Synergy, Redundancy, and Independence in Population Codes, Revisited

Peter Latham and Sheila Nirenberg
J. Neurosci. (2005)

Questions

How redundancies in neural codes are important?

- specifically, "noise correlation"
- in the context of decoding

How to measure its importance?

- $\Delta I_{\text{shuffled}}$
- $\Delta I_{\text{synergy}}$
- ΔI

Definition of “noise correlation”

Response correlations on a stimulus-by-stimulus basis.

Responses are noise correlated if and only if

$$p(\mathbf{r}|s) \neq \prod_{i=1}^{N} p(r_i|s)$$

white board

No noise correlation makes decoding easy

Encoding: $p(\mathbf{r}|s)$

Decoding: $p(s|\mathbf{r}) = \dfrac{p(\mathbf{r}|s)p(s)}{\int p(\mathbf{r}|s)p(s)d\mathbf{r}ds}$

p(r|s) can be high-dimensional

- difficult to estimate
- can we assume independence w/o significant “loss”?

$$p(\mathbf{r}|s) \approx \prod_{i=1}^{N} p(r_i|s)$$

How to measure the effect of independence assumption?

Compare the decoding performance w/ and w/o this simplified assumption:

$$p(\mathbf{r}|s) \approx p_{\text{ind}}(\mathbf{r}|s) \equiv \prod_{i=1}^{N} p(r_i|s)$$

How?

- Measuring mutual information I(s; r) w/ and w/o the independence assumption.
- Decoding neural codes w/ and w/o the independence assumption.

Meaning of mutual information I(s; r) in decoding

On average, how many yes/no questions can you reduce to identify what is s, by knowing r.

- e.g., H(s) = 3 bits, I(s, r) = 2 bits.

Shuffled information

Mutual information between stimuli and responses:

$$I(s;\mathbf{r}) = -\int p(\mathbf{r})\log p(\mathbf{r})d\mathbf{r} + \int p(\mathbf{r}|s)p(s)\log p(\mathbf{r}|s)d\mathbf{r}ds$$

Mutual information using independence assumption:

$$I_{\text{shuffled}}(s;\mathbf{r}) = -\int p_{\text{ind}}(\mathbf{r})\log p_{\text{ind}}(\mathbf{r})d\mathbf{r} + \int p_{\text{ind}}(\mathbf{r}|s)p(s)\log p_{\text{ind}}(\mathbf{r}|s)d\mathbf{r}ds$$

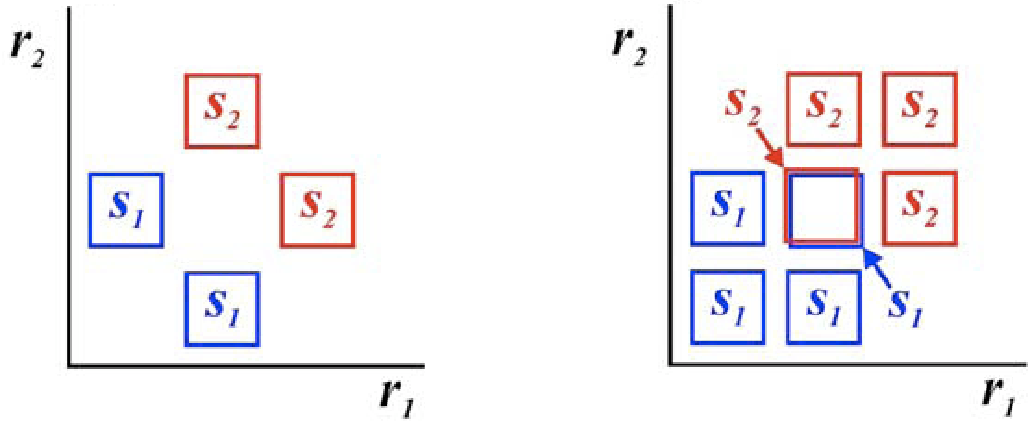
What does these difference tell us?

$$\Delta I_{\text{shuffled}} = I(s;\mathbf{r}) - I_{\text{shuffled}}(s;\mathbf{r})$$

$$p_{\text{ind}}(\mathbf{r}|s) \equiv \prod_{i=1}^{N} p(r_i|s) \qquad p_{\text{ind}}(\mathbf{r}) \equiv \int p_{\text{ind}}(\mathbf{r}|s)p(s)ds$$

$\Delta I_{\text{shuffled}}$ does not tell us anything conclusive

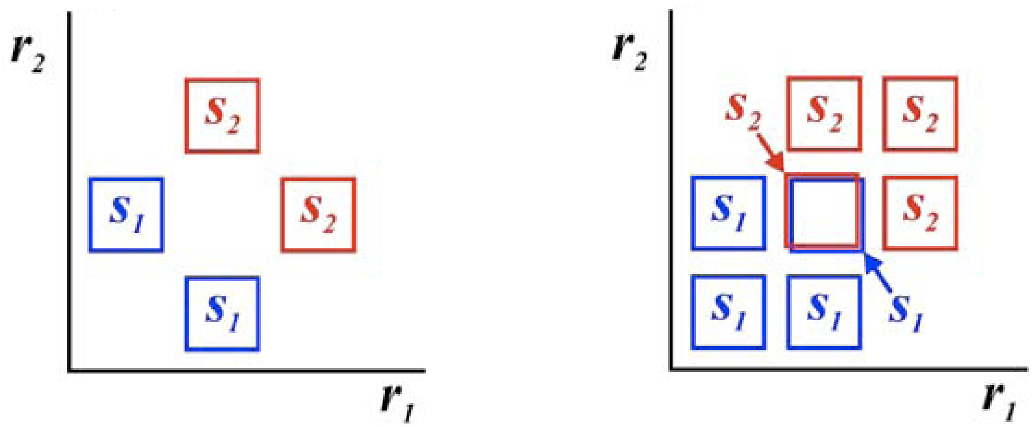
Example 1: $\Delta I_{\text{shuffled}} > 0$



Does the noise correlation play a role in discriminating s1 and s2? ... [y/n]

$\Delta I_{\text{shuffled}}$ does not tell us anything conclusive

Example 1: $\Delta I_{\text{shuffled}} > 0$

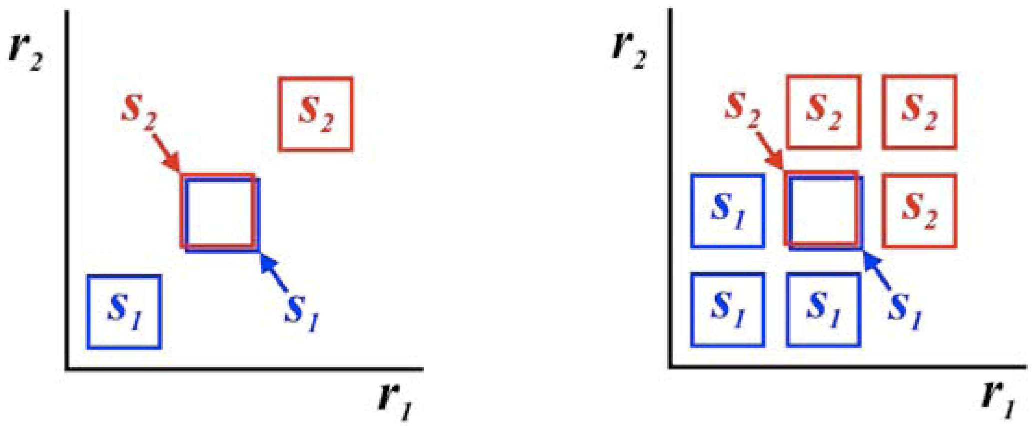


Does the noise correlation play a role in discriminating s1 and s2? ... We may say yes, compared to the case with independence assumption.

But the decoder w/o knowing noise correlation performs as well as the one w/ knowing noise correlation!

$\Delta I_{\text{shuffled}}$ does not tell us anything conclusive

Example 2: $\Delta I_{\text{shuffled}} < 0$



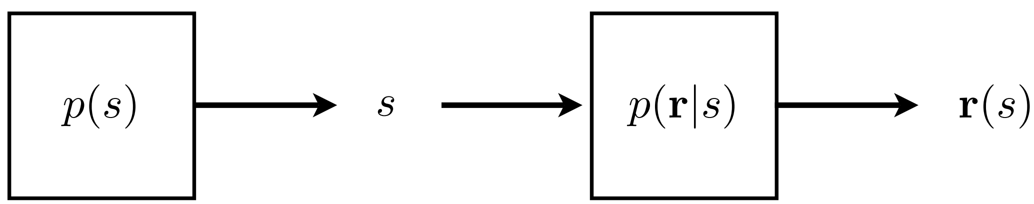
Larger mutual information by ignoring noise correlation!

Decoding performance does not differ.

$\Delta I_{\text{shuffled}}$ does not tell us anything conclusive

Summary:

- $\Delta I_{\text{shuffled}}$ could be positive or negative while the decoding performance does not differ for the ideal and the shuffled decoders.
- It is because $\Delta I_{\text{shuffled}}$ is evaluated over p_ind, instead of the true probability p.



These processes are all given and should not be altered.

- Given r(s) as such, it is appropriate to consider a decoder that does not utilize noise correlation.

Synergy and redundancy

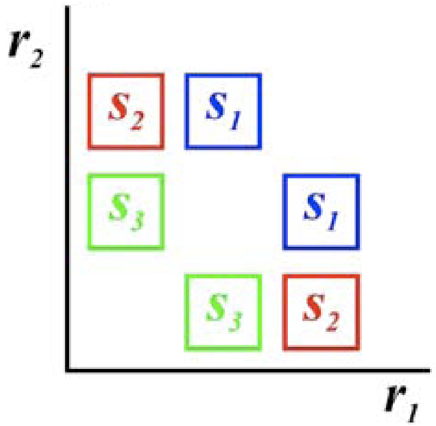
$\Delta I_{\text{synergy}}$ is by definition:

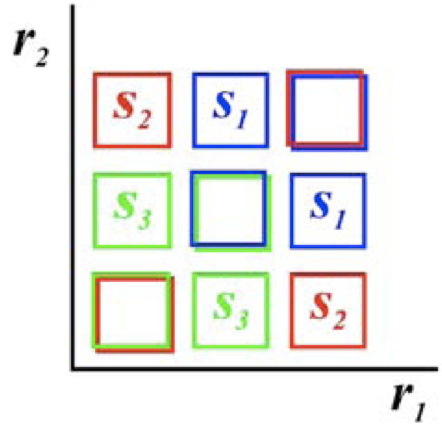
$$\Delta I_{\text{synergy}} = I(s;\mathbf{r}) - \sum_{i=1}^{N} I(s;r_i)$$

- independence of the responses implies $\Delta I_{\text{synergy}} = 0$.
- converse does not hold.
- $\Delta I_{\text{synergy}} > 0$ means synergistic (?).
- $\Delta I_{\text{synergy}} < 0$ means redundant (?).

$\Delta I_{\text{synergy}}$ is also a confounded measure

Example 3: $\Delta I_{\text{synergy}} > 0$





- This does indicate a synergistic code (knowing two responses provides more info about s).
- But, decoders w/ and w/o knowing joint events perform perfectly.

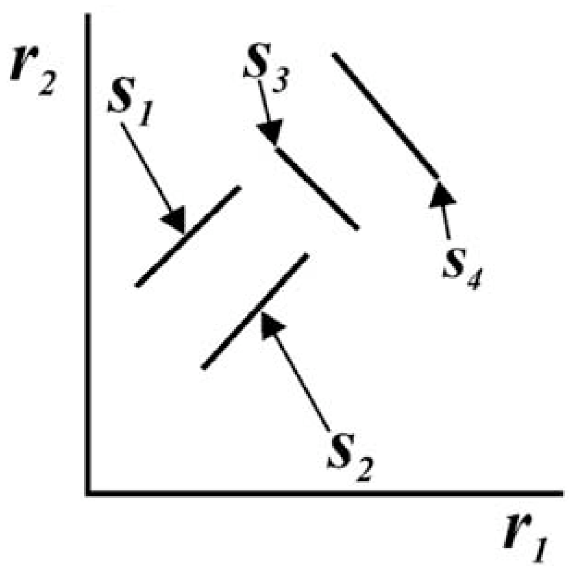
$\Delta I_{\text{synergy}}$ is also a confounded measure

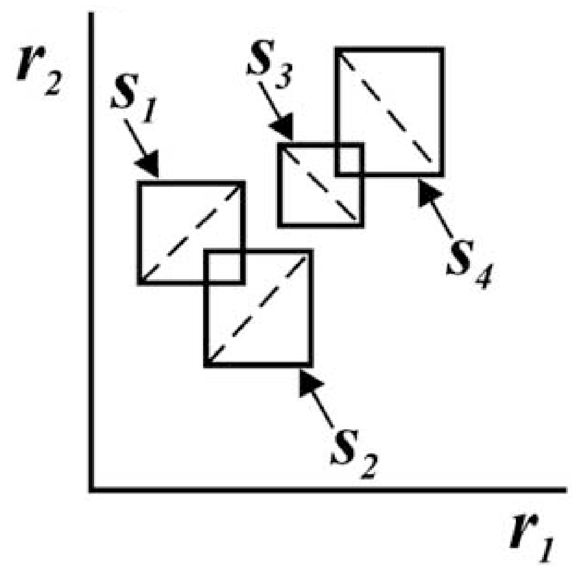
Summary:

- $\Delta I_{\text{synergy}}$ does not indicate the importance of noise correlation on decoding.
- This is because it evaluates data points that never happens with the true distribution but occurs with the independent version.

When “noise correlations” matter?

Example 4:

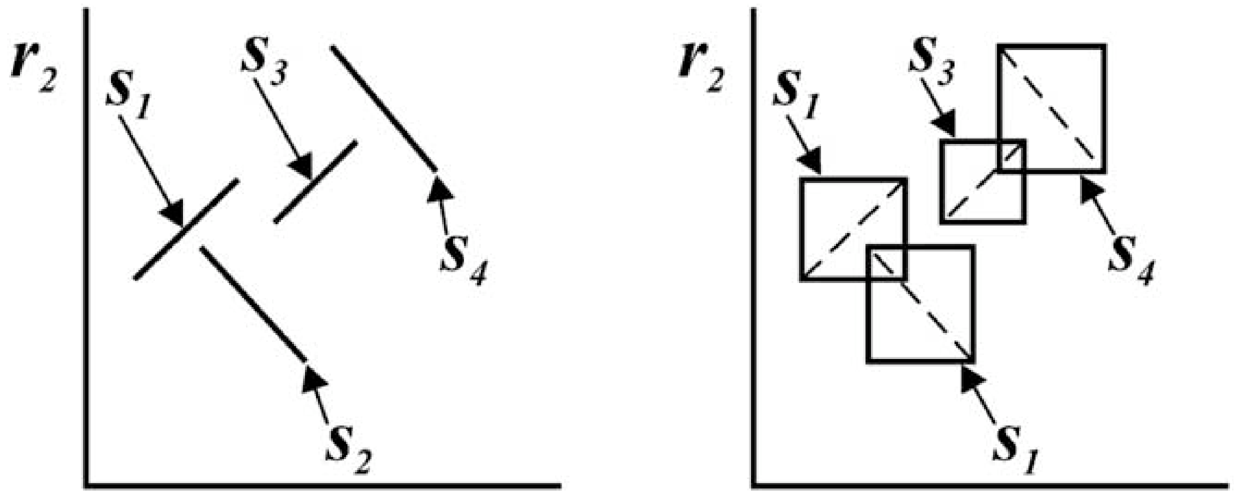




- You can assume “no noise correlation” without any loss of decoding performance.

When “noise correlations” matter?

Example 5:



- You *cannot* assume “no noise correlation” without any loss of decoding performance.

Quantifying the importance of correlations

What properties should the quantity have?

- Be evaluated only with the true responses, not with the responses generated with the independent version.
- Measure a similarity between p_ind(s|r) and p(s|r)

Quantifying the importance of correlations

What properties should the quantity have?

- Be evaluated only with the true responses, not with the responses generated with the independent version.
- Measure a similarity between p_ind(s|r) and p(s|r)

How similar p_ind(s|r) and p(s|r) is, at the data points generated with p(r|s), not with p_ind(r|s), with respect to all stimuli:

$$\Delta I = \int p(s)p(\mathbf{r}|s)\log\frac{p(s|\mathbf{r})}{p_{\mathrm{ind}}(s|\mathbf{r})}d\mathbf{r}ds$$

Quantifying the importance of correlations

Four desirable properties:

1. Weighting the similarity with p(r)
2. Bounded below from zero
3. It is zero if and only if p(s|r) = p_ind(s|r)
4. It is a cost in yes/no questions for not knowing about correlations (Nirenberg et al., 2001; Nirenberg and Latham, 2003).

$$\Delta I = \int p(s)p(\mathbf{r}|s)\log\frac{p(s|\mathbf{r})}{p_{\mathrm{ind}}(s|\mathbf{r})}d\mathbf{r}ds$$

$$= \int p(\mathbf{r})p(s|\mathbf{r})\log\frac{p(s|\mathbf{r})}{p_{\mathrm{ind}}(s|\mathbf{r})}d\mathbf{r}ds = E_{p(\mathbf{r})}\left[KL\left(p(s|\mathbf{r})\|p_{\mathrm{ind}}(s|\mathbf{r})\right)\right]$$

Take-home message

ΔI is a correct quantity to measure how noise correlations are important in decoding.

- $\Delta I_{\mathrm{shuffled}}$ and $\Delta I_{\mathrm{synergy}}$ are not.
- elegant examples

It tells us how many bits you will need in addition to neural responses to identify stimuli when you ignore noise correlations.