

## Supplementary Information

This Supplementary Information contains two sections. Section 1 provides an example in which correlations are important and  $\Delta I \neq 0$ , and Section 2 provides technical details regarding Meister and Hosoya's second comment, the one concerning linear reconstruction.

### 1. Calculating $\Delta I$ when correlations are important

In Meister and Hosoya's two examples, correlations are not important and  $\Delta I = 0$ . Here we show an example in which correlations are important and  $\Delta I$  correctly detects this – by being nonzero. Along the way, we demonstrate that  $\Delta I$  is nonzero *whenever* correlations are important (see comments following Eq. (1)).

The example we show is very similar to Meister and Hosoya's case **a**; the only difference is that we change just one of the entries (see Table I below). Before getting to that example, we outline our approach for calculating  $\Delta I$ . In our paper<sup>1</sup> we give two formulas for this quantity: one in Eq. (2) and one in Methods. The latter is far easier to use, at least in this example, so we reproduce it here,

$$\Delta I = D\left(P(s|r_1, r_2) || P_{IND}(s|r_1, r_2)\right) = \sum_{r_1, r_2} P(r_1, r_2) \sum_s P(s|r_1, r_2) \log_2 \left[ \frac{P(s|r_1, r_2)}{P_{IND}(s|r_1, r_2)} \right] \quad (1)$$

where  $P(s|r_1, r_2)$  is the probability that stimulus  $s$  occurred given that responses  $r_1$  and  $r_2$  were observed, and  $P_{IND}(s|r_1, r_2)$  is the same thing, except that the responses are assumed to be independent (see Eq. (2b) below).

The sum over  $s$  in Eq. (1) is the Kullback-Leibler divergence<sup>2</sup>, which is nonzero unless  $P(s|r_1, r_2) = P_{IND}(s|r_1, r_2)$  for *all* stimulus/response triplets  $(s, r_1, r_2)$ . Thus, whenever  $P_{IND}(s|r_1, r_2)$  differs from  $P(s|r_1, r_2)$  for at least one stimulus/response triplet,  $\Delta I$  is nonzero. This means if the translation of responses using the independent dictionary is at all different from the translation of responses using the correlated dictionary for any observed responses and stimuli, then  $\Delta I$  is nonzero.

The distributions  $P(s|r_1, r_2)$  and  $P_{IND}(s|r_1, r_2)$  come from  $P(r_1, r_2|s)$  and  $P(r_1|s)P(r_2|s)$ , respectively, via Bayes' theorem. Specifically,

$$P(s|r_1, r_2) = \frac{P(r_1, r_2|s)P(s)}{P(r_1, r_2)} \quad (2a)$$

$$P_{IND}(s|r_1, r_2) = \frac{P(r_1|s)P(r_2|s)P(s)}{P_{IND}(r_1, r_2)} \quad (2b)$$

$$P(r_1, r_2) = \sum_s P(r_1, r_2|s)P(s) \quad (2c)$$

$$P_{IND}(r_1, r_2) = \sum_s P(r_1|s)P(r_2|s)P(s) \quad (2d)$$

$$P(r_1|s) = \sum_{r_2} P(r_1, r_2|s) \quad (2e)$$

$$P(r_2|s) = \sum_{r_1} P(r_1, r_2|s). \quad (2f)$$

Here  $P(s)$  is the probability of observing stimulus  $s$ , taken to be  $1/2$  for both stimuli.

If we were given  $P(s|r_1, r_2)$  and  $P_{IND}(s|r_1, r_2)$ , it would be easy to calculate  $\Delta I$  from Eq. (1). However, this is rarely the information we have – more typically (and certainly in Meister and Hosoya’s examples), we are given  $P(r_1, r_2|s)$  and  $P(s)$  and have to grind through the formulas in Eq. (2) before we can get to Eq. (1). So that’s what we do now: grind through those formulas for an example in which correlations are important.

The example we choose is a perturbation to Meister and Hosoya’s case **a**, in which we change  $(0, 0)$  to  $(0, 1)$  but otherwise do not change the probabilities. The resulting response distribution,  $P(r_1, r_2|s)$ , is shown in Table 1. For this and all tables, only responses that have a nonzero probability are shown.

Table I

$s$	$r_1$	$r_2$	$P(r_1, r_2 s)$
0	0	1	$1/2$
0	1	1	$1/2$
1	1	1	$1/2$
1	2	2	$1/2$

The reason that we expect correlations to be important in this example is that the two stimuli have different correlational structures; that is, the degree of correlation depends on the stimuli.

Comparing Eqs. (1) and (2), we see that there are several quantities that need to be calculated before we can get to  $P(s|r_1, r_2)$  and  $P_{IND}(s|r_1, r_2)$ . These are  $P(r_1|s)$ ,  $P(r_2|s)$ ,  $P(r_1, r_2)$ , and  $P_{IND}(r_1, r_2)$ . The first two are given in Table II.

Table II

$s$	$r_1$ OR $r_2$	$P(r_1 s)$	$P(r_2 s)$
0	0	$1/2$	0
0	1	$1/2$	1
1	1	$1/2$	$1/2$
1	2	$1/2$	$1/2$

Because we’ll need  $P(r_1|s)P(r_2|s)$  for Eqs. (2b) and (2d), we compute it and reproduce it in Table III.

Table III

$s$	$r_1$	$r_2$	$P(r_1 s)P(r_2 s)$
0	0	1	1/2
0	1	1	1/2
1	1	1	1/4
1	1	2	1/4
1	2	1	1/4
1	2	2	1/4

Using Tables I and III, Eqs. (2c-2f), and the fact that  $P(s=0) = P(s=1) = 1/2$ , we can now construct  $P(r_1, r_2)$  and  $P_{IND}(r_1, r_2)$ ; these are given in Table IV.

Table IV

$r_1$	$r_2$	$P(r_1, r_2)$	$P_{IND}(r_1, r_2)$
0	1	1/4	1/4
1	1	1/2	3/8
1	2	0	1/8
2	1	0	1/8
2	2	1/4	1/8

Finally, we can use Eqs. (2a) and (2b) to compute  $P(s|r_1, r_2)$  and  $P_{IND}(s|r_1, r_2)$ ; the result is in Table V.

Table V

$r_1$	$r_2$	$P(1 r_1, r_2)$	$P_{IND}(1 r_1, r_2)$
0	1	0	0
1	1	1/2	1/3
1	2	-	1
2	1	-	1
2	2	1	1

When the cells are correlated, responses 1 and 2 never occur together. Thus,  $P(1|1, 2)$  and  $P(1|2, 1)$  are not evaluateable; for these entries in Table V we just put a dash.

From the last table, we see that  $P_{IND}(1|1, 1)$  is *not* equal to  $P(1|1, 1)$ ; this means the correlated and uncorrelated dictionaries are different, and correlations are important for decoding. We can now readily calculate  $\Delta I$ , since it contains only two terms,

$$\Delta I = P(1, 1)P(1|1, 1) \log_2 \left[ \frac{P(1|1, 1)}{P_{IND}(1|1, 1)} \right] + P(1, 1)P(0|1, 1) \log_2 \left[ \frac{P(0|1, 1)}{P_{IND}(0|1, 1)} \right].$$

Plugging in numbers and using  $P(0|r_1, r_2) = 1 - P(1|r_1, r_2)$ , we find that  $\Delta I = (1/4) \log_2(9/8) \approx 0.042$ .

As discussed above,  $\Delta I$  is nonzero whenever correlations are important, where “important” means  $P_{IND}(s|r_1, r_2) \neq P(s|r_1, r_2)$ . Thus, it is a striking finding<sup>1</sup> that for the 498 pairs examined in the mouse retina, many of which were highly correlated,  $\Delta I$  was always close to zero – within 11% of the full information.

## 2. Technical comments regarding linear reconstruction

Meister and Hosoya’s second comment concerns our use of linear reconstruction to assess the role of correlations. We believe that their main point is that linear reconstruction, as we use it, is not able to detect whether correlations are important. They state that:

“To test whether spike pairs that are synchronous on the millisecond scale carry special visual messages, one should use a decoder that can recognize such spike pairs. The method used by the authors, namely linear filtering of the individual spike trains<sup>3</sup>, cannot do that. The filter functions are necessarily broad (probably 0.1 s) such that millisecond shifts of individual spikes cannot be interpreted.”

The above statement seems to assume that we construct one set of filters and then use those filters to reconstruct the stimulus from correlated and uncorrelated spike trains. However, we do not do that. Instead, we build two decoders, one from correlated spike trains and one from uncorrelated trains<sup>4</sup>. We then run both decoders on the same set of spike trains – the true, simultaneously recorded ones. Because both decoders are run on the same spike trains, differences in the reconstructions must be due to differences in the filters.

The remaining question, then, is: are the filters sensitive to correlations? One can show in general that the answer is yes. Just to be sure, we compared the correlated and uncorrelated filters for the four examples shown in Fig. 4 of ref. 1. As can be seen in Fig. 1 below, the difference between the two filters scales with the relative loss of information,  $\Delta I/I$ , where  $I$  is the mutual information between stimulus and response. Thus, the filters *are* picking up the effects of correlations. The reason the two reconstructions in each panel of Fig. 4 are almost identical is that the amount of information that you lose when you treat the cells as independent encoders is small; not that linear reconstruction is incapable of picking up the effects of correlations.

A technical note is in order: while the 10% difference in  $\Delta I/I$  corresponds to an  $\sim 10\%$  difference in the filters, the error in the reconstruction is much smaller than 10%. This is because the linear reconstruction fits the filters to a cost function with a quadratic minimum, so a 10% difference in the filters results in an  $\sim 1\%$  ( $0.1^2$ ) error in the reconstruction. Such favorable scaling is relatively generic: any decoder with a quadratic minimum will have a reconstruction error that scales as  $(\Delta I/I)^2$ .

Still, one might argue that straightforward linear reconstruction from two spike trains is not a good test. Indeed, Meister and Hosoya suggest a different approach:

“Instead, one should identify synchronous spike pairs first, and then assign separate messages to them.”

Given their subsequent reference to a paper that discusses this issue<sup>5</sup>, what we believe Meister and Hosoya are advocating here is the following: Take two spike trains and construct a third, which consists only of synchronous spikes. Then perform linear reconstruction using the two original two spike trains (with the synchronous spikes removed) plus the third, synchronous, spike train. Compare that reconstruction with the one made using the original two spike trains.

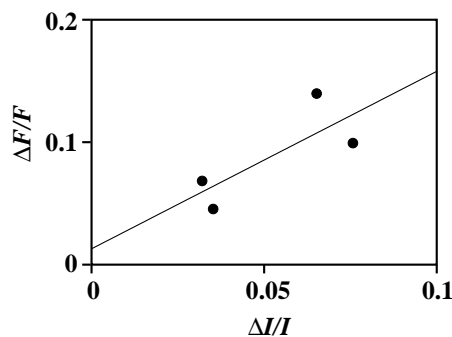


Figure 1: Filled circles: relative error in the filters,  $\Delta F/F$ , plotted versus  $\Delta I/I$  for the four examples shown in Fig. 4 of ref. 1. The relative error is defined via the relations  $\Delta F^2 \equiv \sum_{n=1}^2 \int dt [f_c^n(t) - f_u^n(t)]^2$  and  $F^2 \equiv \sum_{n=1}^2 \int dt f_c^n(t)^2$ , where  $f_c^n(t)$  is the filter for neuron  $n$  constructed from correlated spike trains and  $f_u^n(t)$  is the same thing, except constructed from uncorrelated spike trains. Line: least squares fit to the data. The slope is 1.4 and the intercept is 0.013.

This process sounds totally reasonable, but let’s examine it a little more closely. Since we want to build a near-optimal decoder, we should build our decoder using time bins that are so small that no more than one spike from any one neuron can occur in a bin. In this limit, the transformation from two to three spike trains is invertible – you can go from the original two spike trains to the three spike trains and back without losing any information. Thus, by the Data Processing Inequality, the three spike trains contain *exactly* the same amount of information about the stimulus as the two spike trains. Any difference one finds in the quality of the reconstruction using two versus three spike trains is *entirely* an artifact of the method, and has nothing to do with the importance of correlations.

This highlights the difficulty of testing for the importance of correlations using intuitively reasonable, but non-rigorous, methods: it can be extremely difficult to untangle the effects of correlations from the effects of the algorithm one adopts. It is for this reason that our main analysis used an information-theoretic cost function,  $\Delta I$ , that does not depend on any particular reconstruction algorithm.

Meister and Hosoya also fault the fact that we use linear reconstruction on spatially uniform stimuli:

“The current best guess is that synchronous spike pairs originate in retinal interneurons and thus have different spatial receptive fields from those of the participating ganglion cells<sup>6,7</sup> If so, then the information they convey is about spatial detail, which cannot possibly be tested with the spatially uniform stimulus used by the authors.”

We should emphasize that the “current best guess” is just that – a guess. In addition, while we performed linear reconstruction using a spatially uniform field, our main result (Fig. 3) was based on movies of natural scenes consisting of interacting mice.

Meister and Hosoya's critique ends with the statement:

“Ideally, one would carry such an analysis beyond mere pairwise correlations. Any given pair of cells may have only few synchronous spikes in common, but once all the nearby partners are considered, the fraction of spikes in such firing patterns may exceed 80% (ref. 8).

Thus, the degree to which concerted firing contributes to coding by retinal ganglion cells remains an open issue.”

We agree that “the degree to which concerted firing contributes to coding by retinal ganglion cells” remains an open issue – for populations of neurons. However, given that our analysis was based on 498 cell pairs and a rigorously derived information-theoretic cost function<sup>1</sup>, we feel that the issue has been largely closed for pairs of mouse neurons in the photopic regime. We believe this is an important step, since six years ago Meister et al.<sup>6</sup> appeared to imply that concerted firing *does* contribute to coding by retinal ganglion cells, based on one cell pair and qualitative analysis. The next step – determining whether or not  $\Delta I$  is small for neuronal populations – will have to wait for future studies.

## References

1. S. Nirenberg, S.M. Carcieri, A.L. Jacobs, and P.E. Latham. Retinal ganglion cells act largely as independent encoders. *Nature* **411**:698-701 (2001).
2. T.M. Cover and J.A. Thomas. *Elements of information theory*. John Wiley & Sons, New York, (1991).
3. Warland, D.K., Reinagel, P. & Meister, M. Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.* **78**, 2336-2350 (1997).
4. [http://culture.neurobio.ucla.edu/papers/Nirenberg01\\_si.pdf](http://culture.neurobio.ucla.edu/papers/Nirenberg01_si.pdf)
5. Dan, Y., Alonso, J.M., Usrey, W.M. & Reid, R.C. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat. Neurosci.* **1**, 501- 507 (1998).
6. Meister, M., Lagnado, L. & Baylor, D.A. Concerted signaling by retinal ganglion cells. *Science* **270**, 1207-1210 (1995).
7. Brivanlou, I.H., Warland, D.K. & Meister, M. Mechanisms of concerted firing among retinal ganglion cells. *Neuron* **20**, 527-539 (1998).
8. Mastrorarde, D.N. Correlated firing of cat retinal ganglion cells. I. Spontaneously active inputs to X- and Y-cells. *J. Neurophysiol.* **49**, 303-324 (1983a).