Supplementary Information

S.1 Discussion of coding implications of time-lagged coupling

We propose that a key benefit of time-lagged coupling in a neuronal population is to lengthen its coding timescale. Further, variability in coupling across cortex could allow for a diversity of functions that require distinct timescales. In AC, neurons were weakly coupled with a population coding timescale of hundreds of milliseconds. Timescales in AC may therefore be influenced significantly by external stimulus dynamics relative to the timescales of coupling between neurons. The relative absence of coupling in AC could aid representations of rapidly fluctuating stimuli and high dimensional sensory features. In contrast, in PPC, neurons were strongly coupled with a population coding timescale of approximately one second.

Previous studies have revealed that coupling in the form of noise correlations between time-averaged spike count responses can have a detrimental, information-limiting effect in sensory codes[1-3]. Global modulations, such as changes in arousal and attention, can alter shared fluctuations in a neuronal population, with benefit coming from removing noise correlations[4-6]. The relative lack of coupling in AC fits well with the view of these studies that reducing noise correlation may benefit sensory coding. In contrast, in PPC, we found a potentially beneficial effect of strong, time-lagged coupling and of the resulting high temporal consistency of information. Higher levels of coupling and consistency corresponded to accurate task performance. This effect of coupling was identified by focusing on the dynamics of population activity, codes for choice information, and cell-to-cell coupling rather than single modulatory factors.

In a strongly coupled population, the timescale may be driven more by internal dynamics than by the timing of external stimuli. This type of code could accumulate information because the coupling timescales could be long enough to combine temporally separate inputs. We tested this idea using a simple statistical model in which we modulated coupling between neurons that had temporally offset sequential activity (Supplementary Information S.8; Extended Data. Fig. 9). Time-lagged coupling allowed accumulation of incoming information over successive neurons, leading to higher instantaneous information in the population than in the absence of coupling (Extended Data Fig. 9d). This effect overcame any information-limiting effects of shared fluctuations (Supplementary Information S.8-S.12). Importantly, coupling resulted in a temporally consistent information signal, as we revealed in our experimental PPC data (Supplementary Information S.9; Extended Data Fig. 9c).

A population code with high temporal consistency could be of importance for conveying signals to areas that drive behavior. At any instant in time, a downstream network could instantaneously read out a signal that contains consistent information about the recent estimate of the appropriate choice, without needing to average signals over time to remove moment-by-moment noise. In

our model, coupling resulted in higher instantaneous information values, which could aid correct behavioral choices (Extended Data Fig. 9d). In addition, it is possible that readouts that drive important behaviors could require consistent inputs over time before triggering an action. Our model revealed that such a readout results in higher behavioral accuracy on trials with higher consistency (Supplementary Information S.10; Extended Data Fig. 9e,f).

S.2 Activity patterns of AC neurons

AC activity was heterogeneous and included cells with different response latencies following stimulus onset, cells with adaptation-related effects as the stimulus repeated, and cells with responses that did not directly relate to the sound stimulus[7-9]. When examining the activity of AC neurons during the task, we were initially surprised to find a complex temporal organization of activity across the AC population. Even among neurons responsive to the sound stimulus, activity was transient (Fig. 1d; Extended Data Fig. 1). Additionally, many of these neurons responded robustly only to the first sound stimulus and had diminished responses to subsequent stimulus repeats in the same trial (Extended Data Fig. 1g,o). A minority of sound-responsive neurons showed the opposite trend, becoming progressively more active with each subsequent stimulus repeat in the trial (Extended Data Fig. 1h,o,p). These effects were likely related to sensory adaptation to the repeating stimulus rather than to the performance of the task because individual neurons had similar response dynamics in active and passive listening to identical stimuli (Extended Data Fig. 1g-q). Related effects have been reported by others in AC: an auditory stimulus can have either suppressive or facilitating effects on the responses of single neurons to subsequent stimuli, with effects lasting for hundreds of milliseconds or even several seconds[7,10-14], and AC neurons tend to respond more robustly to rare stimuli than to repeating stimuli[8]. Our stimulus, chosen to activate AC neurons independently of their sound frequency tuning, may also have affected the temporal latencies of auditory responses. The dynamic ripple stimulus[15] was comprised of many repeating, superimposed frequencies spread across the mouse's hearing range. Enhancing or suppressive interactions between frequencies could impact the timing of a given neuron's response[10,12-14].

In the main text, we refer to stimulus information in reference to stimulus category (left vs. right sound location). However, some AC neurons were able to distinguish sound locations in finer detail than left vs. right (see examples in Fig. 1f, Fig. 2b, Extended Data Fig. 1g,h). Using population decoders, we found that AC had significant information for sound location even when decoders were trained on pairs of locations separated by only 30 degrees in azimuth. In comparison, PPC populations had little detectable sound location information (Extended Data Fig. 4 l-n). Still, only a fraction of AC neurons, about 20% of the total population, carried information about the sound's location.

Among the AC neurons were cells that encoded the mouse's upcoming or recently executed choice. This finding was evident when examining the weights attributed by the GLM to choice-

related predictors (Extended Data Fig. 2j, see "Encoding in AC and PPC, below) and when decoding choice information from activity in AC populations and single neurons (Fig. 2e,g). The presence of choice information in AC, a primary sensory region, may seem surprising. However, in animals performing auditory tasks, AC neurons with activity related to behavioral responses[16], auditory perception[17], or cognitive variables[9], rather than to presented auditory stimuli, have been reported. Furthermore, neurons with significant choice probabilities have been described throughout sensory cortex[18-20]. The source of the choice information in AC is unclear and cannot be determined from our results; AC may participate in the generation of the choice signal, or it may receive feedback related to the perceptual decision[19,20] or motor output[21].

We verified that the presence of AC choice information was not an artifact of our analysis procedures. In our analyses, we decoupled stimulus and choice by selecting equal numbers of correct and error trials. We considered that the sharp sound location tuning of AC neurons could, with an uneven distribution of correct and error trials at "easy" (lateral) and "hard" (medial) sound locations (Fig. 1c), lead us to mistakenly report choice information in AC populations. To test for this potential confound, we measured the amount of information about choice present in sound location after choice had been decoupled from stimulus category. This analysis revealed that the sound location carried less information about the choice than that carried by neuronal populations in AC, for each individual experimental session (Extended Data Fig. 4o). Hence, choice information in AC could not be accounted for by its sound location tuning.

S.3 Encoding in AC and PPC

In the main text, we focused on the performance of our single cell and population decoders to test for the presence of stimulus and choice information in AC and PPC. However, many other behavioral variables could affect AC and PPC activity as well. Our GLM incorporated all of the variables that we tracked during the task: sound stimulus location and timing, maze position, reward timing, running velocity, virtual reality view angle, and behavioral choice (Extended Data Fig. 2a). By examining the distribution of weights attributed to these predictors by the GLM fits in AC and PPC cells (Extended Data Fig. 2i), we can also gain insight into which variables are encoded by these regions. We analyzed the GLM predictors belonging to three categories: 1) sound, 2) running, and 3) position/choice. Positive weights related to all three categories of predictors were assigned in both AC and PPC neurons, yet AC had higher weights in the sound category ($p < 0.001$, rank sum test), while PPC had higher weights in the position/choice category ($p < 0.05$, rank sum test).

We also examined the distribution of weights attributed to the three encoding variable categories across single neurons in AC and PPC by comparing the weights of pairs of categories (Extended Data Fig. 2j-k). Overall, in both AC and PPC, a positive correlation existed between tuning for each pair of variables; in particular, sound and choice weights were positively correlated (AC: $r = 0.66$, $p < 0.001$; PPC: $r = 0.69$, $p < 0.001$; Extended Data Fig. 2j-k), in agreement with the

finding that there is intersection between the neurons' stimulus and choice information as described in Supplementary Information, S.4. However, we noticed that in AC the distribution of weights seemed to exhibit a finer, more complex structure that was not captured by a linear correlation coefficient. For instance, neurons with the highest sound weights had lower running weights. To gain more insight into this pattern, we clustered cells according to their weights for sound, running, and position/choice.

Clustering was done independently for AC and for PPC using a variational Bayesian Gaussian mixture model[22] (the maximum number of clusters was set to 5; using larger values did not yield qualitatively different results). A large number of neurons fell within a cluster characterized by low weights in all three categories in both AC and PPC. However, cluster membership of the remaining neurons revealed a different organization of tuning across areas. In AC, while most of the neurons had tuning to all three variable categories, two groups of cells had stronger tuning either to stimulus or to both running and position/choice. In PPC, clusters of neurons only seemed to differ in overall tuning strength. This was again consistent with our analyses of the relationship between stimulus and choice information in AC and PPC, supporting the idea that most of the stimulus information encoded by PPC was related to the computation of the correct choice (see Supplementary Information S.4).

The "gain" of the predictors resulting from the fitted encoding model (the exponential of the convolution of the predictor weight by the predictor value[23]) can give insight into which behavioral variables were most strongly related to the activity in AC and PPC neurons during different epochs of the trial. While sound-related gain was greater than 1 in both AC and PPC neurons during the stimulus presentation, the gain for sound predictors was significantly larger in AC neurons ($p < 0.001$, KS test, Extended Data Fig. 2l). In contrast, position, choice, and running predictor gains tended to be higher in PPC neurons than in AC neurons ($p < 0.001$, KS test, Extended Data Fig. 2m,n).

S.4 Choice information independent of all other measured behavioral correlates

In the main text (Figs 2,4) we presented choice information values obtained after decorrelating choice and stimulus category, whereas in Section S.5 of Supplementary Information, we focused on the intersection between stimulus category and choice information, that is on the relationship between neural tuning for stimulus and choice.

Here, to better investigate whether populations of neurons in AC and PPC carried information about choice that was independent of all other measured behavioral correlates, we performed analyses based on a conditional choice decoder inspired by the work of Park et al[23]. Briefly, similarly to all other decoding analyses, we first fit the GLM as explained in the Methods. Then, we decoded choice in each test trial using the GLM decoder as follows. First, we computed the likelihood of the trial's data under the prediction made by the GLM using all the real behavioral

predictors in that trial, including the predictor describing the animal's choice in that trial. Second, we computed the likelihood using all the real behavioral predictors in that trial, but, importantly, with the choice predictor set to the opposite value of the choice taken by the animal in that trial. From these likelihoods, using Bayes' rule, we computed the posterior probability of both possible choice values, assuming a uniform prior over choices. We then decoded the choice to be the one with the highest posterior probability. If no genuine choice information is present and all choice information is present in behavioral features, then the decoder should not perform better than chance because the behavioral predictors are kept fixed to their real value. In contrast, if there is genuine choice information beyond what can be decoded from behavioral parameters, then the decoder should perform better than chance, as the likelihood computed using the real value of the choice predictor should be higher than that computed using the flipped value. We stress a conceptual difference between this decoder (conditional choice decoder) and the one used in the main text. Both decoders attempt to measure genuine choice information, that is choice information that cannot be explained by neural tuning to other variables. The conditional choice decoder assumes knowledge of the value of all behavioral (non-choice) correlates and uses this knowledge to fully discount their possible effect on neural activity and on decoding choice from it. The decoder used in the main text, in contrast, does not assume knowledge of any correlate and instead integrates away these variables (and the information they may carry) by assuming a data-derived prior distribution (Methods).

From the choice values decoded in each trial, we quantified choice information as the mutual information[24] between the decoded choice and the choice actually taken by the animal. The cumulative choice information conditioned on all behavioral variables was significant in both AC and PPC (AC: $p < 0.05$, PPC: $p < 0.001$, one-tailed t-test on the value of the choice information at the last aligned time point; Extended Data Fig. 3g-i). These choice information values were close in magnitude to those we report in the main text. These results indicate that AC and PPC activity carried substantial amounts of genuine choice information. These results also confirm the effectiveness of integrating away the effect of these behavioral variables when using the marginalized decoders employed in the main text.

To facilitate comparisons of choice information values with previous literature, we additionally quantified choice selectivity using choice probability[25]. We expressed the values of conditional choice information as choice probability values, which we computed as the area under the receiver operating characteristic curve (auROC[26]). These data show choice probability values larger than 0.5 and have a similar trend to that of the conditional choice information (Extended Data Fig. 3j-l).

In addition, we performed another analysis to verify that choice can be decoupled from other behavioral parameters. We computed the correlation coefficients between the GLM choice predictors and the GLM lateral running speed predictors (those that are expected to be most

correlated with choice) across all datasets. We found that these running speed predictors were correlated with choice (as expected), but not very strongly over the full recordings. The correlation coefficients were bounded between -0.16 and 0.41, with most values being substantially smaller in magnitude (10^{th} percentile value: -0.09; 90^{th} percentile value: 0.17). These results confirm that choice information can be separable from other behavioral parameters, such as running speed, using the GLM approach.

S.5 Estimates of the behavioral use of sensory information in AC and PPC

In this section, we used a variety of approaches to investigate whether the sensory information carried by AC or PPC neurons was read out and used for this task.

Since the purpose of the all the analyses presented in this section was to probe the common variations in stimulus and choice signals (in other words, the intersection of these two signals), here we did not decouple stimulus and choice signals by selecting equal numbers of correct and error trials for analysis in each stimulus condition, as this decorrelation prevents the detection of a relationship between the two types of information. In this section, we instead used all available trials for all analyses.

A well-established method to investigate whether sensory information carried by a set of neurons is used for perceptual discrimination is to compare the neuron's selectivity for the sensory stimuli, the so-called neurometric function, to the animal's performance, the so-called psychometric function[18,25,27]. If a neuron's information is used for the task, then stimulus discriminations that are easy for the neuron should be easy for the animal. We computed neurometric curves that quantified decoding performance as function of the distance of the sound location from the midline. We designed a pairwise stimulus decoder that decoded the sound category from neuronal activity in trials in which the sound originated at locations symmetric with respect to the midline. For instance, we compared trials with sound locations of +15 or -15 degrees, or we compared trials with sound locations of +60 or -60 degrees, and so on. Neurometric performance was the percent correct of this pairwise stimulus discrimination. To derive comparable psychometric curves as function of the distance from the midline, we computed behavioral performance for trials with stimuli at a given angle relative to the midline. The psychometric and neurometric curves obtained in each single session are reported in Extended Data Fig. 4. Correlations between psychometric and neurometric curves were significantly larger than chance in 13 out of the 14 sessions analyzed (one-tailed permutation test, $p < 0.05$, 14-fold FDR corrected[28]). Session-averaged psychometric curves were highly correlated with the session-averaged neurometric curves both for PPC and AC (AC: $r = 0.93$, $p < 0.001$; PPC: $r = 0.99$, $p < 0.001$; Extended Data Fig. 3a,d; Extended Data Fig. 4a). The probability that 13 out of 14 sessions could have a significant neurometric to psychometric correlation at $p < 0.05$, under the null hypothesis that we recorded neuronal activity not related to the task, is $p < 10^{-15}$ (Binomial test). These results suggest that sensory information in the

neuronal activity of both PPC and AC is used for behavior and that a relationship between neuronal sensory information and behavior is apparent both on average over the entire dataset and in single sessions.

To further test the association between neuronal activity and behavior with a more statistically powerful test that (unlike the correlation between trial-averaged psychometric and neurometric functions) uses the statistical power of considering all individual trials, we used a test based on the single-trial, single-session association between a neuron's sensory information and behavior. Namely, we computed the area under the receiver operating characteristic curve (auROC) quantifying the separation between the distributions, within a single session, of all single-trial population sensory signals in left stimulus and right stimulus trials. This auROC was computed following the same procedure outlined above and used in Extended Data Fig. 3j, but by applying it to stimulus decoding instead of choice decoding. If this auROC is significantly larger than 0.5 in an individual session, it means that the neuron's sensory information reports a left turn when a left sound is presented and a right turn when the right sound is presented, which is exactly the task that we asked the mouse to perform and is what the mouse actually did as shown in the psychometric curves. We found that the auROC was significantly larger than 0.5 (one-tailed t test; $p < 0.01$, FDR corrected) in all 14 sessions, leading us to conclude that we detected a strong single-trial correlation between the neuron's sensory information and behavior in all 14 individual sessions.

In addition to these measures, we also performed an additional analysis that measures the association between sensory information carried by neuronal activity and behavioral reports in single trials. We computed a recently developed measure called intersection information (*II*). *II* quantifies how much the sensory information in neuronal activity can explain task performance[29,30]. *II* is computed as the probability that in a given trial the stimulus is decoded correctly from neuronal activity and the animal's choice reports the correct percept. This quantity is the fraction of trials in which the correct behavioral performance of the animal can be accounted for by the accuracy of sensory information carried by the considered neuron. Chance values for this quantity were obtained by shuffling choice indices at fixed stimulus and stimulus decoded values, to destroy any dependence between the stimulus decoded from neuronal activity and the animal's choice[29]. An above-chance *II* value of 1% means that the probability of correct behavior increases by 1% whenever that cell carries correct information, with respect to the baseline level of behavioral performance. The amount of *II* per cell above chance level was 0.2% ± 0.1% in AC (larger than zero with $p < 0.0001$, paired signed-rank test) and 0.7% ± 0.2% in PPC ($p < 0.0001$, paired signed-rank test; Extended Data Fig. 3b,e). The fraction of cells with *II* greater than 1% above chance was 13% and 24% in AC and PPC, respectively. These results suggest that sensory information carried by both AC and PPC in single trials influenced task performance in the same trial. The sensory information in PPC had a larger effect on task performance, suggesting a key role of PPC in the translation from sensation to action.

If the sensory information encoded by single neurons is read out to drive behavior, choice probability can be expected to be positively correlated with stimulus selectivity across cells[31]. We measured the amount of stimulus information contained within each cell's activity without discounting the correlation between stimulus and choice. We then calculated the correlation between this stimulus information and the amount of "genuine" choice information computed with the conditional decoder described in the Supplementary Information section S.4 (this information value is the closest one conceptually to choice probability). This procedure mirrors the traditional comparisons between stimulus sensitivity and choice probability in individual cells[32]. We found a significant correlation in both AC and PPC (Spearman $r = 0.33$, $p < 0.001$ in AC; Spearman $r = 0.43$, $p < 0.001$ in PPC; Extended Data Fig. 3c,f), supporting the idea that stimulus information carried by neurons in AC and PPC could be used to perform the task.

S.6 Redundancy of choice information over time – comparison of instantaneous and cumulative decoders

Our results in Fig. 4 show that choice information is temporally consistent within a trial in PPC. This finding suggests that information should be redundant across nearby time points. We tested this idea by comparing the decoders for choice based on instantaneous and cumulative (across time) population activity (Extended Data Fig. 7). The decoder based on instantaneous population activity used a time window of a single imaging frame (~60 ms). The decoder based on cumulative population activity used all time points in the trial prior to and including the time point of interest. For example, for the third time point in a trial, the cumulative decoder was based on three imaging frames (~180 ms). In AC, instantaneous information was low, but the cumulative information increased over time (Extended Data Fig. 7a). This result indicates that the information across time points was largely distinct and thus could be accumulated. In contrast, in PPC the instantaneous and cumulative information increased together during the trial before a choice was reported (Extended Data Fig. 7b). This finding implies that the information from one time point to the next was correlated and thus in part redundant. These results support the findings that information was consistent (Fig. 4), and thus redundant, across time in PPC but not in AC. Furthermore, after the choice was reported, this relationship between instantaneous and cumulative information in PPC did not hold (Extended Data Fig. 7c), indicating that information was no longer redundant across time. This result is in agreement with our findings of decreased coupling in PPC after the choice was reported (Fig. 4g) and that the consistency of choice information decreased after the report as well (Fig. 4h).

S.7 Scaling of the amount of information and its consistency timescale with population size

In the main text, we used a fixed population size of 37 neurons to examine the amount of information about stimulus and choice carried by population activity and the timescales of population coding. This population size was the smallest number of simultaneously recorded cells across the experimental sessions analyzed. To better understand how the diverse timescales

of population coding in each area arise from the interplay between subgroups of cells, we studied how the amount of information and the timescale of its consistency scaled with population size.

We selected random sub-populations of cells of varying sizes for each experimental session. We computed total information (maximum of the cumulative information) and consistency (τ_2 slow time constant as in Fig. 4) using only cells included in these random sub-populations. For each experimental session and sub-population size, we repeated this procedure 48 times with different random sub-populations and averaged the results at a fixed sub-population size.

We found that information about stimulus and choice increased with the size of the sub-populations (Extended Data Fig. 8a). The scaling of information with population size was well fitted by the random-overlap model[33,34], which assumes that each cell covers a random portion of the overall information space needed to encode perfectly the external variable of interest (choice or stimulus), with the average redundancy of the information carried by two different neurons being equal to the random overlap of the information they provide. Here we use the random overlap model because it is known to fit the scaling of information with population size when averaging over subpopulations at fixed size. Thus, it can be used as a descriptive statistical model for accurate extrapolations. However, we do not take it as a model to interpret the coding mechanisms producing the population scaling because the explanatory power of the model arises from the averaging over all subpopulations, rather than from the behaviour of individual subpopulations[33]. This analytical model predicted that the number of cells needed to decode 0.99 bits of information (99% of the total information about choice) was 865 in AC (95% confidence interval [823, 904]; model fitted with matlab's 'lsqcurvefit' function; c.i. computed using matlab's 'nlparci' function on the outcome of the model fit) and 410 in PPC (95% confidence interval [378, 449]). The number of cells needed to decode 0.99 bits of information about stimulus category in AC was 240 (95% confidence interval [231, 250]). Overall, the analysis of scaling with population size predicted that a few hundred randomly selected cells would be sufficient for near-perfect encoding of stimulus and choice.

Using the same approach of averaging over subpopulations of varying size, we next computed the scaling of choice information consistency with the number of neurons analysed (Extended Data Fig. 8b). We focused on populations of at least 10 cells to avoid possible confounds due to the temporal sparseness of the activity of recorded cells. The consistency measure is based on the temporal dynamics of the probabilistic output of the choice decoder, which will not fluctuate in absence of activity across the population analyzed. In PPC, the consistency timescale of choice information in real, simultaneously recorded data increased by ~13% when the population size increased from 10 to 20 cells ($p < 0.01$; z-test) and then saturated at ~0.9 s for population sizes larger than 20. The mild scaling of this time constant with population size suggests that our estimates of population coding time constants (reported in Fig. 4) matched at least the order of magnitude of those that would be measured in a larger population – for instance, a population of

the size that would be sufficient for near-perfect decoding as described above. Interestingly, when shuffling was performed to disrupt cross-cell correlations (as in Fig. 4), the consistency timescale of choice information remained constant with population size at a value of ~0.7 s, which was 20% smaller than that of the real PPC population. The lack of increase with population size for shuffled data suggests that functional coupling among cells was necessary to achieve long timescales in PPC and that these long PPC timescales could not be achieved with larger uncoupled populations. In addition, these timescales could not be correctly estimated by pooling data from non-simultaneously recorded cells. In AC, the consistency of both real and shuffled data did not increase with population size and had a value of about 0.5 s, a much shorter value than in PPC. This finding suggests that the difference in timescales between PPC and AC were conservative and may be even more pronounced in larger populations. It also confirms the result that AC timescales were not enhanced by coupling.

Given that the estimation of consistency may be noisier when considering population signals with lower instantaneous information, these results also suggested that our AC vs. PPC comparison in Fig. 4 was not confounded by a possibly higher signal-to-noise ratio of the consistency estimation in PPC, as a population of size 20 in PPC had similar choice consistency to a population of size 35 (Extended Data Fig. 8b), but only 57% of the choice information (Extended Data Fig. 8a), and thus a smaller signal-to-noise ratio.

S.8 Statistical generative model of sequential population activity in AC or PPC

We built a simple statistical model of coupled sequential neuronal population activity that qualitatively captured the most salient phenomena observed in the experimental data. We used this model to examine the relationship between functional coupling between cells and the timescale of temporal consistency of instantaneous choice information in the population code. We also used this model to generate synthetic datasets to test ideas about the possible behavioral readouts of population codes for choice.

The model was inspired from our real data and was formulated as follows. We assumed that the sequential population activity in a brain region (either PPC or AC) was expressed as a temporal sequence of cell activities at subsequent times. The firing of each cell at the appropriate time in the sequence depended on two parameters: the value of the choice parameter c^i encoded by the population activity (right or left choice, with some cells preferring left choice and some preferring right choice) and the firing of the previous cells in the sequence (Extended Data Fig. 9a). To capture the time-dependent nature of the cell-to-cell coupling observed in our data (Fig. 3f), we assumed that each cell was coupled to the previous two in the sequence. The parameters regulating the strength of this cell-to-cell sequential coupling were called γ_1 and γ_2 (Extended Data Fig. 9a). Model cells had a binary output. We modeled the activation probability of the n-th cell in the sequence on trial i, when population activity encoded either the choice preferred by the cell (c^i = 0) or the opposite choice (c^i =1), with a generalized linear mixed model with the

choice and the activity of the two previous cells in the sequence as predictors (Extended Data Fig. 9a):

$$log\left[\frac{p(r_n^i=1|c^i,r_{n-1}^i,r_{n-2}^i)}{p(r_n^i=0|c^i,r_{n-1}^i,r_{n-2}^i)}\right] = \beta_0 + u_n + \beta c^i + \gamma_1 r_{n-1}^i + \gamma_2 r_{n-2}^i \quad \text{(S1)}$$

where r_n^i represents the activation of cell n in trial i (activation is binarized into 0 and 1 representing active and non-active cells, respectively), $\beta_0 + u_n$ is related to the average activation probability of cell n (β_0 is the average across all cells and u_n its residual specific to cell n), and γ_1 and γ_2 represent the coupling to the previous two cells in the sequence.

We estimated model parameter values from real data. For each experimental session, we considered the 10 most choice-informative cells. We divided these cells into left-choice-preferring and right-choice-preferring cells, and we identified the time in the trial of their peak instantaneous choice information (see Fig 2f-h for a distribution of such times in AC and PPC). For each trial, we averaged the activity of these cells in the 5 time frames centered on this peak and binarized the averaged value (setting the binarized value to 1 if the cell was active in at least one of the 5 time bins, and 0 otherwise). For each experimental session and for each trial, this yielded a binary sequence of activation of cells preferentially encoding left choice and a sequence of cells preferentially encoding right choice. Each element of this binary sequence represented the activation of a cell over the ~200 ms for which it was informative about the choice (see Fig. 2k for the half-width of the average information peak across single cells). To simplify the model and its fit to real data, we assumed the unknown parameters β_0, β, γ_1 and γ_2 to be the same across all cells (across all sessions, including both left-preferring and right-preferring cells), but we allowed u_n to take on different values for each cell, under the hypothesis that across all cells it has a Gaussian distribution with mean zero and standard deviation σ (also an unknown parameter). We separately fit the data in AC and PPC sessions. In practice, this simplified model assumes that all sequentially active cells in the same brain region have the same choice tuning strength for their preferred choice and the same statistical dependency on the activity of the previous cells in the sequence. However, this model still allows each cell to have a different baseline activity rate by including the term u_n. We note that the choice encoded by population activity cannot be directly estimated from our experimental data because we have access only to the animal's choice, and neuronal activity related to choice may be sometimes different from that reported by the animal, for example if the choice encoded by neuronal activity is not properly read out (see Supplementary Information S.9 for discussion of possible models of behavioral readout mechanisms). It was therefore not possible to estimate the tuning of each neuron to the encoded choice directly. We instead measured the beta parameters by quantifying the strength of the encoded choice from real data under the assumption that the encoded choice corresponded to the actual one. We verified in later simulations that all results presented here and in the next section were qualitatively similar for a wide range of beta values,

including values higher or lower than those reported below. The best-fit parameter estimates from fitting the model of Eq. S1 to AC or PPC sessions using the 'glmm' R package[35] were (mean ± sem): AC: β_0 = -2.88 ± 0.10, β = 1.200 ± 0.080, γ_1= 0.47 ± 0.10, γ_2= 0.00 ± 0.10, σ = 0.329 ± 0.083; PPC: β_0 = -3.41 ± 0.61, β = 2.182 ± 0.088, γ_1 = 0.557 ± 0.083, γ_2 = 0.130 ± 0.078, σ = 0.54 ± 0.14.

While, as explained above, the beta parameters cannot be directly interpreted as measures of the tuning to the encoded choice, the gamma parameters can be interpreted as measures of the coupling between the peak activity of cells next to each other in a choice-dependent activity sequence. It was therefore reassuring to find that the coupling parameters γ_i had larger best-fit values in PPC than AC, in agreement with our other findings that report larger time-lagged coupling in PPC than in AC (Fig. 3-4, Extended Data Fig. 5).

We used this model to generate many simulated trials of synthetic population activity. For instance, we simulated a set of trials when the right choice was encoded (Extended Data Fig. 9b). Dividing the cells into left-preferring cells $\{r_n\}$ and right-preferring cells $\{\tilde{r}_n\}$, the model generated stochastic sequences of population activity, with the firing rate of right-preferring cells that was higher than that of left-preferring cells (Extended Data Fig. 9a).

From the data generated above, we constructed the optimal instantaneous decoder of the choice signal encoded by the population activity. This instantaneous decoder of choice was constructed by taking as decoded choice the one preferred by the cell firing the most at that given time in the sequence. In other words, we decoded right choice if at that given time frame the right-preferring cell was more active than the left-preferring cell. At instants in which the activity of the left- and right-preferring cells was the same, we decoded the choice randomly. A second property of the simulated data, visible in the rasters of simulated activity, was that in trials in which right choice was encoded, the coupling between cells generated time sequences with right-preferring cells consistently firing more than the left-preferring cells over several consecutive time steps (green vs purple shading in Extended Data Fig. 9b).

S.9 Using the statistical generative model to analyze the relationship between coupling and information timescale

To understand the effect of coupling between cells on population coding timescales, we simulated these neuronal population responses by systematically varying the coupling parameters γ_1 and γ_2 between zero and their data-derived values. Specifically, we multiplied the best-fit gamma parameters by a scaling factor between 0 and 1 (0 corresponds to no coupling and 1 corresponds to a coupling equal to that of the best fit parameters to real data). This theoretical manipulation of the coupling strength is a useful feature of the modeling study because it was not feasible to modulate coupling in this manner experimentally with the approaches used in this

paper. Such an experimental manipulation would require disrupting the dynamical mechanisms from which cross-cell correlations emerge while performing the experiment.

To quantify the effect of varying the coupling parameter in the model, we took the output of the instantaneous choice decoder for population activity, and we repeated the same analyses that we performed on real data. In particular, we computed choice information and the consistency of the choice signal, following analogous procedures to those we used for the experimental data. Furthermore, we assessed how these quantities differed between simultaneous data generated by the model and shuffled data where the activity of simulated individual cells had been scrambled across trials.

The simulated data exhibited several features that support our findings and conclusions about the real data (Extended Data Fig. 9). In particular, we found that, in this model, coupling was beneficial to instantaneous population coding in several ways. First, coupling lengthened the timescales of consistency of the population choice signal (Extended Data Fig. 9c). When trials were shuffled to disrupt coupling, there was a greater reduction in information timescale when the coupling had a high value, just as we observed in our real data when comparing PPC and AC (Extended Data Fig. 9c; note how shuffled data had zero consistency, matching data generated by a model with no coupling). Second, coupling increased the value of instantaneous information because it allowed choice information to propagate from cell to cell along the sequence. Approximately 28% of the instantaneous information was due to this mechanism in the simulated PPC data, compared with ~18% in AC (Extended Data Fig. 9d). The increase of instantaneous information through accumulation was a straightforward and necessary consequence of the existence of coupling among sequentially informative cells. Importantly, we note that this effect cannot be disrupted simply by trial shuffling because this effect is embedded in the activity of single cells (Extended Data Fig. 9d; note how the shuffled simulated data had the same instantaneous information as the simultaneous simulated data generated straight from the GLM coupled model of Eq S1). In real data, an experimental manipulation to break coupling would be necessary to disrupt this advantage of coupling in increasing the value of instantaneous information. Thus, our analyses that disrupted coupling by shuffling trials revealed a conservative lower bound of the real benefit of coupling on timescales and information.

S.10 A model of the behavioral readout of the choice signal encoded by sequential PPC population activity

We wished to understand why having a temporally consistent signal could be beneficial for behavior, as was suggested by our experimental data (Fig. 4). We reasoned through how the long timescale of choice information consistency that we found in PPC could affect behavior depending on the readout mechanism.

We considered a model scenario where instantaneous information is read out, that is the behavioral readout operates on instantaneous population activity. For our simplified model, "instantaneous" means operating on the short timescale associated with the activation of an individual cell in the sequence. In this case, the long timescale of consistency of instantaneous information could be beneficial to drive or execute behavior. This benefit may occur, for example, if correct execution of the choice is easier when the instantaneous information reports a consistent choice signal during a long stretch of the trial. In this case, the readout would receive consistent, rather than conflicting, information at each time step. In such a readout scenario, a time-lagged coupling between sequentially active cells creates long timescales that can be exploited to improve behavioral performance. In this case, we expect that trials in which PPC population activity is more strongly coupled will have a longer timescale, leading to a higher probability of a correct choice. Hypothesizing this readout mechanism led naturally to the prediction that if we split data into behaviorally correct and incorrect trials, we would find stronger coupling and longer timescales when examining correct trials.

In practice, to implement this readout mechanism in our simulations, we assumed that PPC transforms stimulus information into an appropriate choice (transforming a left sound signal into the encoding of a left turn choice and a right sound signal into a right turn choice). Given that we found in real data that the features of PPC population activity correlated with behavioral performance (Extended Data Fig. 3), in the model, we further assumed that PPC activity was read out to guide behavior. Thus, in our model, we included two sources of behavioral errors. The first source was information that was incorrectly encoded in PPC (encoding error), meaning that the choice signal encoded in PPC did not correspond to the one appropriate for obtaining a reward. We called α the fraction of trials in which PPC model activity encoded choice appropriately. This source of error was not important to the analysis of the readout system that we perform below, in that using different values of α to run the model did not change qualitatively the results in Extended Data Fig. 9e-f. For our simulations, we chose $\alpha = 0.8$, a value that was consistent with the behavioral performance in the experimental data. The second source of behavioral error in our model, and the one most important to the operation of behavioral readout, was the readout error, that is the error made by the readout system that uses the choice signal encoded by population activity to execute behavior. In more detail, we hypothesize that the readout system, to form an effective motor drive, needs choice to be encoded consistently by PPC for a certain amount of time – say L time steps in our simulation. Thus, our readout model, which translates encoded choice into the motor output, reads out and executes the choice encoded by neuronal activity only if the choice signal is expressed consistently for at least L time bins. Otherwise, it implements a random choice. The model parameter L is thus the minimum time length of consistent choice coding needed by PPC to effectively drive behavior. The results shown in Extended Data Fig. 9 correspond to $L = 7$, though, as for α, the choice of a particular value of L did not change the results qualitatively.

We produced a series of simulated correct and error trials from this readout model, and analyzed the consistency of choice information separately in each set of trials. The simulated correct trials had a more consistent choice signal than error trials (Extended Data Fig. 9e-f; $p < 0.001$, permutation test on the consistency value at lag = 2 steps), as expected from the intuition presented above and as found in real PPC data (Fig. 4j). Furthermore, we also computed a coupling index for the synthetic data by re-fitting the simulated data with a simple GLM and comparing the cross-validated predictive power of a coupled version of the GLM against an uncoupled one, similarly to what was done on the experimental data. The coupling index computed for all trials was 0.013 ± 0.003 (estimated from 10^5 trials, each composed by 50 time steps). Importantly, the coupling index computed for behaviorally correct trials was significantly larger than the coupling index in error trials (correct trials: 0.019 ± 0.001; error trials: 0.009 ± 0.004; $p < 0.01$, z-test), again in agreement with what we observed in experimental data (Fig. 4i). It is important to note that in this simulation the coupling weights γ were the same across all simulated trials. Shorter consistency and lower coupling values in error trials were obtained because the readout selected as incorrect trials those that, because of fluctuations of neuronal activity, had shorter sequences of consistent activity. When analyzed in comparison to correct trials, error trials thus had shorter consistency and weaker statistical coupling. This result suggests that one interpretation of our experimental findings, in which consistency and coupling differed between correct and error trials, is that the nature of the readout is such that trials with longer lasting choice information consistency produced correct behavior, rather than implying a change in the strength of functional interactions between cells from one trial to the next.

We also considered a second scenario in which the readout of population activity was based on cumulative PPC activity from the beginning of the trial until the time of the turn (cumulative choice information). In this case, the readout would accumulate population signals, unlike in the first scenario in which the readout only listened to instantaneous population activity. In this scenario, the effect of coupling would be expected to be detrimental to behavioral performance because with cumulative information, coupling would be expected to limit the information levels. This prediction can be understood intuitively in our model. The optimal cumulative choice decoder in our model would be one that compares the difference in the number of time bins, over the course of the trial, that have higher firing in right- vs. left-preferring cells. Having correlations between different time bins would decrease the signal-to-noise ratio of this difference. This intuition is broadly compatible with the finding that in real PPC data, shuffling trials increased cumulative PPC information by 14% (Extended Data Fig. 9, SI section S.10). We implemented this second cumulative readout model on the simulated data. We found that when dividing data into correct and error trials and analyzing them separately, correct trials had weaker coupling ($p < 0.001$, z-test) and shorter consistency ($p < 0.001$, permutation test) than error trials. The prediction of this cumulative readout model seems contrary to our empirical finding that coupling and consistency were higher in correct trials in PPC (Fig. 4). We thus ruled out this

readout scenario from our interpretation of the empirical results and of the benefit of long timescale choice coding.

In conclusion, our investigation of different readout models illustrates qualitatively one possible readout mechanism that could give rise to the difference in coupling index and consistency between correct and error trials that we observed in real data. In particular, these results can be accounted for by a readout that operates on instantaneous information but needs information to be consistent for some time period in order to implement the action associated with the choice.

S.11 Experimental results: Variations in cumulative and instantaneous choice or stimulus information when shuffling trials to disrupt coupling

To evaluate the possible effect of within trial correlations and coupling on choice and stimulus information, we computed population choice and stimulus information, using the methods explained in the main text, either using real simultaneously recorded population responses or pseudo-population responses obtained by shuffling trials to disrupt the effects of coupling. We computed the effect of trial shuffling both on the instantaneous and cumulative information. Instantaneous information was computed at a relevant time point based on the decoded variable (for stimulus information, 1 second after the first stimulus time onset; for choice, the moment of the turn), and cumulative information was the total value computed over at the last aligned time point (where data were aligned to the first sound onset for stimulus information and to the turn for choice information). We found that, although shuffling leads to a moderate (~14%) but not significant ($p = 0.11$, paired signed rank test) increase in the amount of PPC cumulative choice information accumulated over the whole trial length, shuffling to disrupt correlations did not change the amount of instantaneous choice information available from the population activity in a single imaging frame ($p = 1$, paired signed rank test; difference in percentage between shuffled and non-shuffled information < 1%; Extended Data Fig. 9i,j). Thus, in our PPC data, shuffling to disrupt correlations decreased the temporal consistency of the signal without changing the instantaneous decoder performance.

We note that we decided, for clarity and consistency with all other decoding analyses (Fig. 2d,e, Fig. 4, and Extended Data Fig. 3g,j, 4l,m,o, and 7a,b), to perform all analyses plotted in Extended Data Fig. 9g-j by averaging over populations of 37 cells, the minimal amount of cells present across all selected experimental sessions. However, as the effect of cross-cell correlations on information content may increase with population size, we additionally performed the same analyses using the maximal population size in each session (that ranged from 37 to 70). In this case, the increase with shuffling of cumulative information was slightly larger, approximately 20%, and significant ($p < 0.05$, paired signed rank test). Instantaneous information was still not affected by shuffling ($p=1$, signed rank test; average instantaneous information was approximately 4% lower in shuffled data).

There is an intuitive reason why shuffling to disrupt correlations did not increase instantaneous information in our experimental results. It is established that positive noise correlations reduce information when combined with positive signal correlations, whereas positive noise correlations combined with zero signal correlation do not affect information[36-38]. Here, cells carried choice-related activity in a temporal sequence, with only one or a few cells carrying a choice signal in a given time frame. Within a single imaging frame, neurons will typically share some noise correlation but are unlikely to have a signal correlation. Hence, it is natural to expect no negative effect of coupling on instantaneous population information. On the other hand, using cumulative activity may result in some signal correlations across cells and thus a negative, information-limiting effect of noise correlations, leading to a moderate increase of information when shuffling.

S.12 Experimental and Modeling Conclusions: The impacts of coupling on information and its timescale

In real data, we found that one of the main effects of coupling in a dynamic population code was to lengthen the timescale of information. This effect was apparent because coupling and information consistency co-varied, including across regions (Fig. 3d, 4c-f), epochs of the task (Fig. 4g-h), and trial types (Fig. 4i-j). Also, disrupting coupling by shuffling trials to create pseudo-populations resulted in shorter timescales of information consistency (Fig. 4f,h,j). In the model, we were able to modulate the level of coupling in simulated PPC populations to demonstrate that coupling was directly related to the consistency of information in a dynamic population code (Extended Data Fig. 9c).

Coupling also had effects on choice information levels in the population, as it allowed for the accumulation of information across time steps. As a result, populations with higher coupling accumulated information more effectively, resulting in higher levels of instantaneous information. This result was shown directly in our model (Extended Data. Fig. 9d), and was compatible with the experimental finding of higher choice information in PPC, an area with stronger coupling, than in AC (Extended Data. Fig. 9i). Importantly, this specific beneficial effect of coupling cannot be discounted only with analysis techniques, such as disrupting coupling by shuffling of trials to create pseudo-populations. The reason is because these effects become embedded into the activity patterns of single neurons as a trial transpires in real time and thus cannot be removed with post hoc analyses. Such effects would need to be removed with experimental manipulations. When we disrupted coupling by shuffling trials to create pseudo-populations using our experimental data, there was little effect on instantaneous information levels and only a modest effect on cumulative information levels (see above and Extended Data Fig. 9g-j). All these points and effects were made clear when we compared, in our simulated data, the effects of manipulating the level of coupling in the population vs. shuffling trials to disrupt coupling as we did with our experimental data (Fig. 4f,g,j). While the shuffle removed the effects of coupling on the consistency of information across time (Extended Data Fig. 9c), it

did not remove the increased level of instantaneous information in the population that resulted from higher coupling (Extended Data Fig. 9d).

Together, our experimental and modeling results indicate that coupling in a dynamic population code has the key effects of lengthening information timescales and increasing information levels through the accumulation of information over time.

Supplementary Information References

1 Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat Rev Neurosci* **7**, 358-366 (2006).

2 Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience* **12**, 1594-1600 (2009).

3 Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat Neurosci* **14**, 811-819 (2011).

4 Goris, R. L., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nat Neurosci* **17**, 858-865 (2014).

5 Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., Smirnakis, S. M., Bethge, M. & Tolias, A. S. State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235-248 (2014).

6 Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *Elife* **4**, e08998 (2015).

7 Bartlett, E. L. & Wang, X. Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol* **94**, 83-104 (2005).

8 Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nature Neuroscience* 2003).

9 Rodgers, C. C. & DeWeese, M. R. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. *Neuron* **82**, 1157-1170 (2014).

10 Asari, H. & Zador, A. M. Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *J Neurophysiol* **102**, 2638-2656 (2009).

11 Klampfl, S., David, S. V., Yin, P., Shamma, S. A. & Maass, W. A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *J Neurophysiol* **108**, 1366-1380 (2012).

12 Kilgard, M. P. & Merzenich, M. M. Order-sensitive plasticity in adult primary auditory cortex. *Proc Natl Acad Sci U S A* **99**, 3205-3209 (2002).

13 Brosch, M. & Schreiner, C. E. Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* **77**, 923-943 (1997).

14 Brosch, M., Schulz, A. & Scheich, H. Processing of sound sequences in macaque auditory cortex: response enhancement. *J Neurophysiol* **82**, 1542-1559 (1999).

15 Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z. & Shamma, S. A. Dynamics of precise spike timing in primary auditory cortex. *J Neurosci* **24**, 1159-1172 (2004).

16 Yin, P., Mishkin, M., Sutter, M. & Fritz, J. B. Early stages of melody processing: stimulus-sequence and task-dependent neuronal activity in monkey auditory cortical fields A1 and R. *J Neurophysiol* **100**, 3009-3029 (2008).

17 Bizley, J. K., Walker, K. M., Nodal, F. R., King, A. J. & Schnupp, J. W. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Curr Biol* **23**, 620-625 (2013).
18 Hernandez, A., Nacher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vazquez, Y., Camarillo, L. & Romo, R. Decoding a perceptual decision process across cortex. *Neuron* **66**, 300-314 (2010).
19 Nienborg, H. & Cumming, B. G. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* **459**, 89-92 (2009).
20 Yang, H., Kwon, S. E., Severson, K. S. & O'Connor, D. H. Origins of choice-related activity in mouse somatosensory cortex. *Nat Neurosci* **19**, 127-134 (2016).
21 Schneider, D. M., Nelson, A. & Mooney, R. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* **513**, 189-194 (2014).
22 Blei, D. M. & Jordan, M. I. Variational inference for Dirichlet process mixtures. *Bayesian Analysis* **1.1**, 121-143 (2006).
23 Park, I. M., Meister, M. L., Huk, A. C. & Pillow, J. W. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat Neurosci* **17**, 1395-1403 (2014).
24 Quian Quiroga, R. & Panzeri, S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci* **10**, 173-185 (2009).
25 Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of ...* 1992).
26 Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics*. (Wiley, 1966).
27 Cohen, M. R. & Newsome, W. T. Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J. Neurosci.* **29**, 6635-6648 (2009).
28 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* **57**, 289-300 (1995).
29 Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E. & Fellin, T. Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior. *Neuron* **93**, 491-507 (2017).
30 Zuo, Y., Safaai, H., Notaro, G., Mazzoni, A., Panzeri, S. & Diamond, M. E. Complementary contributions of spike timing and spike rate to perceptual decisions in rat S1 and S2 cortex. *Curr Biol* **25**, 357-363 (2015).
31 Haefner, R. M., Gerwinn, S., Macke, J. H. & Bethge, M. Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nat Neurosci* **16**, 235-242 (2013).
32 Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S. & Movshon, J. A. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience* **13**, 87 (1996).
33 Ince, R. A. A., Panzeri, S. & Kayser, C. Neural codes formed by small and temporally precise populations in auditory cortex. *J. Neurosci.* **33**, 18277-18287 (2013).
34 Gawne, T. J. & Richmond, B. J. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* **13**, 2758-2771 (1993).
35 Knudson, C. Generalized Linear Mixed Models via Monte Carlo Likelihood Approximation. R package version 1.1.1. https://CRAN.R-project.org/package=glmm. 2016).

36 Oram, M. W., Foldiak, P., Perrett, D. I. & Sengpiel, F. The 'Ideal Homunculus': decoding neural population signals. *Trends Neurosci* **21**, 259-265 (1998).
37 Abbott, L. F. & Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput* **11**, 91-101 (1999).
38 Panzeri, S., Schultz, S. R., Treves, A. & Rolls, E. T. Correlations and the encoding of information in the nervous system. *Proc Biol Sci* **266**, 1001-1012 (1999).