M2 COGMASTER (ENS, UNIVERSITÉ PARIS–DESCARTES, EHESS)

Major in mathematics and modeling

Internship report

Candidate : **Laureline LOGIACO**

SPIKE–TIME METRICS ANALYSIS OF ANTERIOR CINGULATE CORTEX ACTIVITY IN AN EXPLORATION/EXPLOITATION TASK

Advisors :

Angelo	ARLEO	(CNRS / Université Pierre et Marie Curie)
Emmanuel	PROCYK	(Inserm / Université Claude Bernard)

SPIKE–TIME METRICS ANALYSIS OF ANTERIOR CINGULATE CORTEX ACTIVITY IN AN EXPLORATION/EXPLOITATION TASK

ABSTRACT

Anterior Cingulate Cortex (ACC) single units firing rate modulations correlate with the behavioral policy of monkeys engaged in an exploration/exploitation task. The possible importance of spike timing is a central issue that remains to be investigated. A well–designed tool allowing to address it is Victor and Purpura spike train metrics, which quantifies how dissimilar two spike trains are, as a function of the importance of spike timing. When analyzed with this method, the data show that adequate spike timing sensitivity increases the discrimination between the strategy switch moment and exploitation moments by single units, single trial activity. Further, temporal sensitivity also improved the correlation between single cell, single trial activity at the moment of the behavioral switch, and response times after the switch. Finally, we used the multi–units extension of the Victor and Pupura metrics on pairs of cells and found that the response of the cell whose activity appears more related to the switch is not denatured by the firing of the least switch related cell. Thus, these results suggest that a 'downstream' neural network which would decode ACC single units activity to produce an adapted motor output may be sensitive to spike times.

Aknowledgements

Angelo Arleo, for advising the entire work.

Emmanuel Procyk, for providing the data, and for advising at various stages of the analysis.

Luca Leonardo Bologna, for teaching me how to use the cluster and for general advice with matlab.

Jérémie Pinoteau, for discussions about spike train metrics and statistics, and help with matlab.

Romain Brasselet, for discussions about spike time metrics.

Denis Sheynikhovitch and Mehdi Khamassi, for discussions about reinforcement learning and dopamine.

Louis–Emmanuel Martinet and Jean–Baptiste Passot, for helping with computer problems in general.

Eric Logiaco, for statistical discussions.

Ed Smith, for making me discover spike time metrics.

Contents

1 Introduction and rationale of this work

In experimental neurophysiology, a classical approach consists in establishing correlations between an experimental contingency (e.g. presence of a stimulus), and the neural response, often quantified as a spike count or a spike rate of a single unit or of multiple neurons. During the last 20 years, some researchers have been trying to extend this procedure, to see how and how much the neural responses could be informative about the external world. Indeed, to produce an adapted motor response, the brain should access signals from sensory receptors which can be mapped on the ensemble of stimuli (encoding). The subsequent stages of neural processing must be able to *decode* these signals, which means that they should react differentially to the different activities produced in different contexts.

The methodology initially consisted in reconstructing a visual stimulus from the neural responses [6]. A related approach consists in predicting, based on the neural activity, which stimulus, among a known set, was presented to the animal [16]. In both cases, information theory (and, in particular, Shannon mutual information [36]) can be used to quantify how good the prediction is. This *decoding* process (in reference to machine–based signal processing) relies on the assumption about the ability of a *neural decoder* (a putative neural network) to discriminate between some features of the neural activity which are informative about a stimulus. For instance, in some situations, the precise timing of spikes has been shown to be informative [16]. Moreover, there is evidence that a neural decoder could be sensitive to this timing (but see [21]). Notably, an animal can discriminate two types of electrical stimulations only differing by timing; moreover, the discrimination ability of mice has been shown to be incompatible with a spike count code [26]. However, different codes might be used in different brain areas and/or contexts [22]. In addition to the temporal precision of the code, another issue is how the responses of many neurons combine to produce a perception or a motor output. More precisely, the response of two neurons can provide redundant, cumulative or deleterious information [4, 32].

Spike train metrics (e.g. [39, 4]) is one of the tools designed to test more precisely these questions. It consists in computing a *distance* between two spike trains which is parameterized relatively to the importance of the timing of spikes, and to the importance of the identity of the neurons which fired. This distance measures how different two spike trains are. This approach has several advantages: it does not necessitate to establish arbitrary boundaries for counting spikes into bins, its decoding properties can be mapped on a simplified neural decoder (and thus, it respects at least some of the biological constraints on neural decoding), and it is less sensitive to a bias in information computing.

In previous studies, the so–called *neural coding* approach has been applied in both the sensory or motor parts of the nervous system (visual areas [39]; haptic receptors [33], auditory system [23], gustatory system [8]; olfactory system [24]; motor area [5]), or, more rarely, to perceptual decision making [22, 30]. However, the same questions are certainly relevant to the neural processes underlying action planning, and behavioral policy management (for instance, strategy switching). The temporal accuracy issue is particularly important because of the neural plasticity and learning mechanisms that these tasks necessitate [2]. The difficulty to find a temporal reference from which one could compare the timing of spikes, as well as the difficulty to define different behavioral situations that the neurons might discriminate, might explain why this issue has not been addressed yet, to our knowledge. However, these difficulties might be overcome in the studies addressing the question of strategy monitoring, when the animal has to flexibly adapt between *exploring* new possibilities and repeating a learned action sequence (i.e. *exploiting*), once that a reward has been received, as in Quilodran et al. (2008) [31]. In their study, Quilodran et al. recorded activity of single units and small (up to 5/6 units) clusters in the Anterior Cingulate Cortex (ACC) of awake non-anesthetized monkeys. This area receives

inputs from the basal ganglia [40], which are themselves responsive to salient events as reward delivery [14]. More precisely, neural activities from ventral tegmental area and substantia nigra have been shown to correlate with the difference between expected and received reward (reward prediction error [35]), and with the reward uncertainty [9], which are variables that are arguably important for action learning and planning. The ACC is also connected to the lateral prefrontal cortex, which in particular seems involved in memory maintenance of contextual information necessary to choose an appropriate action plan [28]. Finally, ACC projects onto motor areas, which corroborates its putative role in action monitoring [28].

Numerous studies in macaque monkeys have reported ACC activity linked to the mean reward expectation associated with optimal behavior [3], to reward prediction error or expectancy [15, 37] associated or not with a particular motor sequence [12, 29], and more recently with unsigned reward prediction error (or surprise) [11]. The activity seems to be task dependent in the details [11], but a role in behavioral strategy management (exploration of new possibilities versus exploitation of learned associations) seems to be a constant. Accordingly, lesion studies tend to show impacts on the learning of action value [17] or on non automatic, cognitively demanding behavioral adaptation (rat study, [10]). In humans, lesions lead to an inability to repress automatic actions triggered by external stimuli (as grasping a door knob), and deficits in production of self–initiated actions (as spontaneous speech), which is again in line with the idea that ACC is involved in weighing different actions based on the experience of the animal and the current contingency [28].

Previous work established in this task the relationship between averaged single unit activity and behavioral policy [31]. In another task and another animal, population trial by trial activity [19] has been used to decode different task epochs, some of which corresponding to exploitation *vs.* exploration. Building on this, we hypothesized that ACC activity in single units and in pairs of units can encode the shift between exploration and exploitation, and that different neural codes (differently accounting for spike timing resolution and neural identity) might have different information content. Finally, we hypothesized that an informative neural code that may additionally be exploited by a neural network and thus be causally related to behavior would also better correlate with the behavioral response times of the animal. To test these hypotheses, we used the Victor and Purpura metrics [39], as well as the multi–units extension of this metrics [4]. As it is the oldest one, several papers have demonstrated its ability to segregate between spike train groups in many different experimental situations, and simulations have revealed that its efficiency is often better than, or equivalent to, the efficiency of other metrics (e.g. [25]).

Our analyses show that the activity of a subset of ACC single units discriminate well between two moments that only differ in terms of the behavioral policy adopted by the animal (exploration *vs.* exploitation). They also suggest that the discrimination is more efficient when the temporal structure of ACC spike trains is properly accounted by the metrics. Moreover, the proposed metrics analysis allows *single-cell-single-trial* activity at the moment of the behavioral shift (i.e. from exploration to exploitation) to predict the response latency of the animal at the moment of the following action (6 seconds later). We show that the correlation between single-cell-single-trial activity and behavioral response latency is better captured by the temporal structure of spike trains. Finally, our preliminary results on multi–unit metrical analysis suggest that this correlation is not impaired by mutual interference between two ACC neurons recorded simultaneously.

2 Methods

2.1 Neurophysiology and behavior

Details about the task and the recordings are described elsewhere [31]. Briefly, the task consists in blocks of trials (problems) in which monkeys (Macacca mulatta) need to find by trial-and-error which target among a set of four is rewarded. During the first period of a block (exploration), monkeys search for the correct target in successive trials. After discovery of the goal target they can repeat the correct response for several trials. Indeed data show that monkeys then switch behaviorally to an exploitative state (see Fig. 1, page 9). They will be rewarded at least 3 additional times for touching the same target. In ≈ 10 % of the remaining trials, they will be rewarded 7 additional times. A signal (flashing targets) then indicates that a new problem starts, i.e. a new target will be rewarded. In 90 % of the trials, the following target is different from the preceding one. In 50% of the trials, the reward is big, and it is small in other trials.

The data analyzed here came from monkey M of Quilodran et al. (2008) [31], and consist of 145 ACC single units recorded in clusters of 1 to 5 neurons. Please note that in Quilodran et al. (2008) [31] the first–reward trial is referred to as 'C01 trial', whereas the following rewards are called 'Corr trials'. It is of importance that monkeys have been trained for months or years before the recordings, and thus they perfectly know the task. Consequently, their search strategy is optimal and they are making little if any errors in exploitation. They almost always predict the rewarded target when they have made three errors, and they almost always begin a new problem by choosing a different target from the previously rewarded one.

2.2 Choice of temporal reference

We wanted to study neural activity potentially related to the behavioral switch, which is likely to occur after the monkey received the first reward. The reward time was chosen as a temporal reference for all neural analyses because, as it relies on a simple mechanical push, it is more precise than the touch time recorded on the tactile screen. Additionally, the pump produces a sound which seems relevant for the monkeys and provides a precise "expected reward time" for the animal (unpublished observation by Emmanuel Procyk).

2.3 Spike train distances algorithms

2.3.1 Single unit Victor and Purpura distance

We used the single unit Victor and Purpura algorithm to compute distances between spike train pairs [39]. The distance is taken as the minimal cost to transform the first spike train into the second one. Such a transformation consists in using sequentially one of the three following steps:

- adding a spike, for a cost of 1
- deleting a spike, for a cost of 1
- changing the time of a spike by an amount dt, for a cost $q \cdot dt$, where q is a free parameter that allows the importance of spike timing in information encoding to be continuously varied.

When $q = 0$, there is no cost for changing the timing; consequently, the distance becomes an absolute difference of spike count between the two spike trains. As q increases, changing the timing of spikes is more and more

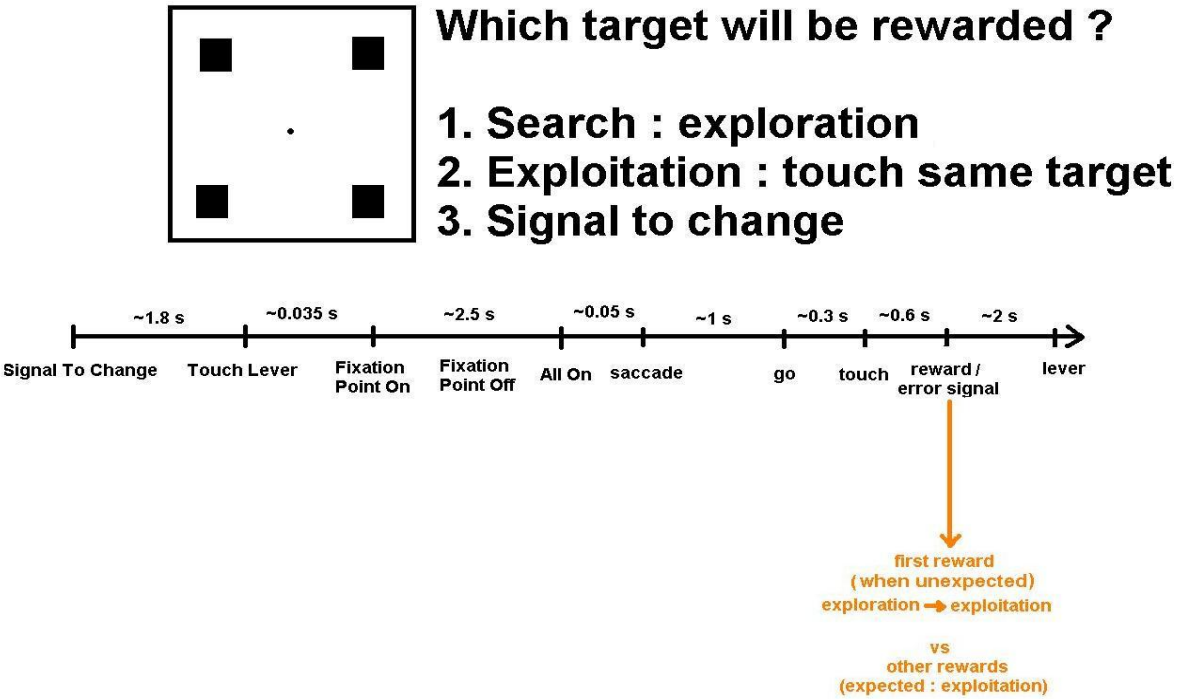

Figure 1: *Structure of the task: Each problem begins by the touch of the lever on the screen. The monkey is then required to fixate a fixation point. The lightening of the targets indicates to the monkey it can choose a target and saccade toward it. The animal then receives a go signal allowing it to touch the chosen target. If this was the good target, it receives a reward and will resume the same actions for generally three trials more. Else, it will have to choose a different target on the next trial in order to discover the rewarding target. After the monkey has received the maximum number of rewards, the signal to change informs it that a new problem (i e probably a new rewarded target direction) starts. The moment at which the analysis focuses (reward time) is indicated.*

costly, and to have a small distance, a pair of spike trains must have spikes that occur close to one another. Spikes may be moved to be matched if they are separated by at most $2/q$ s; if they are further away, it is less costly to delete one spike and then to reintroduce a new spike at the good time.

2.3.2 Multi–units Aronov / Victor and Purpura distance

The distance approach has also been adapted to the comparison of the responses of groups of units between different categories [4]. Two different putative coding parameters are varied: the temporal precision, and the importance of knowing the identity of the neuron which fired a spike. For example, if two neurons receive the same input and fire with uncorrelated noise, then it is better to simply pool their responses to retrieve the signal. On the contrary, if two neurons encode two signals that are deleterious to each other, then it is important to distinguish between them to retrieve a maximum of information. Another case when neural identity could be important is when the activity of neurons is correlated when they respond to the same external input (so termed "noise correlations", in which the deviation of the activity from the mean is correlated between neurons) and

when, additionally, more information might be gained by taking these correlations into account [20]. More detailed discussion about this issue is given in the Discussion (Sec. 4.2, page 37).

The multi–unit metrics will compare the responses of n cells in two different trials, by building two 'multi–unit spike trains' in which each spike has a label corresponding to the identity of the neuron which fired. The distance between them is the minimal cost to transform one of them into the other, by using the following steps:

- adding a spike, with a cost of 1
- removing a spike, with a cost of 1
- changing the time of a spike by an amount dt, with a cost $q \cdot dt$
- changing the identity of the neuron which fired, with a cost k

If $k = 0$, then the neuron identity does not matter at all; if $k = 2$, the responses are never switched between neurons, because removing the spike from neuron 1 in sequence 1 and adding a spike from another neuron at the time we want is less costly (cost 1+1=2). In general, two spikes from two different neurons may be matched if $2 > q\ dt + k \Leftrightarrow dt < \frac{2-k}{q}$.

2.3.3 Algorithms, parameters and computation

The calculations were run on a cluster of 320 nodes (Consorzio interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca – CASPUR), running MATLAB R2007a and MATLAB R2007b.

We used codes freely available at the website of Jonathan Victor (http://www–users.med.cornell.edu/ jdvicto/metricdf.html). For the single units analyses, the c/MEX code of Daniel Reich was used. For the multi–units algorithm, the vectorized MATLAB code by Thomas Kreuz was used, to which a piece of code handling the cases when two spike trains are empty have been added.

For both single and multi–unit analysis, we tested temporal costs $q \in [0, 5, 10, 15, 20, 25, 30, 35, 40, 60, 80]$. For the multi–units analysis, we also considered the identity costs $k \in [0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2]$.

2.4 Formal definitions of the measures used to study encoding

2.4.1 Measure of the discrimination power thanks to the spike time metrics

Distances are computed for each spike train pair into the dataset. A neural decoder could retrieve to which category x a spike train belongs if its 'global distance' to the spike trains of this category x is smaller than its 'global distance' to the spike trains of other categories. This corresponds to a gross physiological approximation, in which a downstream decoding neuron with a given time constant would fire when its input spike train has a low distance relative to a learned 'stereotypical' spike train produced in that situation. If temporal structure matters (temporal cost $q > 0$), the decoder would fire when its synaptic entry (the spike train) would match a learned (more or less precise) temporal pattern (thanks to 'coincidence' detection with other inputs for instance). However, if spike count is more optimal (temporal cost $q = 0$), the 'decoder neuron' would fire when a sufficient number of presynaptic spikes are produced. The firing (or silence) of this postsynaptic neuron is informative about whether the presynaptic spike train was produced in the 'category x' context.

The classical approach [39, 4, 33] was to define the global distance from a spike train to a category of spike trains (i.e. all spike train emitted when a certain stimulus was present, or when the first reward was given) as follows:

$$D(spike\ train\ s,\ category\ C) = \left(\frac{\sum\limits_{spike\ trains\ s_C \in C,\ s_C \neq s} D(s, s_C)^z}{N_{s_C}} \right)^{\frac{1}{z}} \tag{1}$$

where N_{s_C} relates to the number of spike trains (if appropriate, different from s) in category C, and $z < 0$; classically $z = -2$. This negative exponent allows to tackle the issue of the outliers with big distances, by giving more importance to small distances in the group. Hence, the classification is biased toward those categories which contains spike trains that are very close to s, paying less attention to the presence of spike trains that are very different from s. The spike train s is classified in the category for which $D(spike\ train\ s,\ category\ C)$ is minimal; if N_C categories have an identical minimal distance $D(spike\ train\ s,\ category\ C)$, $\frac{1}{N_C}$ spike train is attributed to each category.

In our data set, this methodology seemed partly inappropriate for the spike count. In effect, if two spike trains have exactly the same number of spikes, then their distance is nil, and for each category C in which such a same spike count exists in one trial, $D(spike\ train\ s,\ category\ C)$ will be nil. However, intuitively, if category 1 produces most of the time spike trains with N1 spikes, and category 2 produces most of the time spike trains with N2 spikes but contains one spike train with N1 spikes, intuitively, a spike train with N1 spikes should be classified in category 1, not for 0.5 in category 1 and 0.5 in category 2. This is particularly problematic for $N1 = 0$, because at times cells do not answer at all. For codes different from spike count, nil distances are much rarer, and this is less an issue.

To avoid what seems to be an unfair disadvantage of spike count while addressing the problem of big distances outliers, we also computed the distance of one spike train to a category as follows:

$$D(spike\ train\ s,\ category\ C) = \underset{spike\ trains\ s_C \in C,\ s_C \neq s}{median} D(s, s_C) \tag{2}$$

In the following, we will refer to the first method as the "quadratic classification", and to the second method as the "median classification".

Once each spike train has been classified, one can build a "confusion matrix" in which the entry on line i and column j is given by:

$$N_{i,j} = number\ of\ spike\ trains\ coming\ from\ category\ i\ that\ are\ classified\ in\ category\ j \tag{3}$$

The ability of the neural responses to discriminate between the two categories can then be assessed as follows:

- By computing the mutual information between the true categories and the reconstructed categories, as:

$$I(T,R) = \sum_{i=1}^{N_C} \sum_{j=1}^{N_C} \frac{N_{i,j}}{N_{tot}} \ln \left(\frac{\frac{N_{i,j}}{N_{tot}}}{\frac{\sum_{k=1}^{N_C} N_{i,k}}{N_{tot}} \frac{\sum_{l=1}^{N_C} N_{l,j}}{N_{tot}}} \right) \tag{4}$$

  where N_C is the number of categories, and $N_{tot} = \sum_{i=1}^{N_C} \sum_{j=1}^{N_C} N_{i,j}$ is the total number of trials. I(T,R) is zero in the limit of big samples when the classification is random, regardless of the difference in the number of

trials between categories (see derivation in the Appendix, Sec. A.1.1, page 41). However, the maximum value that I(T,R) can take does depend on the imbalance between the number of trials in each category (see the Appendix, Sec. A.1.2, page 42). Therefore, to allow a fair comparison between different cells with different number of trials available, we computed $\bar{I} = \frac{I(T,R)}{I_{max}(T,R)}$ as:

$$\bar{I} = \frac{I(T,R)}{\sum_{i=1}^{N_C} -\frac{\sum_{j=1}^{N_C} N_{i,j}}{N_{tot}} \ln\left(\frac{\sum_{j=1}^{N_C} N_{i,j}}{N_{tot}}\right)} \tag{5}$$

Where $I_{max}(T,R)$ would be the value of the mutual information if the cells correctly classified all the available trials.

Because of the limited sampling, we cannot estimate accurately the different entries of the confusion matrix. Therefore, the information value computed in that way is upwardly biased [27]. This is a real problem even when comparing between different costs q, because the bias also depends on the probability distribution of the different entries of the matrix, which is cost dependent. However, it should be noted that in our particular case, the bias should be small, because we do not have a big number of time bins with low response probabilities as in the direct method. To address this issue (as in [33]), the bias was empirically estimated as the mean value of $\bar{I}$ computed from an ensemble of 1000 sets (single units analysis) or 100 sets (multi–units analysis) in which spike trains identities were randomly permuted (randperm function of MATLAB), mixing up the spike trains between categories while keeping the same number of trials by category. This mean "chance information" was subtracted from the value obtained in the true data. Intuitively, this is an estimate of the bias because it gives the value of the information one gets because of limited sampling, when no real clustering of the responses occurs. This is a fair estimate of the real bias whenever the true probability distribution is such that none of the entries of the confusion matrix is empty [27], which is a fairly reasonable assumption (meaning that none of the situation is perfectly discriminated by the responses; see the Appendix Sec. A.2.2, page 43 for a more detailed explanation). Moreover, in the cases when it happens, it leads to an overestimation of the bias which would be on the order of $\frac{1}{N_{trials}}$, with $median(N_{trials}) > 92$ in our case, and less than 3 % of the cells or couples of cells had less than 20 trials. Therefore, the possible overestimation is itself reduced to a few percent of the total possible information, which is fairly less than the percentage of information for the highly discriminating cells, which the analysis mainly focuses on.

When subtracting the estimated bias, if slightly negative information values were obtained, the value of 0 was assigned to the given data point. For the very significant cells which will be selected for further analyses, the information bias was typically less than a few percents of the uncorrected information.

- By computing the percentage of correct classification.

  To avoid this measure to be completely dominated by the most numerous category, it is taken as:

$$\%\ correct = \frac{\sum_{N_C} \frac{N_{correctly\ classified(C)}}{N_{tot}(C)}}{N_C} = \frac{\sum_{i=1}^{N_C} \frac{N_{i,i}}{\sum_{j=1}^{N_C} N_{i,j}}}{N_C} \tag{6}$$

	Temporal cost q=0	Temporal cost q=5
Median method, information	49	57
Median method, % correct	52	59
Quadratic method, information	30	39
Quadratic method, % correct	34	47

Table 1: *Number of cells found significant (over 145) at an analysis window length of 600 ms, for two example costs.*

Computation of this measure on shuffled data have revealed its ability to cluster tightly around $\frac{1}{N_C}$, while still being large in highly informative cells, on the contrary of a weighted average of the percentage of correct. However, because it weights similarly all categories regardless of their effective, whereas information weights each category by its sample size, the two indices can in principle give different results. It should also be noted that when requiring a high percentage of correct, one imposes that the classification is allowed by smaller intra–categories distances when compared to between–categories ones. This is in line with the very approximate physiological interpretation of the classification. In contrast, the information can be quite high if spike trains from category 1 are very often classified in category 2, whereas spike trains in category 1 are very often misclassified in category 2. Importantly, in our data, the two measures were mostly consistent, indicating that this assumption was verified.

For each measure, many increasing analysis windows and many costs were tested against permuted data at a 5% level. Therefore, for any cell or couple of cells, the fact that one of the test was significant could occur with a probability much higher than 5%. Moreover, if we assume that one cell was significant by chance at one cost and one window, it is also likely that it will also "appear" significant at close costs and windows (and indeed, this effect was observed in surrogate data). It was therefore difficult to adjust the p–value in a very rigorous way; moreover, lowering a lot this p–value would make it more difficult to correctly evaluate with the permutation method. To select the significant cells, it was therefore required that many tests were significant on several increasing time–windows, as detailed in the Results, Sec. 3. Finally, the significance of the encoding in the population of cells studied was assessed by looking at the number of cells that were significant at a fixed analysis window (0.6 s length, which is compatible with the timing of reward–related activity described in [31]), and a fixed cost (test at 5 % level), and by comparing it to the confidence interval of the expected number of cells at the risk of 5 % (see Table 1, page 13). For example, at cost $q = 5$ for single units, at least 39 cells were found significant.

The expected number of significant cells by chance at one cost (with the approximation that cells are independent, which is untrue but probably reasonable given that cells were often not recorded simultaneously) is $0.05 \cdot 145 = 7.5$, and the 5% confidence limit is $7.5 + 2 \cdot 145 \cdot 0.05 \cdot 0.95 = 21$ cells. Therefore, we find more cells than expected by chance, which globally validates our method in this particular brain area and task.

2.4.2 Measure of information gain with couples of units as compared to single units

To measure if, and to which extent adding some neurons increase the information compared to looking at single units, we used a measure G (for gain) defined as:

$$G(couple_j) = I_{j\ joint}(q_j^*, k_j^*)\ -\ \max_{single\ units\ i}(\ I_i(q_i^*)) \tag{7}$$

Where the costs are chosen to maximize each information independently, i.e.:

$$(q_j^*, k_j^*) = \underset{(q,k)}{argmax}(I_{j\ joint}(q,k)) \tag{8}$$

$$q_i^* = \underset{q}{argmax}(I_i(q)) \tag{9}$$

Therefore, this measure is positive if it is possible to extract more information with the multi–units measure as compared to the single units one.

2.5 Statistical tests

All tests used were non parametric and two sided. They included the rank sum test (equivalent to a Mann-Whitney U test) for testing differences in the median, the sign test to test if a median is different from 0, and the Friedman anova to test for a differential impact between non independent factors (for instance, the temporal and identity costs). Additionally, the function tmcomptest of the MATLAB file exchange was used to compare proportions.

In addition to the tests available in MATLAB, a custom permutation test was built to assess the significance of the difference between two groups for the correlation of two variables. The couples (variable 1 trial 1, variable 2 trial 2) were randomly permuted between the two groups, while keeping the possible imbalance of effective between groups. For each permutation, the correlation between variable 1 and 2 was computed for each "shuffled group", and the absolute value of the difference of correlation between groups was computed. The difference in correlation in true data was considered significant if it was superior to the 95th percentile of the permuted data (the number of permutations was usually 10000 and could be reduced to 1000 if more precision appeared useless).

3 Results

3.1 Correlates of the behavioral switch are detectable in ACC single unit activity

3.1.1 ACC single units activity discriminate better between first reward vs. subsequent rewards of a problem when the temporal structure of the spike trains is taken into account

We used the Victor and Purpura spike train distance based classification to quantify how well the neural responses discriminate between the moment of the first reward of a problem (category 1), and the moment of the second, third and fourth reward (category 2), when the monkey is rewarded for coming back on the same target. In both categories, the monkey is submitted to the same external event: it receives a reward. The two categories only differ by the behavioral state of the animal, dependent on the history of the trial. In category one, the monkey receives an unpredicted reward and it has to change its action policy, i.e. to stop exploring the targets and rather come back to the same target. In category 2, the monkey just receives the confirmation that its choice was successful. As there is evidence that in the rare cases ($\sim$ 10 % of the problems) when the monkey makes three errors during exploration, it can switch behaviorally before the first reward is received, because the solution can be inferred, we only included the trials preceded by 0, 1 or 2 errors in the first category (see the reaction time analysis in [31]; and in [29]). In the second category, any number of errors could have been made during the exploration period. Because of the fact that the problems with 3 errors were only excluded from category one, there is a very slight unbalance between the percentage of trials in each spatial direction and the percentage of big *vs.* small reward trials in both categories. However, if this slight unbalance could be used to discriminate between the two categories, then similar or higher classifications should be found in the permuted data, because during the permutation trials are assigned randomly and similar or higher unbalance arise between the two shuffled categories.

Because we were interested in the activity specific to the behavioral shift, which can only happen after the first reward was given, we always began our analyses windows at the moment of the reward. To see the time course of the discrimination ability of the neural responses, analysis windows of increasing sizes were used. In Fig. 2, page 16, an example is shown for a single unit, with the median classification (see Methods, Sec. 2.4.1, page 10). For each analysis window, many Victor and Purpura temporal costs q were tested. It can be seen that for many different costs q, the discrimination was above chance for several consecutive analyses windows; even though the time course of the discrimination ability and the maximum value reached depends on the cost, with spike count based discrimination (dark blue curve, q=0/s) becoming unreliable for long analyses windows. The results are in good agreement for the two discrimination measures (information and percentage of correct). The maximum classification is around 90 % of correct or 60 % of the maximum possible information.

To assess robustly the influence of the temporal cost q on the discrimination ability, we selected an ensemble of cells which showed high and consistent discriminating abilities the following way. A k–means algorithm (with 20 iterations) was used to cluster the cells into two groups, for each combination of (classification measure, classification method). It was further imposed that for each cell a cost could be found for which the classification was consistently superior to chance for a large (≥ 7) number of subsequently increasing analysis windows. This selection procedure made no assumptions about the cost at which cells were most informative. The number of selected cells is indicated in Table 2. The cells selected were mostly the same between groups, and the cells selected with the information were exactly sub–ensembles of the group of cells selected with the percentage of correct (see Table 2, page 16). We compared the proportion of neurons selected thanks to the

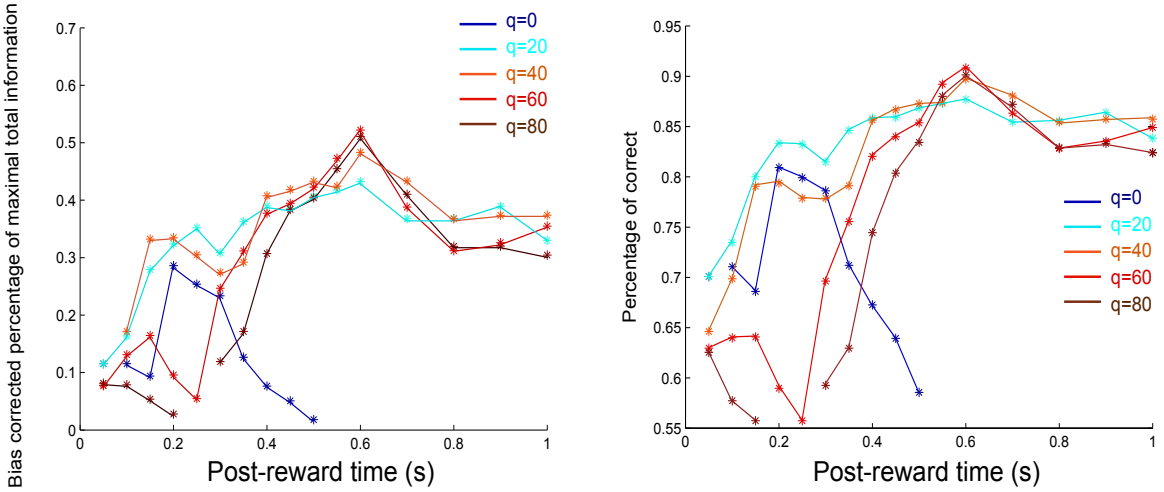

Figure 2: *Discrimination between first and subsequent rewards by a single unit, with the median method, for a subset of Victor and Purpura temporal costs q (per second). The discrimination value is plotted as a function of the analysis window length, with all windows beginning at the reward time. Only values that were higher than the 95th percentile of the distribution of discrimination values in the permuted data are shown. Left: unbiased percentage of maximum information with median method; right, percentage of correct with median method.*

Quadratic method, information	$N = 18$	Common: $N = 16$
Median method, information	$N = 20$	
Quadratic method, % of correct	$N = 32$	Common: $N = 30$
Median method, % of correct	$N = 39$	

Table 2: *Number of cells selected for each classification method/ classification measures. All the cells selected with information were also selected with the percentage of correct when keeping the classification measure constant.*

function tmcomptest of the MATLAB file exchange. No significant differences were found between the median and quadratic method, for either the information or the percentage of correct, whereas for a given method, the percentage of correct selected significantly more cells.

Cells with high and consistent discriminating abilities The effect of the cost was assessed on these subsets of 'well encoding' cells, for an analysis window length at which the maximum information was reached in individual cells: 0.6 s (Fig. 3, page 17). Whatever the measure and the method used, the classification becomes more reliable as one passes from a cost of zero (spike–count based classification) to superior costs (between 5 and 10/15/20), and then decreases when higher costs are used. One main difference between the two classification methods is that, as expected, the quadratic method leads to bad classification with the spike count (see Methods, Sec. 2.4.1, page 10). The time course of the classification was also assessed for different costs (Fig. 4, page 18).

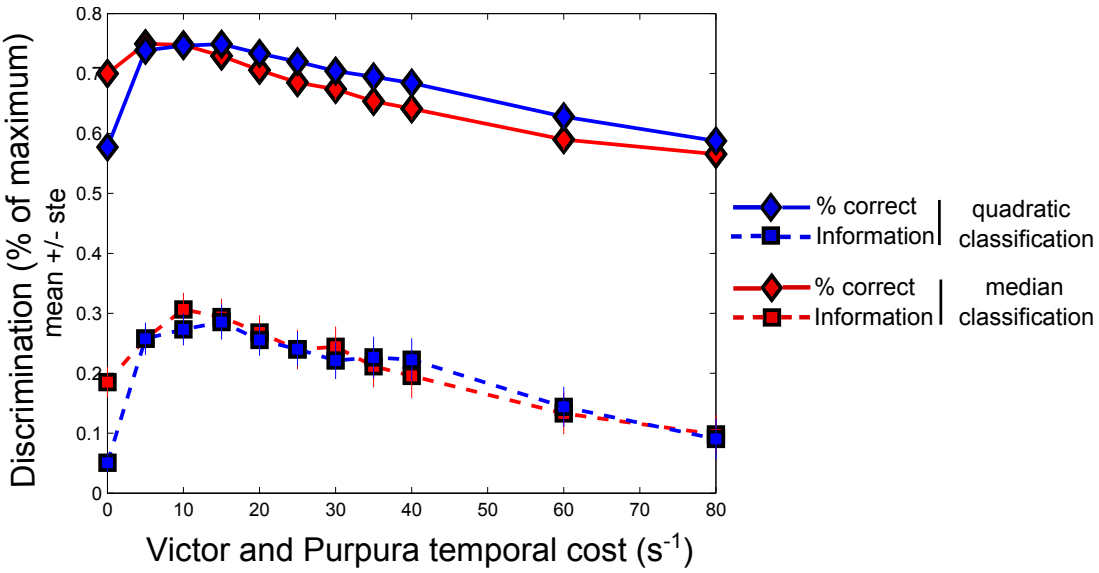

Figure 3: *Means +/- standard errors of the different discrimination measures as a function of Victor and Purpura temporal cost, for an analysis window of [0 0.6s], and among the cells which have high and consistent discriminating power. The standard errors for the percentage of correct are too small to be visible.*

The population–summed discriminability measures had a significantly higher median than the population–summed discriminability measures in permuted data for at least one cost as soon as 100 ms after the reward for the quadratic method, and as soon as 50 ms after the reward for the median method (ranksum test, $p < 0.05$). This is in line with the beginning of the activity burst following the reward (see [31], their figure 3C aligned on CO1 (i.e. first correct) trials).

Globally, for all temporal precision, the classification becomes more and more reliable as the time windows become longer, with an increase in discriminability mostly before 450/500 ms, followed by a saturation.

The two classification methods behave a bit differently: again, the cost of 0 leads to little classification with the quadratic method, whereas it allows considerable classification with the median method. Although quadratic and median classifications tend asymptotically toward similar values of information (for analyses windows of 0.6 s or 1 s, better cost with quadratic classification *vs.* better cost with median classification, ranksum test, $p > 0.05$ for all methods), the median classification slightly tends to increase more quickly (for an analysis window of 0.1 s, median method *vs.* quadratic method information at their best cost, $p = 0.0371$). The differences are less pronounced for the percentage of correct.

It can be seen that the difference between a cost of 0 and superior costs is consistently maintained for all sizes of analysis window. As for all costs the increase of the discriminability power followed the same time course, we used the anova of Friedman to test for an effect of cost while removing the effect of time. There was a very significant global effect of cost (for all methods, the p–values were below the precision limit of the MATLAB functions when all costs $\in [0, 5, 10, 15, 20, 25, 30, 35, 40, 60, 80]$ were included). We used multiple comparisons with Tukey's honestly significant difference criterion to try to assess a difference between spike count based classification and timing sensitive classification. It only leaded to significant differences for the quadratic classification (for both information and percentage of correct), with costs $\in [5, 10, 15, 20]$ significantly different from cost 0.

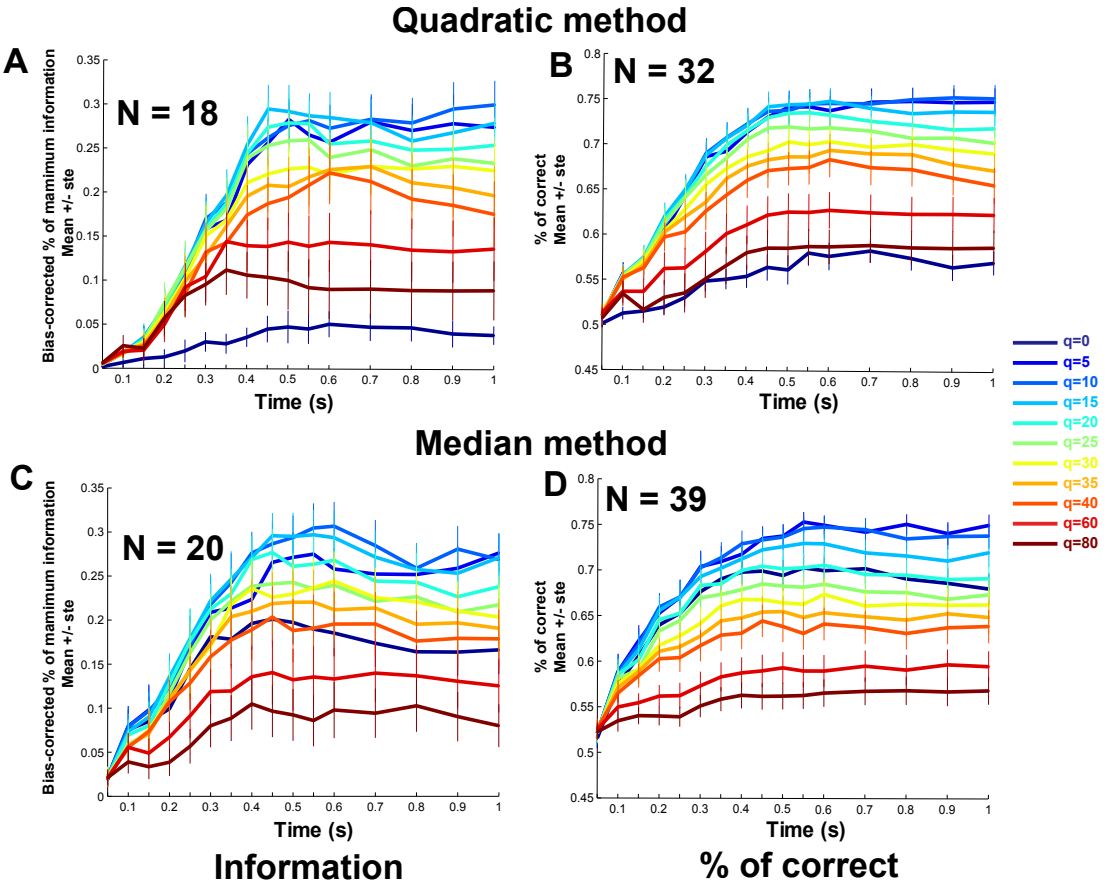

Figure 4: *Discriminability between first and subsequent rewards for well encoding single units. A) Information with quadratic method; B) Percentage of correct with quadratic method; C) Information with median method; D) Percentage of correct with median method. Curves represent the mean discriminability among a subset of cells with high and consistent discrimination abilities. Bars represent standard errors. The different colors represents different Victor and Purpura temporal costs q, as indicated in the legend (unit: / s). The figures on top left of the graphs are the number of units used.*

However, for the median method, when the Friedman anova was realized on the costs $\in [0, 5, 10]$, a very significant effect of the cost was still found (see Table 12 in the appendix for the p–values, page 48). Moreover, when the median discrimination was compared with a ranksum test between cost 0 and the best cost on individual analyses windows, significant differences of median were found for all analyses windows ≥ 0.4 seconds for the information (best cost=10), and for analyses windows ≥ 0.8 seconds for the percentage of correct (best cost =5). Finally, the information (but not the percentage of correct) showed a tendency to increase between cost 5 and cost 10 (Table 12 page 48, Friedman test). Further discussion about possible limits of Friedman anova is developed in the Appendix, Sec. A.3, page 46.

Taken together, the results show that cells which discriminate well and consistently the first reward from the subsequent ones are more informative when the temporal structure of the spike train is taken into account. As high costs are not significantly different from cost 0, this also suggests that the timing precision of the spikes is limited, and that weighting too much the timing blurs the differences between first and subsequent rewards.

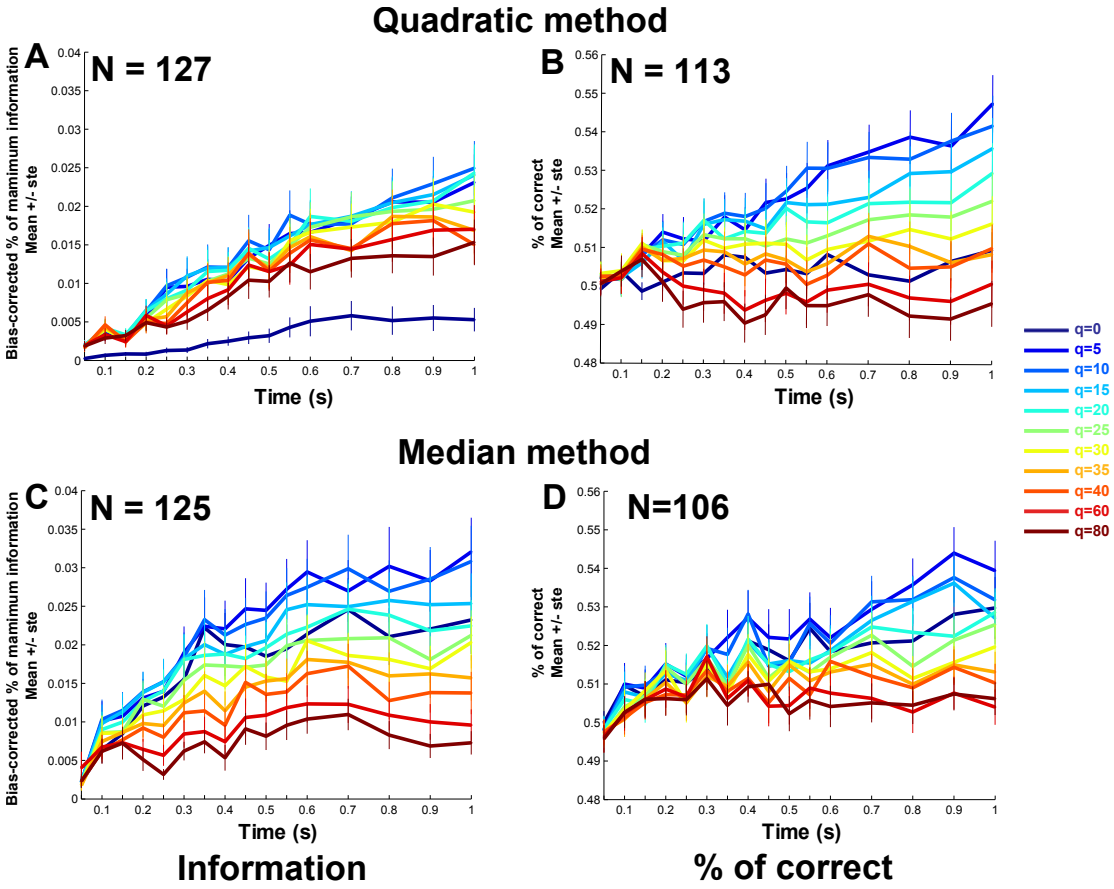

Figure 5: *Discriminability between first and subsequent rewards for mildly encoding single units. A) Information with quadratic method; B) Percentage of correct with quadratic method; C) Information with median method; D) Percentage of correct with median method. Lines represent the mean on a subset of cells with low and/or inconsistent discrimination abilities (see text). Bars represent standard errors. The different colors represents different Victor and Purpura temporal costs q (/s), as indicated in the legend. The figures on top left of the graphs are the number of units used.*

Cells with low and/or inconsistent discriminating abilities For the cells with little and/or inconsistent discriminating abilities (Fig. 5, page 19), the Friedman anovas showed an effect of the cost, but the post hoc comparisons with Tukey's honestly significance criterion were not significant. When the anova of Friedman was tested for temporal costs $q \in [0, 5, 10]$, the effect of costs remained very significant, with one exception were it was but a tendency (see Table 13 in the appendix for the p–values, page 49).

As a conclusion, in this population, the effect of the cost seemed still present, with cost higher than 0 allowing more discrimination than spike count, but the effects were less pronounced.

3.1.2 The correlation between single units activity at first reward and behavioral response time at the next trial increases when temporal structure of the spike trains is taken into account

We hypothesized that if the units that were discriminating well between first and subsequent correct were causally linked to the behavioral shift which occurs after the delivery of the first reward, then the activity

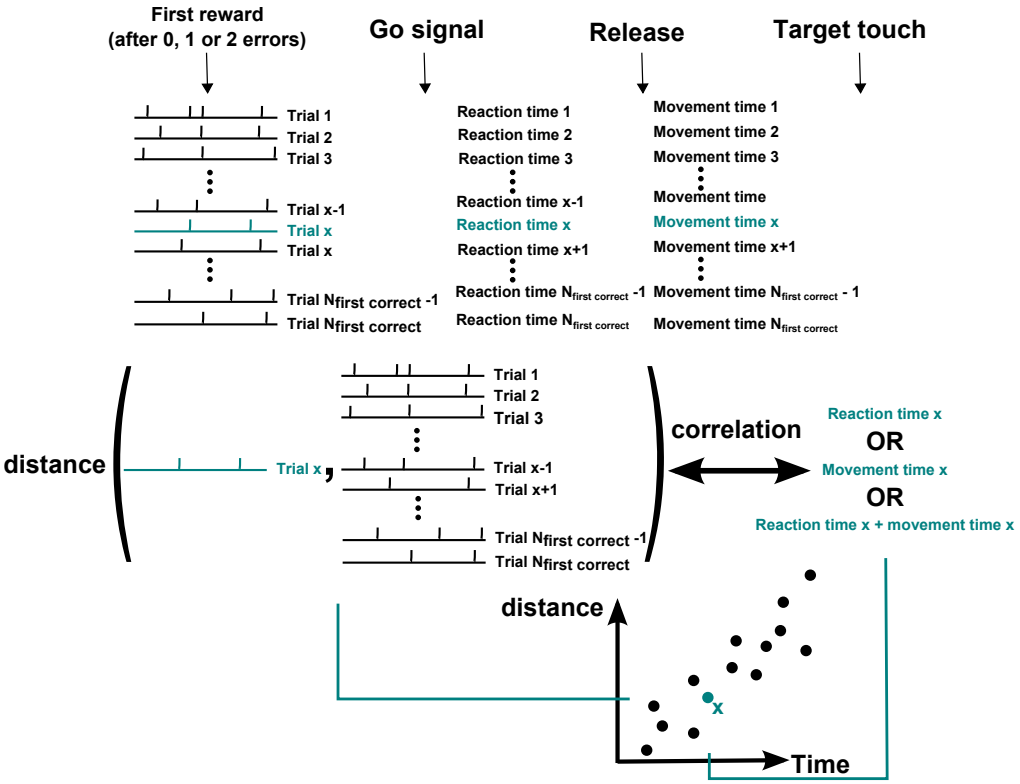

Figure 6: *Illustration of the process linking coding hypotheses and behavior. At each first reward trial, a certain spike train is produced. The first reward is then followed by a go signal, to which the monkey has to answer by releasing the lever and then coming back on the same target to be rewarded. Three different response times can be used to assess the behavioral state of the monkey: the time from the go signal to the release of the lever (reaction time), the time from the release of the lever to the target touch (movement time), or the sum of these two times. At each trial, we try to find a relation between how much a spike train of an individual cell differs from an ideal 'first reward spike train' (measured as the median or quadratic average between this spike train and the ensemble of all other spike trains emitted during the presentation of the first reward), and the behavioral response time.*

following the first reward might be related to the behavior following this first reward. If the monkey was very attentive in a trial, the neurons producing the shift could encode the shift more robustly, and the monkey could be quicker to respond to the subsequent go signal and to touch the target ('response time').

The ensemble of spike trains produced when the monkey received the first reward should give a good estimate of an activity able to produce a behavioral shift, because the monkey is very well trained and behaves almost optimally. Outliers differing from these spike trains are likely to reflect trials in which the shift was not produced 'normally'.

The spike train metrics approach allows to quantify how much a spike train produced in a given first reward trial differs from the other spike trains produced when the first reward is delivered. We computed either the median or the small distances biased average of the ensemble of pairwise distances between a given spike train and the other spike trains emitted during the delivery of the first reward. We were hoping to find a relation between the distance of a spike train to the estimated ideal, prototypical 'first reward response', and the time

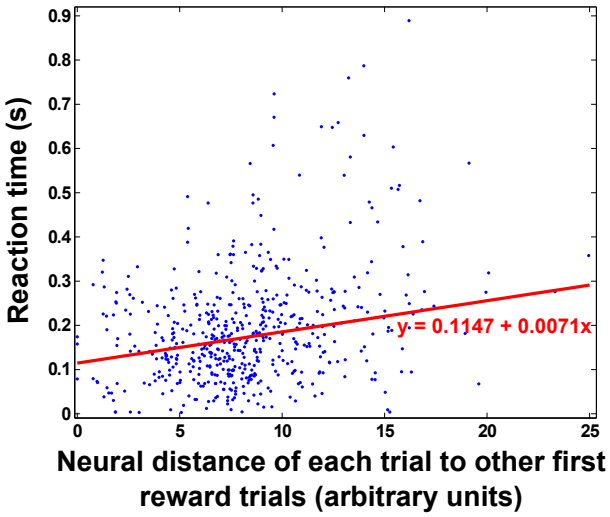

Figure 7: *Each dot represents one trial for one cell. The ordinate shows the reaction time following the (first) reward of this trial. The x axis shows the distance of the spike train emitted in a [0, 1]s post reward window to an estimated "stereotypical first reward spike train" (see text), using the quadratic method, and a temporal cost q=20/s. Data are pulled among cells with high and consistent information for discriminating first* vs. *subsequent rewards. Three outliers for the reaction time, around 5–7 seconds, are not visible but were included for the data analysis. The red line represents a robust linear fit (robustfit function of MATLAB) to all data points (including the outliers). The slope of the fit was highly significant ($p = 2.1881\ 10^{-8}$).*

between the go signal and the release of the lever, called 'reaction time' (which will be followed by the target touch) at the next trial (Fig. 6, page 20). The very rare cases when first rewards were followed by an error were not excluded, because it was also expected that on these trials the neurons activity would be abnormal. However, we only selected first rewards after 0, 1 or 2 errors, because it is thought that there is no behavioral shift after the fourth target touch, the good solution being inferred since the third error.

There is a correlation between distance of a first reward spike train to the prototypical 'first reward response', and reaction time at the next trial. We first used the Spearman rank correlation coefficient to test a possible non linear relationship between the distance of a spike train to the prototypical 'first reward response', and the different behavioral times. We present here the results for the reaction times, but very similar results were found for movement time or the sum of reaction and movement time, as presented in the Appendix, Sec. A.6 page 50.

Cells that discriminate well between first and subsequent rewards show a positive correlation. We pulled the data from all cells of the "high and consistent discrimination power" group. An example of the scatter of reaction time and neural distance to the stereotypic "first reward spike train" is shown for a window of 1 s and a cost of 20 /s, with the quadratic method (Fig. 7, page 21). The results for all costs and analyses windows is shown in Fig. 8, page 22.

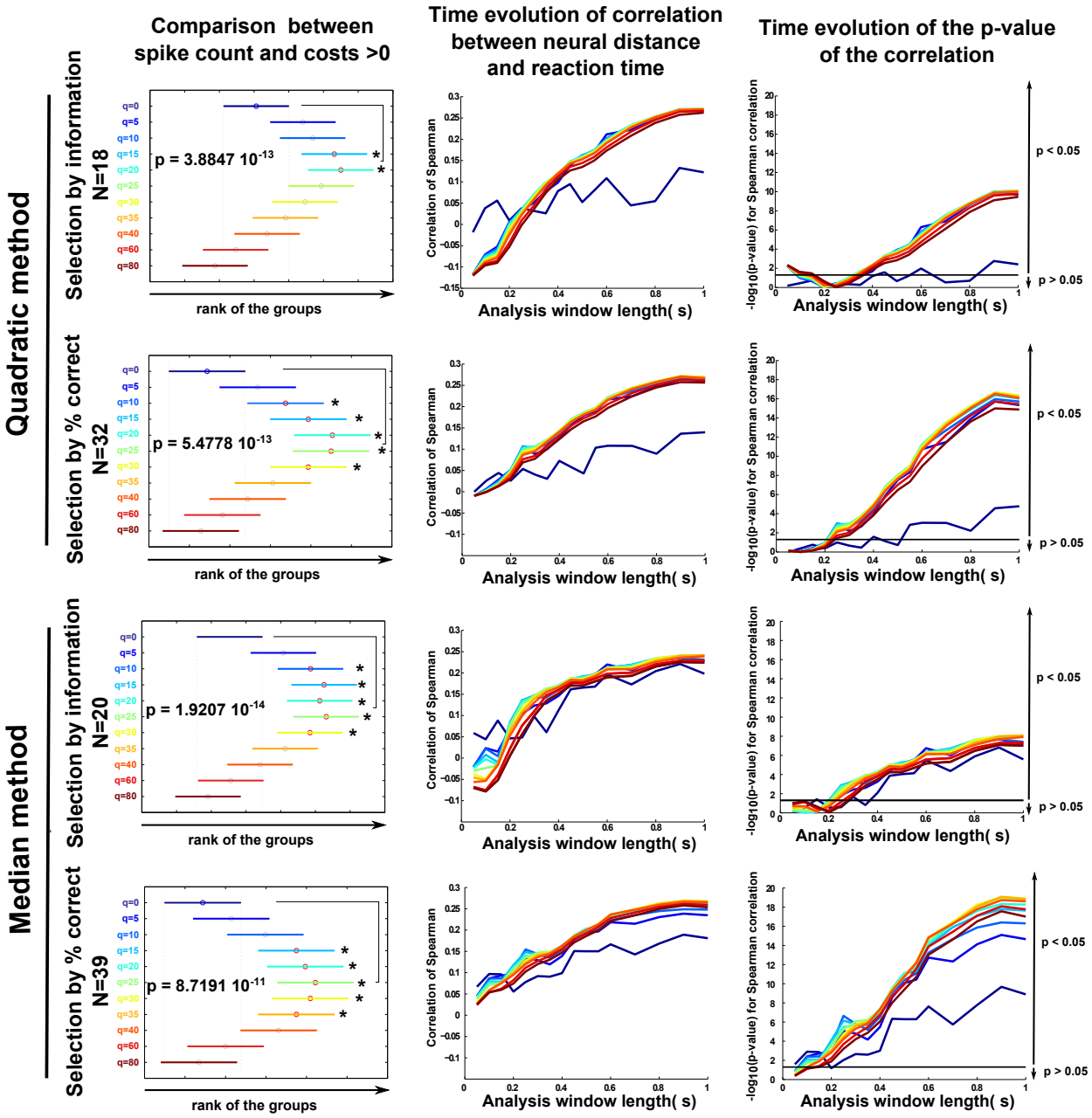

Figure 8: *Correlation between reaction time and distance of a spike train to the 'ideal first reward response', for high discrimination ability cells. The first two rows use the quadratic classification, the last couple of rows use the median classification. The correlation between distance to an evaluated 'ideal first reward' spike train and reaction time was assessed by pooling all the data points from the cells that were selected as highly and consistently discriminative with the information or with the % of correct, as indicated (same groups as in Fig. 4, page 18). The first column shows the result of the post–hoc comparisons with Tukey's honestly significant procedure on an Anova of Friedman comparing the different costs (and removing the time effect). Stars indicate Victor and Purpura temporal costs q that were significantly different from cost q=0 (spike count based distance). Please see the Appendix, Sec. A.3 page 46, for more details on Friedman anova. The figures are the p–values of the Friedman test on all costs. The second column shows the evolution of the correlation when distances are computed on spike trains of increasing lengths. The third column shows the evolution of* $-\log_{10}(p - value)$ *for the correlation. Colors represent different costs as indicated (unit: /s).*

Observation of the second column of Fig. 8 page page 22 shows that the correlation between reaction time and the distance at the reward time to an evaluated 'ideal first–reward spike train' (see Fig. 6, page 20) increases with the length of the analysis window, following roughly the same evolution as the discriminability between first and subsequent rewards, with a slight time–lag (compare with Fig. 4, page 18). The median method leads to significantly positive correlation coefficient slightly quicker than the quadratic method, but both methods lead to very similar maximal correlation coefficient. The results are robust to the precise cells selected, as they are very similar for two sizes of selected groups of neurons (information *vs.* percentage of correct). It should be noted that for cells selected by their information values, for the smaller analysis windows, the correlations are rather slightly negative (while only reaching significance with the quadratic method).

To test for the effect of the cost, an anova of Friedman was used, with the analysis windows as a cofactor. All methods showed a very significant effect (Fig. 8, page 22, inset in column 1). Post hoc comparisons with Tukey's honestly significant difference criterion showed that in all cases, the correlation with a spike count based distance was smaller than the correlation with a distance taking moderately into account the temporal structure of the spike train (costs 15 and 20 per s always correlate better than cost 0). Details about possible limitations of Friedman anova are discussed in the Appendix, Sec. A.3, page 48.

Therefore, using this methodology, single cells, single trial activities were better related to behavior if the temporal structure of the spike trains was taken into account.

The cells which discriminated less well between first and subsequent rewards are not, or little correlated to reaction time when pulled. When the correlations were assessed by pulling all the 'mildly encoding cells', correlations were only significant for the median method, for the 125 neurons with low and or inconsistent information, and only for the [0, 0.15] s analysis window (best cost: q=40/s, c=0.0455, p=0.0249). Note that at this same analysis window length, the correlation was, on the contrary, slightly negative for those cells which had high information.

Thus, the correlation between first-reward activity and behavior was not trivially found in all cells of ACC. However, we would like to stress that the low discriminating group is likely to be composed of subgroups, and additional analyses show that some of these subgroups taken separately would be more correlated to the touch time (see appendix, session A.7, page 54).

3.1.3 The correlation between the reaction times and neural activity at first reward is not (or not only) a side effect of the influence of reward rank on reaction time

Previous studies ([31], their supplementary material; [29]) had shown an effect of the number of failures preceding the first correct on the reaction time of the monkey. Up to 2 failures, for monkey M, the reaction time after the first correct increases with the number of failures (termed 'rank of first correct': 0 failures corresponds to a rank of 0, etc). This has been interpreted as the fact that when the number of errors was high the monkey was taking more time to recall the previous failures and choose the appropriate response, which was thus less automatic. As shown in Table 14 page 55 in the appendix, we confirmed the correlation between the different response times of the monkey and the rank of the first reward (when restricted to [0, 1 or 2]).

Although we tried to approximate an 'ideal' spike train produced during a 'normal' behavioral shift by pulling the responses over all ranks in [0, 1, 2], if responses are more similar at a given rank as compared to between rank, we could not rule out a possible influence of the rank on the distance of a spike train to this prototype, as exemplified in the Appendix, Sec. A.8.2, page 55.

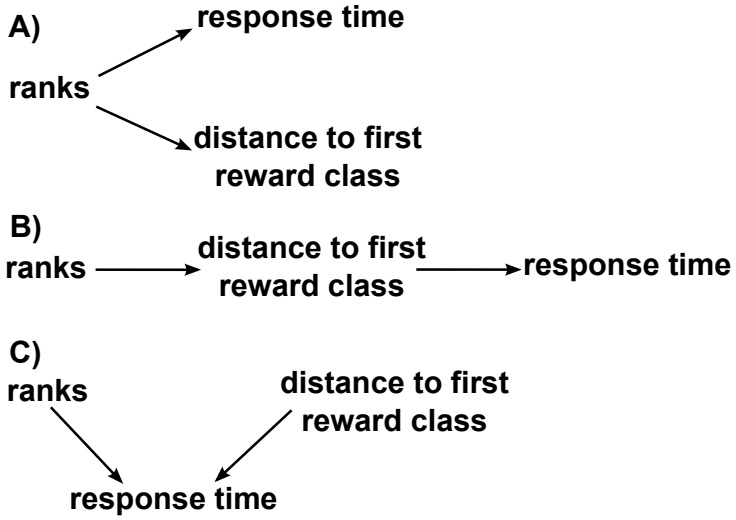

Figure 9: *Three oversimplified models for the relationships between ranks, response times and distance of a spike train to the 'first reward class'. Arrows represent putative causal links.*

Three oversimplified models could account for these two kinds of correlation (see Fig. 9, page 24):

- The rank of the first correct trial influences the distance of this trial to the 'first reward class', and, independently, the rank of the first reward has an influence on the response times of the animal (relying on the activity of different neurons from those we studied) (Fig. 9 A).
- The rank of the first correct influences the distance of this trial to the 'first–reward class', which in turn would be causally related to the responses times of the animal. All the influence of the distance on time comes from the rank effect. Therefore, the activity would more likely reflect the integration of information related to the position of the reward, rather than being generally involved in shifting from one behavior to another (Fig. 9 B).
- The rank of the first correct and the deviation of the activity from an 'ideal first reward' response would have separate and independent influences on the response times (Fig. 9 C).

In the first model, it is required that the correlation between ranks and distance to the prototypical first reward response is stronger than the correlation between the distance and the response times. In the second model, it is required that the correlation between ranks and distance is as great as the correlation between distance and response time (when assuming similar level of noise between ranks and distance and between distance and response time).

Therefore, we tested the significance between the difference of the correlation between ranks and distance (for an analysis–window length l_0 and the temporal cost q_0 that were maximizing it) and the correlation between distance and response times (at the same analysis window length l_0 and cost q_0). The results show that at the analysis window and cost at which the correlation between neural distance and spike train is maximized, there is no significant correlation between the neural distance and the rank, and the correlation between neural distance and response times was always very significantly higher than the correlation between ranks and distances. The detailed p–values are presented in Table 15, page 57, in the appendix.

Thus, when decoded in a way that maximizes the correlation with behavior, the neural responses were little related to the rank of the first correct (situation closer to model C) in Fig. 15 page 57). However, at some different costs and analyses window lengths, the neural response could correlate significantly with the ranks (see the Appendix, Sec. A.8.3, page 56); which shows that the rank of the first correct has effectively an influence on some aspects of the neural response, even though the measure of the distance of one spike train to all "first reward spike trains" at all ranks should rather minimize the impact of this variable on the neural response measure. Accordingly, when the maximal ranks *vs.* distance correlation was compared to the maximal response times *vs.* distance correlations, the latter was still higher (which was significant or was a tendency; all p–values for the permutation test <0.1).

Finally, we compared the correlations between neural distance and response time and between ranks and response time (Table 16, page 57 of the appendix). The influences of ranks and neural distance were generally of comparable strength, with the exception of the correlation with movement times, for the less stringent selection of neurons (neurons selected with percentage of correct). In the latter case, the rank was a more related factor than the neural response.

3.1.4 Summary of the results for the single units analysis

As a general conclusion, regardless of the methodology used, single units which discriminated well and consistently between the first and subsequent rewards could be found. For these units, the discrimination was better when the temporal structure of the response was –moderately– taken into account. The deviation of these units' activity from their usual response also correlated with the time the monkey took to act at the next trial, 6 seconds later. Additionally, it is always found that taking into account the temporal structure of the spike train maximizes the link between their activity and the touch time. Finally, this correlation was largely independent from the effect of the ranks on the response times.

An interesting point is that the activity of the 'well–discriminating' cells was not generally restricted to the first reward (see the Appendix, Sec. A.9, page 58 for further details). In effect, an analysis correlating the deviation of a 'second reward spike train' to an approximated 'stereotypic second reward spike train', and reaction times at the third touch, gave also positive correlations. Moreover, the activity of these cells could also discriminate rather well between the beginning of the first trial of a problem, when the monkey returns to an exploration strategy, and the beginning of a trial in the exploitation phase. However, detailed results rather suggests that the first reward time and, possibly, the behavioral switch to exploitation, is the most relevant factor for the activity of at least some of these cells.

3.2 Correlates of the behavioral shift in the most informative neurons were robust to the superposition of the response of a least informative neuron

We analyzed the activity of couples of simultaneously recorded cells in the anterior cingulate cortex. These cells could come from two different electrodes of the array, or could have been recorded on a single electrode and separated thanks to spike sorting. 122 couples of cells composed of 126 different single units could be extracted from the data set of 145 single units (23 units were recorded alone).

Quadratic method, information	$N = 19$	Common: $N = 17$
Median method, information	$N = 20$	
Quadratic method, % of correct	$N = 39$	Common: $N = 35$
Median method, % of correct	$N = 48$	

Table 3: *Number of couples selected for each classification method/ classification measures. All the couples selected with information were also selected with the percentage of correct when keeping the classification measure constant.*

	Median method		Quadratic method	
	[0, 0.35] s	[0, 0.6] s	[0, 0.35] s	[0, 0.6] s
All couples	$median = 0.9904$; $mean = 0.8097$	$median = 0.8258$; $mean = 0.7152$	$median = 0.9346$; $mean = 0.7989$	$median = 0.8601$; $mean = 0.7287$
Best couples	$median = 0.9308$; $mean = 0.8080$	$median = 0.8222$; $mean = 0.6886$	$median = 0.9488$; $mean = 0.7887$	$median = 0.9201$; $mean = 0.8270$

Table 4: *The median and mean absolute difference in information between two cells of a couple, normalized by the information of the best encoding cell of the couple. Note that the mean is lower than the median, indicating the presence of outliers that have information more equally distributed within the couple. Couples for which both neuron had (after bias correction) zero information were discarded.*

3.2.1 Discriminability between first and subsequent reward for couples of cells is slightly increased compared to the 'best' single unit

We used the Aronov/Victor and Purpura method [4] to quantify how well couples of two units discriminated between first and subsequent rewards, as a function of two parameters: the temporal precision q, and the neuron identity cost k (see Methods, Sec. 2.3.2, page 9). Due to the increased computation time, we only considered two different analysis window lengths: 0.35 s (when the single unit information was increasing, and when the correlation between neural distance and responses time was already significantly positive); and 0.6 s (when the single unit information was maximal).

Similarly to the single units analysis, we selected an ensemble of couples with high and consistent discrimination power by using a k–means algorithm on the maximal discrimination measure (over all costs q and both analyses windows), and by additionally requiring that couples would have discrimination values above the 95th percentile of the distribution for permuted data for both analyses windows. The number of selected couples is shown in Table 3 page 26; again, the percentage of correct selected those couples selected by information, plus other couples; and the median method and the quadratic method selected largely overlapping groups of cells. The difference of proportion was significant between information and % of correct, but not between the two methods (at a risk of 5%, tmcomptest of the MATLAB file exchange).

There was a small increase in discriminability for couples of cells when compared to the best cell. In general, it was very rare that two well discriminating cells were recorded simultaneously. In effect, the percentage of highly encoding cells was roughly $\frac{30}{145} \approx 21\%$, which gives a probability of less than 5 % to have two very informative cells recorded in the same time if they were independent. For the 45 sessions analyzed, the expected number of sessions is 2 or 3. Accordingly, the differences in information between two cells of a couple were quite high (see Table 4, page 26; the difference in information is normalized by the information

	First reward		Subsequent rewards	
	[0, 0.35] s	[0, 0.6] s	[0, 0.35] s	[0, 0.6] s
All couples	$median = 0.5515$; $mean = -0.0921$	$median = 0.4722$; $mean = -0.7495$	$median = 0.3864$; $mean = -0.9173$	$median = 0.4186$; $mean = -0.9070$
Best couples selected with information, for quadratic method	$median = 0.6400$; $mean = 0.1854$	$median = 0.7391$; $mean = 0.2688$	$median = 0.3811$; $mean = -3.2193$	$median = 0.4583$; $mean = -1.2578$
Best couples selected with information, for median method	$median = 0.6148$; $mean = 0.2094$	$median = 0.7161$; $mean = 0.2777$	$median = 0.3864$; $mean = -3.0284$	$median = 0.4410$; $mean = -1.1671$

Table 5: *Median and mean values for* $\frac{(spike\ count\ most\ informative\ neuron)-(spike\ count\ least\ informative\ neuron)}{(spike\ count\ most\ informative\ neuron)}$, *computed separately for the first reward and subsequent rewards, and among the whole population or groups of well–discriminating cells.*

for the best neuron of the couple).

In the same time, the less informative neurons were not silent as is shown in Table 5, page 27. We computed $\frac{(spike\ count\ most\ informative\ neuron)-(spike\ count\ least\ informative\ neuron)}{(spike\ count\ most\ informative\ neuron)}$ separately for the first reward category and the subsequent reward category; and we got median values around 0.5, indicating that the spike count of the least informative neuron was often on the order of half the spike count of the most informative neuron. Moreover, the means were much lower than the median and often negative, indicating the presence of outliers which discharged a lot without being informative.

Consequently, little informative neurons potentially have a strong impact on the multiunits distance, which by construction (and at the contrary to what a neural network can do) weights the two neurons equally in the decoding process. Therefore, it was interesting to ask whether in general more information could be gained by taking into account two neurons, or if it would be more "optimal" for a decoder to ignore the cells that have a low encoding. For this purpose, we computed the difference between the maximal information given by the couple of cells and the maximal information the best cell could give, a measure that we called G (see Methods, Sec. 2.4.2, page 14). Results are given in Table 6, page 28, and show that in the whole population, the gain was small but significantly positive. Among the groups of 'better encoding couples', the effect was of similar or higher magnitude, but it only reached significance for one analysis window, possibly because of the small number of couples and thus the low power of the test. Also, the mean was always superior to the median, indicating that outliers rather tended to be with high information gain.

We also tested a possible correlation between the imbalance in information content between the two cells and the gain G for the couple. Although the correlation were often negative as expected; they rarely reached significance (Table 7, page 28).

As a conclusion, even though there was very often a high imbalance between the discrimination ability of the two cells, the global tendency was for a slight (on the order of a few percent) improvement in the information when both cells were taken into account, as compared to the best cell considered independently.

Taking into account the temporal structure and weighting neural identity generally improved the information. We were interested in knowing which temporal and identity parameters allowed the slight improvement in information observed. Probably because the percentage of correct method was less selective (and also

	Median method		Quadratic method	
	[0, 0.35] s	[0, 0.6] s	[0, 0.35] s	[0, 0.6] s
All couples	$mean = 0.0215$; $median = 0.01$; $p = 4.7459\ 10^{-5}$	$mean = 0.0231$; $median = 0.0189$; $p = 1.8715\ 10^{-6}$	$mean = 0.0271$; $median = 0.0123$; $p = 1.1720\ 10^{-8}$	$mean = 0.0290$; $median = 0.0207$; $p = 5.4817\ 10^{-6}$
Best couples	$mean = 0.0517$; $median = 0.0518$; $p = 0.2632$	$mean = 0.0411$; $median = 0.0255$; $p = 0.0414$	$mean = 0.0613$; $median = 0.0486$; $p = 0.0192$	$mean = 0.0546$; $median = 0.0432$; $p = 0.0636$

Table 6: *The median and mean information gain G as well as the p–value for a two sided, non parametric sign test (H_0: the median is null).*

	Median method		Quadratic method	
	[0, 0.35] s	[0, 0.6] s	[0, 0.35] s	[0, 0.6] s
All couples	$c = -0.1199$; $p = 0.2101$	$c = -0.0397$; $p = 0.6698$	$c = -0.1682$; $p = 0.0686$	$c = -0.0364$; $p = 0.6918$
Best couples	$c = 0.0202$; $p = 0.9326$	$c = -0.2392$; $p = 0.3098$	$c = -0.6011$; $p = 0.0065$	$c = 0.0088$; $p = 0.9741$

Table 7: *Spearman correlation coefficient between G and information imbalance and its p–value.*

perhaps less sensitive), it resulted in less differences for the discrimination power between the interesting variables q and k, and we focus here on the most stringent selection of cells: the selection relative to information values. The results for the information selected cells are presented in Fig. 10, page 29 for the quadratic method, and Fig. 11 page 30 for the median method.

Qualitatively, it can be seen that the mean information among the best couples is maximized for a temporal cost of 15–20, and for a neural identity cost superior to 1. To test for the significance of the difference between costs, the medians of distribution of information values were compared between (q_{opt}, k_{opt}) that maximized the median information and other (q, k), individually for each analysis window. The p–value of the ranksum test was plotted as a function of k and q. Five interesting values are summarized in Table 22, page 64 of the appendix:

- The comparison between (q_{opt}, k_{opt}) and $(q = 0, k = 0)$, which allows to see if mere pooling was deleterious when compared to taking into account the temporal structure and neurons identities
- The comparison between (q_{opt}, k_{opt}) and $(q = q_{opt}, k = 0)$, which allows to test whether the improvement at optimal cost is mainly due to the effect of temporal structure
- The comparison between (q_{opt}, k_{opt}) and $(q = 0, k = k_{opt})$, which allows to test whether the improvement at optimal cost is mainly due to the effect of neural identity
- The comparison between (q_{opt}, k_{opt}) and $(q = q_{opt\ at\ k=0}, k = 0)$, which allows to test whether a temporal cost can be found at which neural identity does not matter (significantly)
- The comparison between (q_{opt}, k_{opt}) and $(q = 0, k = k_{opt\ at\ q=0})$, which allows to test whether an identity cost can be found at which temporal precision does not matter (significantly).

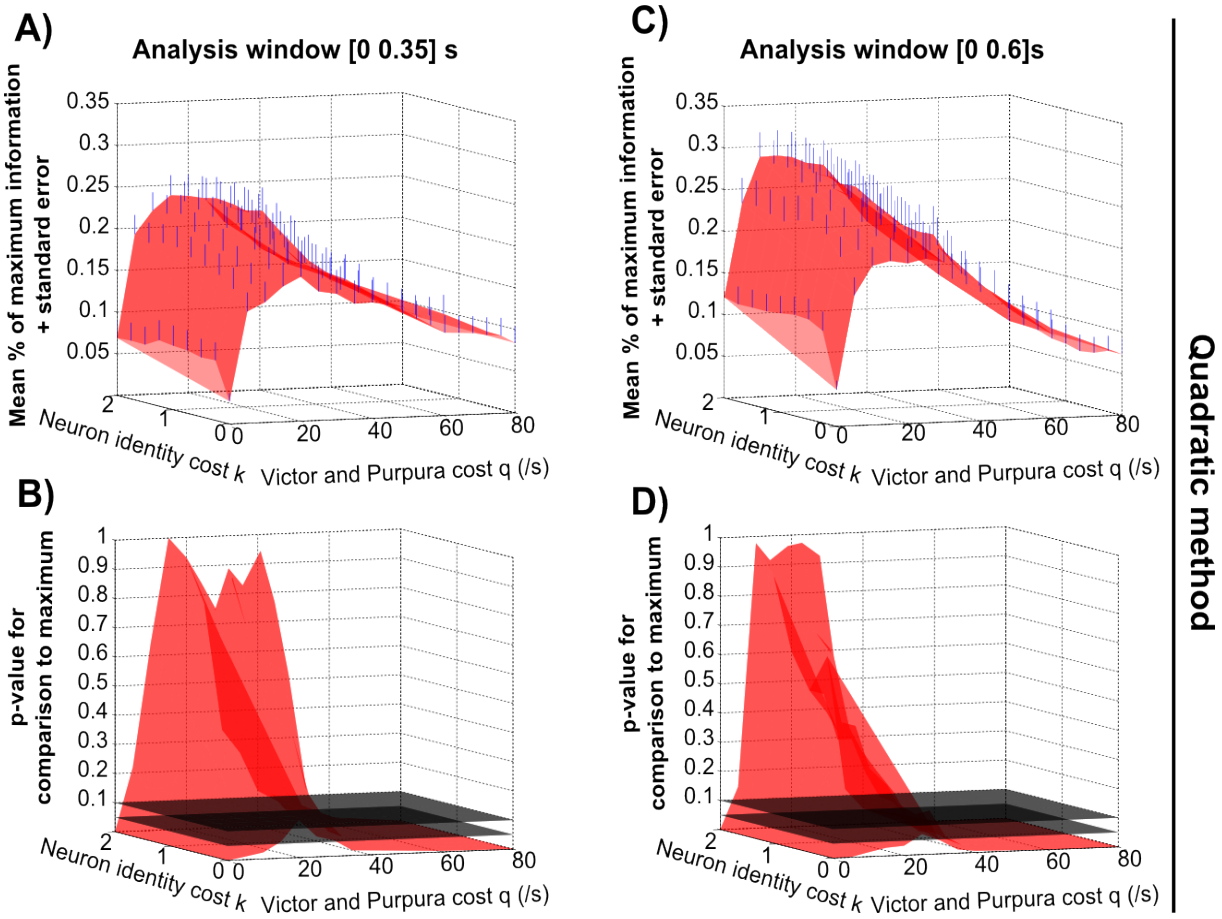

Figure 10: *(A) and (C) Information with quadratic method: mean + standard error; (B) and (D) p–value for the comparison of the median between the optimal distribution and the distribution of information at the indicated cost. For clarity, two flat surfaces at p=0.05 and p=0.01 are drawn.; (A) and (B): Analysis window of* $[0, 0.35]$ *s; (C) and (D): Analysis window of* $[0, 0.6]$ *s.*

For the longer analysis window, all of these tests were significant, indicating an importance of both neural identity and temporal structure for decoding. At the smaller analysis window, even though the results are qualitatively very similar, some tests failed to reach significance.

A more detailed analysis was conducted to better understand the influence of the cost k. At best temporal cost q, its major effect was to improve the correct classification of the spike trains of the "subsequent reward" group. At a temporal cost 0, it had a similar positive effect on the classification of "first reward spike train" and subsequent reward spike train. More detailed results and discussion are presented in the Appendix, Sec. A.10.2, page 64.

3.2.2 Correlation between response times and neural distance were unchanged between ‘best’ couples of cells and ‘best’ single units

Similarly to the single unit analysis, we computed the correlation between the distance of a multi–units spike train to a ‘reward–stereotypic’ multi–units spike train and response times at the following touch. Results were

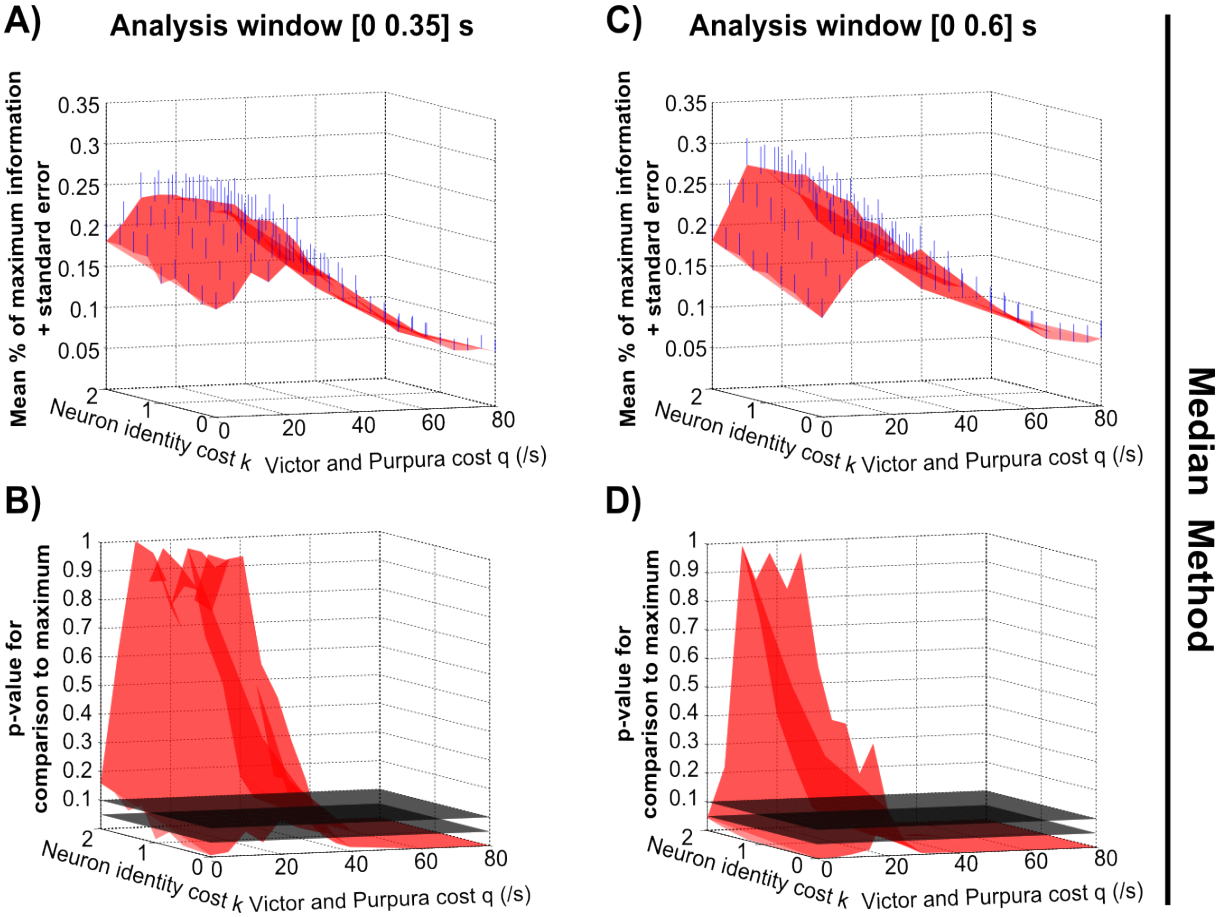

Figure 11: *(A) and (C) Information with median method: mean + standard error; (B) and (D) p–value for the comparison of the median between the optimal distribution and the distribution of information at the indicated cost. For clarity, two flat surfaces at p=0.05 and p=0.01 are drawn.; (A) and (B): Analysis window of* $[0, 0.35]$ *s; (C) and (D): Analysis window of* $[0, 0.6]$ *s.*

globally consistent for reaction times, movement times and the sum of them; we report here mainly results for reaction times. A more detailed analysis, including remarks on small differences found with movement times, is detailed in the Appendix, Sec. A.6, page 50.

There was no evidence for a change in correlation between activity and behavior for couples of cells as compared to individual cells. As the correlation could not be assessed on individual units (because of a lack of statistical power), it is more difficult to compare it between couple of cells *vs.* single units. Two comparisons were however tempted:

- We compared the correlation among the N best single units selected by a k–means algorithm on information values, plus significance at both analyses window lengths (0.35 and 0.6 s), with the N best couples extracted from the N1 ($N1 > N$) best couples selected the same way as the single units. Note that because all the single units were not recorded together, most of the time the best single units were associated with other, less well discriminating cells in the couples (method 1 of Table 8, page 31).

		[0, 0.35]s (RT, MT, sum)	[0, 0.6]s (RT, MT, sum)
Method 1	Quadratic method ($N = 13$)	singles: (0.27,0.26,0.34); couples: (0.28, 0.19,0.28); p value: (0.92, 0.32,0.44)	singles: (0.30,0.30,0.37); couples: (0.36, 0.24,0.36); p value: (0.43,0.43,0.84)
	Median method ($N = 12$)	singles: (0.29,0.29,0.34); couples: (0.31, 0.21,0.31); p value: (0.66, 0.30,0.66)	singles: (0.31,0.32,0.36); couples: (0.29, 0.25,0.31); p value: (0.87,0.36,0.50)
Method 2	Quadratic method ($N = 19$)	singles: (0.26,0.13,0.25); couples: (0.23, 0.12,0.23); p value: (0.72, 0.95,0.74)	singles: (0.37,0.28,0.37); couples: (0.29, 0.22,0.30); p value: (0.17,0.48,0.19)
	Median method ($N = 20$)	singles: (0.29,0.19,0.29); couples: (0.23, 0.18,0.23); p value: (0.39, 0.88,0.29)	singles: (0.38,0.28,0.39); couples: (0.32, 0.22,0.32); p value: (0.28,0.37,0.23)

Table 8: *Comparison of the maximal correlations between response times and neural distances to a "stereotypic first reward", between best couples and best single units, using the two methods as described in the text.*

- We compared the correlation with the N1 best couples and the correlation obtained when taking for each couple the cell with the highest information (method 2 of Table 8, page 31)

In either case, no significant differences were found (even though with the second method correlation values were always lower for the couple as compared to the best single units). The results are summarized in Table 8, page 31.

The correlation between neural activity and behavior increases if temporal structure is taken into account. We compared the correlation at optimal costs (q_{opt}, k_{opt}) with the correlation at other costs. Here again, for the reaction times and at the longer analysis window, the correlation at (q_{opt}, k_{opt}) always significantly exceeded (all p values<0.05) the correlation at $(q = 0, k)$, for any k (see Fig. 12 page 32 for the quadratic method and Fig. 13, page 33 for the median method). Additionally, there was little evidence for a significant decrease in the correlation when temporal precision increased; but the decrease for very high temporal costs q was also very small for individual analyses windows with the single units analysis. For the sum (movement time+reaction time), the results were very close though sometimes the significance became a tendency. For the movement times, the differences were far less pronounced and reduced to the geometric method (see the Appendix, Sec. A.6, page 50).

Neural identity was less important for our measure of correlation between neural activity and behavior. At optimal temporal costs, no significant effect of taking into account neural identity could be found. Although seemingly at odds with the results for the information, a more detailed study of the effect of neural identity at optimal temporal cost (see Appendix, Sec. A.10.2, page 64) reveals that it increased information mostly by improving the classification of spike trains in the 'second, third and fourth rewards' category, but not of spike train in 'first reward' category. In contrast, the behavioral correlation analysis only relies on the distance of first reward spike train to the 'first reward category'. Therefore, the firing of the low informative cell was not deleterious to assess the deviation of the activity of the high informative cell relative to a 'stereotypical' first reward spike train, even though the firing of the low informative cells was little related to behavior (shown in

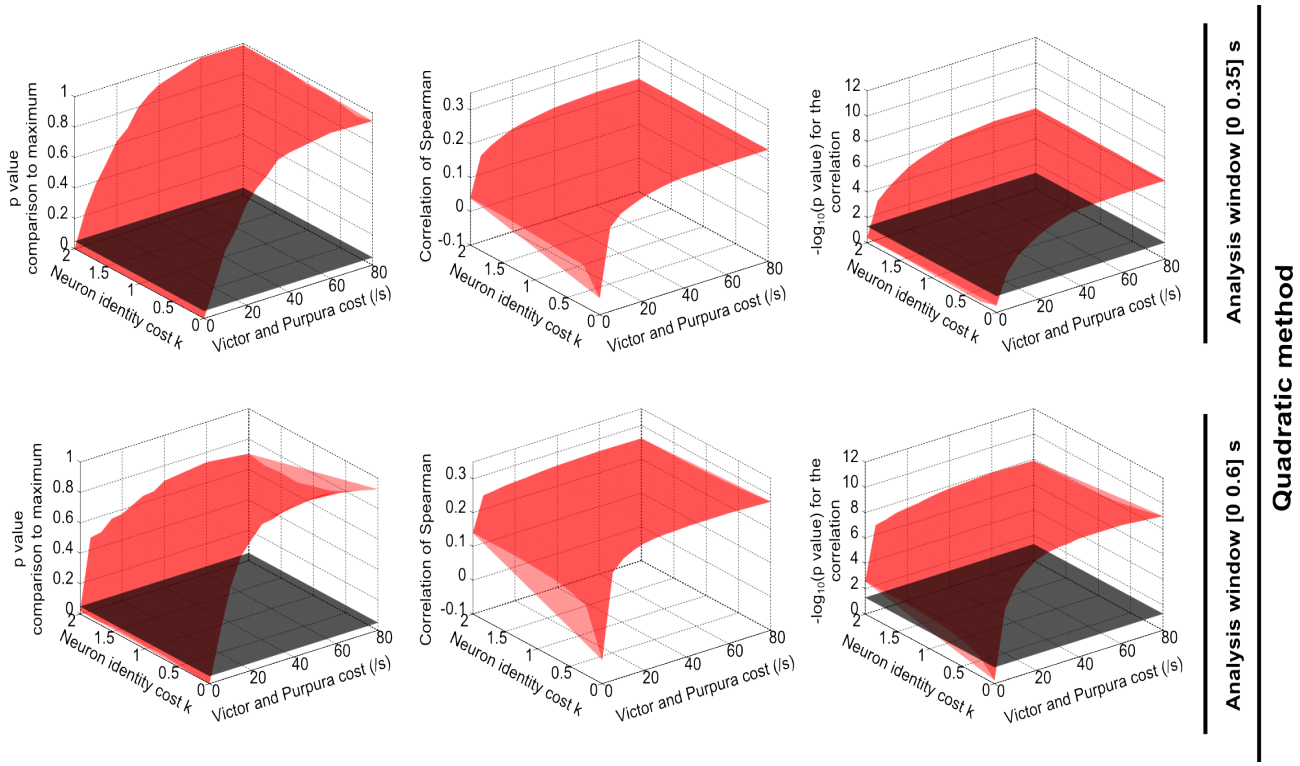

Figure 12: *Correlation between neural distance and reaction times, for the best couples of cells, with the quadratic method ($N = 19$ couples); as a function of the temporal cost q and the neural identity cost k. First row: analysis window of $[0, 0.35]$ s; second row: analysis window of $[0, 0.6]$ s. First column: p value for the comparison of the correlation with the correlation at optimal costs. The value of 1 indicate the position of the optimal costs; and a black flat plane at p=0.05 is drawn. Second column: value of the Spearman correlation coefficient. Third column: $-\log_{10}(pvalue)$ for a test of significance of the correlation coefficient (H_0: the correlation coefficient is null); a black flat plane at p=0.05 is drawn.*

Table 9, page 32). Because of this low correlation for least discriminating cells, taking each cell of the couples independently and then pulling them also generally lead to less correlation (see Table 10, page 33).

	[0, 0.35]s (RT, MT, sum)	[0, 0.6]s (RT, MT, sum)
Quadratic method ($N = 19$)	singles: (0.06,0.03,0.06); couples: (0.23, 0.12,0.23); p value:(0.008,0.18,0.009)	singles: (0.03,0.05,0.04); couples: (0.29, 0.22,0.30); p value:($< 0.0001, 0.01, < 0.0001$)
Median method ($N = 20$)	singles: (0.06,0.04,0.06); couples: (0.31, 0.21,0.31); p value: (0.01, 0.09,0.01)	singles: (0.12,0.16,0.14); couples: (0.29, 0.25,0.31); p value: (0.002,0.33,0.007)

Table 9: *Comparison of the maximal correlations between response times and neural distances to a "stereotypic first reward", between couples of cells and the least discriminating cell of each couple. In each entry, the first two lines report the value of the Spearman correlation coefficient, and the third line report the p–value for the permutation test comparing single and multi–units.*

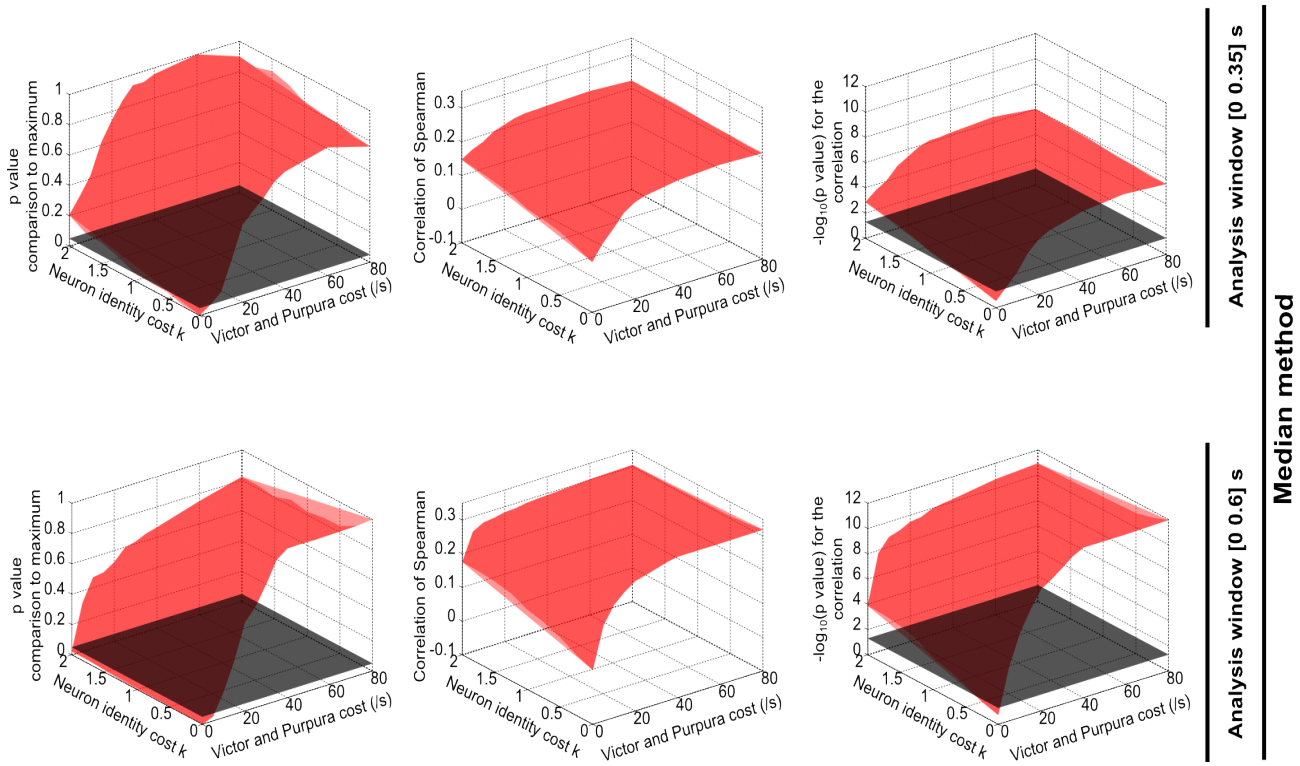

Figure 13: *Correlation between neural distance and reaction times, for the best couples of cells, with the median method ($N = 20$ couples); as a function of the temporal cost q and the neural identity cost k. First row: analysis window of $[0, 0.35]$ s; second row: analysis window of $[0, 0.6]$ s. First column: p value for the comparison of the correlation with the correlation at optimal costs. The value of 1 indicate the position of the optimal costs, and a black flat plane at p=0.05 is drawn. Second column: value of the Spearman correlation coefficient. Third column: $-\log_{10}(pvalue)$ for a test of significance of the correlation coefficient (H_0: the correlation coefficient is null); a black flat plane at p=0.05 is drawn.*

	[0, 0.35]s (RT, MT, sum)	[0, 0.6]s (RT, MT, sum)
Quadratic method ($N = 19$)	singles: (0.13,0.07,0.13); couples: (0.23, 0.12,0.23); p value: (0.06, 0.41,0.07)	singles: (0.14,0.13,0.15); couples: (0.29, 0.22,0.30); p value: (0.01,0.10,0.01)
Median method ($N = 20$)	singles: (0.14,0.11,0.14); couples: (0.31, 0.21,0.31); p value: (0.08, 0.20,0.13)	singles: (0.16,0.16,0.17); couples: (0.29, 0.25,0.31); p value: (0.004,0.26,0.01)

Table 10: *Comparison of the maximal correlations between response times and neural distances to a "stereotypic first reward", between couples of cells and all the cells taken independently and pulled. In each entry, the first two lines report the value of the Spearman correlation coefficient, and the third line report the p–value for the permutation test comparing single and multi–units.*

4 Discussion

Overall, we showed that the temporal structure of spike trains in ACC contains significant information about behavioral shifts. This is in agreement with, and extends previous spike-count related observations by Quilodran et al. (2008) [31], and argue in general for the existence of two modes of action control in exploration and exploitation. Moreover, this analysis also brings more details about how ACC neurons activity may participate in the production of the behavioral shift between exploration and exploitation in this task.

4.1 Behavioral shift markers better correlate with ACC single units activity when temporal structure is taken into account

4.1.1 Single units activity discriminate better between first and subsequent reward when temporal structure of the spike train is taken into account

As a first marker of the behavioral shift between exploitation and exploration, we used the discriminability of single unit spike trains between two situations: the first reward (trials where deduction was possible excluded) *vs.* the subsequent rewards. The rationale for this was that the external event that occur is the same between the two groups: the monkey receives a reward, while the two groups differ by the behavioral strategy that the monkey is deemed to use. In the first group, there is evidence that the monkey shifts between a strategy of exploration to a strategy of exploitation. In the second group, the monkey pursues its exploitation strategy. A neural activity differentiating between the two groups is thus possibly linked to the behavioral switch.

The results show that when using the Victor and Pupura spike train metrics, single units activity discriminates better between the two situations when temporal structure is taken into account enough (Fig. 4, page 18, costs 5/10 *vs.* cost 0), an effect that was more pronounced for the units which discriminate better between first and subsequent reward (Fig. 4 *vs.* 5, page 19). This in turn suggests that to capture the activity deemed to produce the behavioral shift, a neural decoder toward which ACC single units would project would be more efficient if it was sensitive to spike times.

However, this approach suffers from multiple limits. The most obvious one is that pooling the activity of many cells is very probably sufficient to discriminate perfectly between the two situations. Another limit is that the activity allowing the discrimination between the two situations is not necessarily related to the behavioral shift. It could for instance reflect (or also reflect) differences in reward expectation, or 'surprise' [15, 37, 11]. Finally, the two situations occur at different moments of the task. Therefore, they cannot be simply discriminated by a neural decoder, and this discrimination is probably not functionnal, it can only be a cue that the cell studied could be related to the behavioral shift.

Consequently, we complemented the analysis with an other approach based on the correlation between single units activity and behavior.

4.1.2 Single units first reward activity is more correlated to behavioral response latency when the temporal structure of spike trains is taken into account

If the single units that are discriminating well between first and subsequent reward are causally implied in the behavioral switch, then it is expected that their activity correlates with the monkey's behavior when the switch is realized. Previous studies have shown that the response times of the animal (reaction time and movement time) are modulated between exploration and exploitation, being higher for our monkey during exploitation

than during exploration, an observation which can be interpreted as higher control during exploitation. It was therefore hypothesized that:

- There is variability in how well the monkey realizes its behavioral switch, which is due to the fact that the neurons that cause the behavioral switch may discharge in an efficient way in some trial and in a least efficient way in some other trials
- When the monkey realizes less well its behavioral switch after receiving the first reward, it is slower to touch the target at the next trial (i.e. it is slower to decide of the good strategy or, in the rare cases when it makes an error, it hesitates during a very long time)
- Because most of the trials are successful, the efficient switch–producing activity can be approximated by the ensemble of spike trains produced in all first–reward trials (excluding trials for which the solution could be inferred earlier). Consequently, a measure of how much a spike train deviates from the "efficient switch producing" spike train can be given by the measure of the global distance of this spike train to all other spike trains produced during the presentation of the first reward.

Following these hypotheses and assuming that the single units which discriminate well between first and subsequent reward are causally involved in the behavioral shift, a positive correlation between the global distance of a first reward spike train to all other first reward spike trains and the response times at the next touch was expected. It was indeed the case (Fig. 8, page 22). This suggests that the neural activity of these cells is causally related to the behavioral switch. Moreover, the correlation was significantly higher and more consistent in time for reasonably high temporal costs as compared to spike count based classification ($q = 0$). This result is consistent with the hypothesis that the neural network toward which these ACC neurons might project, and which would produce the adapted behavioral output, is moderately sensitive to the spike times of ACC neurons.

The correlation between the subset of cells activity at the first reward, and subsequent trial response times is not likely to reflect a purely motor involvement of these neurons, for three main reasons:

- This activity occurs $\simeq 6$ s before the movement
- The activity discriminates between first and subsequent rewards
- The correlation between the distance of a second reward spike train to all other second reward spike train, and reaction times at the third touch tended to be smaller (see the Appendix, Sec. A.9.1, page 58).

Another interesting observation is that the rank of the first reward trial is also strongly correlated with the following behavioral response times (see the first figure of [31], as well as Table 15, page 57) which more likely reflects an influence of (updated) reward expectancy.

This could at first sight suggest that the correlation between neural distances and response times was only a reflection of a stronger, 'more causal' relationship between rank of the first reward and subsequent response times (models A) and B) in Fig. 9, page 24). In that case, single units ACC activity would be rather related to the integration of previous information about the possible rewarded target (thanks to the outcome of errors) than to a 'behavioral switch' per se, which should be independent of reward rank. Such a correlate would be closer to the previously observed 'reward prediction error', 'reward proximity' and/or 'surprise' effects previously observed [37, 15, 11].

However, at the analyses windows and costs which maximize the correlation between neural distance and behavior, the neural distance was not significantly correlated with reward rank (Table 15, page 57). Moreover, for the reaction time and the sum reaction time plus movement time, the correlation between rank and behavioral times was of similar strength to the correlation between neural distance and behavioral times (Table 16, page 57). This suggests that in this task, there is a way to "read" single units activity which reflects more strongly the behavioral switch per se rather than some of its correlates. This effect, although expected given that the 'first reward' category mixes the different ranks, was not however a trivial consequence of the methodology we used (see the Appendix, Sec. 21, page 56).

The fact that in other tasks where the behavioral switch is less clear, ACC single units activity has been mainly globally related to 'reward expectation' supports the hypothesis that ACC activity can be task-dependent, an effect already observed by Haydn and collaborators [11] when they recorded from the same monkeys, in the same sites, during different tasks.

Moreover, movement times variations seemed more linked to ranks than to a 'general switch related neural distance' (see Table 16, page 57). This is not surprising when considering that as the monkey accumulates error trials, it becomes less uncertain about the direction which will be rewarded, and thus the reaching movement to do, whereas the first removal of the monkey's hand from the lever touch (which is measured by the reaction time) is the same for all directions.

Interestingly, the movement time is also the one which was maximized for a smaller temporal cost in the single units analysis (see Fig. 19, page 52 of the appendix), for which the differences between costs were less prominent in the multi–units analysis (see Fig. 20, page 53 of the appendix), and for which less differences in correlations were observed between well discriminating and mildly discriminating units (Sec. A.7.1 page 54 of the appendix). This suggests a differential influence of ACC activity on the reaction times and on the movement times.

Another caveat with the movement times is that it only relies on times measured by the tactile touch screen, which is rather imprecise (sampling frequency of 50 Hz, which gives a maximal precision of 20 ms), whereas the onset of reaction time and of (reaction + movement) times is recorded as the apparition of a visual signal, which is more precise.

4.1.3 Agreement and discrepancies between neural discriminability and behavioral correlation analysis

Discriminability between first and subsequent rewards and correlation with response times are two complementary methods, because neither of them, nor even the conjunction of both, is sufficient to assess the implication of a single unit in the behavioral shift. In effect, a neuron activity could correlate well with the response times if it was motor–related, for instance. Notably, moderate correlation values were observed in subsets of very low discriminating cells (see Appendix, Sec. A.7, page 54), which activity is not likely to be linked with the behavioral switch.

Moreover, the discriminability between first and subsequent rewards can occur because of differences in reward expectancy between the two situations, for instance, which could occur before the switch–related activity. This might explain why the discriminability is already significant for small analyses windows (see Fig. 4 page 18, analyses windows $\geq 0.05s$ and analyses windows $\geq 0.1s$ for median and quadratic methods respectively), whereas the correlation is non significant or even slightly negative at similar analyses windows. This suggests that the latency of the activity which might cause the behavioral switch is closer to the latency at which the correlation becomes positive (0.2 to 0.3 s post first reward, see Fig. 8 page 22). Furthermore, it is possible

that for early analyses windows, a large neural distance reflects an advanced response relative to the more common case, rather than a 'non optimal spike train' (as it seems to be the case for longer analyses windows), which in turn could cause a smaller reaction time and explain the negative correlations observed (see the Appendix, Sec. A.5, page 49, for further argumentation).

The two analyses globally agree on the fact that spike count based metrics performs less well than metrics taking into account spike timing. However, the correlation analysis tended to find higher best costs than the discrimination analysis. This might be because increasing further the importance of spike timing tended to increase more the distance intra 'first–reward category' than the distance inter–category. However, again, the discrimination between the two categories cannot really occur at the moment of the shift, and therefore it is possible that less optimal costs for discrimination are actually used when the shift–related activity is read out.

Thus, globally, the single units analysis suggest that spike timing has an importance in the 'encoding' of the behavioral shift by ACC single units, although the determination of the exact temporal precision is probably impossible because different methods (for instance, different metrics; e g [38, 34]) are likely to give slightly different results, and because of technical limitations to measure accurately the responses times and the reward time. Finally, it is also possible that the internal reference which putatively allows neurons to be sensitive to spike timing does not exactly corresponds to the reward time, and this would probably lead to an underestimation of the temporal precision in the present studies. Other temporal references, for example based on the phase of spikes relative to the theta or the gamma rhythms rhythms in local field potentials, remain to be tested.

More speculatively, it can be noted that this relative importance of spike timing is compatible with the hypothesis that a "downstream" decoding neuron would be sensitive to the temporal structure of the 'well discriminating cell', for example because it would receive as an other input a spike train corresponding to the 'stereotypic, ideal' first reward spike train of this well discriminating cell. This (oversimplified) framework is close to the concept of cortical neurons as 'coincidence detectors' [18]. However, in our case, the relevant temporal precision of the spike trains is probably very rough. 'Summation' of depolarizations between the ideal pattern and the 'best neuron' spike train would thus need a smoothing mechanism, for instance a rather high membrane time constant in the decoding neuron, adapted to the temporal precision of its input. This hypothesis is made more explicitly with the Van Rossum [38] distance (which uses an exponential, 'synaptic–like' kernel to convolve two spike trains to be compared, before taking the distance as the difference between the two continuous functions obtained), as well as with the SRM distance [7] (which computes the distance as the difference between the spiking probability functions of a Spike Response Model 'downstream' decoding neuron when it receives one *vs.* the other spike train).

4.2 Correlates of the behavioral switch encoded by the best single unit activity are robust to the superposition of the firing of a less 'switch–correlated' simultaneously recorded cell

In general, due to the low proportion of well–encoding cells, there was a very high unbalance between the discriminatory power of two simultaneously recorded cells. Consequently, the recording did not allow to see reliably what would happen when the activities of two cells of comparable (and rather high) discriminatory power would be considered jointly, because pooling the activity of two cells recorded independently neglects the presence of noise correlations, which may either result in an overestimation (because the noise of the two cells would vanish by averaging when considered independent) or an underestimation (if the joint deviation of the cells' activities from their mean activity is informative about the stimulus) of the gain of information when

the two cells are considered jointly [20].

In an attempt to see if the 'worst cell' activity (which is non zero, see Table 5 page 27) would blur the 'best cell' activity, or if on the contrary the activity of the two cells would positively combine, we used the Aronov multiunits metrics to compute the informative power of couples of simultaneously recorded cells, as a function of the importance of temporal precision (cost q) and of the importance of knowing which cell fired which spike (cost k).

This method weights both cells equally. Therefore, if a decrease in information is observed for the couple as compared to the best cell, then it would be more optimal for a downstream decoded to ignore (i.e. to decrease the synaptic weight of) the worst cell to discriminate between first and subsequent reward. If no difference between the couple activity and the best cell activity is found, then it is equally good for a downstream neural decoder to ignore one cell or to take both cells into account; it suggests that the 'best neuron' activity is robust to the firing of the least discriminating neuron, (for example because it does not fire at the same time as the 'best cell', or because weighting neural identity and implicitly assuming a divergence between the projection of the two cells on two independent decoders is sufficient to conserve the information, due to the fact that the worst cell does not indicate the opposite choice to the best cell). Finally, if a gain is found for a couple of cells as compared to a single cell, then it is either more optimal to have the two cells converging toward a same "decoder neuron" (case when k_{opt}=0), or to have two independent neural decoder, each one decoding the input from only one neuron, and indicating to which extent its input neuron indicates the presence of one of the situations. The activity of the two 'decoders' (case $k_{opt} \simeq 2$) would be combined afterwards, for instance by convergence to a third neuron. These different (oversimplified) cases are presented in Fig. 14, page 39.

For the information, the results show there was globally a very slight but significant increase in the discriminability between first and subsequent reward in the couple compared to the best cell considered alone (even though the gain might be either positive or negative in an individual couple). Moreover, in the best couples, the couple information increased when temporal precision and neural identity were taken into account (see Fig. 10, page 29 and 11, page 30).

However, at the optimal temporal cost, the effect of taking into account neural identity was mostly to improve the correct classification of spike trains belonging to the '2nd, 3rd and 4rth rewards' category. Please refer to appendix A.10.2, page 64 for further discussion about possible reasons why the effect was most sensitive for this category.

For the first reward category, the results thus suggest rather a robustness of the firing of the best informative neurons relatively to the superposition of spikes from the least informative neuron at optimal temporal cost. This was indeed found in the correlation analysis, which only relies on the 'first reward category' spike train (see Fig. 12, page 32 and 13, page 32). In effect, the correlation with behavior relying on couples' activity was found to be statistically equivalent to the correlation relying on the 'best single units' activity (Table 8, page 31), whatever the neural identity cost k, provided that temporal structure was taken into account, and even though the 'worst cells' activity considered separately was little correlated to behavior (Table 9, page 32).

Much more speculatively, this robustness also suggests that in a '(rough) temporal pattern matching' framework, in which the neuron's activity are 'decoded' by coincidence detection of the activity of one cell with a 'stereotypical', 'optimal' first reward spike train, 'bad cells' spikes would rarely coincide with this 'stereotypical' spike train. If a spike time dependent plasticity mechanism is added, the synaptic weight from the 'bad cell' to the neural decoder may be expected to decrease. It would thus be interesting to investigate further this question thanks to simulations and modeling.

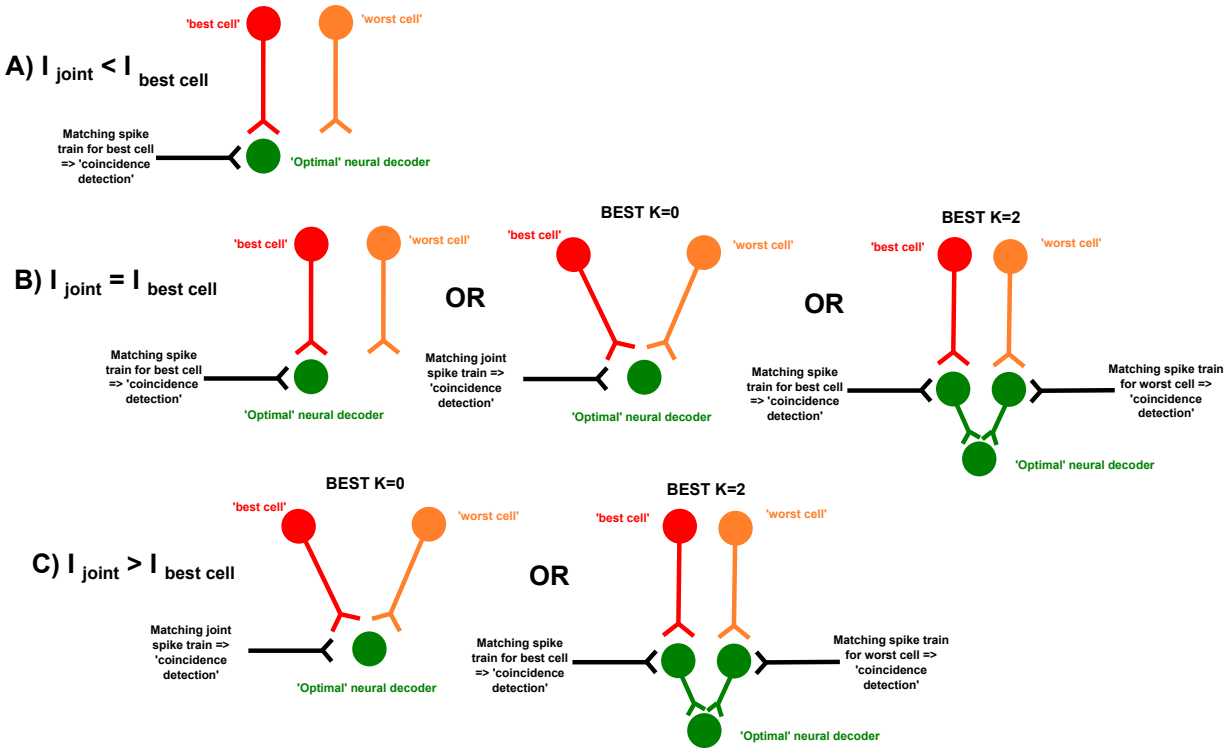

Figure 14: *Three oversimplified 'optimal' decoding hypotheses compatible with the three different results of the couples of cells analysis, at the best temporal cost q (which is* > 0*, suggesting a moderate importance of spike timing). A) The best information for the couple is smaller than the best information for the best cell. B) The best information for the couple is equivalent to the best information for the best single cell. C) The best information for the couple is superior to the best information for the best single cell. 'Coincidence detection' is here taken in a very large sense, and does not imply ms matching, but rather temporally approximate coincidence of input spikes which still sum up because of a post–synaptic membrane time constant compatible with the timing accuracy of the pre–synaptic spikes.*

Of course, these results do not generalize easily to higher numbers of cells, even though this question is certainly relevant because neurons usually receive a great number of inputs [1]. This question needs to be investigated further. It is possible that a mere pooling of spike count from a great number of cells would finally correlate with behavior. Another possibility is that activities from a rather small number of very informative cells, possibly situated in distant areas [13] and thus less correlated, would be 'decoded more optimally', using the temporal structure of the spike trains and a 'coincidence detection like' mechanism. This is not assuming, of course, that the other, less informative cells are useless, because they could have an important contribution in the network's dynamic which is responsible for the activity of well encoding cells, and because less informative cells might be involved in other task–related mechanisms, and/or in other tasks. The idea that all neurons of an area might not be used in a given task was already pointed out by Purushotaman and colleagues [30], who found that only a subset of MT neurons correlated with the behavior on a trial by trial basis.

5 Conclusion

Converging evidence from different studies, using different animals and different tasks, points toward a role of ACC in performance monitoring and behavioral strategy management. The study of Quilodran et al. (2008) showed that in an exploration-exploration paradigm, some neurons' firing rate was differentially modulated between the first reward, when the monkey switches from exploration to exploitation, and the subsequent rewards, when the monkey pursues its exploitation policy. The work presented in this report extends Quilodran et al.'s observations by showing that the activity of a subset of ACC single units discriminate well (up to 90 % correct) between first and subsequent reward. In addition, deviation of spike trains from their estimated 'ideal first reward spike train' correlates positively with the following target touch response time. This reinforces the hypothesis that ACC neurons are involved in strategy switch at first reward.

Importantly, we used a method that allowed us to vary the spike–timing sensitivity of a putative 'downstream' neural network. We suggest that a properly tuned neural decoder (i.e. capable of best exploiting the temporal structure of ACC spike trains) can optimise discrimination ability and mediate the activity/behavior correlation of single-unit-single-trial processes.

Finally, analysis of simultaneously recorded pairs of ACC units showed that both the discrimination and the correlation abilities of 'best cells' were not impaired by the interfering action of least discriminative cells. In future investigations, we will adopt a modeling approach to test whether or not this robustness property coupled with plasticity mechanisms could shape the dynamics of a downstream neural network toward neglecting least informative afferents, and/or would combine activity from many highly informative neurons.

A Appendix

A.1 Properties of the mutual information between true and reconstructed classes

A.1.1 For big numbers of trials, under the hypothesis of chance clustering, I(T,R) is 0, regardless of the differences in the number of trials between categories

To verify that the information measure behaves correctly in our case where the number of trials differs between the classes, a purely theoretical approach can be used. It is sufficient to remark that :

$$I(T,R) = H(true\ categories) - \langle H(true\ categories/reconstructed\ categories)\rangle_{reconstructed\ categories} \tag{10}$$

Where H is Shannon's entropy, and $<>$ denotes averaging. By definition, if true and reconstructed categories are independent,

$$H(true\ categories/reconstructed\ categories) = H(true\ categories) \tag{11}$$

and hence I(T,R)=0.

To see how this occurs practically, with Romain Brasselet, we derived the value of the information in the case when there are 2 categories, the first one with N_1 trials (N_1 very big), and the second one with N_2 trials (N_2 very big), under the null hypothesis that the distances are uniformly distributed within and between categories, i e when there is no segregation of the two categories by the neural responses.

For the median classification, for any spike train s, the median of the distances between s and the spike trains of category 1 should be the same as the median of the distances between s and the spike trains of category 2. Therefore, the asymptotic values of the confusion matrix are given by :

$$\begin{bmatrix} \frac{N_1}{2} & \frac{N_1}{2} \\ \frac{N_2}{2} & \frac{N_2}{2} \end{bmatrix}$$

Straightforward algebra shows that I(T,R) is null in this case.

For the quadratic classification, let us consider a further approximation, by which the quadratic classification reduces to the classification in the class containing the closer neighbor. Then, the probability to be classified in category one will tend toward $\frac{N_1}{N_1+N_2}$, i.e. toward the probability of being close to a spike belonging to category 1 by chance, whereas the probability to be classified in category 2 tend toward $\frac{N_2}{N_1+N_2}$. Therefore, the expected number of spikes belonging to category x and classified into category y is $N_x \ \frac{N_y}{N_1+N_2}$. Hence, the asymptotic values of the confusion matrix are given by :

$$\begin{bmatrix} N_1 \ \frac{N_1}{N_1+N_2} & N_1 \ \frac{N_2}{N_1+N_2} \\ N_2 \ \frac{N_1}{N_1+N_2} & N_2 \ \frac{N_2}{N_1+N_2} \end{bmatrix}$$

Straightforward though somewhat tedious algebra also leads to I(T,R)=0.

A.1.2 The maximum value of I(T,R) computed on finite samples depends on the repartition of the trials between categories

Let us take the case when we were able to perfectly classify the spike trains. We have :

$$H(true\ categories/reconstructed\ categories) = 0 \tag{12}$$

Therefore,

$$I_{max}(T,R) = H(true\ categories) \tag{13}$$

A very famous general property of Shannon's entropy is that it peaks for uniform distribution. Therefore, I_{max} peaks for the following confusion matrix :

$$\begin{bmatrix} \frac{N}{2} & 0 \\ 0 & \frac{N}{2} \end{bmatrix}$$

where N is a positive natural integer. The larger the imbalance of the number of trials between categories is, the smaller I_{max} is. Here are three numerical examples :

$$\begin{bmatrix} 57.5 & 0 \\ 0 & 57.5 \end{bmatrix}$$

$$I = \ln(2) \simeq 0.6931$$

$$\begin{bmatrix} 45 & 0 \\ 0 & 70 \end{bmatrix}$$

$$I \simeq 0.6693$$

$$\begin{bmatrix} 25 & 0 \\ 0 & 90 \end{bmatrix}$$

$$I \simeq 0.5236$$

A.2 Additional considerations about the bootstrapping method

A.2.1 Justification of the use of a bootstrapping method to evaluate the bias in information measure

Treves and Panzeri [27] have established an analytical formula for the bias term as a function of the number of trials, when the information between stimuli and responses is computed thanks to the "direct method". This consists in dividing each trial by a number of bins N_{bins}, and assigning to each bin a discretized neural response, for example spike count. In this case in which the responses can be considered as independent between bins, the information is the sum of the informations between the probability distribution of the spike count in one bin and the probability distribution of the stimuli. Our situation is equivalent to having only one response bin with $N_{categories}$ possible neural responses (trial classified as belonging to category one, or trial classified as belonging to category two), and two "stimuli" or situations (eg : first reward vs subsequent rewards). In these

conditions, and with the assumption that if an infinite number of trials was available no cell would perfectly (or, equivalently, not at all) classify one stimulus (no entry of the confusion matrix would be zero, which is the parallel of non zero p(s, response i) in [27], page 93, beginning of the page), then the shuffled information provides an accurate estimate of the bias. Finally, even if the last assumption was not verified, as argued in the main text, the shuffling procedure would lead to an overestimation of the bias on the order of $\frac{1}{2N_{trials}\ln(2)}$, i e for most sessions less than a few percent of the information of the very informative cells/couples, which seems a reasonable inaccuracy.

A.2.2 Evaluation of the significance and of the bias thanks to a bootstrapping method

Single units analysis We evaluated the significance of an information or of a percentage of correct value using a Monte Carlo method to sample the values expected by chance, when the neural responses do not discriminate at all between the categories. To do so, for each cell and each analysis window, we built 1000 surrogate data sets, each of which corresponding to a random permutation of the spike trains, mixing the two categories. The value of the information or of the percentage of correct was computed, and the data was considered as significantly different from chance when its value was higher than the 50 highest values (0.05*1000). However, this test is only accurate if 1000 values allow a good estimate of the distribution. The total number of permutations is, indeed, much higher : on the order of $N_{tot}!$, where N_{tot} is the total number of trials, typically 100/150. The following section tries to argue that 1000 is reasonable.

To try to assess the convergence of the distribution of the permuted data information, we increased the size of the number of permuted data sets from zero to 1000, and for each number of permuted data sets we computed the mean information (or percentage of correct) as well as the value which would be used as a threshold for significance at 5 % (the 0.95*number of permuted data highest value). We did this procedure a hundred times, and represented the mean +/- 2 standard deviations (among these 100 repetitions) of the mean and the threshold values as a function of the number of permutations. We want to verify that with 1000 permutations, we have an accurate estimation of the mean, which will be used as a measure of the bias, and of the threshold value. An example is shown in figure 15 page 44, for a cell in a window covering the 0.6 s after the reward time, and with two categories corresponding to a)first reward at a rank 0, 1 or 2 ; b) second, third or fourth reward.

As expected, the mean information and the 95th percentile tend to be badly estimated for samples less than 500 permutations, as indicated by a high variance of the estimations. This was probably due to the fact that the distribution of information in permuted data sets has a very long tail (not shown). For higher numbers of permutations, the variance of the estimations saturates, while being non zero, probably because the confusion matrix is made of discrete numbers, and so the information measure is also discrete.

It was verified visually that, for some other cells, the values of the mean information and of the 95th percentile did not vary too much between 600 and 1000 permutations.

Additionally, we noted that the distribution of percentage of correct may not be exactly symmetric and may be slightly biased (here, for the median classification, the median is 0.4986, significantly different from 0.5 : signtest, $p = 1.022210^{-10}$). This is probably due to the fact that clusters of small distances spike trains can emerge even with permuted data, for example because of the unbalanced number of trials between classes.

Multi–units analysis Due to the (greatly) increased time for the computations when using the multi–units distance, the number of random permutations on which we computed the discriminability was reduced to 100. To assess the goodness of the evaluation of the bias and the significance, we computed 5*200 permutations for

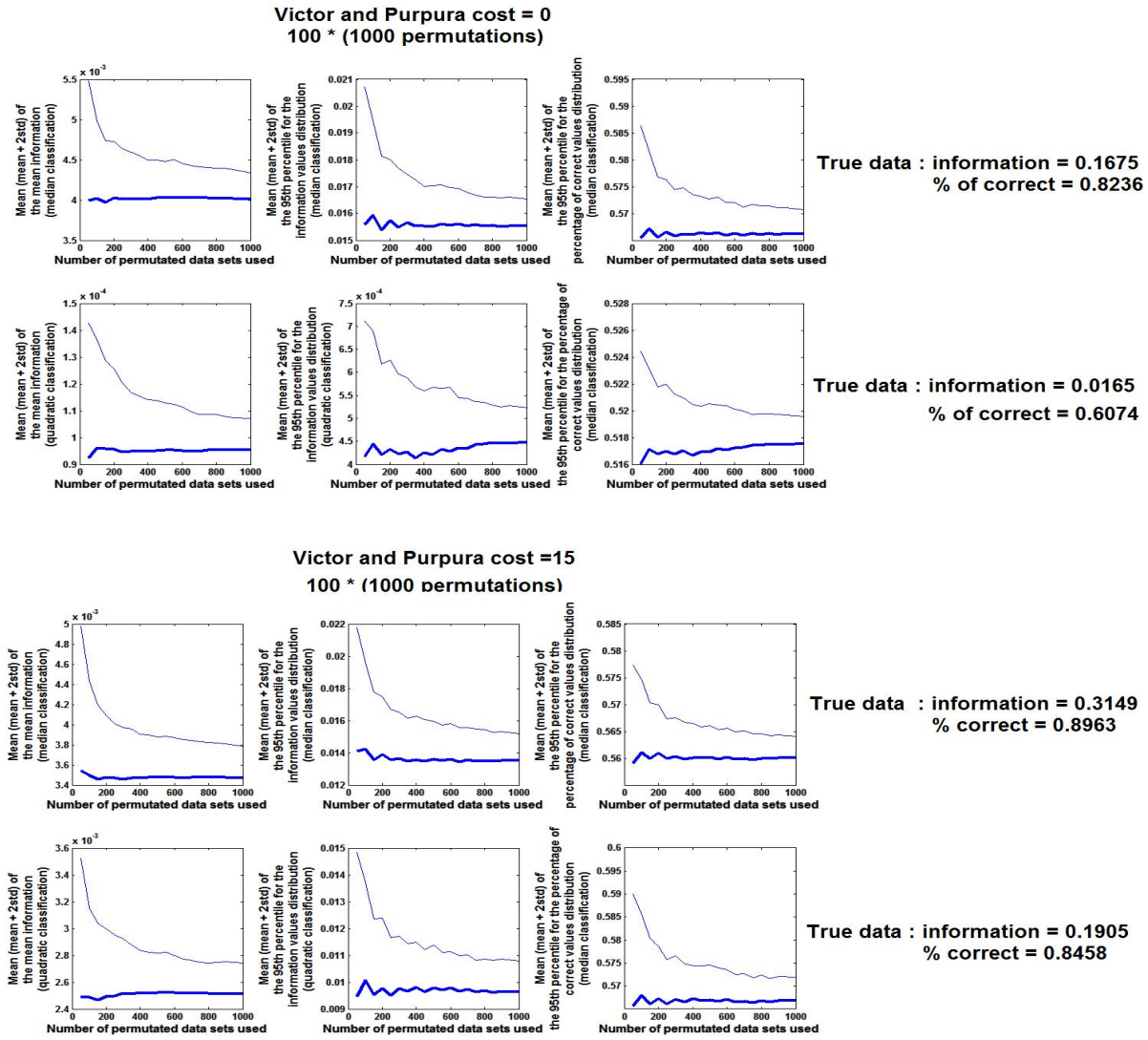

Figure 15: *Convergence of the estimates of the information and the percentage of correct in an ensemble of single unit permuted data sets. Top: Victor and Purpura cost of 0; bottom: Victor and Purpura cost of 15. For each cost, the first row represent the median classification, and the bottom row represents the quadratic classification. Each row ifs composed of the mean information, the 95th percentile of the information distribution, and the 95th percentile of the percentage of correct distribution as a function of the number of permutations used. Bold line: mean of 100 sets of 1000 permutations. Thin line: mean + 2 standard deviation among these 100 sets.*

two example pairs of cells, one which discriminated well between the first and subsequent rewards, and another which did not. We computed the mean and the standard deviation among the 5 repetitions of both the mean information and the 95th percentile, as a function of how many permutations were used. The informative pair is shown in figure 16 page 45; one of the cell is the example in the previous section. The shape of the curves was not different with the other pair, but the asymptotic value of the mean and of the 95th percentile could be

higher (possibly due to the fact that there were less trials), respectively equal to 0.01 and 0.07 for instance.

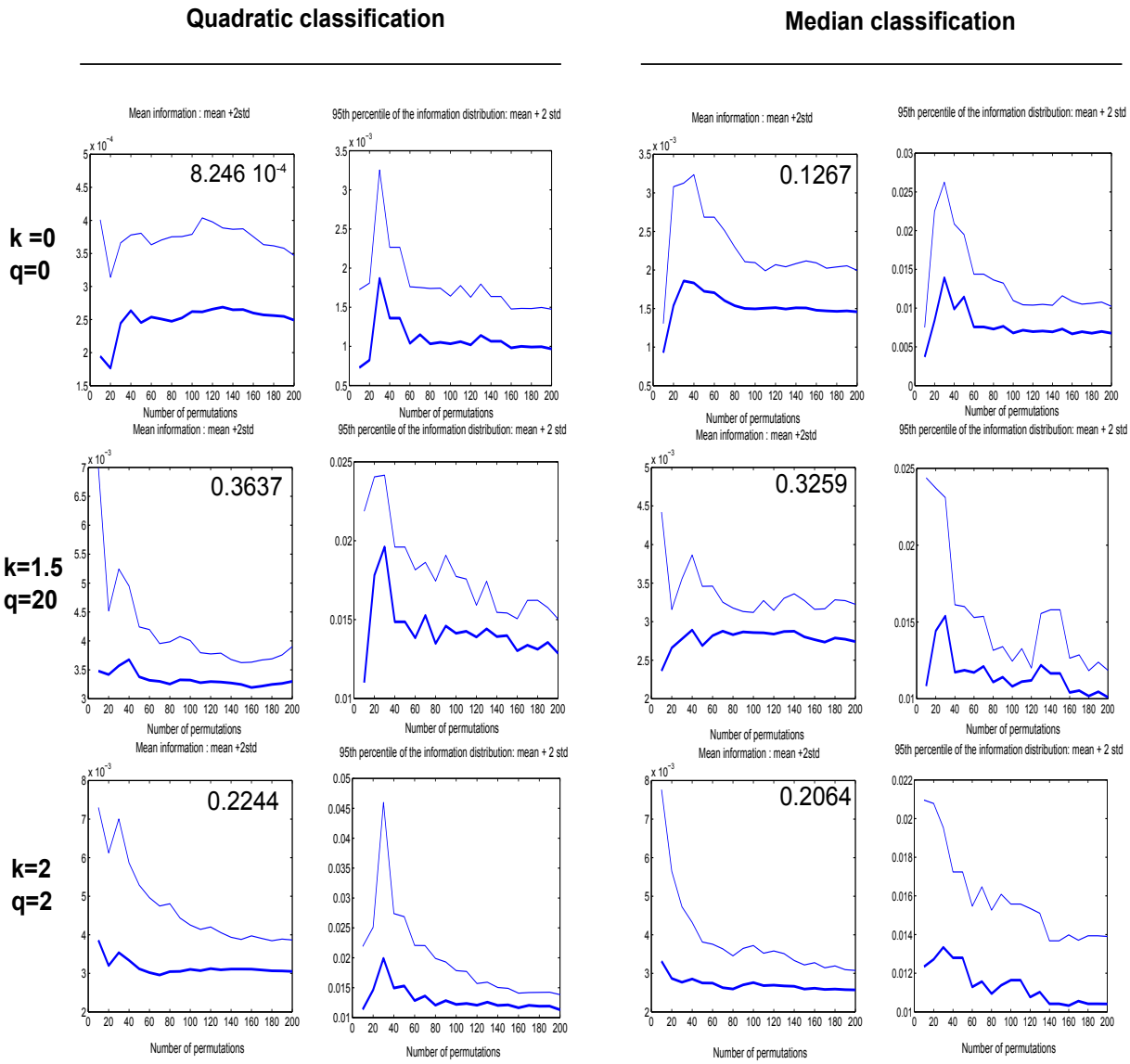

Figure 16: *Convergence of the estimates of the information in an ensemble of permuted data sets from a couple of cells. The left side is the obtained with the quadratic method, and the right side with the median method. For each method, the left column is the mean, and the right column is the 95th percentile. Each row is a different combination of the temporal (Victor and Purpura) cost q, and the neuron identity cost k. Bold line: mean of 5 sets of 200 permutations. Thin line: mean + 2standard deviation among these 5 sets. The value of the information in the true (non–permuted) data is shown as an inset.*

It can be seen that in most cases, a hundred permutations were sufficient to reduce by an considerable amount the uncertainty on the estimates. When very little permutations are used, apparently, the five repetitions we computed were not sufficient to sample that variability (e g, first line of the figure). The decrease seems quicker than for single units, possibly due to a reduced variability of the total response in a trial. However, it is true that for cells with a low information, in some cases, the significance at a single analysis window might

not be very accurate. On the contrary, this is not a problem for the very informative cells on which the main analysis focuses.

In general, this analysis reveals that for low informative cells or couples, the assessment of significance is imperfect (there is still some variance on the estimation of the 95th percentile, which can be as high as 25% of its asymptotic value in the examined examples). This is why we required that more than one analysis window was significant, and this is why we generally focused on cells that had a high information.

A.3 Possible limitations of Friedman anova

Friedman Anova assumes that the data depend on two main factors, F_1 which can take values $f_1 \in [1, 2, ..., n_{f_1}]$ and F_2 which can take values $f_2 \in [1, 2, ..., n_{f2}]$.

Each k^{th} observation O for the combination of factor (f_1, f_2) is modeled as :

$$O_{(f_1,f_2,k)} = \mu + \alpha_{f_1} + \beta_{f_2} + \epsilon_{(f_1,f_2,k)} \tag{14}$$

where μ is a location parameter, and ϵ is an error.

The test thus assumes that all data come from distributions with the same (continuous) shape, but with different locations due to the effect of the factors F_1 and F_2. It determines whether the data is in agreement with the null hypothesis that $\alpha_{f_1} = 0$, and thus whether the variability in the data likely comes from error and/or from differences in the factor F_2.

In our analyses, we used the "time" (measures done with increasing analyses window lengths) as the factor F_2, and the factor F_1 that was tested is the Victor and Purpura cost q.

A general caveat to using Friedman anova is that as the analyses windows are of increasing length, each measurement is not independent from one another. This is violating one of the assumptions of the test. For the information analysis, this is not a real problem because it can be shown additionally that on individual analyses windows, there are significant differences between costs. For the correlation analysis, we had only one measurement for each time point (the coefficient of correlation), and therefore classical tests could not be done at each individual analysis window.

In the following, we will elucidate how this fact could harm our conclusions, and why we think the test is still probably accurate.

The problem with time–related measurement is that if one measurement is significant at one time point, then it is likely to propagate the significance at a following time point. For instance, we would say that the fact that the $q = 5$ curve in figure is superior to the curve $q = 0$ at an analysis window of 450 ms is only caused by the inheritance of a difference already present at an analysis window of 400 ms. Consequently, pooling among these analyses windows when computing the statistics is not relevant, because of the fact that the difference is conserved does not mean that it is more significant.

However, we would like to argue that such cases are probably not occurring in our particular situation. Indeed, each data point comes from the pooling of many trials from many (mainly) independent cells which have all been shown to be independently very well discriminating between first and subsequent rewards ; therefore, a high stochastic deviation seems unlikely.

Moreover, it can be seen that the correlation changes a lot with increasing analyses windows, with even a sign change in some cases. This suggests that the spikes that are added in time are strongly changing the

		reaction times	movement times	reaction plus movement times
median method	information	$p = 0.0019$; $q \in [30, 35, 80]$	$p = 0.054$; $q \in [35]$	$p = 0.0203$; $q \in [35, 80]$
	% correct	$p = 9.2\ 10^{-8}$; $q \in [30, 35, 40, 60, 80]$	$p = 0.1469$	$p = 2.1\ 10^{-8}$; $q \in [30, 35, 40, 60, 80]$
quadratic method	information	$p = 1.3\ 10^{-12}$; $q \in [30, 35, 40, 60, 80]$	$p = 3.6\ 10^{-9}$; $q \in [35, 40, 60, 80]$	$p = 1.0\ 10^{-12}$; $q \in [30, 35, 40, 60, 80]$
	% correct	$p = 5.7\ 10^{-13}$; $q \in [30, 35, 40, 60, 80]$	$p = 3.4\ 10^{-11}$; $q \in [35, 40, 60, 80]$	$p = 5.8\ 10^{-13}$; $q \in [30, 35, 40, 60, 80]$

Table 11: *p values for the Friedman anova on values of correlation between neural distances and response times. Analyses windows from [0.2, 0.3] to [0.9, 1] were included. The conclusions are unchanged if for each group, the Friedman anova is realized on all analyses windows where at least one cost is significantly positive. The values of q are those that are significantly higher than cost* $q = 0$ *(post hoc with Tukey's honestly significant procedure).*

correlations relative to previous smaller analyses windows. Therefore, it is likely that a 'chance deviation' at the beginning of the spike train would be 'turned down' by the new spikes.

Additionally, it should be possible to further reduce the bias by taking analyses windows of very different lengths. The results were quite robust to a decrease in the number of analyses windows used. We tried to use three analyses windows at which the correlations were already significantly positive, and which were as different from each other as possible. For instance, one can only use the analyses windows of 0.25, 0.5 and 1 s and still get a significant effect of the cost with the median method. In that case, each successive window doubles the window length, thus making the probability of chance inheritance very weak.

Finally, a supplementary control analysis was made : the spike trains were binned every 0.1 s from the beginning of the reward, and the correlation was assessed on each of these individual, separated analyses windows. Correlations were always significant and positive for some costs for analyses windows [0.2, 0.3] s and longer analyses windows. An anova of Friedman was realized with the correlation on each of the windows ([0.2, 0.3] s , [0.3, 0.4]s , [0.4, 0.5] s, [0.5, 0.6] s, [0.6, 0.7] s, [0.7, 0.8] s, [0.8, 0.9] s, and [0.9, 1] s) as the F_2 factor. The significance of the effect of the cost remained for all analyses but the correlation of movement times with the median method, which was already less marked in the preceding analyses, and for which the number of significant analyses windows were reduced (table 11, page 47).

Further, post hoc comparisons always showed that some costs $q > 0$ were significantly different from cost 0. It might be argued that there are correlations remaining, because the fact that a neuron emits a spike at time t has an impact on the probability that it emits a spike at time $t1 > t$. However, the particular correlation type that would invalidate our analysis is that if by chance the spike train during the first analysis window correlate with behavior, then it increases the chance that the spike train emitted on the next analysis window correlates by chance with behavior at the same cost, and this effect remains when different trials coming from different cells are pooled. This seems rather unlikely. Finally, of course, this approach also has its own limits (arbitrariness of the analysis window length, and of its boundaries ; 'border effects'), but they are different from the method described in the main text and give globally consistent results. Notably, for this analysis, it is fair to argue that spike count classification (and low–cost classifications) are unfavored because it is only possible to integrate or

match spikes over less than 0.1 s. Accordingly, no decrease of correlation was observed with higher costs in the 'individual analyses windows' analysis. This hypothesis is however not consistent with the main text result, which allows integration of the spikes during up to 1s. Conversely, the main text analysis could be biased if the maintenance of the difference between spike count and higher costs for many analyses windows was only due to an inheritance effect of the difference from a smaller, previous analysis window ; but this possibility is made unlikely by the complementary analysis.

Taken together, these analyses suggest that higher costs were consistently above $q = 0$, for many analyses windows. This has some relevance notably because it is very likely that the cells that mostly participate in the correlation for early analyses windows are not the same as those which have more impact on longer analyses windows (see notably examples of single cells raster plots in [31], which shows that individual cells firing is not continuous over the whole post–reward second).

However, these analyses do not state whether at one analysis window the difference between $q = 0$ and higher costs is big or significant.

Of course, the best case would be to be able to observe significant differences between two costs using an analysis that requires only one window per cost. We can do it by shuffling the data points between two costs and look if a similar or higher difference in correlation coefficients can occur by chance. In our data, it is not likely to be the case for all groups, because the difference in correlations can be small.

Finally, we would like to clarify why in figure 8, page 22, the ranks of different costs (first column of the figure) may appear at odds with the differences in the curves for different costs. Notably, cost 0 appears to have a higher rank than cost 80, even though the curve at cost 0 looks well below the curve at cost 80. This is because on the very first analyses windows, cost 0 is just above the other costs, and thus gets the highest rank, which has a strong effect on its cumulative rank, whereas cost 80 is always either last or second to last, and thus it gets a smaller cumulative rank.

A.4 Additional p–values tables for comparison between temporal costs q of single units discrimination abilities

A.4.1 Best discriminating cells

	Information, median method	Percentage of correct, median method
All cells, $q \in [0, 5, 10]$	$p = 1.0494\ 10^{-11}$	$p = 1.1551\ 10^{-9}$
One cell / session, $q \in [0, 5, 10]$	$p = 3.05036\ 10^{-11}$	$p = 5.79605\ 10^{-9}$
All cells, $q \in [5, 10]$	$p = 0.0701$	$p = 0.7172$
One cell / session, $q \in [5, 10]$	$p = 0.0508$	$p = 0.5451$

Table 12: *p–values for the anova of Friedman for the median method and restricted subsets of Victor and Purpura temporal cost q, as indicated. Either all cells with high and consistent discrimination values are included, or only one cell per session is included, to respect better the independence assumption of Friedman anova.*

A.4.2 Remaining cells

	Information median method	% correct median method	Information quadratic method	% correct quadratic method
All cells, $q \in [0, 5, 10]$	$p = 6.40399\ 10^{-5}$	$p = 0.0013$	$p = 0$	$p = 0$
One cell / session, $q \in [0, 5, 10]$	$p = 0.0591$	$p = 0.001$	$p = 9.5479\ 10^{-15}$	$p = 2.2204\ 10^{-16}$
All cells, $q \in [5, 10]$	$p = 0.6491$	$p = 0.3445$	$p = 0.2172$	$p = 0.8601$
One cell / session, $q \in [5, 10]$	$p = 0.9256$	$p = 0.6796$	$p = 0.1811$	$p = 0.8485$

Table 13: *p–values for the anova of Friedman for the median method and restricted subsets of Victor and Purpura temporal cost q, as indicated. Either all cells with low and/or inconsistent discrimination values are included, or only one cell per session is included, to respect better the independence assumption of Friedman anova*

A.5 Additional discussion about the slightly negative correlations between early first–reward neural activity and subsequent response times

On the contrary to long analyses windows, where the correlations between distance to a 'stereotypic' first reward spike train and subsequent response times were positive, for very small analyses windows, smaller (and at times significant) negative correlations were observed (see figure 8, page 22, for instance). In the main text, we propose that for long analyses windows, spike trains that are very different from the approximated 'canonical' first reward spike train are produced in trials in which the activity was not efficient to produce the switch (because the monkey is very trained), whereas when the analysis is restricted to the very first spikes, some of the outliers might be those for which the neural response was quicker and could thus potentially allow the monkey to answer quicker. To test this hypothesis, we tried to correlate the distance of one spike train from the ensemble of other first reward spike train with either the latency of the first spike, the median latency of all emitted spikes, and the mean latency of all emitted spikes. When no spikes had been emitted, we arbitrarily attributed the value of the length of the analysis window plus 1 ms to the minimal, median or mean latency. We used the groups of 'well discriminating cells' selected by their information values (which showed the negative correlations in figure 8, page 22), and we computed Spearman's coefficient of correlation.

Very significant negative correlations were found for all three measures for the two first analyses windows, strengthening the idea that 'big neural distance' spike trains matched with early responding spike trains for small analyses windows. Additionally, the correlations were more negative for high temporal costs q than for $q = 0$, in agreement with the fact that $q = 0$ did not lead significant correlations for small analyses windows (see figure 8, page 22).

Finally, when the same correlations were computed for longer analyses windows ([0 0.6]s and [0 1]s), the median and mean latency correlated positively with the neural distance, whereas the minimal latency (latency of the first spike) kept being negatively correlated with the neural distance, though with a smaller absolute coefficient (and it was not always significant). Thus, for longer analyses windows, the main effect was rather that for spike trains with high distance most of the spikes occurred late.

A.6 Correlation between first-reward neural activity and movement times or (reaction + movement) times

A.6.1 Correlation between distance to an estimated 'ideal first reward spike train' at first reward and following reaction+ movement times

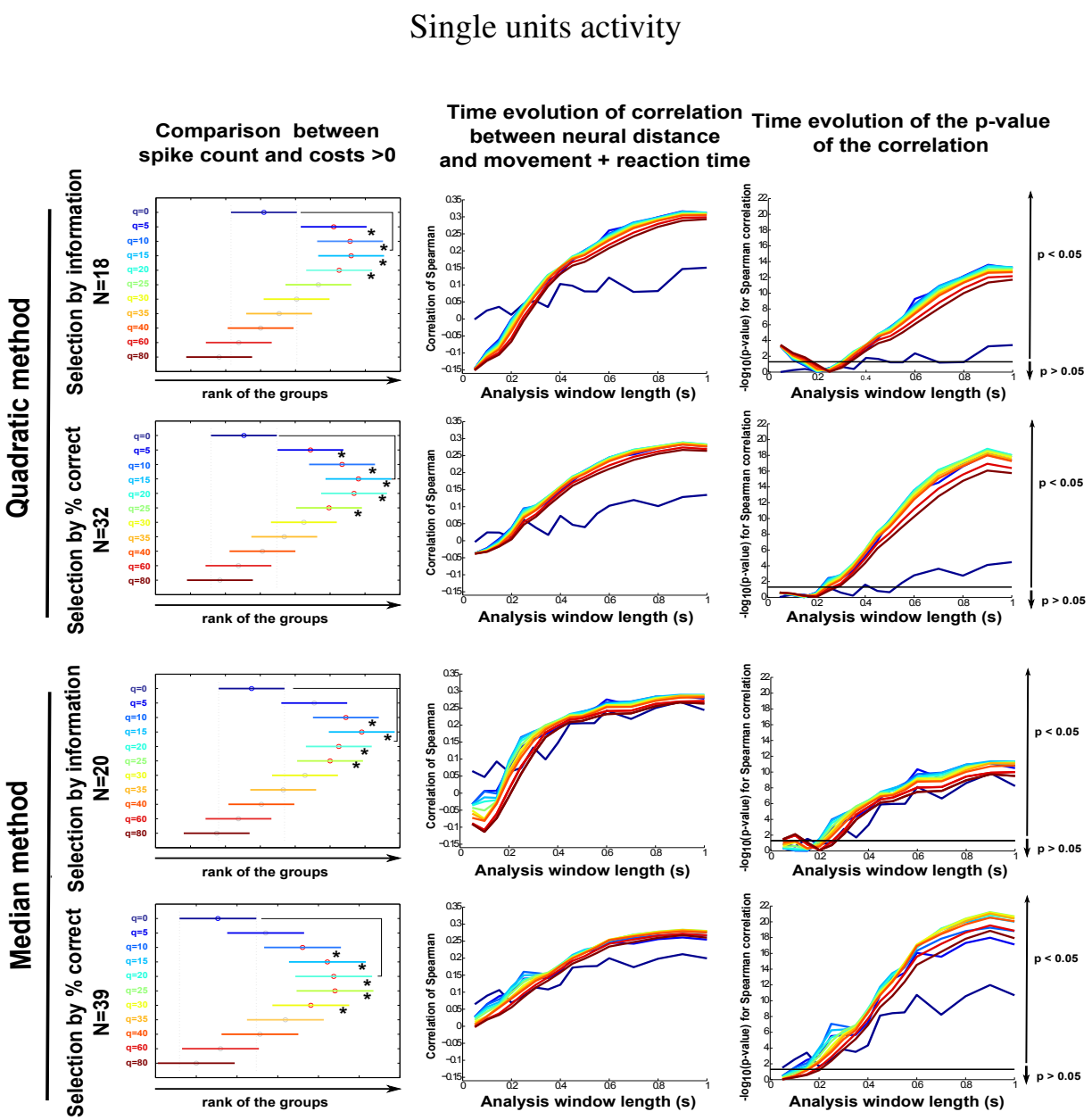

Figure 17: *Correlation between (reaction + movement) times and distance of a spike train to the "ideal first reward response", for high discrimination ability cells. The first two rows use the quadratic classification, the last couple of rows use the median classification. The correlation between distance to an evaluated 'ideal first reward' spike train and reaction + movement time was assessed by pooling all the data points from the cells that were selected as highly and consistently discriminative with the % of correct or with the information, as indicated (same groups as in figure 4, page 18). The first column shows the result of the post–hoc comparisons with Tukey's honestly significant procedure on an Anova of Friedman comparing the different costs (and removing the time effect). Stars indicate Victor and Purpura temporal costs q that were significantly different from cost q=0 (spike count based distance). Please see the appendix, section A.3 page 46, for more details on possible limits of Friedman anova. The figures are the p–values of the Friedman test on all costs. The second column shows the evolution of the correlation when distances are computed on spike trains of increasing lengths. The third column shows the evolution of* $-\log_{10}(p-value)$ *for the correlation. Colors represent different costs as indicated (unit: /s).*

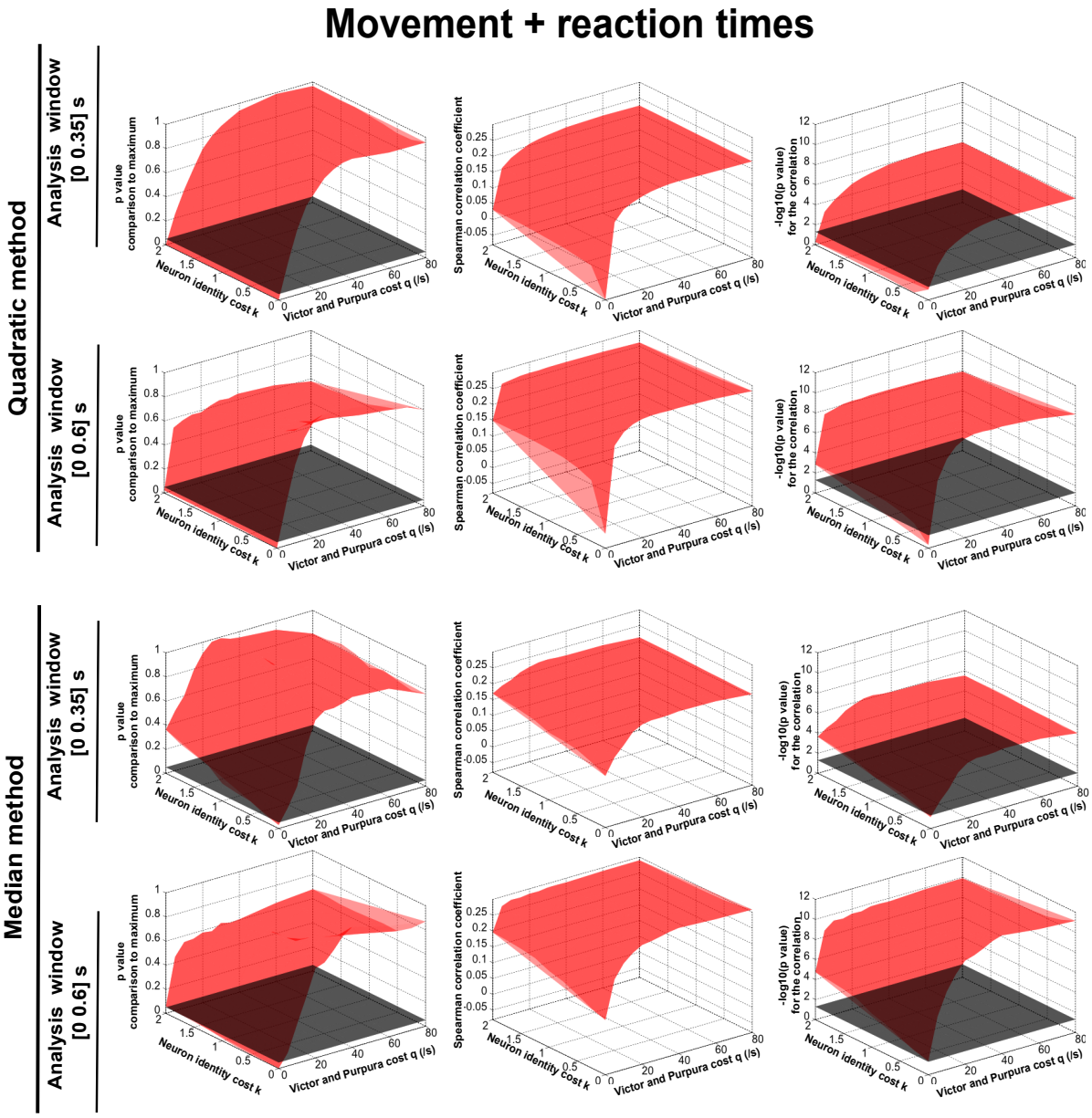

Figure 18: *Correlation between neural distance and reaction plus movement times, for the best couples of cells, with the quadratic method ($N = 19$ couples, first two rows) and the median method ($N = 20$ couples, first two rows); as a function of the temporal cost q and the neural identity cost k. First row: analysis window of [0 0.35] s; second row: analysis window of [0 0.6] s. First column: p value for the comparison of the correlation with the correlation at optimal costs. The value of 1 indicate the position of the optimal costs ; and a black flat plane at p=0.05 is drawn. Second column : value of the Spearman correlation coefficient. Third column: $-\log_{10}(pvalue)$ for a test of significance of the correlation coefficient (H_0 : the correlation coefficient is null); a black flat plane at p=0.05 is drawn.*

A.6.2 Correlation between distance to an estimated 'ideal first reward spike train' at first reward and following movement times

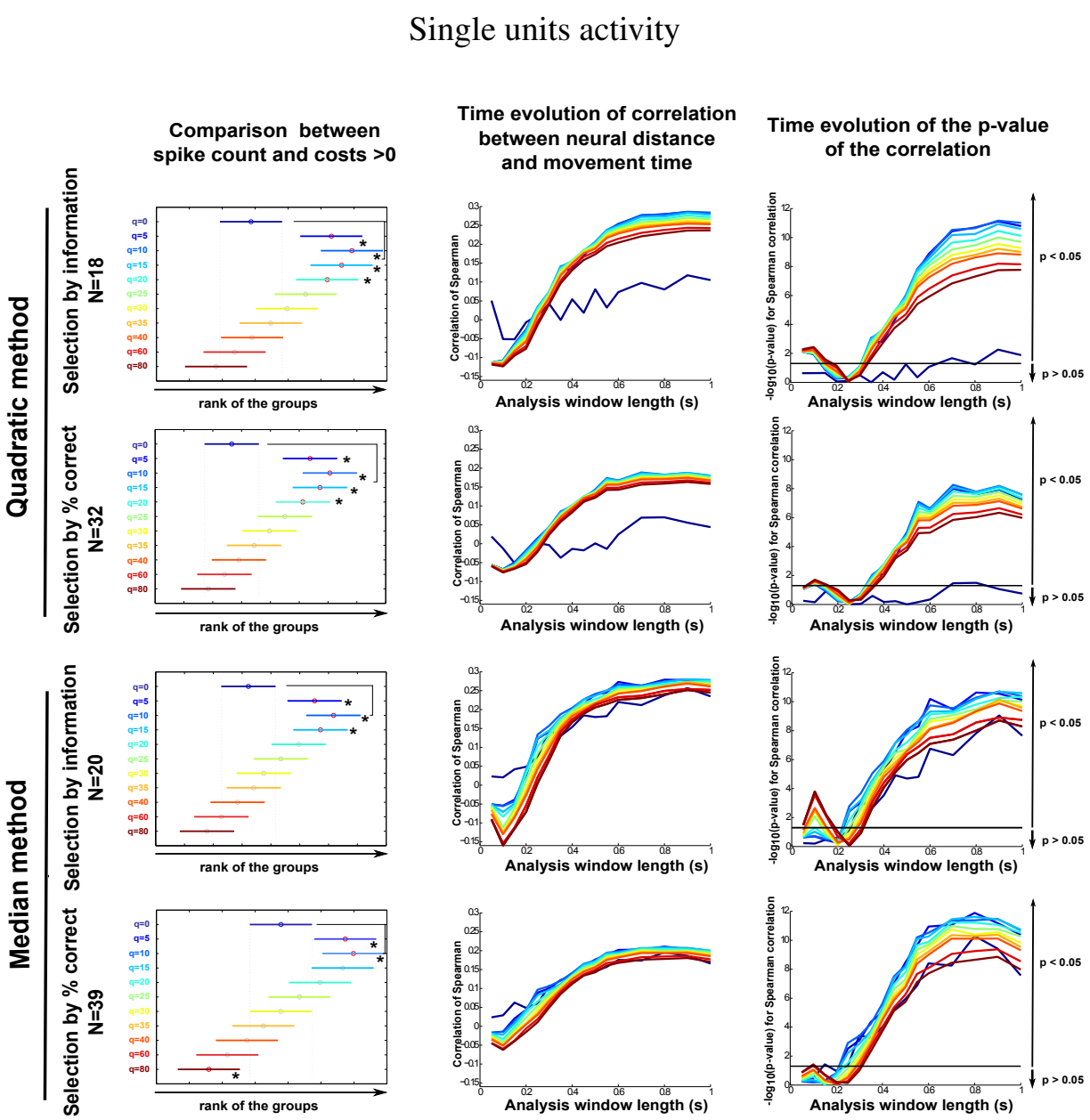

Figure 19: *Correlation between movement time and distance of a spike train to the 'ideal first reward response', for high discrimination ability cells. The first two rows use the quadratic classification, the last couple of rows use the median classification. The correlation between distance to an evaluated 'ideal first reward' spike train and movement time was assessed by pooling all the data points from the cells that were selected as highly and consistently discriminative with the % of correct or with the information, as indicated (same groups as in figure 4, page 18). The first column shows the result of the post–hoc comparisons with Tukey's honestly significant procedure on an Anova of Friedman comparing the different costs (and removing the time effect). Stars indicate Victor and Purpura temporal costs q that were significantly different from cost q=0 (spike count based distance). Please see the appendix, section A.3 page 46, for more details on possible limits of Friedman anova. The figures are the p–values of the Friedman test on all costs. The second column shows the evolution of the correlation when distances are computed on spike trains of increasing lengths. The third column shows the evolution of* $-\log_{10}(p - value)$ *for the correlation. Colors represent different costs as indicated (unit: /s).*

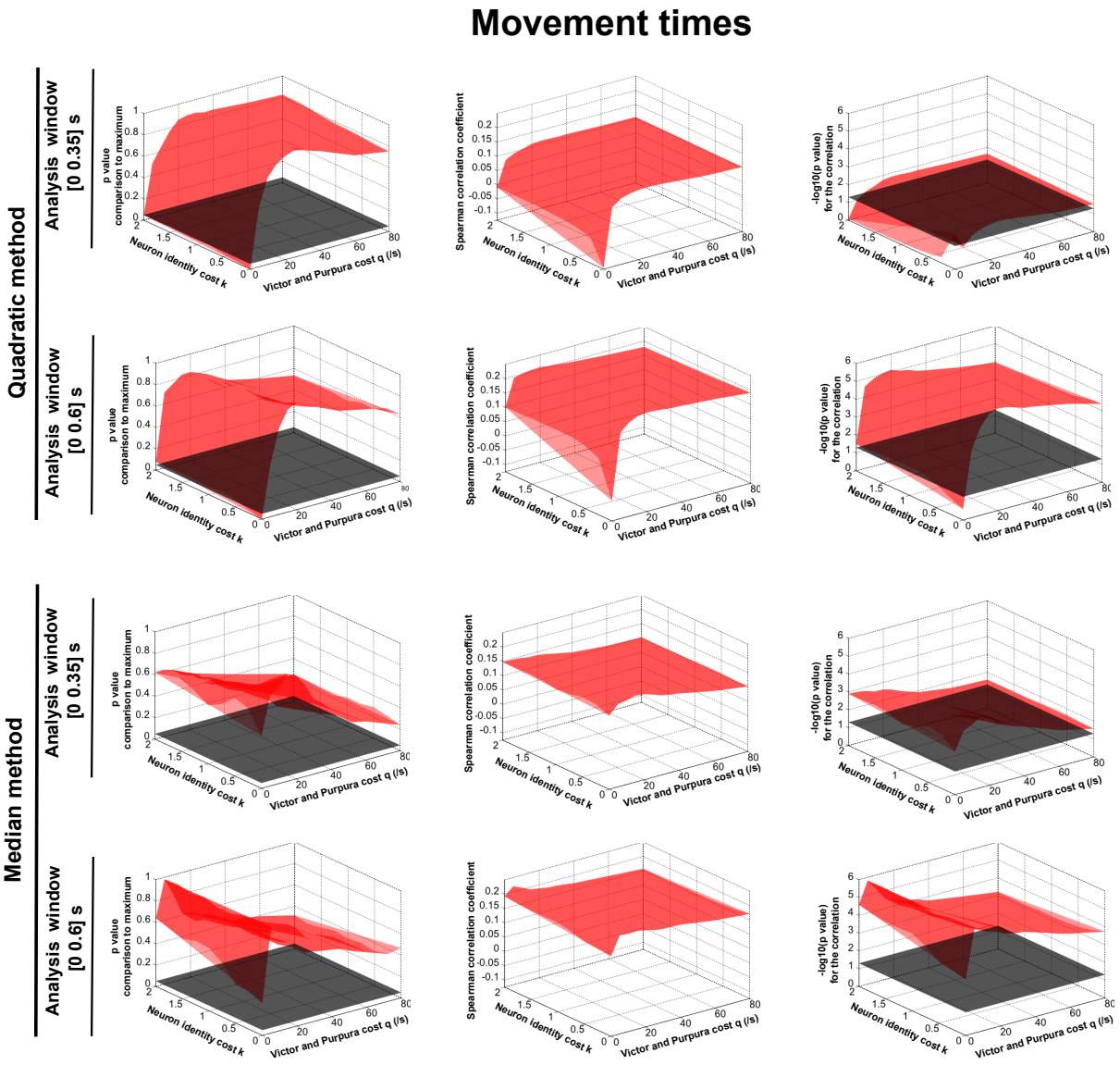

Figure 20: *Correlation between neural distance and movement times, for the best couples of cells, with the quadratic method ($N = 19$ couples, first two rows) and the median method ($N = 20$ couples, first two rows); as a function of the temporal cost q and the neural identity cost k. First row: analysis window of [0 0.35] s; second row: analysis window of [0 0.6] s. First column: p value for the comparison of the correlation with the correlation at optimal costs. The value of 1 indicate the position of the optimal costs ; and a black flat plane at p=0.05 is drawn. Second column: value of the Spearman correlation coefficient. Third column: $-\log_{10}(pvalue)$ for a test of significance of the correlation coefficient (H_0: the correlation coefficient is null); a black flat plane at p=0.05 is drawn.*

As stated in the main text, results for the movement times are far less clear than results for the reaction

times or for reaction + movement times. First, the difference between correlations at different costs are often not statistically different. Second, for the smaller analysis window and the quadratic method, correlations are rather small in absolute value, but similarly significantly negative at $(q = 0, k = 0)$, and significantly positive at (q_{opt}, k_{opt}). This strengthens the idea developed in the main text that cells which activities appear more 'switch–related' are less strongly and consistently related to movement times than to reaction times or to the total action times (reaction plus movement times).

A.7 Some subpopulations of 'bad' discriminating cells can produce a first reward activity which correlates with behavior

We also tested possible correlations between deviation of a first reward spike train from a 'canonical' first reward spike train and following response times in the following populations of cells :

- the cells which maximal (among costs and windows) discrimination values lied above the first quartile and below the third quartile of the whole population maximal discrimination values distribution ('medium' group)
- the cells which maximal (among costs and windows) discrimination values lied below the first quartile of the whole population maximal discrimination values distribution ('low' group)

The discrimination value was either the information or the percentage of correct, computed with either the quadratic or the median method.

We will only give a brief overview of the results

A.7.1 Correlations with movement times

For the 'medium' group, the correlation increased in time up to values quite similar to those of the 'best cells' (presented in section 19, page 52). For the 'low' group, the correlations were either positive or non significant depending on the methodology used.

These results show that the correlation with movement times was globally not very specific to the cells which response discriminate well between fist and subsequent rewards.

A.7.2 Correlations with reaction times

For the 'medium' group, correlations were always very weak (around 0.05), though sometimes slightly significantly positive or negative.

For the 'low' group, results were strongly dependent on the methodology used. When cells were selected thanks to their percentage of correct discrimination, with the median or quadratic method, the correlations were generally decreasing in time and becoming slightly negative. When cells were selected thanks to their information, correlations were also slightly negative with the quadratic method, whereas with the median method they were positive and rather high (around 0.18) for an analysis window of [0 0.15]s, and then decreased. Indeed, the two methods did not lead to the same subsets of 'low encoding' cells (only 20 cells over 36 were common), but differences were still present when the same cells were analyzed with median as compared to quadratic method.

This puzzling result may show that the population of 'low discriminating' cells is heterogeneous, and may be formed of subpopulations which we did not separate well when looking at the discrimination ability between first and subsequent rewards. Because they don't discriminate well between first and subsequent rewards, these cells might are not likely to be involved in switch production per se ; however, they could still correlate with behavior if they were motor or excitation related for instance. It is also possible that they would discriminate between first and subsequent reward if a different method was used.

A.8 Details on the relative influence of ranks and neural distance on reaction times

A.8.1 The response times at the second rewarded touch significantly correlate with the rank of the first reward trial

We confirmed the results of [31] and found significant correlations between the rank (i. e. the number of preceding errors) of the first reward trial and the reaction time at next touch. Results are detailed in table 14 page 55.

	Reaction time	Movement time	Reaction time + movement time
Selection by quadratic method, information	$c = 0.2170$, $p = 2.50710^{-7}$	$c = 0.3665$, $p = 4.66310^{-19}$	$c = 0.31112$, $p = 6.58310^{-14}$
Selection by quadratic method, % of correct	$c = 0.2032$, $p = 3.18110^{-10}$	$c = 0.3443$, $p = 1.38510^{-15}$	$c = 0.2904$, $p = 9.52110^{-20}$
Selection by median method, information	$c = 0.2078$, $p = 8.23710^{-6}$	$c = 0.3617$, $p = 1.543^{10^{-18}}$	$c = 0.3005$, $p = 5.30610^{-13}$
Selection by median method, % of correct	$c = 0.2006$, $p = 1.5610^{-10}$	$c = 0.3254$, $p = 9.16310^{-29}$	$c = 0.2748$, $p = 1.16110^{-19}$

Table 14: *Spearman correlation values between rank of the first correct and response times on the next target touch, for trials pooled on the same cells and trials as in 8, and their p–value*

A.8.2 It was theoretically possible to have an influence of the rank of a first reward spike train on its global distance to a category composed of 'first reward spike trains' of all ranks

This section aims at giving a proof a principle to the apparently counterintuitive following statement : even though the group of first reward spike trains is composed of equal number of trials preceded by 0, 1 or 2 errors, the number of errors can still have an impact on the distance of the spike train to the 'first reward category' (computed as in figure 6, page 20). This is a toy example ; we do not argue that the data looks like it.

To illlustrate this effect, we consider the following (oversimplified) case in which the three classes (corresponding to 0, 1 or 2 errors preceding the trial) are each composed of the three points of aligned equilateral triangles (figure 21, page 56).

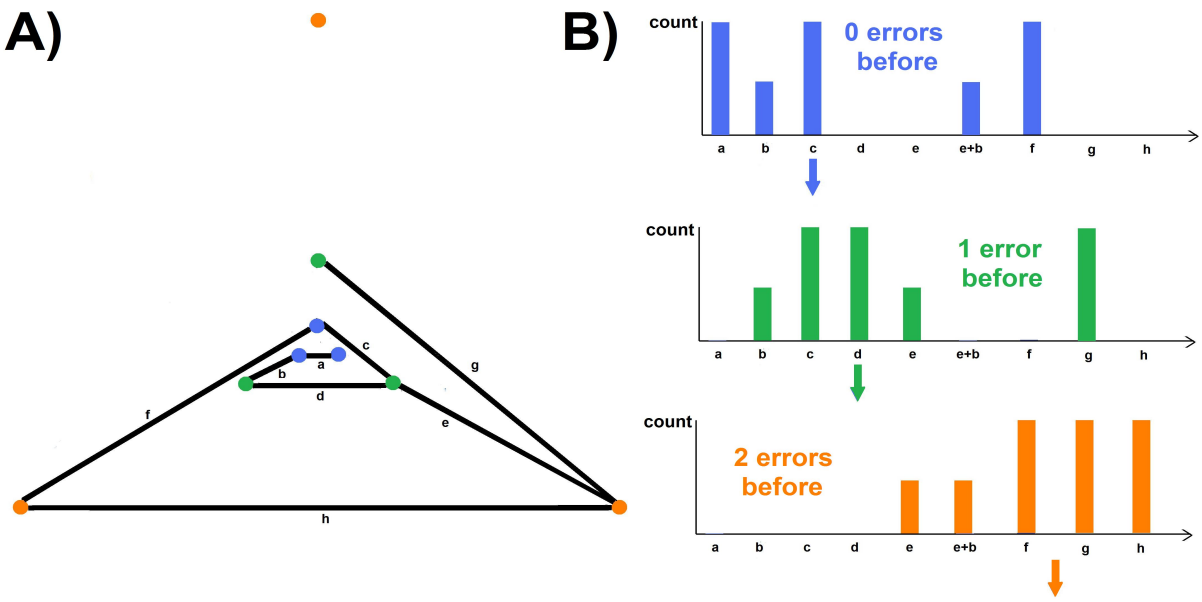

Figure 21: *A)Three different subcategories, each one composed of three points forming the angles an equilateral triangle, are considered in one category. The three triangles are aligned. B) Repartition of the pairwise distances between any "spike train" of the category of the corresponding color, and all other spike trains in the mixture group. The arrow represents the median. As can be seen, the value of this median is consistently influenced by the category to which the "spike train" belongs.*

For quadratic classification, when taking the limit case in which the distance of one spike train to the ensemble of other spike trains is the distance to the nearest neighbor, it is very apparent that for spike train of the blue group, this distance (a) is smaller than the distance for those of the green group (b), which is itself smaller than the distance for those of the orange group (e).

For the median classification, the effect of the group is still visible, as illustrated by the histograms of the pairwise distances between spike trains of a given group, and all other spike trains (figure 21 B , see the monotonic increase of the median, as indicated by the arrow).

A.8.3 Ranks and distance to an ideal 'first reward spike train' could slightly correlate, but at different costs and with a different timing than distance and behavior

Small (<0.13) but significant correlations between rank of the first reward and distance of a 'first reward spike train' to all other 'first reward spike trains' were indeed found, but generally for Victor and Purpura temporal costs $q \in [0, 5, 10]$ and small analyses windows (<0.3 s), so at different costs and times than those at which the correlation between behavior and neural distance is maximized. The difference between the maximal correlation rank/distance and response times/distance was however always significant or a tendency ($p < 0.1$), as stated in the main text.

A.8.4 At the cost effective to correlate with behavior, there was no correlation between ranks and our definition of first reward spike train distance

Results showing that when decoded in a way that maximizes the correlation between neural distances and behavior, the first reward rank had no significant impact on the neural distance are detailed in table 15, page 57.

	Best correlation between (reaction, movement, sum) times and distances	Correlation between ranks and distances (same cost and analysis window length) and its significance	p–values for comparison of the correlations (permutation test)
Selection by quadratic method, information	$c = (0.2715, 0.2860, 0.3161)$	$c = (0.0060, 0.0309, 0.0326)$, $p = (0.8875, 0.4685, 0.4431)$	$p = (< 10^{-4}, < 10^{-4}, < 10^{-4})$
Selection by quadratic method, % of correct	$c = (0.2717, 0.1887, 2890)$	$c = (0.0370, 0.0290, 0.0400)$, $p = (0.2568, 0.3747, 0.2202)$	$p = (< 10^{-4}, 6\ 10^{-4}, < 10^{-4})$
Selection by median method, information	$c = (0.2419, 0.2802, 0.2890)$	$c = (0.0106, 0.0446, 0.0355)$, $p = (0.8044, 0.2947, 0.4045)$	$p = (< 10^{-4}, p < 10^{-4}, 2\ 10^{4})$
Selection by median method, % of correct	$c = (0.2688, 0.2109, 0.2836)$	$c = (0.0243, 0.0360, 0.0271)$, $p = (0.4181, 0.2313, 0.3681)$	$p = (< 10^{-4}, < 10^{-4}, < 10^{-4})$

Table 15: *Spearman correlation values on trials pooled on the same cells and trials as in 8, between distances and ranks and between distances and the response times, as indicated. The significance of the difference between correlation is computed thanks to a two sided permutation test.*

A.8.5 With the exception of movement times, ranks and neural distance correlated equally well with behavior

	Best correlation between (reaction, movement, sum) times and distances	Correlation between ranks and response times	p-values for comparison of the correlations (permutation test)
Selection by quadratic method, information	$c = (0.2715, 0.2860, 0.3161)$	$c = (0.2170, 0.3665, 0.3112)$	$p = (0.5830, 0.1680, 0.8510)$
Selection by quadratic method, % of correct	$c = (0.2717, 0.1887, 0.2890)$	$c = (0.2032, 0.3443, 0.2904)$	$p = (0.0910, 0.0050, 0.8390)$
Selection by median method, information	$c = (0.2419, 0.2802, 0.2890)$	$c = (0.2078, 0.3617, 0.3005)$	$p = (0.3470, 0.1800, 0.9350)$
Selection by median method, % of correct	$c = (0.2688, 0.2109, 0.2836)$	$c = (0.2006, 0.3254, 0.2904)$	$p = (0.1330, < 10^{-3}, 0.9770)$

Table 16: *Spearman correlation values on trials pooled on the same cells and trials as in 8, for different response times and two possible explanatory variables : the neural distance and the rank of the first correct. The significance of the difference between correlation is computed thanks to a two sided permutation test.*

We compared the maximal correlation between neural distance to an ideal 'first reward spike train' and following response times, versus the correlation between rank of the first reward and following response times. Results are presented in table 16, page 57. It can be seen that ranks and neural distances correlated similarly with reaction times and with (reaction + movement) times, whereas ranks correlated better than neural distances with isolated movement times for the less stringent selection of neurons (with percentage of correct).

A.9 Cells which discriminate better between first and subsequent reward are probably also involved at other stages of the task, but probably less strongly

A.9.1 There is a medium correlation between discrimination ability between first and subsequent reward and discrimination ability between beginning of a new problem vs beginning of a new exploitation trial

A previous study (Quilodran et al [31], their figure 6) had shown that in a subset of cells showing modulations of the firing rates in the first vs subsequent rewards, there was also a modulation of activity at the beginning of each trial, when the monkey had to come back to the same target, but not at the beginning of a new problem, after the four rewards had been delivered. To quantify how robust this difference of modulation was in the whole ACC population, we aligned all spike trains at the time of reward delivery, and began our analysis windows 1.6 seconds after reward delivery, just before the modulation of activity that had been observed in Quilodran et al. Analysis windows were increased up to 3.6 s by 0.2 s. We contrasted the activity occurring after the 1st, 2nd and 3rd rewards (produced before the monkey comes back to the learned rewarded target, category 1), with the activity produced after the last 4th reward (before and during the beginning of a new problem, category 2). Problems with more than four rewards were discarded. It should be emphasized that the external physical events occurring are no longer the same between the two situations. In category 1, a lever touch appears on the screen around 2 seconds after the reward, followed by a fixation point. In category 2, a signal to change indicating the start of a new problem appears around 2.5 seconds after the reward.

ACC single cells also discriminate between the continuation of a problem and the beginning of a new problem The same strategy as the one of 'first-reward analysis' was followed : a k–means algorithm on the maximum discrimination value (among all costs and analyses windows) was used to separate the cells into a group of low and a group of high discrimination ability cells. Finally, the selection was further refined by imposing that there exists a cost for which the discrimination was superior to the 95th percentile of the permuted data, for at least 5 consecutive analysis windows. The results are presented in figure 22, page 60 for the 'best cells'.

Even though the choice of the reward time as the temporal reference is somewhat more arbitrary at this moment of the task, slightly taking into account temporal structure of the spike trains still improved the classification as it significantly increased for $q = 5$ as compared to $q = 0$; however, the 'optimal' temporal accuracy appeared lower than at the reward time, because the classification tended to be worst at $q = 10$ as compared to $q = 5$ (see table 17, page 59), whereas the opposite tendency was found at the reward time.

The correlation between discriminability power at the reward time and at the beginning of trials is significant but small We tested whether the cells which discriminated well between first and subsequent rewards were also these which discriminated well between the beginning of an exploitation trial and the beginning of the first exploration trial, by computing the correlation between the maximum discrimination measures (information or % of correct, with the median or quadratic method) among all costs and analyses windows.

For all ACC single units, the correlation was rather weak but significant (table 18, page 59, first line).

Moreover, table 19 page 61 reports the proportion of cells in the 'well discriminating groups' at the reward time that were also in the'well discriminating groups' at the beginning of trials. These proportions are arguably small, though generally significantly superior to the proportions of 'well discriminating cells' at the beginning

		quadratic method, information	quadratic method, % correct	median method, information	median method, % correct
All costs	All cells	0	0	0	0
	1 cell/session	0	0	0	0
$q = 5 > q = 0$	All cells	0	0	$9.6\ 10^{-6}$	0.0081
	1 cell/session	0	0	$4.3\ 10^{-6}$	0.0341
$q = 5 > q = 10$	All cells	0.0678	0.0006	0.0391	$7.3\ 10^{-6}$
	1 cell/session	0.0965	0.001	0.0747	0.0004

Table 17: *Results for the Friedman anova for an effect of the costs, for groups of well-discriminating cells (selected thanks to the classification method/measure indicated in the columns) at the approximate moment of the beginning of the trial.*

	quadratic method, information	quadratic method, % correct	median method, information	median method, % correct
All 145 cells	$c = 0.4923$; $p = 4.65\ 10^{-10}$	$c = 0.3806$; $p = 2.33\ 10^{-6}$	$c = 0.3703$; $p = 4.55\ 10^{-6}$	$c = 0.3808$; $p = 2.29\ 10^{-6}$
Best cells at reward time for each (measure, method)	$c = 0.0506$; $p = 0.84$	$c = 0.0165$; $p = 0.9286$	$c = 0.0301$; $p = 0.9013$	$c = 0.2787$; $p = 0.0859$

Table 18: *Correlation between the maximal discrimination ability value at the reward time, and the maximum discrimination ability value at the trial start moment, within groups of best discriminating cells selected by the classification method/discrimination measure indicated by the column.*

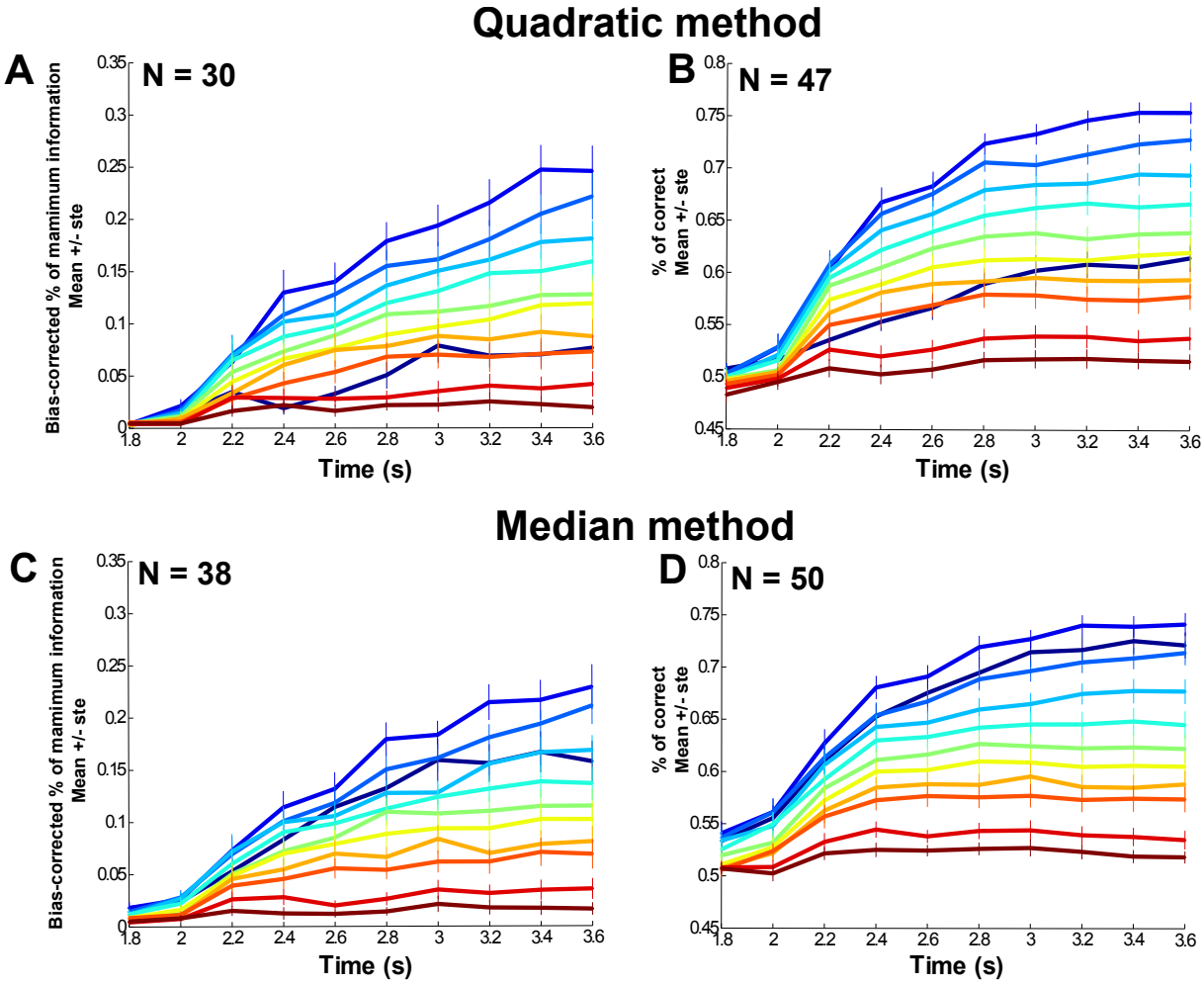

Figure 22: *A) Information with quadratic method; B) Percentage of correct with quadratic method; C) Information with median method; D) Percentage of correct with median method. Curves represent the mean discriminability among a subset of cells with high and consistent discrimination abilities. Bars represent standard errors. The different colors represents different Victor and Purpura temporal costs q, as indicated in the legend (unit: / s). The figures on top left of the graphs are the number of units used. Please note that all analyses windows start 1.6 s after the reward and end at the time indicated on the x axis.*

of trials among the whole ACC single units population.

Therefore, a 'reward time well discriminating cell' had a higher probability to be also a 'beginning of trials well discriminating cell' than a cell taken at random in the whole ACC single units population.

Finally, among the populations of 'best cells' at the reward time, there was no significant correlation between the discrimination ability at the reward time and at the trial start moment, suggesting that some cells which were rather exclusively 'good' at the reward time existed (table 18, page 59, second line).

Globally, the results show that the populations of cells discriminating well at the reward time and at the beginning of trials are only partly overlapping, and rather argue for a 'specialization' of at least some cells in one or the other encoding.

	quadratic method, information	quadratic method, % correct	median method, information	median method, % correct
$\frac{N_{best\ reward\ time\ \cap best\ trial\ start}}{N_{best\ reward\ time}}$	$\frac{9}{18}$	$\frac{20}{32}$	$\frac{10}{20}$	$\frac{20}{39}$
$\frac{N_{best\ trial\ start}}{N_{tot}}$	$\frac{30}{145}$	$\frac{47}{145}$	$\frac{38}{145}$	$\frac{50}{145}$
p value comparison of proportions	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p > 0.05$

Table 19: *Comparison of the proportion of cells with good discrimination ability at the two moments among the 'reward time best discriminating cells', and the proportion of cells with good trial start discrimination among the whole ACC population.Results are the output of the tmcomptest function of MATLAB*

A.9.2 There was a significant but arguably smaller correlation between 2^{nd} reward activity and responses times at the third touch

On the second and third reward, the monkey receives the confirmation that it chose the good target and that it has to go on touching the same target. If the ACC cells are generally implied each time the animal maintains its exploitation policy, then their activity at the second reward time should correlate with their behavior at the third touch. If, on the contrary, ACC single units are really specialized in the production of a behavioral switch from exploration to exploitation, then no or little correlation is expected after the first reward. We will focus on the most selective groups of cells : these that have been selected thanks to their information measure. As a preliminary analysis, we verified that the distributions of behavioral times were statistically indistinguishable between the second and third rewarded touch, which controlled that differences in correlations were unlikely to be due to differences in behavioral variability, for instance. Results are presented in table 20 ; page 62.

		$2^{nd}touch$	$3^{rd}touch$	p value comparison (see legend)
same trials and cells as 'well discriminating' with median information group	reaction times	$median = 0.1729s$; $std = 0.5135s$	$median = 0.1669s$; $std = 0.2901s$	$p_{ks} = 0.6675$; $p_{rksum} = 0.8288$; $p_{ansbdl} = 0.5851$
	movement times	$median = 0.206s$; $std = 0.8598s$	$median = 0.204s$; $std = 0.4291s$	$p_{ks} = 0.5851$; $p_{rksum} = 0.2881$; $p_{ansbdl} = 0.8372$
same trials and cells as 'well discriminating' with quadratic information group	reaction times	$median = 0.1702s$; $std = 0.5136s$	$median = 0.1691s$; $std = 0.2899s$	$p_{ks} = 0.8373$; $p_{rksum} = 0.8376$; $p_{ansbdl} = 0.9974$
	movement times	$median = 0.206s$; $std = 0.8591s$	$median = 0.204s$; $std = 0.4833s$	$p_{ks} = 0.967$; $p_{rksum} = 0.2282$; $p_{ansbdl} = 0.9475$

Table 20: *Comparison between second and third touch of the reaction times and the movement times for the cells and trials selected with the median and quadratic method, for the information measure. Medians and standard deviations (std) are reported, as well as the p values for a two sample Kolmogorov Smirnoff test (kstest2) comparing the global distribution shapes (p_{ks}), a rank sum test comparing more specifically the medians (p_{rksum}); and an Ansari Bradley test which compares the dispersion (e g variance) of the distributions.*

The data show positive significant correlations between the deviation of the second reward spike train to an 'ideal' second reward spike train, and reaction times, or with the sum reaction + movement times.

For movement times, positive correlations could also be found for long analyses windows, but stronger negative correlations were found on small analyses windows. This means that reaction times and movement times were not consistent anymore (indeed, the neural distances are the same between the two groups). Future research will have to investigate further this question. A hypothesis may be that, once the monkey is in the middle of its exploitation phase, reaction times and movement times taken separately may not reflect the confidence of the monkey into its choice; instead, when the reaction time is low, the monkey might 'compensate' with a longer movement time.

Moreover, in any case, the positive correlations were smaller than at the first reward moment. We finally tested the significance of this smaller correlation, by comparing the maximal (over analyses windows and costs) positive correlation distances vs response times, between the first and second reward, thanks to a permutation test (see methods, section 2.5, page 14). We can see in table 21 page 63 that the difference failed to reach significance for the median method and the reaction times, but was significant for the three other groups.

1 : spike trains at the first reward moment ; response times at the second touch

2 : spike trains at the second reward moment ; responses times at the third touch

	reaction times : 1 vs 2	reaction + movement times : 1 vs 2
same trials and cells as 'well discriminating' with median information group	$c = 0.2419$ vs $c = 0.1614$; $p = 0.1670$	$c = 0.2890$ vs $c = 0.1571$; $p = 0.0212$
same trials and cells as 'well discriminating' with quadratic information group	$c = 0.2715$ vs $c = 0.1551$; $p = 0.0405$	$c = 0.3161$ vs $c = 0.1505$; $p = 0.0023$

Table 21: *Maximal (over windows and costs) correlations between reward activity and behavioral times at next touch, for the first reward (first column) and the second reward (second column), along with the p–value of a permutation comparison test between first and second reward.*

Taken together, these results suggest that at least the group of 'best cells' relative to information at the first reward correlated differently and less well with behavior at the second reward. This suggests that at least some cells have an activity which correlates better with behavior at the moment of the behavioral switch, and which activity may be functionally more important at the moment of the behavioral switch.

A.10 Relative effects of k and q on the discrimination ability of couples

A.10.1 Comparison of information for best couples between the optimal costs and other costs

We report here (table 22, page 64) the results for the comparison of the information distributions among 'best couples of cells', for different outstanding couples of costs.

	(q_{opt}, k_{opt})	$(q = 0, k = 0)$	$(q_{opt}, k = 0)$	$(q = 0, k_{opt})$	$(q_{opt\ at\ k=0}$, $k = 0)$	$(q = 0$, $k_{opt\ at\ q=0})$
Selection with information, median method , [0 0.6] s	$I_{max} = 0.2883$; $q_{opt} = 15/s$; $k_{opt} = 1.25$	$p = 2.19\ 10^{-4}$	$p = 0.0239$	$p = 0.0036$	$p = 0.0499$; $q_{opt\ at\ k=0} = 25$	$p = 0.0438$; $k_{opt\ at\ q=0} = 2$
Selection with information, quadratic method , [0 0.6] s	$I_{max} = 0.3162$; $q_{opt} = 15/s$; $k_{opt} = 1.5$	$p = 5.0978\ 10^{-7}$	$p = 0.0285$	$p = 1.8327\ 10^{-4}$	$p = 0.0285$; $q_{opt\ at\ k=0} = 15$	$p = 2.0845\ 10^{-4}$; $k_{opt\ at\ q=0} = 0.5$
Selection with information, median method , [0 0.35] s	$I_{max} = 0.2706$; $q_{opt} = 15/s$; $k_{opt} = 1$	$p = 0.006$	$p = 0.0438$	$p = 0.0128$	$p = 0.0834$; $q_{opt\ at\ k=0} = 10$	$p = 0.1404$; $k_{opt\ at\ q=0} = 1.75$
Selection with information, quadratic method , [0 0.35] s	$I_{max} = 0.2751$; $q_{opt} = 25/s$; $k_{opt} = 1$	$p = 2.0198\ 10^{-6}$	$p = 0.0702$	$p = 3.2699\ 10^{-4}$	$p = 0.1290$; $q_{opt\ at\ k=0} = 20$	$p = 4.5611\ 10^{-4}$; $k_{opt\ at\ q=0} = 1.75$

Table 22: *Summary of the comparison of median information among well-discriminating couples of cells. The second column shows the maximal information, as well as the optimal costs q and k. The other columns give the result of the rank sum test between the optimum and the median information at indicated costs.*

A.10.2 Influence of the identity cost k on the classification of first vs subsequent reward

To better understand the influence of the identity cost k on the classification, we looked at how the change from k=0 to $k = k_{opt}$ improved the classification for best couples for the longer analysis window [0 0.6] s at which the differences were clearer. For practical reasons, this was done on the distances that were extracted for the behavioral times analysis and excluded a few data points compared to the information analysis (those for which one of the response time could not be retrieved). The results are presented in table 23, page 64 at $q = q_{opt}$, and in table 24, page 65 at $q = 0$. We can see that at optimal temporal cost, the main effect was to improve the classification for the 'subsequent reward category', whereas the improvement was more balanced between categories at cost $q = 0$.

	(q_{opt}, k_{opt})		$(q_{opt}, 0)$	
	$1^{st} reward$ s	other rewards	$1^{st} reward$	other rewards
% correct ; median method	$\frac{302}{473}$	$\frac{1482}{1583}$	$\frac{295}{473}$	$\frac{1386}{1583}$
% correct ; quadratic method	$\frac{313}{458}$	$\frac{1433}{1544}$	$\frac{299}{458}$	$\frac{1348}{1544}$

Table 23: *Improvement of the number of correct classification at optimal temporal cost q when neural identity is optimally weighted (k_{opt}) as compared to when it is not ($k = 0$).*

	$(q=0,\, k_{opt\ at\ q=0})$		$(q=0, k=0)$	
	$1^{st}reward$ s	other rewards	$1^{st}reward$	other rewards
% correct ; median method	$\frac{336.5}{473}$	$\frac{1222}{1583}$	$\frac{327.5}{473}$	$\frac{1206}{1583}$
% correct ; quadratic method	$\frac{281.5}{458}$	$\frac{1210.5}{1544}$	$\frac{239}{458}$	$\frac{1013}{1544}$

Table 24: *Improvement of the number of correct classification for spike count ($q = 0$), when neural identity is optimally weighted ($k_{opt\ at\ q=0}$) as compared to when it is not ($k = 0$).*

Adding a neural identity cost will either increase the distance between spike trains, or keep it unchanged if changing neural identity would never reduce the distance.

We could find evidence that for the optimal temporal cost q_{opt}, the improved classification for the 'subsequent rewards' spike train was due to a higher global increase in the distance from a 'subsequent reward' spike train to the 'first reward' category, as compared to a smaller increase in the distance from a 'subsequent reward' spike train to the 'subsequent reward' category (see the two first rows of table 25, page 65). Such global effects were not observed or were substantially smaller for first reward spike train or for spike count classification (table 26, page 66). This suggests that for spike count based classification, the improvement due to increasing k occurred in a more subtle, case by case way.

$$A = [(distance\ to\ category\ `1^{st}\ reward')]_{(q_{opt},k_{opt})} - [(distance\ to\ category\ `1^{st}\ reward')]_{(q_{opt},k=0)} \quad (15)$$

$$B = [(distance\ to\ category\ `other\ rewards')]_{(q_{opt},k_{opt})} - [(distance\ to\ category\ `other\ rewards')]_{(q_{opt},k=0)} \quad (16)$$

$$C = B - A \quad (17)$$

	median(A)	median(B)	median(B-A) ; p value of signtest
'other rewards' spike trains, median method	1.6269	1.1985	-0.3021 ; $p = 6.235\ 10^{-36}$
'other rewards' spike trains, quadratic method	1.6446	1.1120	-0.379 ; $p = 3.1605\ 10^{-59}$
'1_{st} reward' spike trains, median method	1.6996	1.7517	0 ; $p = 0.8164$
'1_{st} reward' spike trains, quadratic method	1.5945	1.7246	0.0266 ; $p = 0.3265$

Table 25: $q = q_{opt}$*: Increase in the distance of spike trains coming from categories indicated in the rows to categories indicated in the column, and comparison of the increase to category 'other reward' vs to category 'first reward', when k is increased from 0 to* k_{opt}.

$$A = [(distance\ to\ category\ `1^{st}\ reward')]_{(q=0,k_{opt\ at\ q=0})} - [(distance\ to\ category\ `1^{st} reward')]_{(q=0,k=0)} \quad (18)$$

$$B = [(distance\ to\ category\ `other\ rewards')]_{(q=0,k_{opt\ at\ q=0})} - [(distance\ to\ category\ `other\ rewards')]_{(q=0,k=0)} \quad (19)$$

$$C = B - A \quad (20)$$

	median(A)	median(B)	median(B-A) ; p value of signtest
'other rewards' spike trains, median method	0	0	0 ; $p = 0.6301$
'other rewards' spike trains, quadratic method	0.8664	0	-0.0973 ; $p = 2.6338\ 10^{-65}$
'1_{st} reward' spike trains, median method	0	0	0 ; $p = 0.4437$
'1_{st} reward' spike trains, quadratic method	1.1003	0.804	0 ; $p = 0.0984$

Table 26: $q = 0$*: Increase in the distance of spike trains coming from categories indicated in the rows to categories indicated in the column, and comparison of the increase to category 'other rewards' vs to category 'first reward' when k is increased from 0 to* $k_{opt\ at\ q=0}$.

A final analysis was conducted to try to understand why, at best temporal cost q, the increase in the distance was so small when k was increased for the intra–category 'subsequent rewards' distance. The number of spikes was smaller in this category (table 27, page 66), which probably decreased the likelihood that two spikes from two different neurons would be close enough to be matched when comparing two spike trains from this category. In contrast, this likelihood increased when comparing a spike train from the 'subsequent reward' category to a spike train from the 'first reward' category, which contains more spikes; as a consequence, weighting neural identity can have a stronger impact on the distance. Finally, when comparing two first reward spike trains, as the number of spikes is important, there would be a big probability of 'mixing spikes' by chance, but as the firing of the 'best cell' is supposed to be rather precise and as the timing of spikes is weighted, it is probable that even at $k = 0$ the responses would not be very mixed, resulting in a medium increase in the distance for higher ks.

	'best cells'	'worst cells'
'other rewards' spike trains, median method	$median = 4.6526$; $mean = 4.4876$	$median = 2.1297$; $mean = 3.5173$
'other rewards' spike trains, quadratic method	$median = 4.0552$; $mean = 4.6817$	$median = 1.5128$; $mean = 3.4956$
'1_{st} reward' spike trains, median method	$median = 6.5364$; $mean = 8.0301$	$median = 2.7222$; $mean = 4.1435$
'1_{st} reward' spike trains, quadratic method	$median = 6.2727$; $mean = 7.9969$	$median = 2.3333$; $mean = 4.0582$

Table 27: *For the longer analysis window, spike mean and median spike count reported separately for 'best discriminating cells', 'worst discriminating cells', and for the two categories.*

References

[1] L. F. Abbott, J. A. Varela, Kamal Sen, and S. B. Nelson. Synaptic depression and cortical gain control. *Science*, 275(5297):221–224, 1997.

[2] Larry F. Abbott and Sacha B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183, 2000.

[3] C. Amiez, J.P. Joseph, and E. Procyk. Reward encoding in the monkey anterior cingulate cortex. *Cerebral Cortex*, 16(7):1040–1055, July 2006.

[4] D. Aronov, D. S. Reich, F. Mechler, and J. D. Victor. Neural coding of spaatial phase in v1 of the macaque monkey. *Journal of Neurophysiology*, 89:3304–3327, 2003.

[5] Bruno B. Averbeck and Daeyeol Lee. Neural noise and movement-related codes in the macaque supplementary motor area. *The Journal of Neuroscience*, 23(20):7630–7641, 2003.

[6] W Bialek, F Rieke, RR de Ruyter van Steveninck, and D Warland. Reading a neural code. *Science*, 252(5014):1854–1857, 1991.

[7] Romain Brasselet. Neural coding in the ascending somatosensory pathway : A metrical information theory approach. *PhD Thesis*, pages 166 – 170, 2010.

[8] Patricia M. Di Lorenzo, Jen-Yung Chen, and Jonathan D. Victor. Quality time: Representation of a multidimensional sensory domain through temporal coding. *The Journal of Neuroscience*, 29(29):9227–9238, 2009.

[9] Christopher D. Fiorillo, Philippe N. Tobler, and Wolfram Schultz. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614):1898–1902, 2003.

[10] Stan B. Floresco, Annie E. Block, and Maric T.L. Tse. Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure. *Behavioural Brain Research*, 190(1):85 – 96, 2008.

[11] Benjamin Y. Hayden, Sarah R. Heilbronner, John M. Pearson, and Michael L. Platt. Surprise signals in anterior cingulate cortex: Neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *The Journal of Neuroscience*, 31(11):4178–4187, 2011.

[12] Benjamin Y. Hayden and Michael L. Platt. Neurons in anterior cingulate cortex multiplex information about reward and action. *The Journal of Neuroscience*, 30(9):3339–3346, 2010.

[13] Adrian Hernandez, Veronica Nacher, Rogelio Luna, Luis Lemus, Manuel Alvarez, Yuriria Vasquez, Liliana Camarillo, and Ranulfo Romo. Decoding a perceptual decision process across cortex. *Neuron*, 66(2):300 – 314, 2010.

[14] J. C. Horvitz. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96(4):651 – 656, 2000.

[15] Shigehiko Ito, Veit Stuphorn, Joshua W. Brown, and Jeffrey D. Schall. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*, 302(5642):120–122, 2003.

[16] R. S. Johansson and I. Birznieks. First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Nature Neuroscience*, 7:170–177, 2004.

[17] Steven W. Kennerley, Mark E. Walton, Timothy E. J. Behrens, Mark J. Buckley, and Matthew F. S. Rushworth. Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, 9:940–947, 2006.

[18] Peter König, Andreas K. Engel, and Wolf Singer. Integrator or coincidence detector? the role of the cortical neuron revisited. *Trends in Neurosciences*, 19(4):130 – 137, 1996.

[19] Christopher C. Lapish, Daniel Durstewitz, L. Judson Chandler, and Jeremy K. Seamans. Successful choice behavior is associated with distinct and coherent network states in anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, 105(33):11963–11968, 2008.

[20] Peter E. Latham and Sheila Nirenberg. Synergy, redundancy, and independence in population codes, revisited. *The Journal of Neuroscience*, 25(21):5195–5206, 2005.

[21] Michael London, Arnd Roth, Lisa Beeren, Michael Häusser, and Peter E. Latham. Sensitivity to pertubations in vivo implies high noise and suggests rate coding in the cortex. *Nature*, 466:123–127, 2010.

[22] R. Luna, A. Hernandez, C. Brody, and R. Romo. Neural codes for perceptual discrimination in primary sensory cortex. *Nature Neuroscience*, 8:1210–1219, 2005.

[23] Christian K. Machens, Hartmut Schütze, Astrid Franz, Olga Kolesnikova, Martin B. Stemmler, Bernhard Ronacher, and Andreas V. M. Herz. Single auditory neurons rapidly discriminate conspecific communication signals. *Nature Neuroscience*, 6:341–342, 2003.

[24] Katrina MacLeod, Alex Bäcker, and Gilles Laurent. Who reads temporal information contained across synchronized and oscillatory spike trains ? *Nature*, 395:693–698, 1998.

[25] António Paiva, Il Park, and José Príncipe. A comparison of binless spike train measures. *Neural Computing amp; Applications*, 19:405–419, 2010. 10.1007/s00521-009-0307-6.

[26] S. Panzeri, N. Brunel, N. K. Logothetis, and C. Kayser. Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, 33(3):111–120, 2009.

[27] Stefano Panzeri and Alessandro Treves. Analytical estimates of limited sampling biases in different information measures. 1995.

[28] Thomas Paus. Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews Neuroscience*, 2:417–424, 2001.

[29] Emmmanuel Procyk, Y. L. Tanaka, and Joseph J. P. Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nature Neuroscience*, 3:502–598, 2000.

[30] Gopathy Purushothaman and David C. Bradley. Neural noise and movement-related codes in the macaque supplementary motor area. *Nature Neuroscience*, 8(1):99–106, 2004.

[31] René Quilodran, Marie Rothé, and Emmanuel. Procyk. Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron*, 57(2):314–325, 2008.

[32] Daniel S. Reich, Ferenc Mechler, and Jonathan D. Victor. Independent and redundant information in nearby cortical neurons. *Science*, 294(5551):2566–2568, 2001.

[33] Hannes P. Saal, Sethu Vijayakumar, and Roland S. Johansson. Information about complex fingertip parameters in individual human tactile afferent neurons. *The Journal of Neuroscience*, 29(25):8022–8031, 2009.

[34] S. Schreiber, J.M. Fellous, D. Whitmer, P. Tiesinga, and T.J. Sejnowski. A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52-54:925 – 931, 2003. Computational Neuroscience: Trends in Research 2003.

[35] Wolfram Schultz, Peter Dayan, and P. Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.

[36] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[37] Munetaka Shidara and Barry J. Richmond. Anterior cingulate: Single neuronal signals related to degree of reward expectancy. *Science*, 296(5573):1709–1711, 2002.

[38] M. C. W. Van Rossum. A novel spike distance. *Neural Comput.*, 13:751–763, April 2001.

[39] Jonathan D. Victor and Keith P. Purpura. Independent and redundant information in nearby cortical neurons. *Science*, 294(5551):2566–2568, 2001.

[40] S M Williams and P S Goldman-Rakic. Widespread origin of the primate mesofrontal dopamine system. *Cerebral Cortex*, 8(4):321–345, 1998.