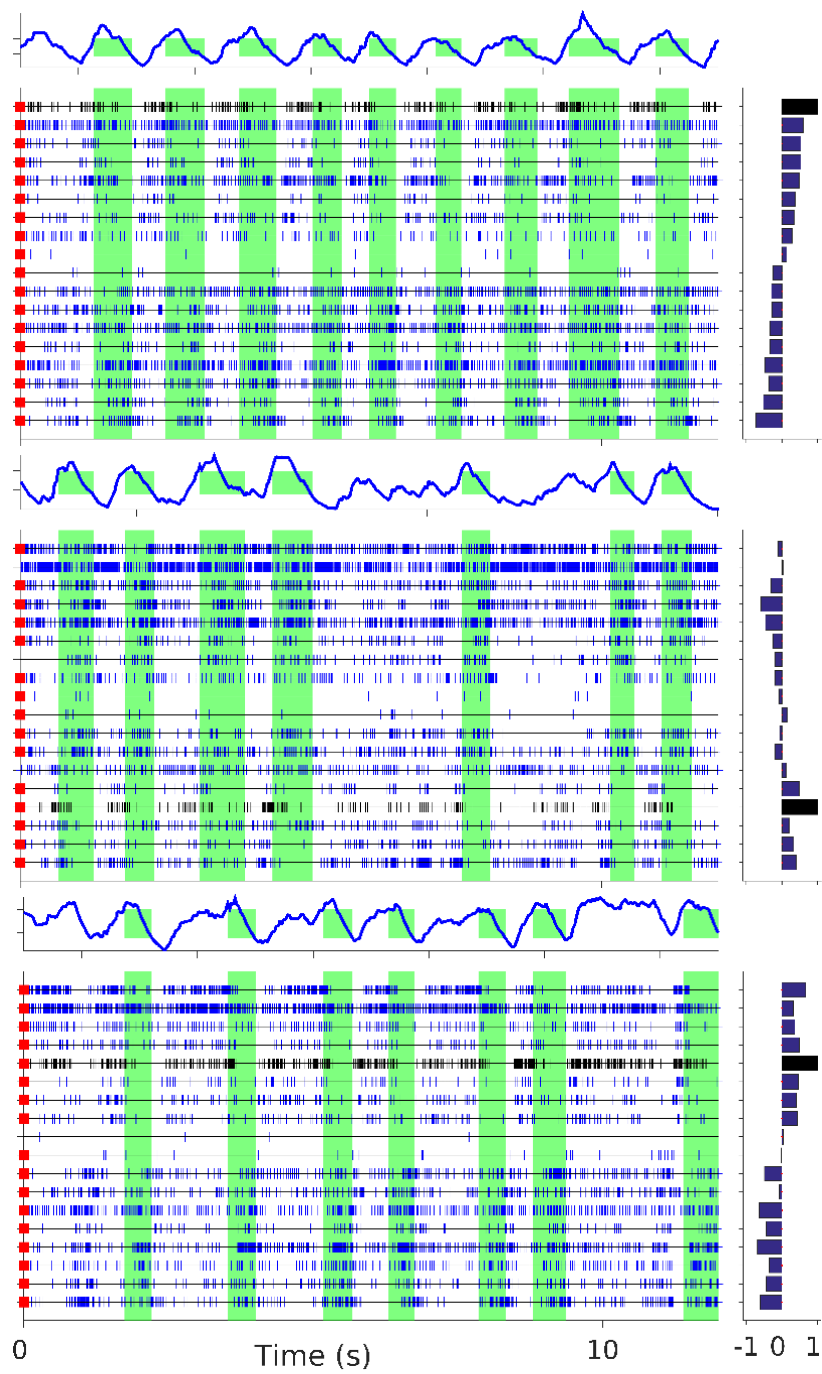


Figure 9: Raster plots for 3 different single unit conditioning configurations for NHP-C. See Figure 8, and text, for details.

could vary dramatically, ranging from 4 of 43 (9%) to 33 of 42 (79%). The average number of modulated units was 19.6, and the average percentage of units modulated was 51%. The number of modulated units does not appear to be a function of the total number of sorted units. The patterns of population engagement seems to depend on the unit being conditioned.

3.3 NON-CONTINGENT COVARIATION: PAIRWISE CORRELATION

Having established that many units co-modulate during the task, and that these patterns of co-modulation vary between task conditions, we sought to characterize similar changes in pairwise relationships between units. For each rings condition, we computed the Pearson correlation coefficient between the firing rates of unit pairs. This calculation included firing rates at both Target A and Target B. Conditioned units were excluded from the analysis. We then compared the pairwise correlations in every task condition with the corresponding correlations for every other condition that was tested, within the same experimental session. Figures 11 and 12 show scatter plots of these paired-pairwise correlation statistics, for NHP-A and NHP-C, respectively. In each plot, the abscissa represents pairwise correlations between recorded units in one rings task condition (i.e., one single unit conditioning configuration), and the ordinate represents pairwise correlation for the same two units in a different rings task condition. The color code communicates significance at $\alpha = 0.05$. Approximately 43% of paired, pairwise correlations were significant in at least one condition.

There are two principle features to observe in these plots. First, the number of unit pairs that were significantly correlated in one rings task condition, but not the other, is substantial. This is made evident by the clouds of red and blue points along either axis. Thus, the significance of correlation could depend on the behavioral contingency. Second, data points tend to cluster nearby the positive diagonal (the axis of maximal variation, as determined by principle components analysis, lies at an angle of approximately 44° , and captures more than 70% of the variance, for each data set). This indicates that pairs of unit tended to maintain similar correlation relationships between rings task conditions.

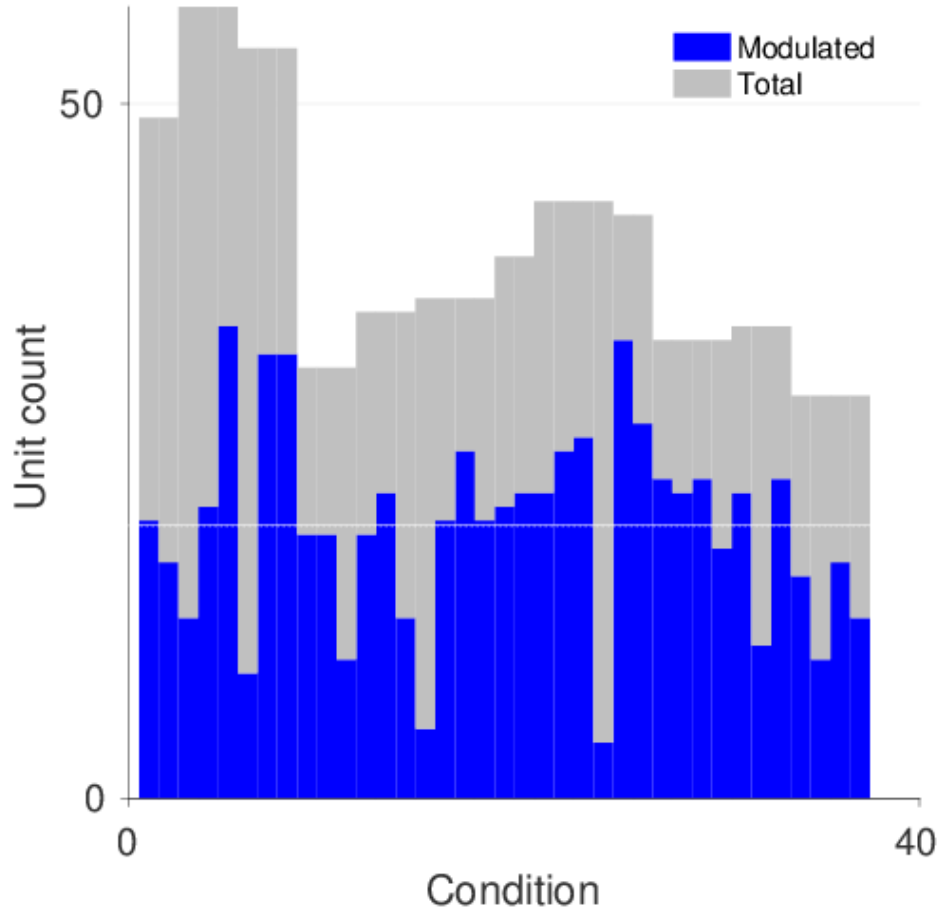


Figure 10: Engagement of the population, across single unit conditioning sessions. Bar graph of the number of units modulated versus the number of units recorded, for each rings condition. The contingent unit is included in the number of modulated units, but the number of non-contingent units clearly dominates. Data for both NHP-A and NHP-C are included in this plot.

The modulation test and the correlation analysis are related, but distinct. The modulation test describes the relationship between each non-contingent unit and the contingent unit, and/or the behavioral goal. The correlation analysis describes the behaviorally-relevant relationship between pairs of non-contingent units. Whereas the modulation test considers only a change in the average difference between firing rates, the correlation analysis can also identify trial-to-trial covariation. Of the unit pairs that were considered, across all rings task conditions, the percentage of significantly correlated pairs was 20% for NHP-A and 42% for NHP-C. Of these correlated pairs, both units were significantly modulated in 53% and 81% of cases. The percentage of unit pairings – correlated or not – in which both units were significantly modulated was 22% and 53%. When the pairwise correlation coefficients computed for different rings task conditions were paired, correlations were significant in both conditions for 63% and 76% of cases. Of those pairings for which both correlations were significant, both units were modulated in both rings task conditions in 72% and 86% of cases. Altogether, these numbers show that significant modulation does not necessarily imply significant correlation, and vice-versa.

3.4 COMPARISON WITH A POPULATION-BASED BEHAVIORAL CONTINGENCY

Is such variation in patterns of population activity typical of cortically-controlled cursor tasks? Are patterns of activity similar, when the behavioral contingency is derived from models of cortical activity during natural movement? When the entire population is included in the behavioral contingency? To assess whether single unit conditioning induced activity patterns similar to other brain-computer interface paradigms, we compared these patterns to those observed during the more standard center-out task. For this experiment, the motion of a 2D cursor was computed using the population vector algorithm (PVA) of Georgopoulos et al. [5, 3].

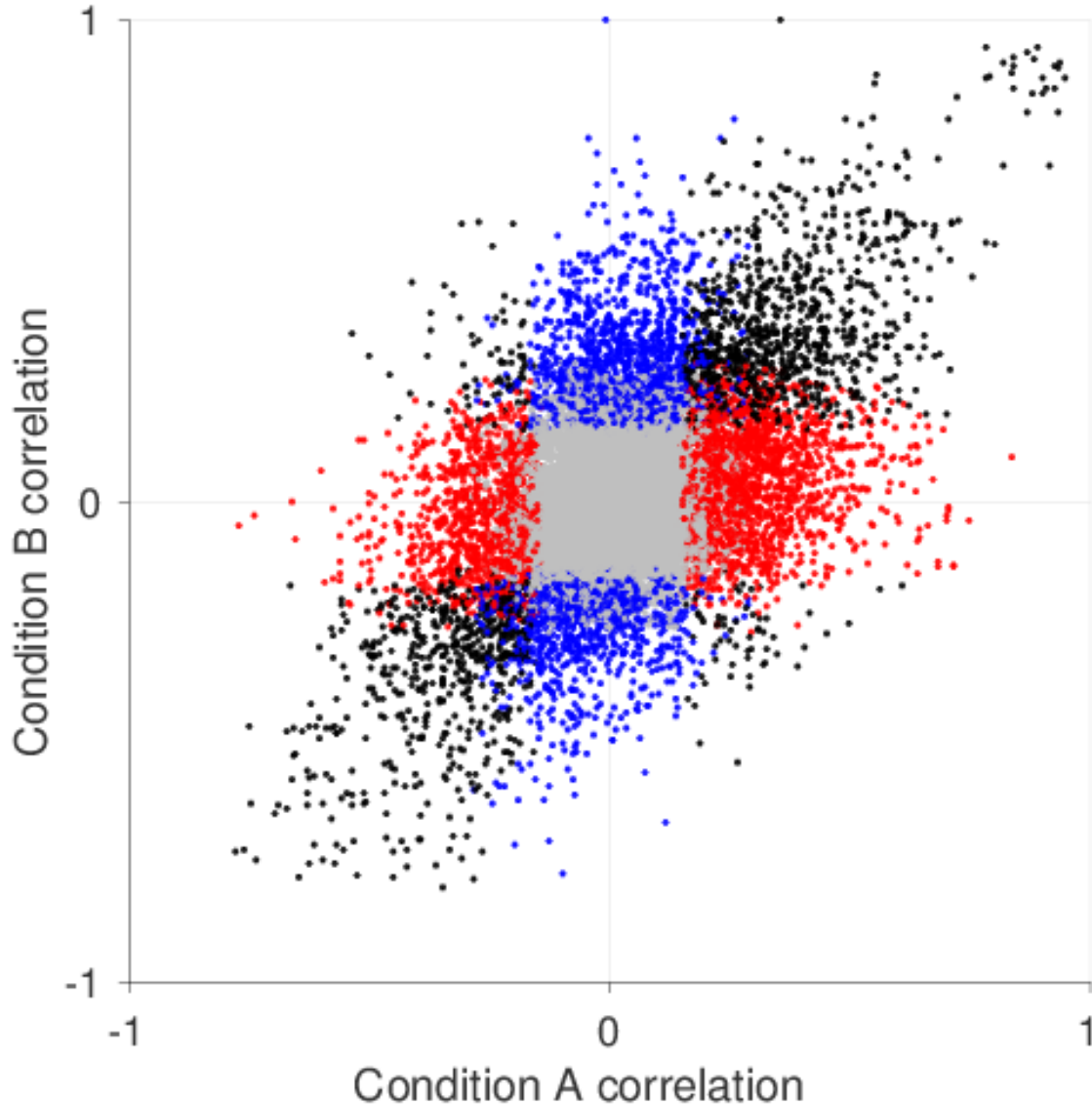


Figure 11: Scatter plot of Pearson correlation coefficients between the same pairs of units in pairs of rings task conditions, for NHP-A. In each plot, the abscissa coordinate of each point represents pairwise correlations between recorded units in one rings task condition, and the ordinate coordinate represents pairwise correlation for the same two units in a different rings task condition. The color code communicates significance at $\alpha = 0.05$. The points at the center of the cloud (gray) are not significant in either condition. The points along the vertical (blue) and horizontal (red) coordinates are significant in one or the other condition. The points along the diagonals (black) are significantly correlated in both conditions.

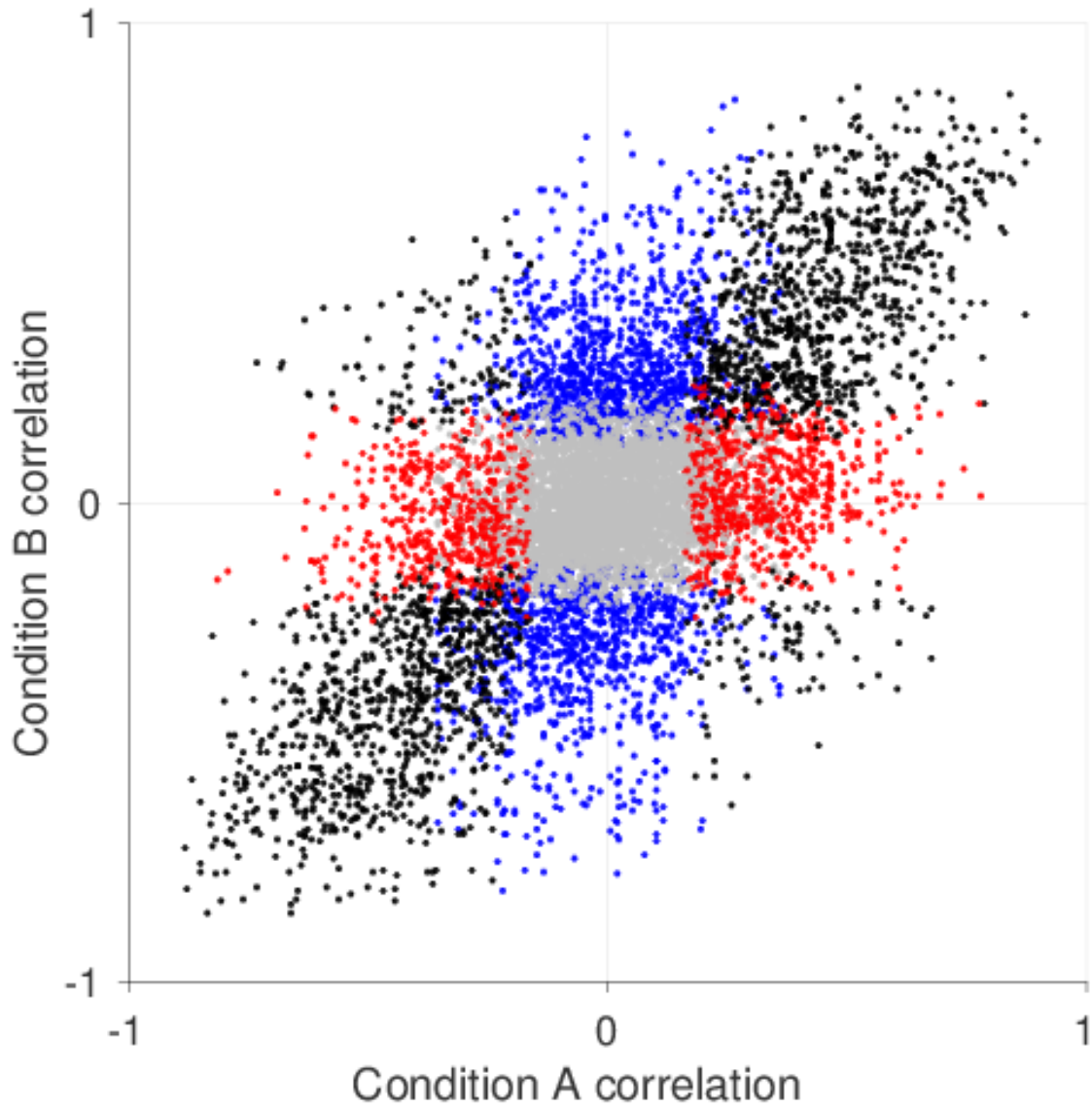


Figure 12: Scatter plot of Pearson correlation coefficients between the same pairs of units in pairs of rings task conditions, for NHP-C. Compare with Figure 11.

Figure 13 shows the number of units modulated for each center-out condition (pair of diametrically-opposed targets). Compare this with Figure 10. The number of modulated units ranged from 7 of 35 (20%) sorted units to 30 of 54 (56%) sorted units. Although the average number of units modulated per condition (19.6) was approximately equal in the two tasks, the variance was greater in the rings task ($\sigma = 7.6$ versus $\sigma = 5.1$). This is apparent in Figure 14, where the number of modulated units is shown as a percentage of the population total, by condition or target (top row) and by task (bottom row). Here, the top row is an alternate view of the data in Figures 10 and 13. Across conditions, we can see that the number of modulated units seemed to vary more in the rings task, but that there is a tendency (i.e., the mode is higher) for a greater percentage of units to modulate. Nonetheless, the distribution of modulation percentages are roughly similar. This latter point is also true when considered across the entire experiment: the histograms are roughly distributed about the same peak. However, the histogram suggests slightly more engagement of modulated units in the center-out task. Taken together, these plots suggest different patterns of modulation between conditions in each task, and between the two tasks.

The magnitude of modulation also varied by task. Figures 15 and 16 show the maximum modulation depth for each unit, across all conditions and for all sessions. There are two main points to take from these plots. First, we see a tendency of the points to skew toward the abscissa, indicating that modulation depths tended to be higher in the center-out task. However, it is important to note that this is a trend and not a rule: there are a substantial number of points for which the maximum modulation depth is higher in the rings task. The second point to note is that – although the majority of units were modulated in both tasks – there could be a substantial difference in the degree of modulation between the two tasks. Several units were modulated twice as much, or more, in one task versus the other.

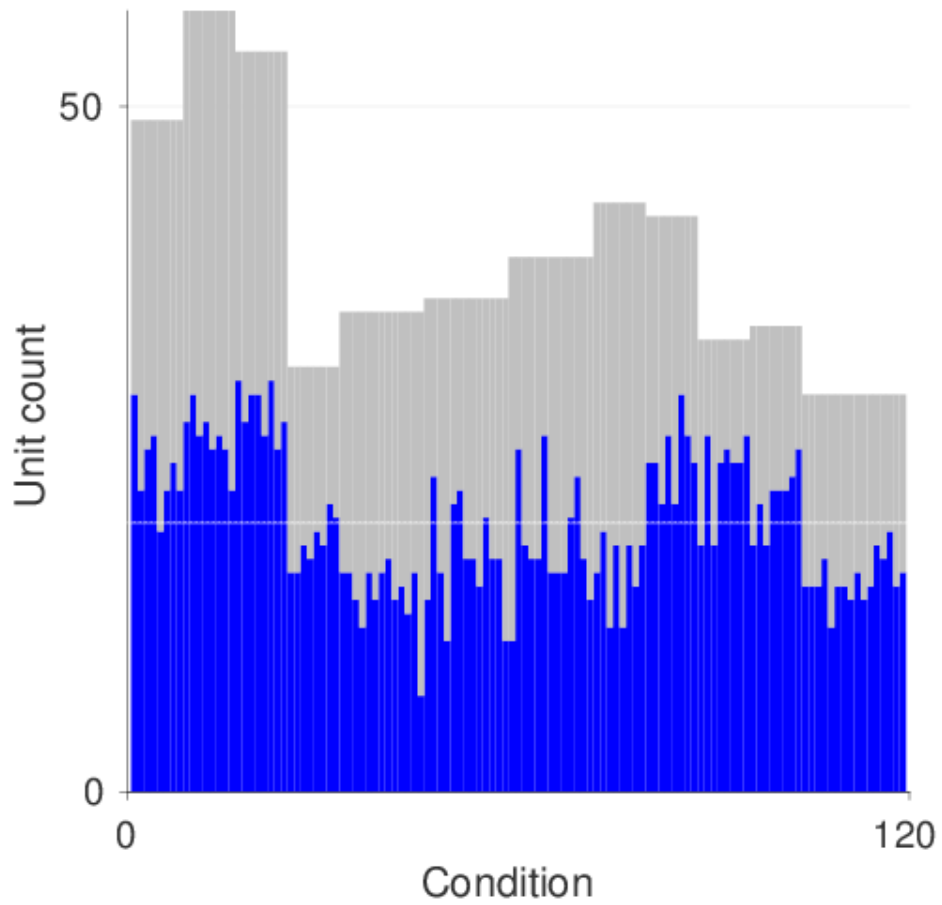


Figure 13: Bar graph of the number of units modulated versus the number of units recorded, for each center-out condition. Compare with Figure 10. Note that the mean number of modulated units are approximately equal (19.6 units) in the two plots.

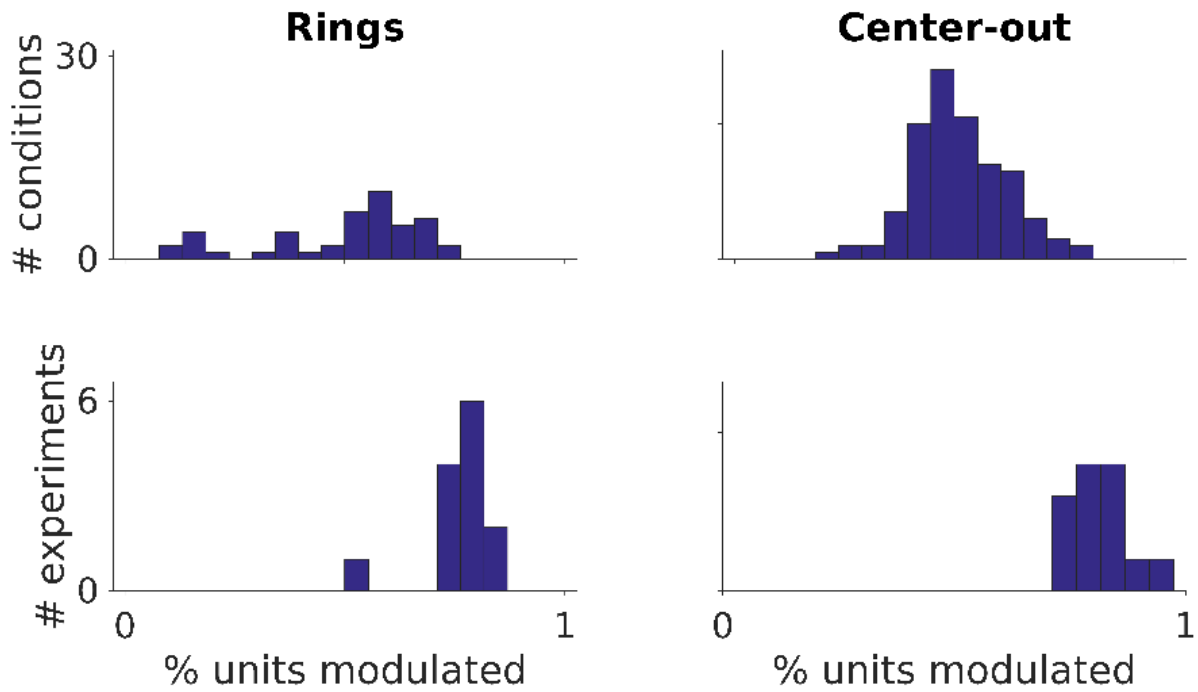


Figure 14: Summary of population engagement in the two tasks. Histograms of the number of modulated units, as a percentage of the population total, by condition (top row) and by task (bottom row).

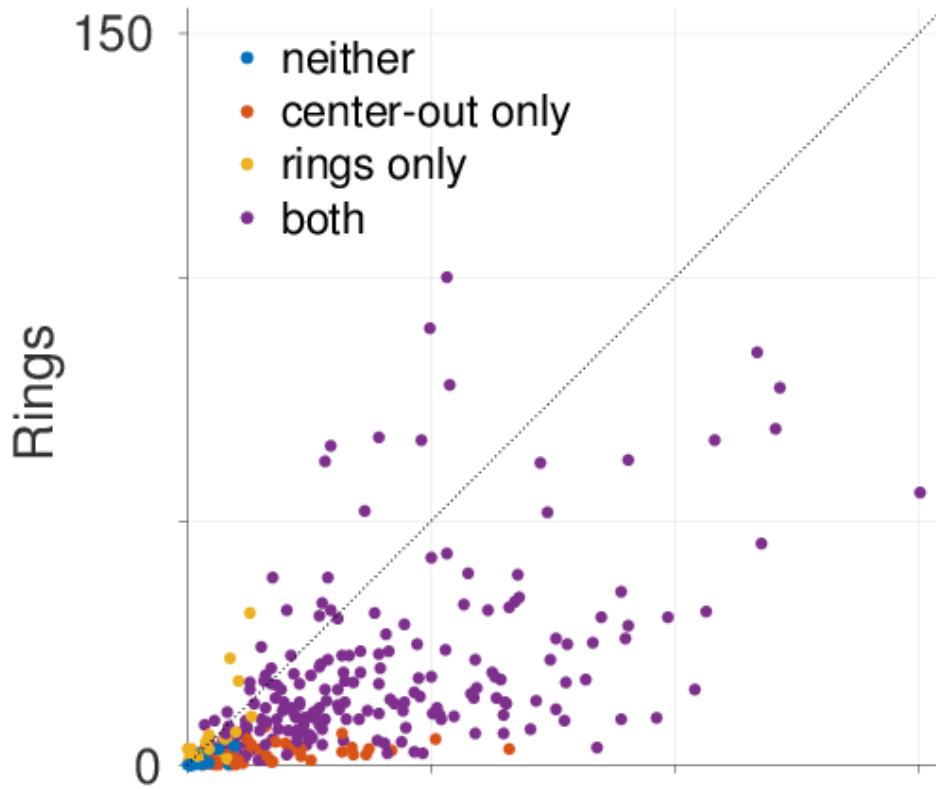


Figure 15: Maximum modulation depths across all conditions tested, within a given experimental session, for each of the two tasks. This plot shows data for NHP-A. Each plotted point represents a single unit. The plotted points are color-coded by modulation test significance ($\alpha = 0.01$): units could be modulated only in the center-out task, only in the rings task, in both tasks, or in neither.

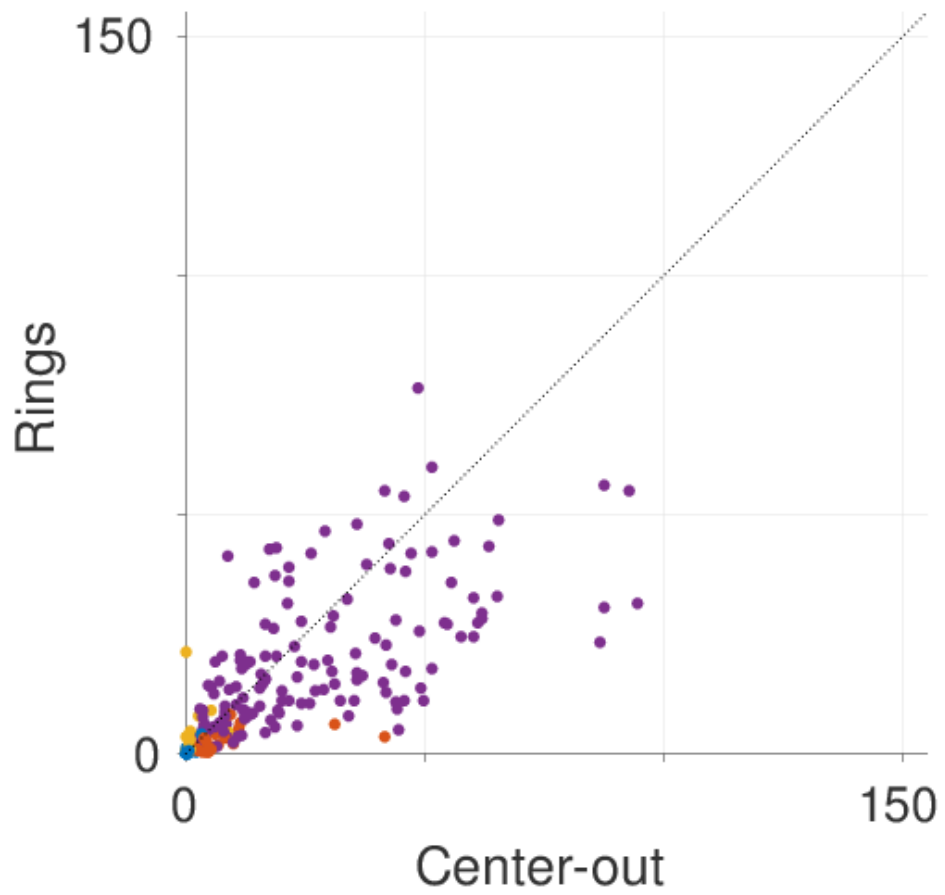


Figure 16: Maximum modulation depths for NHP-C. Compare with Figure 15.

3.5 POPULATION COVARIANCE STRUCTURE

How do patterns of covariation compare within and between the two tasks? We found that the covariance structure in each task tended to be distinct, with some commonality. Figures 17 and 18 show scatter plots of principle components analysis (PCA) scores for 2 of the most significant PCA eigenvectors, for both the center-out task (blue) and the rings task (red). For each subject, the PCA decomposition was computed for a merged data set, consisting of all trials from both the center-out and rings tasks. In each case, the 2D center-out scores tended to assume a roughly planar, toroidal structure in the first 3 eigen-dimensions. This is not entirely surprising, as it mirrors the structure of the task. Interestingly, there tended to be at least two clusters of rings task scores: one cluster intermingled within some sub-set of the volume occupied by the center-out scores, and one cluster lying distinctly outside of this volume, usually perpendicular to the approximate plane of the torus. This can be seen in Figures 17 and 18, both of which are oriented with views along the approximate plane of the torus structure. Importantly, the rings score clusters tended to correspond to rates measured at the high and the low targets, indicating that performance of the rings task corresponded to "moving" in and out of the volume containing the center-out scores.

Figures 17 and 18 illustrate both similarities and differences between the covariance structure in the most significant dimensions of the two data sets, but what does the complete covariance structure look like? To answer this question, we again applied PCA – this time segregating the data by task. That is, we computed a pair of decompositions, using the data from the same two experimental sessions used to generate the score scatter plots. The resulting principle component eigenvalues are represented by pie charts in Figures 19 and 20. The upper-left panel in each figure shows the center-out variance decomposition, and the lower right shows the rings decomposition. Each of the decompositions represents a summary, or model, of the complete covariance structure in each task context. Unsurprisingly, greater than 50% of variance is explained by the first 2 to 3 principle components, in each case. For NHP-C, the first eigenvector accounts for 66% of the variance, but it is important to note that only one unit was conditioned during this session. The off-diagonal pie charts (lower-left and upper-right) represent the covariance decomposition of each data set, taken

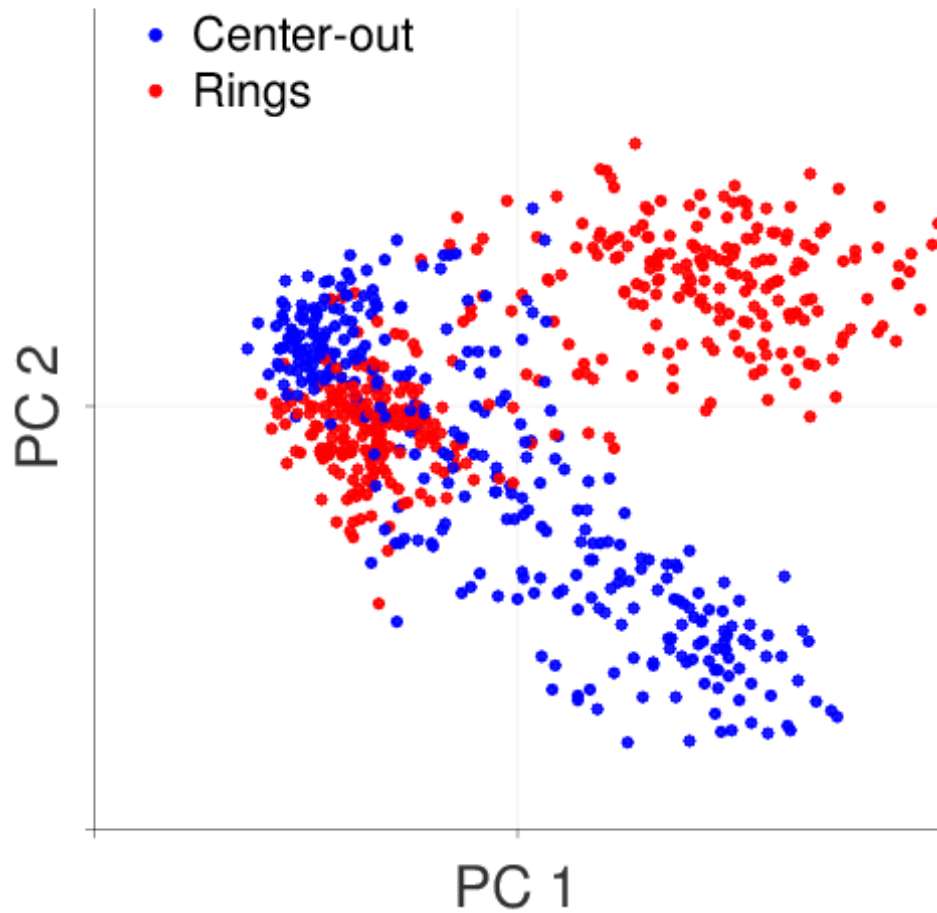


Figure 17: Principal component scores for NHP-A, for all trials within an experimental session. Only two of the higher-order principal component dimensions are plotted. See text for details.

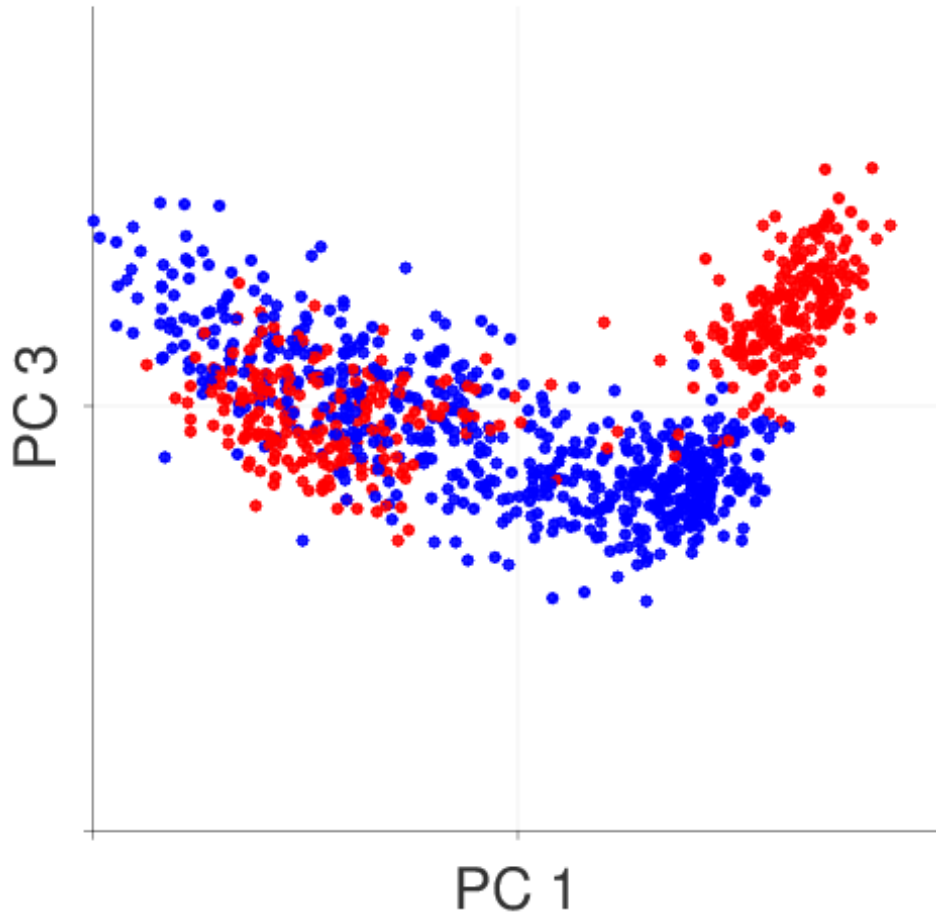


Figure 18: Principal component scores for NHP-C. Compare with Figure 17

with respect to the model fitted to the other data set. So, for example, the slices of the lower left plot correspond to the variance of the rings data explained by each of the eigenvectors of the center-out model (i.e., the PCA decomposition computed from the center-out data). The order of the eigenvectors matches that of the diagonal plots. From these plots, it can clearly be seen that patterns of covariation can differ substantially in each task. Although the first principal component still explains a significant percentage of the variance in each case, particularly for NHP-C, substantial variance is also distributed among lower-order eigenvectors.

Does this tendency for the dominant covariance structure to vary between tasks hold up across experiments? To answer this, we plotted the variance-explained, with respect to the center-out PCA eigenvector decomposition. Pareto plots for this variance decomposition are shown in Figures 21 and 22. For each plot, we performed 20 cross validation iterations, within which PCA was applied to half of the center-out data set. The expected variance explained by each PCA eigenvector was then computed for the held-out center-out data (the remaining half), as well as a selection of the rings data of equivalent size. For each experimental session, the mean and standard deviation of these statistics were computed across cross-validation iterations. Figures 21 and 22 show the average of these statistics across sessions. For each subject, it can clearly be seen that the variance in the rings session is spread away from the dominant center-out PCA eigenvectors, and that there is significant space between the cumulative variance curves. This is consistent with the characteristics of the pie charts of Figures 19 and 20, and indicates substantial differences in the covariance structure of the two data sets, across all sessions.

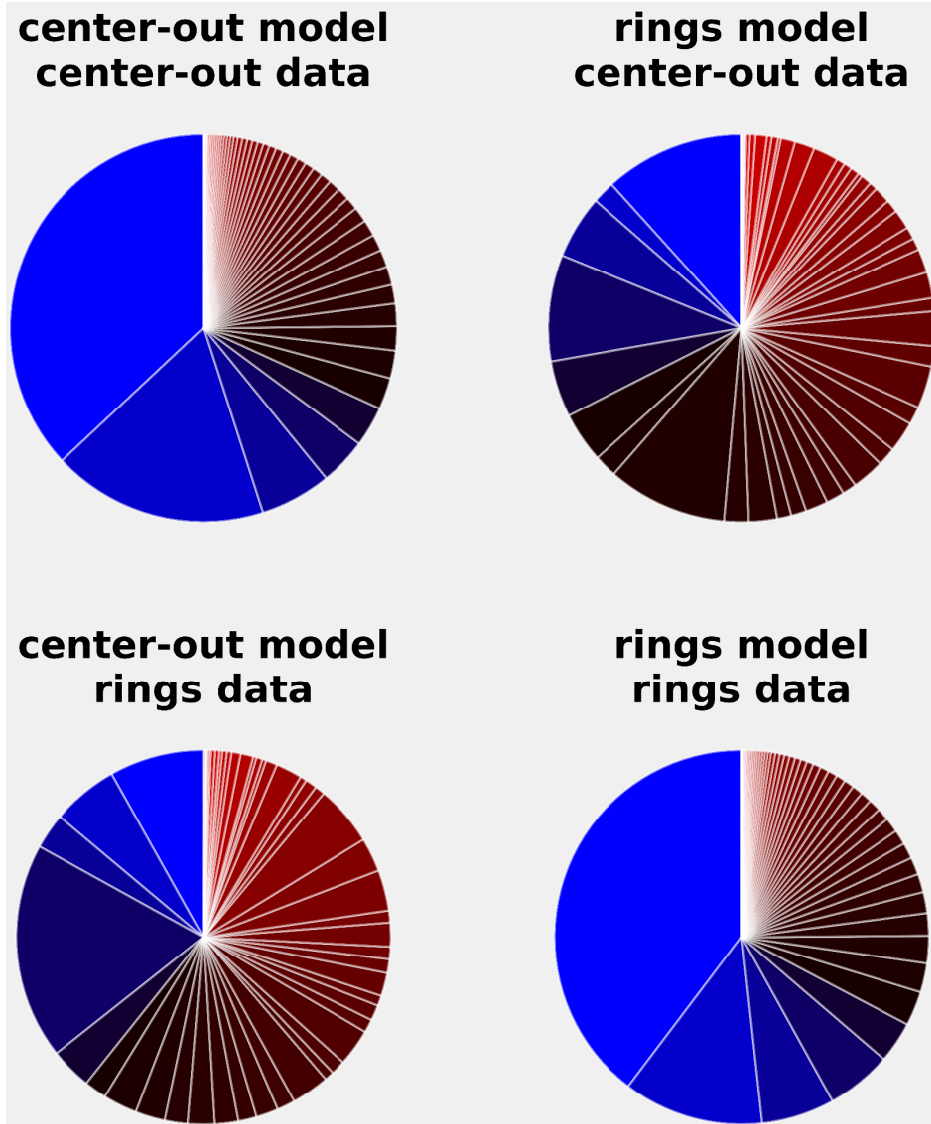


Figure 19: Pie charts of percentage variance-explained by each model, for each data set, for one experimental session with NHP-A. The dataset is identical to that of Figure 17. The upper-left and lower-right pie charts represent the PCA covariance decomposition for the center-out and rings datasets, respectively. That is, each slice of the pie corresponds to the variance explained (i.e., an eigenvalue) by an eigenvector of the sample covariance matrix. The off-diagonal pie charts represent the cross-over decompositions; that is, the slices correspond to the percentage variance-explained for one data set (e.g., center-out), when computed with respect to the model / eigenvectors fitted to the other data set (e.g., rings). See text for further details.

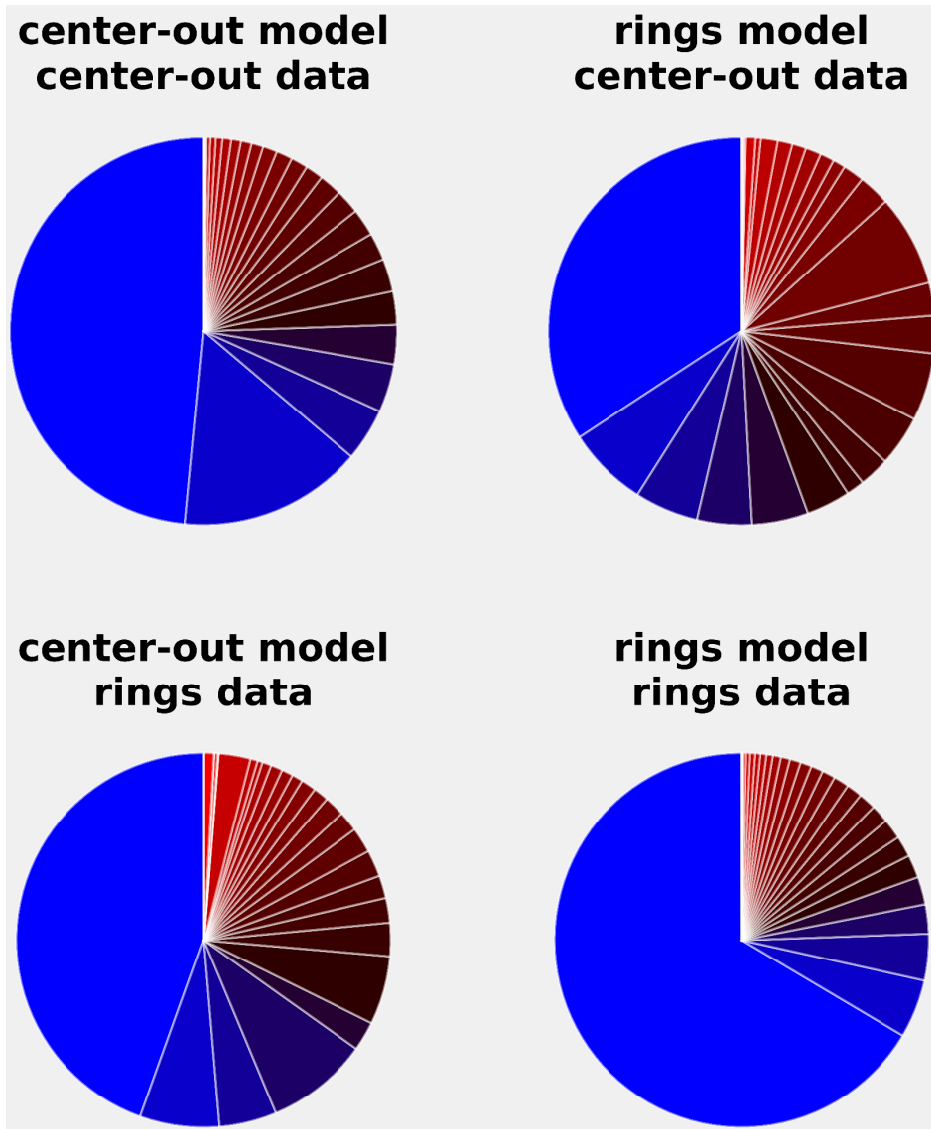


Figure 20: Pie charts of percentage variance-explained by each model, for each data set, for one experimental session with NHP-C. The dataset is identical to that of Figure 18. The interpretation is the same as for Figure 19.

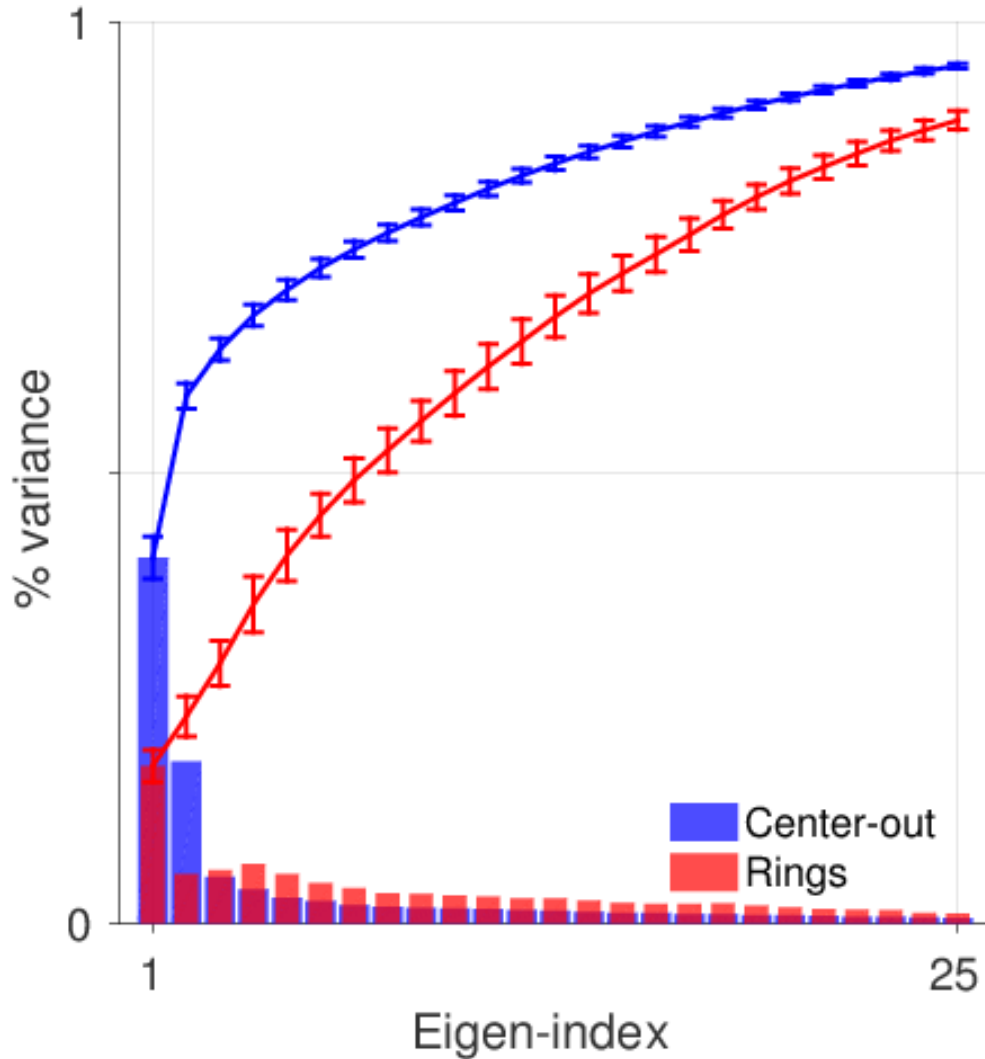


Figure 21: Pareto plot for NHP-A. This plot shows the variance distribution, for each of the tasks, with respect to the eigen-structure of a PCA model fitted to center-out data. The blue lines and bars show data for the center-out task, and the red for the rings task. Bars rising from the bottom of the plot show the variance explained by each principle component eigenvector. These monotonically decrease for the center-out data, but not for the rings data. Ascending lines show the cumulative variance explained. Error bars indicate the standard deviation of the variance cumulative estimates. All statistics are cross-validated across all experimental sessions (see text for details).

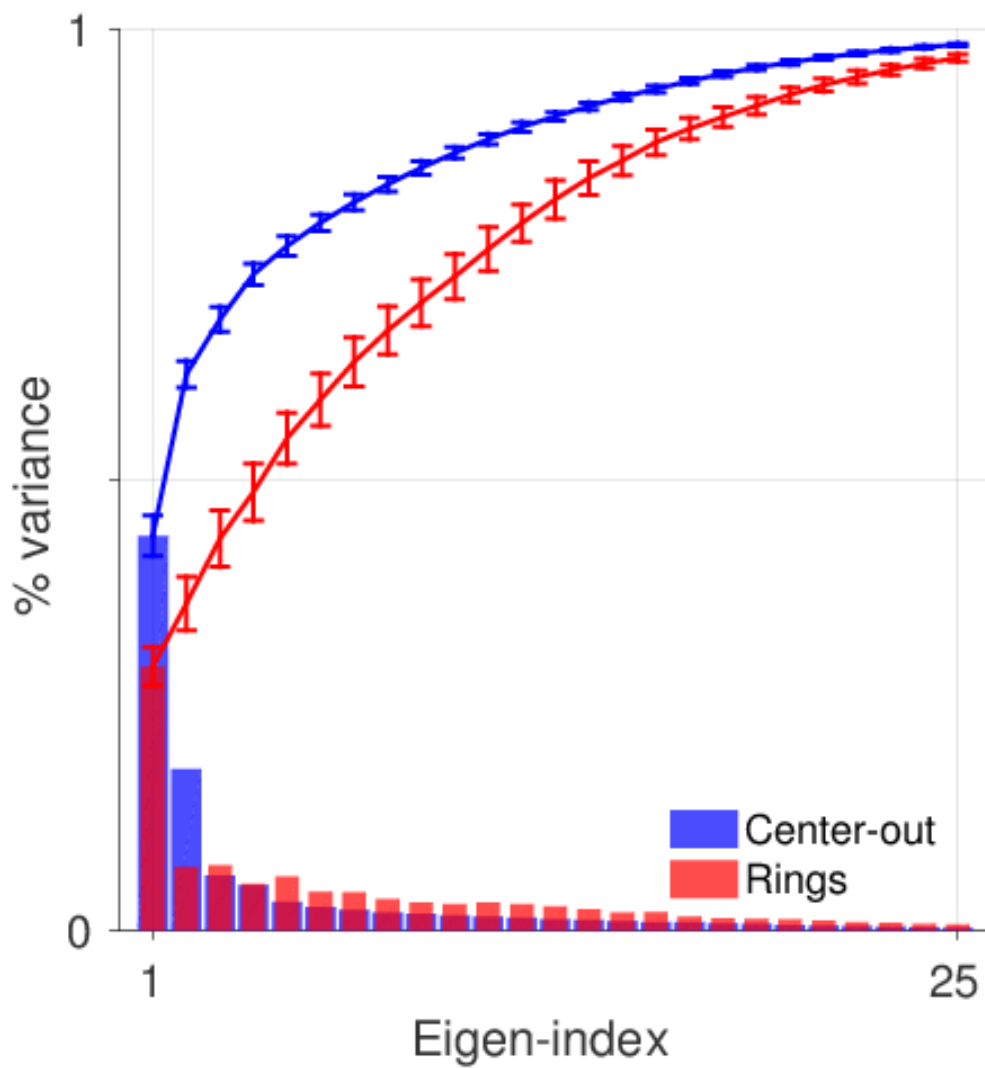


Figure 22: Pareto plot for NHP-C. The interpretation is the same as for Figure 21.

4.0 DISCUSSION

How neurons of the primate motor cortex contribute to intentional behavior is not well-understood. A particular matter of controversy is the degree to which populations of cortical neurons subserve a singular function in the control of behavior, and are constrained by this specific role. For example, it has been suggested that populations of M1 neurons encode the kinematics of hand motion in space [12]. If this is the case, then it could imply that population activity is limited to only those patterns that are observed during hand motion. At the opposite extreme, motor cortical activity might be readily dissociable from hand motion [45] – if, for example, hand kinematics were only one of many variables encoded by the population. In this case, patterns of observed population could be much more diverse, and might be expected to flexibly adapt to changing behavioral priorities.

The current literature paints an uncertain picture of the flexibility of interactions among motor cortical neurons. Here, by flexibility, we mean the extent to which a cortical population is capable of realizing different patterns of activity, as dictated by behavioral demands. On one hand, it has been suggested that motor cortical populations are profoundly flexible, and perhaps capable of realizing arbitrary patterns of activity [21, 25, 26, 27, 29, 34]. On the other hand, it has been suggested that motor cortical neurons covary in a systematic fashion [3, 2], that such patterns of population activity are strongly-linked to natural behavior [6, 4], and that such intrinsic structure might constrain the flexibility of cortical activity [9, 10, 43].

We sought to explore this issue by characterizing patterns of population activity during operant conditioning of the firing rates of single motor cortical neurons. The single unit operant conditioning paradigm dates back to the early work of Fetz and colleagues [17, 21, 20]. Historically, however, the simultaneous analysis of cortical activity that is not relevant to behavioral goals (i.e., non-contingent activity), has focused on pairs of units

recorded on the same electrode [21, 32, 33], on pairs of nearby neurons [40], or on small populations of 4 or fewer neurons [34]. Our work is distinguished by quantitative analysis of simultaneous activity of small populations ($N > 20$) of neurons, during reinforcement of single unit firing rates – and by a particular focus on the covariance structure of population activity. Previously, only Chase and Schwartz [10] have reported related results. The present work may be considered to be an extension and refinement of aspects of that report.

4.1 COVARIATION OF NON-CONTINGENT CORTICAL ACTIVITY

During single unit operant conditioning, it was previously shown that the activity of non-contingent neurons tends to covary with that of the contingent neurons. Such behaviorally-irrelevant covariation has been characterized both as having lower intensity [35, 34], and far greater intensity [21], than the activity of contingent neurons. Depending on analysis parameters, either a minority [40, 32] or a majority [21, 33] of non-contingent neurons were deemed correlated. Correlations seem to be more consistent within local regions of cortex [21, 40, 33] than distant [31]. In one study of a local population, the strongest correlations were observed on electrodes 3.8 mm apart [40]. Although not exclusively [21], covariation tends to be considered more meaningful at fine timescale [40, 33], rather than coarse [32].

We confirm the observation of non-contingent covariation, and extend the characterization. Patterns of covariation – and the relative consistency of those patterns – are most clearly exemplified in the rasters of Figures 8 and 9. The extent of covariation differed by condition – that is, according to which neuron the behavioral contingency was defined in terms of. Across all single unit contingencies, we observed as few as 4 non-contingent units that co-modulated with behaviorally-relevant activity, and as many as 34 (see Figure 10). Although the number of covarying neurons changed with the contingency, it was not uncommon for a large percentage of the population to co-modulate. In no case did the contingent neuron modulate alone. Like the number of modulated units, the intensity and significance of pairwise correlation also varied by condition. There was, however, a tendency toward maintaining a consistent relative pairwise relationship – as is evident by the positive

diagonal trend in Figures 11 and 12. The magnitude of pairwise correlations tended to be less than 0.50, but did exceed 0.90, in some cases. Correlation coefficients exceeded 0.30 for a considerable number of sampled pairs – more than the 5 cases reported by Schmidt et al. [40]. When considered as a population, patterns of pairwise covariation were rather consistent, within conditions.

This observation recalls the concept of the "motor field" suggested by Fetz and Finocchio, and can be considered in terms of the early interpretations of Fetz et al. The motor field of a neuron was defined as the set of muscles that co-varied during reinforcement of variation in the firing rate of that neuron. Fetz et al. noted that, although the intensity of muscle activation could fluctuate, the rank order was relatively consistent, for a fixed single unit contingency. If we analogously define the "cortical field" of a neuron to be those neighboring neurons that co-vary during reinforcement, then the same property seems to hold true here: even if the magnitude of modulation varies, the pattern of co-variation seems to be consistent, within each condition. This can be picked out in Figures 8 and 9, by comparing patterns of activity across trials. The flexibility of the cortical population, then, would be directly dependent on the extent of the overlap in cortical fields: if the cortical fields are highly redundant, then the set of possible patterns of discharge would be limited. Fetz et al., however, consider the "strength" – or meaningfulness – of correlation to be defined in terms of both the intensity of the correlation, and the consistency across behavioral conditions.

We asked if any such consistency could be identified. In this case, patterns of neuron-neuron covariation changed with the contingency, and it is not immediately apparent to what extent pairwise relationships are conserved. How can we uncover patterns that are, in fact, conserved? As Chase and Schwartz suggest [10], dimensionality reduction methods offer a means for identifying dominant covariance structure, which could represent the influence of intrinsic factors that shape population activity – such as synaptic connectivity. Our own principal components analysis (PCA) – as pictured, for example, in the lower right panel of Figures 19 and 20 – supports this idea. That a majority of variance in population activity, across contingencies, is explained by the first 2 to 3 principal components suggests that some facets of population activity are indeed conserved.

The observed populations – consisting of dozens of neurons – discharged patterns of activity that varied in a restricted range. This might be interpreted to indicate a relatively constrained capability for variation. As judged by this criterion, the degree of flexibility of the cortical population would then be considered to fall somewhere between ”profoundly flexible” and ”profoundly rigid”. In other words, the number of significant dimensions – or degrees-of-freedom – appears to be non-trivial, but somewhat less than the number of observed neurons.

4.2 RELEVANCE TO ”NATURAL” BEHAVIOR

It has been suggested that such consistent patterns of covariation could reflect ensemble interactions or intrinsic structural constraints induced by the ”natural” role of the motor cortical network [25, 28, 41, 34]. While we did not address this question directly, we did compare patterns of activity observed during the single unit conditioning paradigm with those observed during reinforcement of a behavioral contingency that is believed to be more ”natural”. Specifically, the population vector algorithm (PVA) [5, 3] was calibrated to map patterns of M1 population activity to the velocity of intended cursor motion, in a manner modeled after the relationship between M1 neurons and hand motion. In the literature, such an approach is often contrasted with the operant conditioning (or ”biofeedback”) approach [25, 28, 10]. The former – sometimes referred to as the ”biomimetic approach” – is considered to be more ethologically appropriate than the latter. Consequently, it is thought ”to confer immediate intuitive control without undue cognitive load” [28].

We observed two prominent differences in first order patterns of population activity, between the two paradigms. First, the number of modulated units varied more during the single unit task than it did during the PVA-based task (compare Figures 10 and 13). Second, the modulation depths of non-contingent units was greater, on average, in the PVA-based task (see Figures 15 and 16). Quite possibly, these differences reflect a natural predisposition of the cortical population for the PVA / center-out task. However, we consider it equally likely that they reflect differences in task construction and experimental design.

This is an important consideration, since little effort was made to match the parameters of the behavioral contingencies between the two tasks, in terms of how they related to cortical activity patterns. Thus, the increased modulation depth in the PVA / center-out task might have occurred simply because the task parameters required more modulation for reward. Similarly, the more consistent engagement of units in the center-out task could be a consequence of the fact that the structure of the task was more consistent and predictable than that of the single unit task – that is, different conditions were related to each other in a fixed and systematic fashion, such that transition between conditions might have been facilitated.

Our comparison of pairwise relationships among cortical neurons was arguably more interesting. Altogether, we found both substantial similarities and substantial differences between the two paradigms. This is most completely represented in the Pareto plots of Figures 21 and 22. In those plots, the first several principal components clearly account for a high percentage of variability in both data sets, indicating that some patterns of pairwise covariation were conserved. However, the percentage of variability of the rings task data associated with lower-order principal components was non-trivial. This indicates that some population activity patterns observed in the rings task were unlikely to be observed in the center-out task – that they were unique to the rings task – and vice-versa. These differences are emphasized and summarized by the gap between the cumulative variance curves in each plot. Specific examples of these similarities and differences are represented by the area of pie slices in the off-diagonal (top-right and bottom-left) plots of Figures 19 and 20.

Such significant differences between the patterns of population activity observed in the two tasks seems to indicate that the underlying drivers of the population were at least somewhat distinct – or that they were applied to the tasks in distinct ways. This is effectively visualized in the principal component eigen-space plots of Figures 17 and 18. It is instructive to interpret this observed structure – and its potential behavioral implications – in terms of the concept of re-aiming [9]. The re-aiming model suggests that the single unit contingency would most naturally be satisfied if the subject were to imagine "moving" between surrogate targets within the PVA / center-out workspace. This is at-odds with what we observed. In terms of the score clusters of Figures 17 and 17, moving between center-out targets would

approximately translate to moving within a 2D plane, oriented roughly perpendicularly to the plane of each plot. However, the patterns of activity during single unit conditioning are shown to modulate outside of this plane – at an angle of 45° or more. This is somewhat surprising, in light of evidence that re-aiming dominates the process of optimizing cortical activity to match a particular behavioral contingency [9], and evidence that a re-aiming mechanism can explain modulation of single unit firing rates in parietal cortex [41]. Movement outside of the plane might, in fact, indicate re-aiming in a higher dimensional space (3D or more), but the question remains as to why different dimensions are preferred in the single unit task. The data we present are not sufficient to answer this question. In any case, that patterns of population activity observed in the single unit conditioning task do not exclusively and preferentially fall within the planar structure, suggests that the population vector approach is not more "natural" or "intuitive" than the single unit contingency. If it were, then the patterns of population activity observed during single unit conditioning would tend to be entirely subsumed by the space associated with the higher-dimensional center-out task.

As well as for considering the differences in population activity, Figures 17 and 18 are effective examples for illustrating the similarities. Although the rings task principal component scores deviate from the PVA / center-out scores, in each plot, there are also substantial regions of overlap. This reflects a preference for particular patterns of population activity. In terms of the re-aiming concept, for example, this could mean that an increase in the radius of the ring cursor might have been accomplished via a behavioral mechanism that was equivalent to that used to move the spherical cursor particular targets in the center-out task. Alternatively, the preferential patterns might also simply correspond to some resting or inactive state, representing lower effort for the subject in both tasks. In either case, the tendency of population activity toward these patterns at least indicates a preference for, if not a hard constraint on, particular population patterns. This suggests limits on the flexibility of motor cortex, and at least calls into question the notion that cortical populations are capable of realizing arbitrary patterns of activity, with equal effort.

Ultimately, our results are more suggestive than conclusive. Based on what we have observed, we are inclined to reject the interpretation of Hwang et al. [41] that cortical activity patterns tend to be restricted to by the "natural movement repertoire". However, we are

equally inclined to reject the suggestion by other groups [26, 27, 29, 34] that "arbitrarily-selected motor cortex neurons" are capable of varying independently, "regardless of any previous association to movement", or other variables. Instead, the answer likely lies somewhere in the middle. The most appropriate interpretation is likely that there are similarities between the processes underlying single unit conditioning and the "biomimetic" control scenario, but that they are not equivalent.

4.3 DISSOCIATION AND DIFFERENTIAL CONDITIONING

The idea of designing behavioral contingencies in terms of previously-observed patterns of activity dates back to the early differential conditioning experiments of Fetz et al. [21, 20]. Fetz and Finnochio suggested that a "powerful test of the consistency of an observed correlation is to directly reinforce its dissociation" [20]. In a similar manner, such differential conditioning is a powerful test of the flexibility of a cortical population. More recently, it has been suggested that even existing patterns of covariation can easily be trained-away, and that the firing rates of cortical neurons can be dissociated from one another, with only modest practice [26, 27, 29, 34]. However – for both the old work and the new – the evidence for successful dissociation is primarily anecdotal and qualitative, with small sample sizes.

We did not attempt differential conditioning, or any sort of reinforced dissociation of correlated activity, for two primary reasons. First, there is an added technical burden, and a more substantial learning curve. Very likely, this is the unstated cause of the shortage of samples and repetitions, in prior work. Second, we were primarily interested in characterizing the baseline "cortical fields" of M1 neurons; that is, the set of non-contingent neurons, among the local population that co-vary with a contingent neuron. We were interested in what such fields looked like (e.g., how many neurons co-vary), and how they varied by contingency. We did not aim to characterize how such fields change with learning.

4.4 CONCLUSIONS

Based on our results, we suggest three main points. First, neurons of the motor cortex are not independent: they vary in coordinated patterns, even when only a subset are relevant to behavioral goals. Second, patterns of population activity change with the behavioral contingency – even when only a subset of the population is relevant to behavioral output – but some neuron-neuron relationships are preserved across behaviors. Such preserved features could be indicators of stable network structure, such as synaptic connectivity. Third, although ”biomimetic” approaches to brain-computer interfaces and cortical conditioning may confer practical, learning-related advantages, we found no strong evidence for preferred – or ”natural” – patterns of population activity. Altogether, our results suggest that cortical population activity is neither entirely labile nor entirely fixed, but that it is instead governed by complex constraints – which have yet to be carefully characterized.

Although our conclusions are limited, our observations do suggest a path forward. The underlying form and function (i.e., network structure and computational role) of a cortical population is reflected in the patterns of activity that the population discharges, and in how those patterns relate to goal-directed behavior. Whether or not such patterns represent transient cell assemblies or stable anatomical structures, they can be applied as ”a language with which to phrase new questions for the next generation of experiments” [37]. By testing hypotheses expressed entirely in terms of observed patterns of population activity – summarized, for example, via principal components analysis – we can make indirect inferences about the unobserved structures and processes that drive the population, while avoiding a priori assumptions about the nature of those drivers. Effectively, this is the approach advocated by Chase and Schwartz [10], and realized by Sadtler et al. [43]. We suggest that it represents the most promising avenue for future research.

BIBLIOGRAPHY

- [1] Moran DW, Schwartz AB. Motor cortical representation of speed and direction during reaching. *J Neurophysiol.* 1999 Nov;82(5):2676-92
- [2] Schwartz AB. Direct cortical representation of drawing. *Science.* 1994 Jul 22;265(5171):540-2.
- [3] Georgopoulos AP, Kettner RE, Schwartz AB. Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *J Neurosci.* 1988 Aug;8(8):2928-37.
- [4] Schwartz AB, Kettner RE, Georgopoulos AP. Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement. *J Neurosci.* 1988 Aug;8(8):2913-27.
- [5] Georgopoulos AP, Schwartz AB, Kettner RE. Neuronal population coding of movement direction. *Science.* 1986 Sep 26;233(4771):1416-9.
- [6] Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci.* 1982 Nov;2(11):1527-37
- [7] Taylor DM, Tillery SI, Schwartz AB. Direct cortical control of 3D neuroprosthetic devices. *Science.* 2002 Jun 7;296(5574):1829-32.
- [8] Schwartz AB. Useful signals from motor cortex. *J Physiol.* 2007 Mar 15;579(Pt 3):581-601.
- [9] Jarosiewicz B, Chase SM, Fraser GW, Velliste M, Kass RE, Schwartz AB. Functional network reorganization during learning in a brain-computer interface paradigm. *Proc Natl Acad Sci U S A.* 2008 Dec 9;105(49):19486-91.
- [10] Chase SM, Schwartz AB. Inference from populations: going beyond models. *Prog Brain Res.* 2011;192:103-12.
- [11] Kalaska JF. From intention to action: motor cortex and the control of reaching movements. *Adv Exp Med Biol.* 2009;629:139-78.

- [12] Georgopoulos AP, Carpenter AF. Coding of movements in the motor cortex. *Curr Opin Neurobiol.* 2015 Aug;33:34-9.
- [13] Evarts EV. Activity of pyramidal tract neurons during postural fixation. *J Neurophysiol.* 1969 May;32(3):375-85.
- [14] Evarts EV. Relation of pyramidal tract activity to force exerted during voluntary movement. *J Neurophysiol.* 1968 Jan;31(1):14-27.
- [15] Evarts EV. Pyramidal tract activity associated with a conditioned hand movement in the monkey. *J Neurophysiol.* 1966 Nov;29(6):1011-27.
- [16] Evarts EV. Relation of discharge frequency to conduction velocity in pyramidal tract neurons. *J Neurophysiol.* 1965 Mar;28:216-28.
- [17] Fetz EE. Operant conditioning of cortical unit activity. *Science.* 1969 Feb 28;163(3870):955-8.
- [18] Fetz EE, Finocchio DV. Operant conditioning of specific patterns of neural and muscular activity. *Science.* 1971 Oct 22;174(4007):431-5.
- [19] Fetz EE, Finocchio DV. Operant conditioning of isolated activity in specific muscles and precentral cells. *Brain Res.* 1972 May 12;40(1):19-23.
- [20] Fetz EE, Finocchio DV. Correlations between activity of motor cortex cells and arm muscles during operantly conditioned response patterns. *Exp Brain Res.* 1975 Sep 29;23(3):217-40.
- [21] Fetz EE, Baker MA. Operantly conditioned patterns on precentral unit activity and correlated responses in adjacent cells and contralateral muscles. *J Neurophysiol.* 1973 Mar;36(2):179-204.
- [22] Fetz EE. Operant control of single unit activity and correlated motor responses. *Operant Control of Brain Activity* (Ed.: MH Chase). Brain Research Institute, UCLA. 1974:61-89.
- [23] Fetz EE. Biofeedback and differential conditioning of response patterns in the skeletal motor system. In *Biofeedback and behavior 1977* (pp. 413-425). Springer US.
- [24] Fetz EE. Role of motor cortex cells in control of operantly conditioned muscle activity. In *Conditioning 1982* (pp. 281-291). Springer US.
- [25] Fetz EE. Volitional control of neural activity: implications for brain-computer interfaces. *J Physiol.* 2007 Mar 15;579(Pt 3):571-9.
- [26] Moritz CT, Perlmutter SI, Fetz EE. Direct control of paralysed muscles by cortical neurons. *Nature.* 2008 Dec 4;456(7222):639-42.

- [27] Moritz CT, Fetz EE. Volitional control of single cortical neurons in a brain-machine interface. *J Neural Eng.* 2011 Apr;8(2):025017.
- [28] Jackson A, Fetz EE. Interfacing with the computational brain. *IEEE Trans Neural Syst Rehabil Eng.* 2011 Oct;19(5):534-41.
- [29] Milovanovic I, Robinson R, Fetz EE, Moritz CT. Simultaneous and independent control of a brain-computer interface and contralateral limb movement. *Brain Comput Interfaces.* 2015;2(4):174-185.
- [30] Wyler AR. Operant conditioning of epileptic neurons in monkeys and its theoretical application to EEG operant conditioning in humans. *Pavlov J Biol Sci.* 1977 Jul-Sep;12(3):130-46.
- [31] Wyler AR, Lange SC, Robbins CA. Operant control of precentral neurons: bilateral single unit conditioning. *Brain Res.* 1980 Aug 18;195(2):337-44.
- [32] Wyler AR. Interneuronal synchrony in precentral cortex of monkeys during operant conditioning. *Exp Neurol.* 1983 Jun;80(3):697-707.
- [33] Wyler AR. Synchrony between cortical neurons during operant conditioning. *Brain Res.* 1985 Aug 19;341(1):66-72.
- [34] Law AJ, Rivlis G, Schieber MH. Rapid acquisition of novel interface control by small ensembles of arbitrarily selected primary motor cortex neurons. *J Neurophysiol.* 2014 Sep 15;112(6):1528-48.
- [35] Ganguly K, Dimitrov DF, Wallis JD, Carmena JM. Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nat Neurosci.* 2011 May;14(5):662-7.
- [36] Pennycook G, Cheyne JA, Barr N, Koehler DJ, Fugelsang JA. On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making.* 2015. 10(6):549.
- [37] Stephens GJ, Osborne LC, Bialek W. Searching for simplicity in the analysis of neurons and behavior. *Proc Natl Acad Sci USA.* 2011 Sep 13;108 Suppl 3:15565-71.
- [38] Scott SH. Neuroscience: converting thoughts into action. *Nature.* 2006 Jul 13;442(7099):141-2.
- [39] Buzski G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron.* 2010 Nov 4;68(3):362-85.
- [40] Schmidt EM, Bak MJ, McIntosh JS, Thomas JS. Operant conditioning of firing patterns in monkey cortical neurons. *Exp Neurol.* 1977 Mar;54(3):467-77.
- [41] Hwang EJ, Bailey PM, Andersen RA. Volitional control of neural activity relies on the natural motor repertoire. *Curr Biol.* 2013 Mar 4;23(5):353-61.

- [42] Golub MD, Yu BM, Chase SM. Internal models for interpreting neural population activity during sensorimotor control. *Elife*. 2015 Dec 8;4.
- [43] Sadtler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-Kabara EC, Yu BM, Batista AP. Neural constraints on learning. *Nature*. 2014 Aug 28;512(7515):423-6.
- [44] Humphrey DR, Schmidt EM, Thompson WD. Predicting measures of motor performance from multiple cortical spike trains. *Science*. 1970 Nov 13;170(3959):758-62.
- [45] Schieber MH. Dissociating motor cortex from the motor. *J Physiol*. 2011 Dec 1;589(Pt 23):5613-24.