The most informative neural code accounts for population heterogeneity

Elizabeth Zavitz[1,2,3], Hsin-Hao Yu[1,2], Marcello GP Rosa[1,2,3], Nicholas SC Price[1,2,3]

(1) Department of Physiology, Monash University, Melbourne, Victoria, Australia
(2) Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia
(3) ARC Centre of Excellence in Integrative Brain Function

Corresponding Author:
Elizabeth Zavitz
elizabeth.zavitz@monash.edu

Competing interests:
The authors have declared that no competing interests exist.

Funding:
This work was funded by Australia's National Health and Medical Research Council, the Human Frontier Science Program, and the Australian Research Council.

Author Contributions:
EZ and NSCP designed the study, EZ, HHY, MGPR, and NSCP ran the experiments, EZ performed the analysis and modelling, EZ and NSCP wrote the paper.

Abstract

Sensory perception depends on the accurate representation of the external world by neuronal populations. Each neuron's contribution to a population code depends on its tuning preferences, but neuronal populations are heterogeneous: their spiking rate, spiking variability and tuning selectivity are all different. To understand how this heterogeneity impacts a neuron's contribution to the population code, we recorded responses to moving stimuli from motion-sensitive neurons in anaesthetized marmosets (*Callithrix jacchus*) and trained linear decoders to discriminate direction using these responses. The relationship between a neuron's preferred direction and the discrimination boundary was the strongest predictor of its decoding weight, and highly direction selective neurons are the most useful in the population code. Millisecond-timescale changes in neuronal responses mean that optimal weights change rapidly; however, perceptual readouts do not have this fine-grained temporal flexibility, and must perform sub-optimal decoding at each point in time.

Author Summary

It is widely understood that perception depends upon interpreting, or decoding, the pattern of activity across a neuronal population. To do this, firing rates from hundreds of sensory neurons must be weighted and combined to accurately represent stimulus properties. Despite the reality of a diverse, heterogeneous population of neurons, all we know is that their weighting is a function of the relationship between their tuning preference, and the decoding decision or perceptual task under consideration. In this manuscript we ask: which neurons in a heterogeneous population are the most informative, can their usefulness can be predicted by their response properties, and to what degree can we expect them to be optimized? We addressed this question using a machine learning approach to learn how to weight each neuron's response in order to perform a range of discrimination tasks, and fit a series of decoding models to show that models based on preferred direction alone are significantly improved if neurons that are more selective are weighted more strongly. These

results show that we can perform better decoding without machine learning if we address the false assumption that all members of the neural population are equally informative.

Introduction

Each of our perceptions is not a reflection of the activity of a single neuron, but an interpretation of the aggregate activity across a population of neurons. Firing rates from ensembles of hundreds of neurons must be weighted and combined to accurately represent stimulus properties. We understand some aspects of how a perceptual 'readout' may be produced from this population activity, such as how neurons with different tuning preferences can be weighted and evaluated in order to make discriminations (1–3) or to identify where a stimulus lies on a continuous scale (4,5). Often, the neural population is idealized as a sequence of identical bell-shaped tuning curves whose peaks are offset from one another to uniformly tile the stimulus space. In this case it is reasonable for every neuron to contribute equally to the readout. However, cortical neurons are diverse: within a single brain area they vary in sharpness of tuning, selectivity, overall firing rates, and trial-to-trial variability (6). Here we address the problem of optimal coding in a heterogeneous population by asking which neurons are the most informative, whether their usefulness can be predicted by their response properties, and to what degree we can expect them to be optimized.

Within the brain, the firing rate of a given neuron depends on the firing rate of its inputs, and the relative weighting of each input. This weighting is governed by the strength of synaptic connections from each input neuron. In the context of sensory decision making, the magnitude of a neuron's synaptic weight indicates the amount to which it contributes to a decision. A computational decoder emulating a decision making process thus represents the activity of readout neurons that receive inputs from a large population of sensory neurons, and its weights are analogous to synaptic weights (7). To decode neural activity computationally, each neuron may be assigned multiplicative weights *a priori*, based on known attributes of the neurons such as preferred direction. This is done often in theoretical

contexts: all neurons are assumed to be identical aside from their tuning preference, and they are assigned weights based on that preference alone. Alternately, weights may be derived with machine learning approaches. As expected (8), the weights that individual neurons are assigned by decoders trained to perform fine discrimination tasks are highest when the individual tuning curves have the most discriminatory power because their tuning curves are steep or have relatively low inter-trial variability (1,9). Even after accounting for this relationship, there is still a significant amount of variability in the weights that neurons in heterogeneous populations are assigned.

How a decoding model's performance varies over time can provide meaningful insight into how optimal neural coding may be. When decoders are trained and tested at matching time points throughout the stimulus response, their weights are optimized to the responses at each time (9,10). However, spike-rate adaptation following strong stimulus drive is associated with changes in firing rates and tuning properties. These stimulus-specific changes in firing rate result in perceptual errors (11). If the weights that neural circuits use to decode population activity are continuously and rapidly updated, adaptation would not affect perception because the readout would be continually optimized to produce an accurate percept. That we observe adaptation-related effects in a multitude of contexts and modalities suggests that readout weightings are stable over short timescales (12), and thus that coding is not optimal over the same time scales.

In this paper, we used the responses from heterogeneous populations recorded in area MT of the non-human primate visual cortex to examine how the direction of moving patterns is encoded by neural populations. To determine which neurons are most informative, we derived weights for each neuron using a machine learning model. We compared these weights against the tuning properties of each neuron in order to predict which neurons will be useful. We found that direction selectivity both predicted the weight of the recorded neurons, and enhanced the performance of *a priori* decoding models (in which weights are pre-defined, rather than learned), independent of the readout decision (identification or

discrimination), and population size. Our results suggest that those neurons that are highly selective contribute the most to the population code, and that little optimization in weights is needed for the predicted perceptual errors between trials to equal the degree of error that is tolerated over time.

Results

The activity across a population of neurons is weighted and interpreted to produce perceptual experiences. In the context of motion processing, theoretical and empirical methods have demonstrated that a neuron's utility in a population code depends primarily on its preferred direction, but how other response characteristics impact population coding is unknown.

To assess how response properties other than preferred direction influence coding weight, we recorded extracellular responses from motion-sensitive neurons in the middle temporal area (MT) of four anaesthetized common marmosets (*Callithrix jacchus*) while we displayed moving dot patterns. The stimulus dots moved coherently in one of 12 directions for 500 ms before a 500 ms blank period, and each direction was repeated 120 times. We combined trial-shuffled responses for 136 single units, then applied a supervised learning algorithm to decode them based on one of three schemes. Fine and coarse two-alternative discriminations were decoded based on a binomial generalized linear model (Figure 1A), and 12-alternative identification was decoded based on a multinomial generalized linear model (Figure 1B).

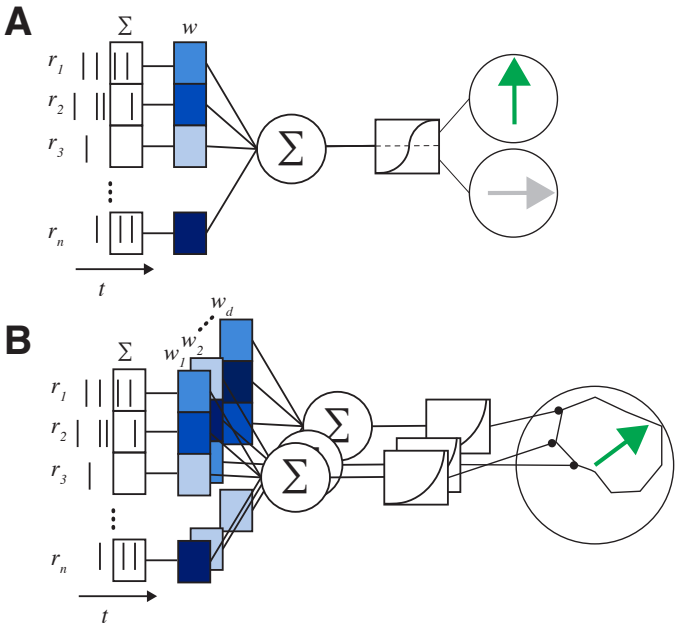


Figure 1: Decoding neural population activity. (A) Schematic of the two-alternative bimodal generalized linear model. The spiking responses (*r*) of a population of *n* neurons are integrated in a spike counting window, multiplied by a weight (*w*), summed, transformed, then compared to a criterion. Therefore every pattern of spike counts predicts one of the two possible directions. **(B)** Schematic of the identification multinomial generalized linear model. In this model, spike counts are weighted, integrated and transformed multiple times: once for each possible stimulus direction category. These responses are subsequently combined using a vector average to produce a direction prediction on a continuous scale. Continuous estimates within ±15° of the stimulus were considered 'correct'.

The purpose of the supervised learning process is to choose weights to apply to each neuron's spike count which maximize decoding performance at each point in time. In both models, the weights learned for each neuron are representative of synaptic connection strength, but we use them here as a proxy for how important that neuron is to making each perceptual decision. More broadly, these models allow us to examine how the importance of a neuron varies depending on its response properties, the time since stimulus onset, and which decision is being made.

Decoding weights generalize well, but not perfectly, over time

To see how variable the neural code is over the 500 ms stimulus duration, we can repeat training and decoding at every time point and measure how weights generalize from one time point to another. The generalization matrix for the 12-AFC decoder (Figure 2A) shows decoding performance for every pairing of training and test times, for spikes counted in

sliding 50 ms time windows. In this case, decoding performance peaks at around 40% correct (chance is 8.3%). Note that we are primarily interested in how changing aspects of the decoder affects relative decoding performance, rather than trying to maximize absolute performance. Along the diagonal (solid box), weights are trained and tested on matching time points. The horizontal boxes show how weights from training early (dotted) and late (dashed) within a trial after motion onset generalize to all other decoding times.

We resampled our sub-populations in this and subsequent analyses because a neuron's weight depends on what information is available from other neurons. We could not assume that the sub-population of 136 neurons we recorded is an accurate reflection of the population that contributes to each readout in the whole animal. We thus resampled smaller sub-populations in order to estimate the sample-related variability in performance and weight.

In both the 12- and two-alternative identification decisions (Figure 2B & C), the performance of the decoders that were trained only at one time depends on the time during which they were trained. Training early (dotted line, 200 ms) only performs as well as the decoder trained at every time point in the earliest part of the response. Training late (dashed line, 400 ms) performs as well as the decoder optimised at every time point through both the mid and late parts of the response. In both cases, training only once leads to below-chance performance following stimulus offset (Figure 2 B&C, dashed lines, 550 ms onward). This reflects transient adaptation effects that only the temporally-matched decoder can take advantage of (10,13,14).

These results show at least three distinct epochs in decoding performance: an early, contrast driven code (50-250 ms), a late sustained code (250-550 ms), and an adaptation-driven code (550 ms onward). The effects of adaptation following stimulus offset mean that models trained only once produce systematic errors that are consistent with perceptual aftereffects. Because we have evidence in the form of aftereffects that these coding errors affect perception, it is unlikely that the brain benefits from decoding weights that are

optimized at each time point after stimulus onset. This implies that whatever weights the brain uses, they will not be optimized at all time points, so the weaker decoding performance when the model fails to generalize perfectly may represent a good estimate of how much error the brain tolerates in a given perceptual code.

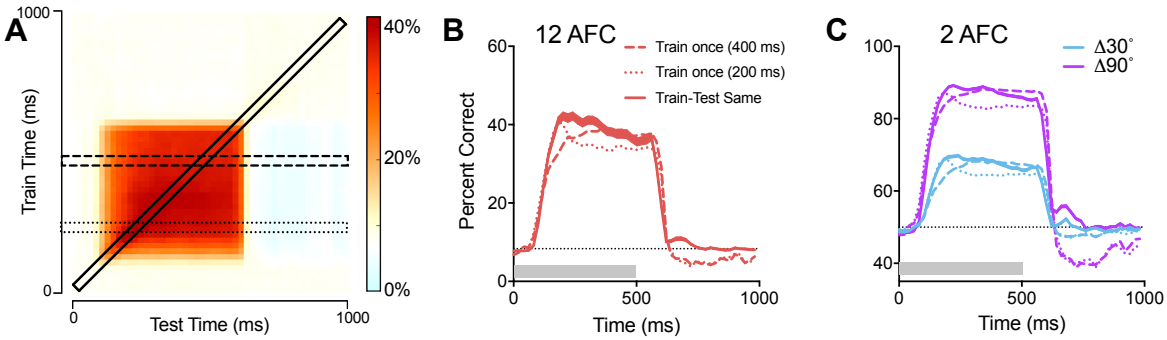


Figure 2: Decoding weights generalize well, but not perfectly, over time. (A) Full generalization matrix for 12-alternative decoding (chance = 8.3%). Percent correct is indicated by color for each combination of training and testing time. **(B)** Twelve-alternative decoding performance for three slices of the generalization matrix: train-test same (solid line), train 200 ms (dotted line), train 400 ms (dashed line). **(C)** as in B for fine and coarse two-alternative discrimination (chance = 50%).

Population size affects decoding performance, not weight structure

Decoding performance as a measure of encoded information depends on many factors, including the size of the spike-counting window for each time point (the integration window), and the size of the neural population (*n*) from which responses are integrated. These factors affect the structure of the recovered decoding weights, because when decoding is easy enough that performance is at ceiling, weights cannot be further optimized. In other words, the weighting that allows a population to solve an easy task perfectly may not extract enough information from the population to solve a more difficult task, even if the information is there. To circumvent this, we varied the integration window in order to match different population sizes for decoding performance, so that the weights could be examined without the number of neurons contributing to the population confounding our results. For simplicity, we examined the weights learned at a representative time point (200 ms) after stimulus onset.

To measure the effect of population size on how the weights were assigned to different neurons, we first calculated decoding performance for integration windows ranging from the prohibitively small to slightly longer than the duration of the stimulus (4-512 ms), and for sub-populations of 20, 40 or 130 neurons (Figure 3A-C). We fit cumulative Gaussian functions to the decoding performance, and interpolated to find an integration window that produced a level of performance well above chance (lower dashed lines), but below the saturation points of any of the functions. This corresponded to 60% correct for the 12-alternative decoder and 80% for the two-alternative decoders (upper dashed lines).

We compared the weight profiles obtained for the different population sizes (Figure 3D-F). Each data point is the average weight for a neuron across 100 resampled sub-populations. Furthermore, as we tested twelve directions, each neuron contributes twelve weights per sub-population, either simultaneously in the 12-alternative condition, or individually for each direction-pairing in the two-alternative condition. Here, for clarity, we show all the weights (dots) for the n = 40 population, and rolling averages for the larger and smaller populations. For all population sizes, the weights are distributed across preferred directions in the same way. In all cases there are relatively few extreme weights, with most weights close to zero. Notably, the weight profiles follow the same shape for the fine and coarse discrimination, unlike previous observations in primate V1 (1). Our simulations (below) suggest that this is because of the broader tuning curves in area MT.

To examine the stability of a neuron's weight in the context of different population sizes, we correlated the weights learned for a neuron in the full population against the weights it received in the 20 and 40-neuron sized subpopulations (Figure 3G-I). For all three decoder types, there is a strong positive correlation between decoding weights, but the regression lines for the subpopulations have slopes lower than 1, with scatter most likely to occur in the bottom left and top right quadrants. This means that while a neuron’s precise weight may vary, that variation is restricted so that it almost never changes sign. This is reasonable,

because a given neuron may contribute more or less to a prediction, but it is unlikely to switch from providing evidence for a stimulus condition to providing evidence against it.

These results show that, the principles that govern weighting are consistent across variations in population size when populations are matched for decision difficulty.

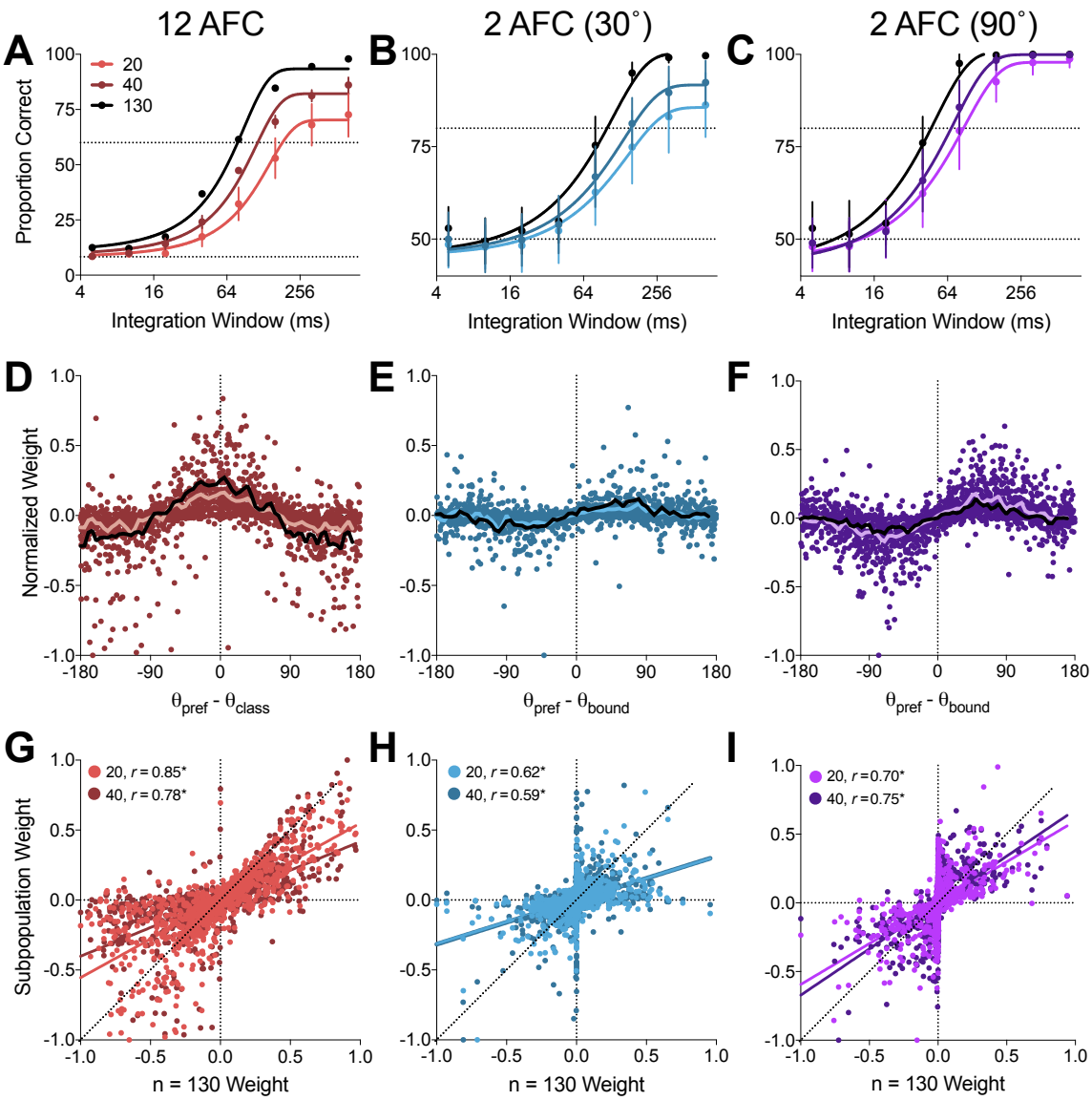


Figure 3: Role of population size in decoding performance and optimal weights. (A) Psychometric functions for a 12 AFC identification decoder for three different population sizes (20, 40, and 130 neurons). Decision difficulty was parameterized by changing the size of the integration window for spike counting. Percent correct data were fit with a cumulative Gaussian function. The upper dashed line indicates the 60% threshold that we used to choose an integration window that would match the performance across population sizes. **(B)** As in A for a coarse discrimination (90º). Here and in C, we used a cutoff of 80% correct to determine a spike counting window. **(C)** As in B, for a fine 2 AFC discrimination (30º). **(D)** Weights for each neuron depended on the separation between their preferred direction and

the stimulus class. Decoding was performed at the threshold spike counting window determined in A. **(E)** As in D, for coarse 2 AFC discrimination relative to the discrimination boundary. **(F)** As in E, for fine 2 AFC discrimination. **(G-I)** Correlation between weights for each neuron across population sizes. Weights are strongly correlated, with a tendency for weaker (closer to zero) weights in the smaller populations. Lines are regression lines for the two subpopulation sizes. Asterisks indicate $p < 0.01$.

Decoding weight Is related to the response properties of the neuron

We and others have shown that decoding neural populations requires setting weights that depend on the relationship between the neuron's preferred direction and the task. However, in populations that vary in a range of response properties (such as selectivity, firing rate, trial-to-trial variability, or spontaneous rate), the weights learned across the population are still difficult to predict based on preferred direction alone. To account for some of this variability, we explored how the weights correlated with the response properties of the neurons, specifically for conditions in which the preferred direction of the neuron predicted that its weight should be at its highest. For example, in a twelve-alternative decision, this means when its preferred direction is within 15 degrees of the presented class.

The weights have a nonlinear relationship with the preferred direction of the neuron (Figure 3D-F). To account for this, we further examined weights only when the preferred direction predicted that they should be most informative (colored weights, Figure 4A). In 100 subpopulations of 20 neurons, we measured the partial Spearman's correlation between the weights and each of selectivity, gain, Fano factor, and bandwidth. The direction selectivity (Figure 4B), and, to a lesser extent, the gain (Figure 4C) of each neuron correlates positively with the weight it is assigned across all types of decoding decisions. Fano factor (Figure 4D) correlates negatively with weight. Low variability, as measured by a low Fano factor, appears to be necessary but not sufficient: all of those neurons with large weights have low Fano factors, but many neurons with low Fano factors are not weighted strongly. Bandwidth of the direction tuning curve (Figure 4E) does not appear to correlate systematically with weight, which suggests that the correlation between selectivity and bandwidth ($r_s = 0.24$, $p < 0.01$) contributes to the strong appearance of a relationship between bandwidth and weight. These results are consistent across population sizes (Supplementary Table 1).

These results suggest that the best predictor of a neuron's decoding weight, after preferred direction, is its selectivity. Fano factor, gain, and bandwidth appear to have progressively weaker and less systematic correlations with decoding weight.

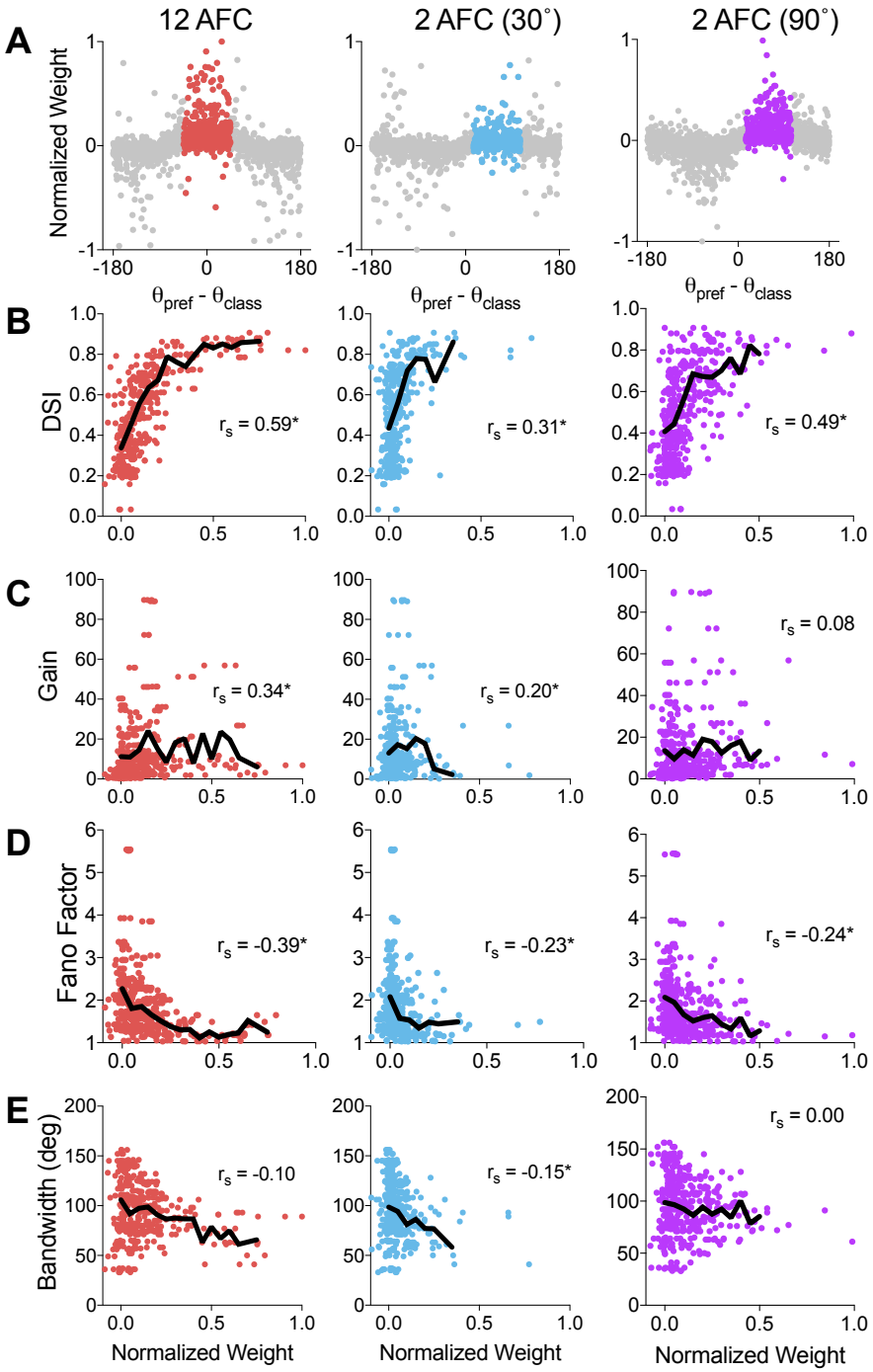


Figure 4: Direction selectivity and Fano factor predict decoding weight. (A) Left: Weights assigned to neurons in 100 subpopulations trained to perform 12-alternative

identification, plotted against the different between each unit's preferred direction and the class of the weight. We further analyzed weights for those classes within the range of +/- 45º of each unit's preferred direction (colored points). Middle, Right: As in A, for two-alternative, 90º and 30º discriminations. The weights we went on to analyze were those assigned when the preferred direction of the neuron was between 10º and 110º from the discrimination boundary. Black lines show running means. **(B)** Direction selectivity index against the weight of each neuron for an identification decision. For clarity, only weights greater than 0 are shown, but negative weights were still included in correlation analyses. Inset *r*-value is the Spearman partial correlation between direction selectivity and weight (discounting variance shared with gain, Fano factor, and bandwidth), asterisks indicate $p < 0.01$. **(C-E)** As in B, for gain, Fano factor, and bandwidth.

Weights assigned by optimal decoders are scaled by selectivity

We aim to describe general encoding principles, but we have recorded from a limited population of cells. Therefore, we generated homogeneous synthetic neural populations that allowed us to define a weighting function that describes the relationship between the decision and the preferred direction. Then, to sample a wider variety of populations and fit a model based on physiological properties, we simulated unique heterogeneous populations of neurons that conformed to the distributions of tuning properties we observed in our recordings. In the following sections we:

1) **Define generative models** that can ascribe weights to simulated neurons in homogeneous populations based on preferred direction alone.
2) **Extend the generative models** to heterogeneous populations by fitting them to subsets of simulated neurons grouped by their Fano factor and direction selectivity.
3) **Define an *a priori* decoding model** that gives a generative model for neurons based on their preferred direction, direction selectivity and Fano factor.
4) **Compare the *a priori* decoding performance** to performance achieved by models defined through machine learning in heterogeneous populations of recorded neurons.

By using two layers of models, were able to dissect which of a neuron's tuning properties are generically important for decoding accuracy, regardless of the instantaneous responses of each neuron.

Defining generative models by simulating and decoding a homogenous population

As shown in this paper and previously, for a given relationship between the preferred direction and the class or decision boundary, weights in heterogeneous populations are variable (e.g. Figure 4A). This variability precludes us from fitting a model that describes the relationship between the preferred direction of the neuron, the class or decision boundary,

and the weight the neuron is assigned in a decoder. To overcome this, we first simulated a homogeneous population.

The properties of the recorded population (Figure 5A-D) were first characterized using inverse Gaussian distributions (Fano factor and gain), an exponential function (spontaneous rates) and a Gaussian distribution (bandwidths). To simulate a homogeneous population, the tuning properties were fixed to the means of the distributions in Figure 5A-D, and preferred directions were sampled randomly from a uniform distribution.

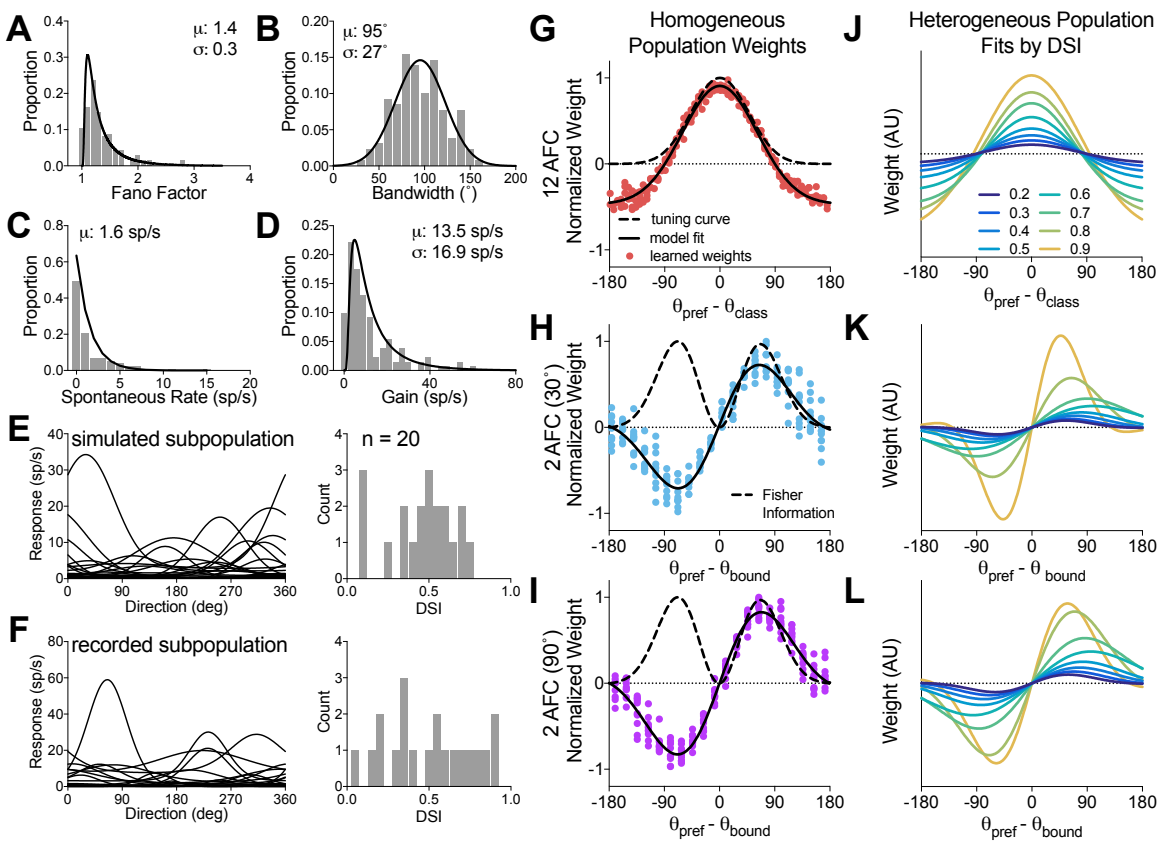


Figure 5: In simulated populations the distribution of weights is also predicted by selectivity. (A) Distribution of the Fano factor for 136 recorded neurons (bars), and the inverse Gaussian probability density function (PDF) used to generate simulated populations (line). The mean and standard deviation for the simulated populations are matched to the recorded population. **(B)** As in A for bandwidth, and a Gaussian PDF. **(C)** As in A, for spontaneous rate and an exponential PDF. **(D)** as in A for gain and an inverse Gaussian PDF. **(E)** example of tuning curves generated by the model for a population of 20 neurons, and the distribution of their measured DSIs. **(F)** A subpopulation of 20 recorded neurons and their DSIs for comparison. **(G)** Weights learned in a simulated homogeneous population with the same mean spontaneous rate, gain, and bandwidth as in the recorded population. The grey dots are the learned weights for each neuron, in each class, for each of 5 randomly generated populations. Plotted with these weights is the Gaussian model fit to the data (solid

line), and the shape of the tuning curve centered at 0 (dashed line). **(H)** As in G, for a fine two-alternative discrimination. The solid line shows the Gabor model fit to the data, and the dashed line shows the Fisher information. **(I)** As in H for a coarse two-alternative discrimination. **(J)** Model fits to weights binned by DSI and Fano factor, bins are non-overlapping 0.1 units wide, and begin at the DSI indicated in the legend. Only the Fano factor bin 1.25-1.5 is shown. See Supplementary Figure 1 for summary of R^2. **(K&L)** As in J for fine and coarse two-alternative discrimination.

To fit models that can account for the relationship between preferred direction and weight in a homogeneous population, we used the same machine learning approach to weight the simulated neurons for decoding. The weights that the decoders learned for neurons in these populations (Figure 5G-I, circles) were very similar to what would be expected given the tuning curve (dashed line, Figure 5G), or the Fisher information (Figure 5H, I). There is some variability in the weights, even in homogeneous populations, caused by Poisson spiking variability over the limited number of trials used to train each decoder. We defined two generative models based on the homogeneous simulation weights that capture the expected weight given the preferred direction of the neuron: a Gaussian model (Equation 1) for the twelve-alternative condition, and a Gabor model (Equation 2) for the two-alternative conditions (Figure 5G-I, black lines) (see Methods). Both models have two free parameters: amplitude and spread.

Extending the generative models by simulating and decoding a heterogeneous population

Using a homogeneous population allowed us to define models that account for the role of preferred direction in decoding. We expected that the shape of those models (amplitude and spread) would vary for neurons that tended to have higher weights. To fit the amplitude and spread parameters used to predict a neuron's decoding weight, we simulated 150 heterogeneous populations of 20 neurons each (e.g. Figure 5E) with the same population-wide distributions of properties (Figure 5F). Neurons were created with preferred directions sampled randomly from a uniform distribution, and other properties drawn independently from the relevant distribution. An exemplar heterogeneous population shows how the simulated neurons (Figure 5E) closely resembled those we recorded (Figure 5F) – including the wide range of observed selectivities – even though selectivity was never set explicitly.

We learned decoding weights for each neuron then partitioned the weights for the 3000 simulated neurons into evenly spaced bins based on their selectivity (8 bins) and Fano factor (9 bins), and fit the generative models to the weights in each bin. Figure 5 J-L illustrates how these models vary with selectivity, for a single Fano factor bin (1.25-1.5). This approach allows us to examine how neuronal weights are best allocated to neurons with specific selectivities and spiking variability. Supplementary Figure 1 shows the model parameters (amplitude and spread) and goodness of fit for all subsets of neurons. Not all combinations of Fano Factor and selectivity matched neurons in the simulation, so we fit a total of 51 models in the 12-alternative decision, and 38 models in the two-alternative decisions.

Defining the *a priori* model

In the Gaussian model used to describe the optimal weights for a 12-alternative decision, we found that the amplitude of the model is higher for more direction-selective neurons with lower Fano factors (Figure 5J, Supplementary Figure 1A, top). Given how well the two-parameter generative models describe neuronal weightings for bins of neurons with similar selectivity and Fano factor, we then tried to predict the amplitude and spread parameters of the weighting functions for any arbitrary neuron, given only its DSI and Fano factor. The amplitude parameter was well defined by a second-order polynomial regression surface (R^2 = 0.97), however, the relationship between the Gaussian model's spread, Fano factor and selectivity is more complicated (Supplementary Figure 1B, top). It appears as though the spread is narrowest for selective neurons with higher Fano factors, but the data are noisy and the range is small (64º-80º), so for further analysis we fixed this parameter at the median: 72º.

Likewise, in the Gabor model used to account for weights in the 30 and 90 degree discriminations, we were able to capture the relationship between model amplitude, Fano factor, and selectivity with a second-order polynomial (R^2_{90}= 0.94, R^2_{30}= 0.93), but the surface that described the model spread was too complex to fit with a reasonable number of parameters. In these cases the ranges were large, but not systematic, and many of the most

365 extreme values were associated with lower R^2. For further analyses, we used a spread with
366 the median value: 98º in the 30 degree discrimination and 107º in 90 degree discrimination.

367 These results show that larger weights are learned for more selective, less variable neurons,
368 and outline quantitatively what weight is expected for a given preferred direction, selectivity
369 and Fano factor. This quantitative description allowed us to set weights for our recorded
370 population of neurons based on these properties alone instead of using a supervised
371 learning algorithm.

372 Comparing *a priori* decoding performance

373 We have shown that weights learned to decode real and simulated populations depend
374 systematically on direction selectivity, but this does not mean that these weights directly
375 improves decoding performance – those neurons with high weights may instead facilitate
376 learning. To address the direct impact of weights on performance, we extended the
377 generative Gabor and Gaussian models described above to define weights *a priori* and
378 compared the decoding performance of several different *a priori* decoding models to
379 decoders trained at each time point. In brief, the learned weights produce the best decoding
380 for the given set of neurons, and defining weights based on the preferred direction alone
381 (PD-only) gives the worst performance. We can compare how other *a priori* models including
382 selectivity (DSI) and Fano factor (FF) perform within this range to give insight into which
383 neuronal properties affect decoding.

384 We trained decoders on the responses of 100 subpopulations of 20 neurons each, sampled
385 with replacement from the 136 neurons we recorded from marmoset MT. For each
386 subpopulation, we compared performance with the learned weights (Figure 6A) to
387 performance with weights set according to four different models. The preferred direction only
388 (PD-only) model (Figure 6B) is the model typically considered in population decoding.
389 Regardless of any other properties, neurons are weighted based on the relationship
390 between their preferred direction and the class being decoded.

To determine the potential benefit of accounting for additional response properties, we modified the weighting function for each neuron based on the fits regression surface. We added selectivity to this model by setting the amplitude of the Gaussian based on where the measured selectivity of the neuron fell on the regression surface with Fano factor = 1.5 (Figure 6C), and added Fano factor in the same manner with selectivity = 0.8 (Figure 6D). High values of Fano factor and selectivity were chosen so that they did not handicap the cell's potential influence. Precise values did not matter – we tested several variants of fixed values and the results were unchanged. Finally, we included a model that accounts for selectivity, Fano factor, and the interaction between them (Figure 6E).

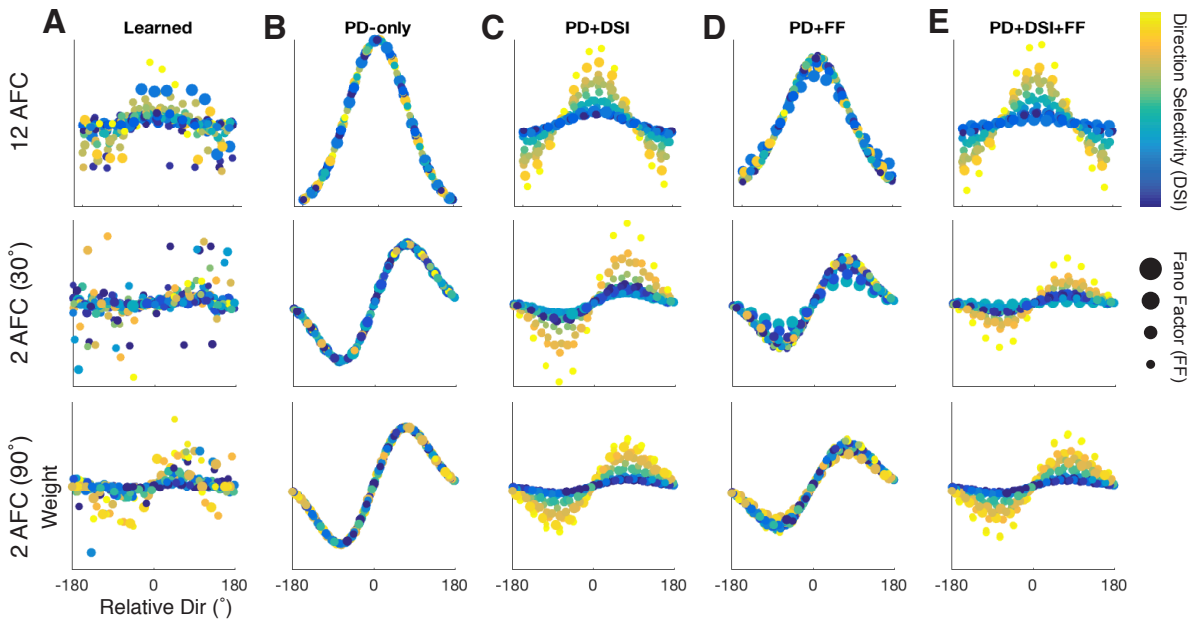


Figure 6: Weight models based on machine learning and fits to simulated weights. (A) Five models of weights used to identify stimulus direction in a 12-alternative decision. From left, the learned weights model uses those learned for each subpopulation of 20 neurons (one example population is depicted here). The preferred direction (PD-only) model weights each neuron based on its preferred direction. The PD+DSI and PD+FF models are scaled based on each property alone, and the PD+DSI+FF model scales weights based on all three properties. **(B)** As in A, for a fine (30º) two-alternative discrimination. **(C)** As in A, for a coarse (90º) discrimination. Color scale shows direction selectivity index (DSI).

We used these weighting functions to decode the stimulus direction as before, and compared the average performance of each of the *a priori* models to the model with learned weights at all time points (Figure 7A-C), and then specifically compared the distributions of performances at a single time point (400 ms) (Figure 7D-F). We found that the *a priori* model

performance was indistinguishable from the learned model performance in the coarse (90º) two-alternative discrimination for two models: PD+DSI (Figure 7F, dotted) and PD+DSI+FF (Figure 7F, solid purple) (Wilcoxon Signed Rank Test, $p > 0.01$, $n = 1200$). The Learned model significantly outperformed all other models (Wilcoxon Signed Rank Test, $p < 0.01$).

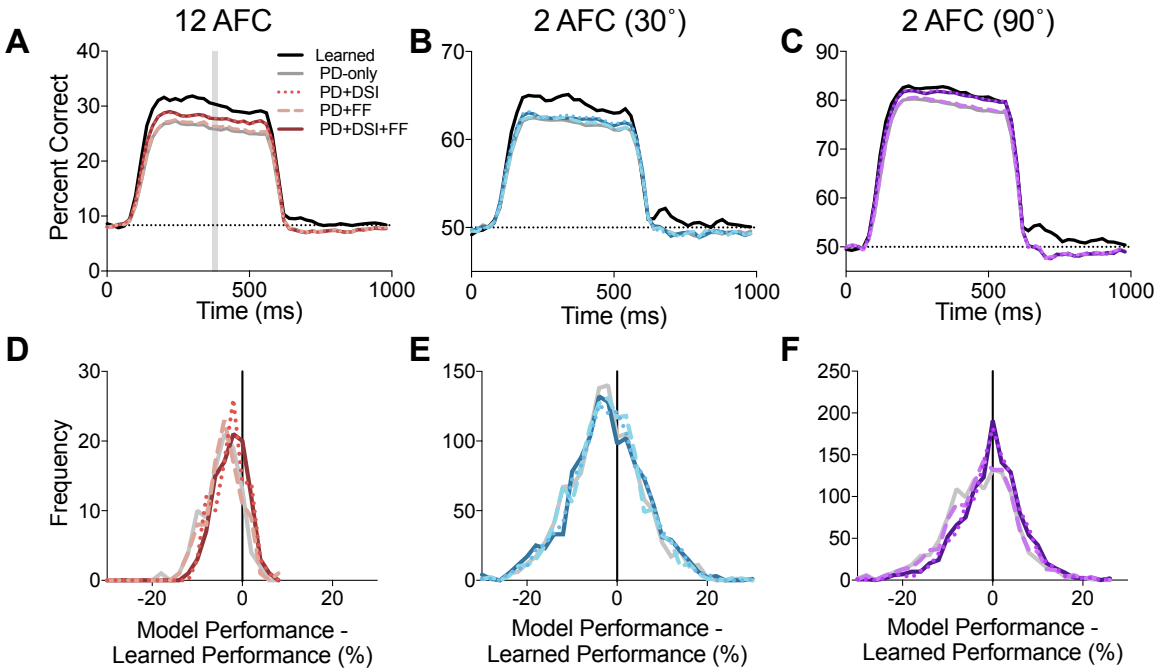


Figure 7: Weight models that account for preferred direction and selectivity account for much of learned decoding performance. (A) The performance of the models in Figure 6 over time. The learned weights outperform the modelled weights, but the models including selectivity and variability perform at least as well as the model that accounts for preferred direction alone. **(B)** As in A for a fine two-alternative discrimination. **(C)** As in A for a coarse two-alternative discrimination **(D)** Comparison between models and learned weights for each of the 100 subpopulations examined 400 ms after stimulus onset (shaded region in A). The learned performance is subtracted from the model performance so negative values are subpopulations where the model was worse than the learned weights, and positive values indicate that the model was better than the learned weights. **(E)** As in D for a fine two-alternative discrimination. Note that the high number of samples compared to D is because each subpopulation was decoded multiple times: once for each pair of directions. Thus, there are 1200 instead of 100 samples. **(F)** As in E for a coarse two-alternative discrimination.

Amongst the four *a priori* models, there was a significant main effect of model on decoding performance (Friedman test, $p < 0.01$). Including Fano factor alone (PD+FF, Figure 7, dashed) significantly improved performance above PD-only in only the coarse two-alternative discrimination (Dunn's multiple comparisons, $p > 0.01$). On the other hand, including selectivity alone (PD+DSI, Figure 7, dotted), or with variability (PD+DSI+FF, Figure 7, solid color) consistently improved decoding performance for all types of discrimination ($p <$

0.01). We found no evidence that including variability along with selectivity (PD+DSI+FF) could improve performance over selectivity alone (PD+DSI) ($p > 0.01$).

These results show that accounting for direction selectivity improves the representation of motion in a population code beyond what is attainable by accounting for preferred direction alone. Furthermore they show that while neurons with a low Fano factor and high direction selectivity are likely to be weighted strongly by machine learning techniques, incorporating the influence of Fano factor on weight does not measurably improve performance in most cases, while accounting for the influence of selectivity on weight does. This may be because selectivity has a stronger relationship with the optimal weight, or it may be because a low Fano factor is useful in adjusting weights during learning, but negligible once the optimal weightings are already known. Because different schemes perform similarly, they also suggest that the brain has a high tolerance for sub-optimal weighting in even small populations, particularly in the case of easier perceptual discriminations.

Discussion

Decoding is one way to assess the amount of information contained in the firing rates of a population of neurons. In its most basic form, decoding applies weights to neurons in a population vector scheme that depends only on each neuron's tuning preference (2,4). At its most flexible, decoding uses machine learning to derive the weights for each neuron that give the best estimates of the stimulus (1,9). This has the advantage of being completely agnostic about which neurons are most informative, and does not require knowing anything about each neuron in advance. Here, we have combined these techniques to show that neurons are not equally informative members of the population. We found that those neurons that are more selective should be weighted more heavily in order to improve the population representation of the stimulus.

Although a machine learning approach to assessing population coding is advantageous overall, it does have limitations. When we use a neuron's learned weight as a proxy for how

informative it is, we are confounding its contributions to learning and its contributions to the ultimate discrimination. Our *a priori* modeling only partially addresses this confound, because the absence of a statistically significant effect (i.e. for Fano factor) does not mean that no effect exists. Further, applying weights multiplicatively means that neurons with very low firing rates may be weighted strongly because those weights are then multiplied by values close to zero. Machine learning approaches are also limited by the need for a large number of trials to train the decoder before it may be tested on held-back trials. Techniques such as leave-one-out testing (15) help to make the most of the available trials while still performing cross-validation, but there is no substitute for larger, more varied training and testing data sets. In particular, while leave-one-out validation approaches cross-validate each test trial, they are not able to assess the variability that comes from changing the training corpus. We were able to measure the variability to an extent by resampling both the subpopulations of neurons and how trials were distributed in an 80-20 training-testing cross-validation regime, but this approach is still limited by the total number of available trials.

Knowing that with a mismatch between the weighting scheme and a population biased by adaptation we can predict observed perceptual aftereffects (Zavitz et al., 2016), it is unlikely that any weighting scheme varies on short timescales. That weights fail to generalize completely from one time to another implies that the perceptual readout is often sub-optimal, even setting aside adaptation-induced biases. However, our results suggest that there is likely some tolerance for sub-optimal weighting schemes. Using the *a priori* models we have assigned weights that were substantially different from those the models learned, and found that as long as they accounted for the preferred direction, the models could be used to discriminate and identify stimulus directions within a median performance drop of less than 4% correct.

The easiest perceptual discrimination we tested (90º two-alternative) was also the least sensitive to the precise weights applied to each neuron. Performance in this discrimination dropped by a median of 2% correct with preferred-direction-only model, and was equal to

performance for the learned weights with any model including selectivity. This is interesting, because performance was not at ceiling with respect to the percent correct, but it may have been at ceiling given the information available in 50 ms of spiking information from 20 neurons. This can happen if a proportion of trials are ‘unsolvable’, either by being entirely ambiguous, or by having firing rates that, on most other trials, represent a different stimulus.

Despite their temporal rigidity, neurons are dynamically re-weighted for different perceptual tasks in both dramatic (identification vs. discrimination), and subtle (fine vs. coarse discrimination) ways. This suggests that the relatively flexible readout scheme evident across perceptual tasks is the result of top-down influences switching between stereotyped weighting functions, or refining a weighting scheme over much longer timescales than investigated here. Similarly, perceptual learning can correct for biases in the population code, so it is likely that over longer time scales errors like those we see in adaptation-related aftereffects are ultimately corrected.

Previous studies have reported that the pattern of weights learned in fine discrimination tasks is steeper in the vicinity of the category boundary, which corresponds to a narrower spread in our model (1). However, these findings were reported for orientation discrimination with primate V1 neurons, which have a much narrower bandwidth than neurons in MT. A model that rapidly changes from negative to positive weights near the discrimination boundary can take advantage of neurons with narrow tuning, but such cells were a minority in our recorded, and therefore our simulated, populations. Still, there is some evidence that in our fine discrimination task the most selective cells had weights described by a narrower model (Figure 5K, Supplementary Figure 1B), and that cells with lower bandwidths tended to be weighted more heavily ($r = -0.15$, $p < 0.01$).

Fisher information determined for a single neuron’s tuning curve predicts decoding weights in two-alternative discrimination tasks for homogeneous populations, and this function can be extended to describe decoding weights in heterogeneous populations. The relationship between the Fisher information and the weights depends on the relationship between each

neuron and the discrimination boundary of the task. For example, when discriminating 0º vs 90º, the boundary is 45º. In an identification task there is no single discrimination boundary. The same can be said for very coarse (i.e. 180˚) discriminations: when the tuning curves don't overlap, there are many, equally valid, discrimination boundaries. In these cases it is less informative to know the Fisher information of a single neuron or the population.

We examined only how our neurons might contribute to direction perception, but it is likely that the same neurons contribute to multiple population codes. In MT alone, it is likely that the population simultaneously represents motion direction, speed (16–18), eye position (19), and disparity (20,21). How each neuron's selectivity interacts with the population to code multiple stimulus properties simultaneously is an open question.

Conclusions

In this work, we found that direction selectivity and Fano factor predicted the weight assigned to a neuron with a given preferred direction. In some ways, this is not surprising as neurons with high selectivity and low variability are also going to be the most useful to the sort of iterative, supervised learning algorithm commonly used to decode neural responses. We addressed this concern by using the generative models fit to the learned weights in simulated populations (Figure 5) to produce weights for the recorded neurons based on their preferred direction, Fano factor, and direction selectivity (Figure 6). These modeled weights underperform the learned weights but, importantly, models that include selectivity as a factor outperform models based on preferred direction alone. This suggests that while high selectivity may improve learning, it also improves a neuron's contribution to a population code.

Methods

Ethics Statement

Our electrophysiology experiments were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes. All procedures were approved by the Monash University Animal Ethics Experimentation Committee.

Electrophysiology

We performed extracellular recordings in four anesthetized, adult New World monkeys (3 male, 1 female; common marmoset, *Callithrix jacchus*). Our procedures are slightly modified from the procedure in (22), as described by (23). Briefly, anesthesia was induced with alfaxalone (Alfaxan, 8 mg/kg), and a tracheotomy and vein cannulation were performed. The animal was artificially ventilated with a gaseous mixture of nitrous oxide and oxygen (7:3). Maintenance solution with opiate anesthetic (sufentanil) and paralytic (pancuronium bromide) was infused intravenously. Atropine and phenylephrine hydrochloride eye drops were applied; then the cornea was protected, and the eyes were fitted with contact lenses to bring a monitor at 350 mm into focus. The ipsilateral eye was occluded. A craniotomy and durotomy were performed over area MT, and a 10x10, 96 channel, “Utah” array (Blackrock Microsystems) was implanted using a pneumatic insertion tool over area MT (localized using gross anatomical landmarks, verified postmortem using receptive field maps and histology). Recordings were collected at 30 kHz using a Cerebus system (Blackrock Microsystems). The raw voltage signal was high pass filtered at 750 Hz, and spikes were detected based on threshold crossings. Spikes were manually segregated offline into isolated single units using custom MATLAB software. Responses from each single unit were trial shuffled, and units were combined across animals. Non-selective units (direction selectivity index < 0.2 in all 50 ms spike bins) were excluded from further analysis. A total of 136 direction selective single units from four animals were used throughout the study.

Visual Stimulation

Stimuli were presented on a VIEWPixx 3D (1920x1080 pixels; 520x295 mm; VPixx Technologies) positioned at a viewing distance of 350 mm. The stimulus filled the display

and consisted of a sheet of white dots on a black background. The field consisted of 0.5 dots per degree square, and each dot subtended 0.3 degrees of visual angle. Dots moved coherently with no noise in 1 of 12 directions and at 8, 10, or 20 cycles per degree. Data were collapsed across stimulus speeds. The stimulus moved in one direction for 500 ms and was followed by a blank (black) screen before another, random, direction was presented. Each direction was presented 120 times.

Characterizing neurons

We characterized our recorded neurons by their preferred direction, bandwidth, Fano factor, gain, and direction selectivity. A raw tuning curve was constructed based on the mean response to each stimulus direction. Gain was calculated as difference between the maximum and minimum responses. Preferred direction was calculated as the vector sum of the tuning curve, and direction selectivity as its vector magnitude. The bandwidth was computed based on the full width at half height of a Von Mises function fit to the tuning curve. Fano factor was calculated as the spike count variance divided by the mean across all directions.

Decoding

The stimulus direction was decoded at multiple time points throughout the 500 ms duration based on spike counts measured in time windows of 4-512 ms. These spike counts were weighted, then integrated across neurons and nonlinearly transformed. In the discrimination model, the sum is compared to a criterion to determine which direction the model reports. In the identification model, each spike count is weighted multiple times – once for each of the possible stimulus directions. These weighted spike counts are integrated and transformed as before, then used to create a direction vector, which the model provides as a continuous estimate of the stimulus direction.

We used two logistic regression models to decode neuronal responses: a multinomial generalized linear model (GLM) to make identification decisions, and a binomial GLM to make discriminations. For optimization, we used the glmnet statistical software package in

Matlab (24,25). The models were trained on 80% of the stimulus trials by refining weights to optimize performance via penalized maximum likelihood. The performance we report is the decoding performance on the remaining 20% of trials. Which trials were training trials were randomly selected each time a decoder was trained. In different experiments we varied the sizes of the populations and the size of the spike integration window, but unless otherwise specified we used subpopulations of 20 neurons, and integrated spikes for 30 ms. We decoded at a temporal resolution of 10 ms.

To perform two-alternative discriminations on a data set collected for 12 stimulus directions, we selected out only the trials in each possible pairing. For example, in the coarse two-alternative discrimination, we trained models to decode 0º vs 90º, 30º vs 120º, etc.. This means that the 12-alternative decoder has 12 times as many trials to train and test on, but that 12 times more decoders were trained for the two-alternative discriminations (i.e., 100 subpopulations for each decoding task, times 12 different direction pairs). For this reason, the weights tend to be noisier in the two-alternative decoders, but their performance distributions are smoother. Note also that the 12-alternative decoder is training a weight for every class, while the two-alternative decoders train only one weight, so the ratio of parameters to training data is the same in all cases.

To create *a priori* decoding models, we used the same procedure, except instead of learning the optimal weights for each neuron and class, we applied the weights predicted by the relevant model.

In the Gaussian decoding model, for direction identification, the relevant weight is given by the difference between a neuron's preferred direction and the class of the weight (θ_{pref}-θ_{class}). The Gaussian was centered at θ_{pref}-θ_{class} = 0, and had three free parameters: gain, standard deviation, and an additive shift. After fitting these parameters to the simulated data, we used selectivity as the gain parameter, and fixed the standard deviation and shift (1 and -1 respectively), to decode our neuronal populations.

$$w(\theta) = \left(ae^{-\theta^2/2\sigma^2}\right) - a/2 \quad (1)$$

The twelve-alternative model was a Gaussian function centered at θ_{pref} - θ_{class} = 0, with free parameters for amplitude and spread. The two-alternative model was a Gabor function that crossed zero at the decision boundary (θ_{pref} - θ_{bound} = 0) between the two possible directions. The amplitude and the spread of the model were free parameters.

$$w(\theta) = ae^{-\theta^2/2\sigma^2} \sin(x(90/\sigma)) \quad (2)$$

In the Gabor decoding model, for direction discrimination, the relevant weight is given by the difference between a neuron's preferred direction and the discrimination boundary between the two classes (θ_{pref}-$\theta_{decisionBound}$). The zero-crossing of the Gabor was fixed at θ_{pref}-$\theta_{decisionBound}$ = 0, and there were two additional free parameters: the gain, and the spread of the Gabor. The frequency of the sine wave was set as a ratio of the Gaussian's standard deviation, so that one full cycle fit within the Gaussian envelope. After fitting these parameters to the simulated data, we used selectivity as the gain parameter, and fixed the standard deviation at 1 to decode our neuronal populations.

Simulation

Populations of neurons were simulated based on the distributions of gain, bandwidth, Fano factor, and spontaneous activity that we observed in our recorded populations. A Gaussian tuning curve was produced with the specified gain, bandwidth and spontaneous level, and spike counts were generated for a given trial based on the tuning curve, Fano factor, and a Poisson spike generator. These spike counts were then decoded in the same manner as our recorded populations of neurons.

Acknowledgements

This work was supported by National Health and Medical Research Council Project Grants APP1008287, APP1066588 and APP1120667, Human Frontier Science Program Career

Development Award to N.S.C.P., the ARC SRI in Bionic Vision, and the ARC Centre of Excellence for Integrative Brain Function. We thank Janssen-Cilag Pty Limited for the donation of sufentanil citrate

References

1. Graf ABA, Kohn A, Jazayeri M, Movshon JA. Decoding the activity of neuronal populations in macaque primary visual cortex. Nat Neurosci. 2011 Feb;14(2):239–45.

2. Seung HS, Sompolinsky H. Simple models for reading neuronal population codes. PNAS. 1993 Nov 15;90(22):10749–53.

3. Shamir M. Emerging principles of population coding: in search for the neural code. Current Opinion in Neurobiology. 2014 Apr;25:140–8.

4. Georgopoulos AP, Schwartz AB, Kettner RE. Neuronal population coding of movement direction. Science. 1986 Sep 26;233(4771):1416–9.

5. Salinas E, Abbott LF. Vector reconstruction from firing rates. J Comput Neurosci. 1994 Jun 1;1(1–2):89–107.

6. Ecker AS, Berens P, Tolias AS, Bethge M. The Effect of Noise Correlations in Populations of Diversely Tuned Neurons. J Neurosci. 2011 Oct 5;31(40):14272–83.

7. Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. Nat Neurosci. 2006 May;9(5):690–6.

8. Pouget A, Dayan P, Zemel RS. Inference and Computation with Population Codes. Annual Review of Neuroscience. 2003;26(1):381–410.

9. Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS. A Fast and Simple Population Code for Orientation in Primate V1. J Neurosci. 2012 Aug 1;32(31):10618–26.

10. Mendoza-Halliday D, Torres S, Martinez-Trujillo JC. Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. Nat Neurosci [Internet]. 2014 Aug 10 [cited 2014 Aug 12];advance online publication. Available from: http://www.nature.com.ezproxy.lib.monash.edu.au/neuro/journal/vaop/ncurrent/full/nn.3785.html

11. Zavitz E, Yu H-H, Rowe EG, Rosa MGP, Price NSC. Rapid Adaptation Induces Persistent Biases in Population Codes for Visual Motion. J Neurosci. 2016 Apr 20;36(16):4579–90.

12. Schwartz O, Hsu A, Dayan P. Space and time in visual context. Nat Rev Neurosci. 2007 Jul;8(7):522–35.

13. Glasser DM, Tsui JMG, Pack CC, Tadin D. Perceptual and neural consequences of rapid motion adaptation. PNAS. 2011 Nov 8;108(45):E1080–8.

14. Van Wezel RJAV, Britten KH. Motion Adaptation in Area MT. Journal of Neurophysiology. 2002 Dec 1;88(6):3469–76.

15. Chen SC, Morley JW, Solomon SG. Spatial precision of population activity in primate area MT. Journal of Neurophysiology. 2015 Aug 1;114(2):869–78.

16. Krekelberg B, Wezel RJA van, Albright TD. Adaptation in Macaque MT Reduces Perceived Speed and Improves Speed Discrimination. Journal of Neurophysiology. 2006 Jan 1;95(1):255–70.

17. Price NSC, Born RT. Adaptation to Speed in Macaque Middle Temporal and Medial Superior Temporal Areas. J Neurosci. 2013 Mar 6;33(10):4359–68.

18. Priebe NJ, Cassanello CR, Lisberger SG. The Neural Representation of Speed in Macaque Area MT/V5. J Neurosci. 2003 Jul 2;23(13):5650–61.

19. Morris AP, Bremmer F, Krekelberg B. Eye-Position Signals in the Dorsal Visual System Are Accurate and Precise on Short Timescales. Journal of Neuroscience. 2013;33(30):12395–12406.

20. Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. Journal of Neurophysiology. 1983 May 1;49(5):1148–67.

21. Uka T, DeAngelis GC. Contribution of Middle Temporal Area to Coarse Depth Discrimination: Comparison of Neuronal and Psychophysical Sensitivity. J Neurosci. 2003 Apr 15;23(8):3515–30.

22. Bourne JA, Rosa MGP. Preparation for the in vivo recording of neuronal responses in the visual cortex of anaesthetised marmosets (Callithrix jacchus). Brain Res Brain Res Protoc. 2003 Jul;11(3):168–77.

23. Yu H-H, Rosa MGP. Uniformity and diversity of response properties of neurons in the primary visual cortex: selectivity for orientation, direction of motion, and stimulus size from center to far periphery. Vis Neurosci. 2014 Jan;31(1):85–98.

24. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.

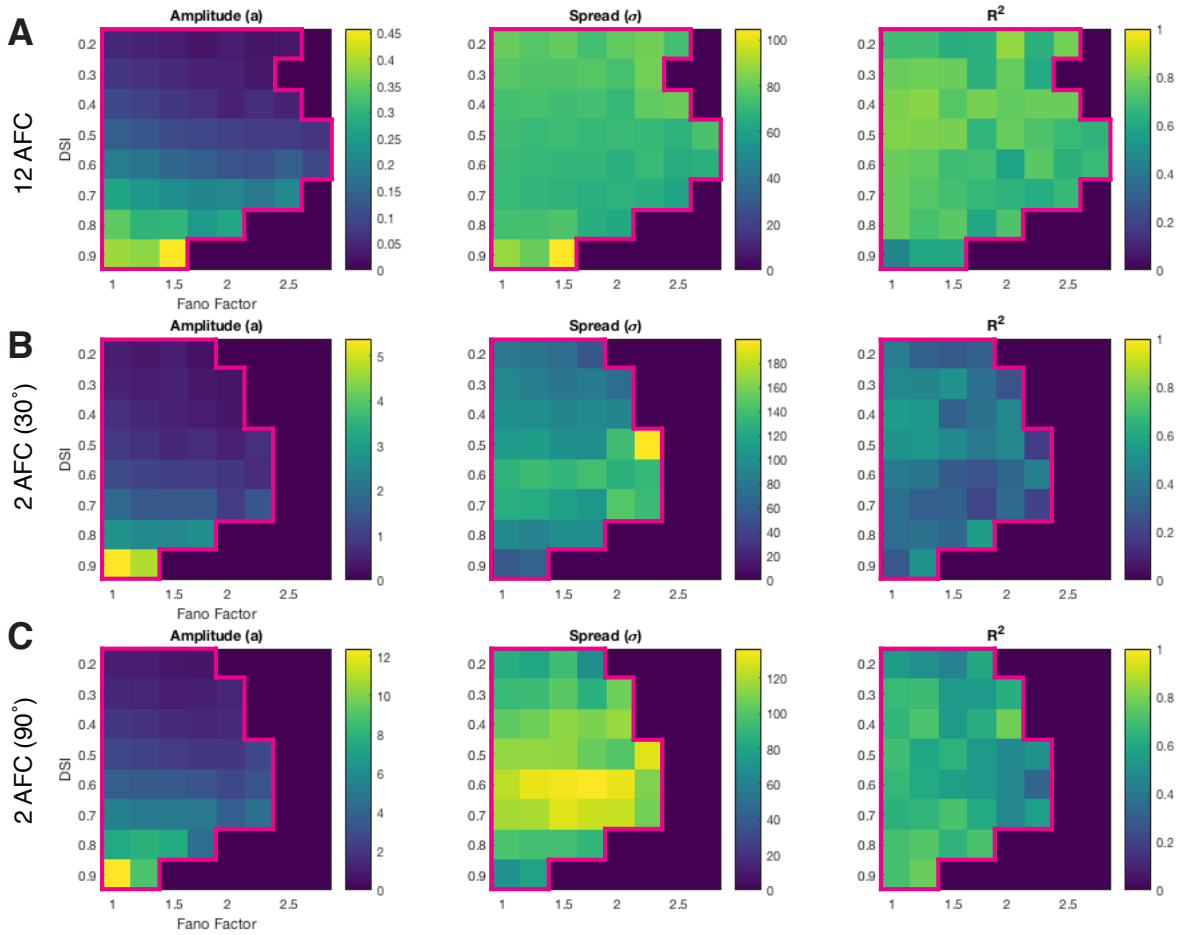
25. Qian J, Hastie T, Friedman J, Tibshirani R, Simon N. Glmnet for Matlab. http://www.stanford.edu/~hastie/glmnet_matlab/; 2013.

Supplementary Information

	12 AFC	**2 AFC 90**	**2 AFC 30**
130			
DSI	0.72	0.12	0.22
Gain	-0.10	0.37	0.12
FF	-0.24	-0.15	-0.11
BW	-0.23	-0.04	-0.14
40			
DSI	0.62	0.32	0.26
Gain	0.27	0.21	0.09

FF	-0.36	-0.21	-0.15
BW	-0.23	-0.13	-0.17
20			
DSI	0.60	0.40	0.31
Gain	0.32	0.12	0.14
FF	-0.37	-0.23	-0.21
BW	-0.10	-0.07	-0.15

Supplementary Table 1: Summary of Spearman's partial correlation coefficients for all decoders and population sizes.



Supplementary Figure 1: Selectivity and Fano factor are reflected in the amplitude of the weighting functions. (A) Left: The amplitude parameter of the Gaussian model (intensity) for neurons with each pairing of Fano factor and selectivity. There is a smooth amplitude gradient with respect to both Fano factor and DSI. Centre: The spread parameter of the Gaussian model. The relationship between spread and the parameters of interest is not as systematic as the amplitude, but there seems to be a narrower spread in conditions where Fano factor and selectivity are both moderate. Areas outside of the pink boundary indicate pairings of properties where there were fewer than thirteen weights (and so only one or no simulated neurons). Right: R^2 for the models from which amplitude and spread were

derived. **(B)** As in A for the amplitude and spread of the Gabor model in the 30º two-alternative discrimination. **(C)** As in A for the 90º two-alternative discrimination model. Fits were achieved with a five-parameter second-order surface where $f(x,y) = a + bx + cy + dxy + ey^2$. The quadratic term was required for direction selectivity (y), but not for Fano factor.