Why do neurons fire spikes?

- Neuronal populations use a rate code, but spikes are a necessary evil to send signals quickly, reliably, and for long distances
  - Coding via spike count
  - Coding via rate envelope

OR

- Individual spikes play an intrinsic role in neural coding and computation
  - Coding via patterns across time
  - Coding via patterns across a population

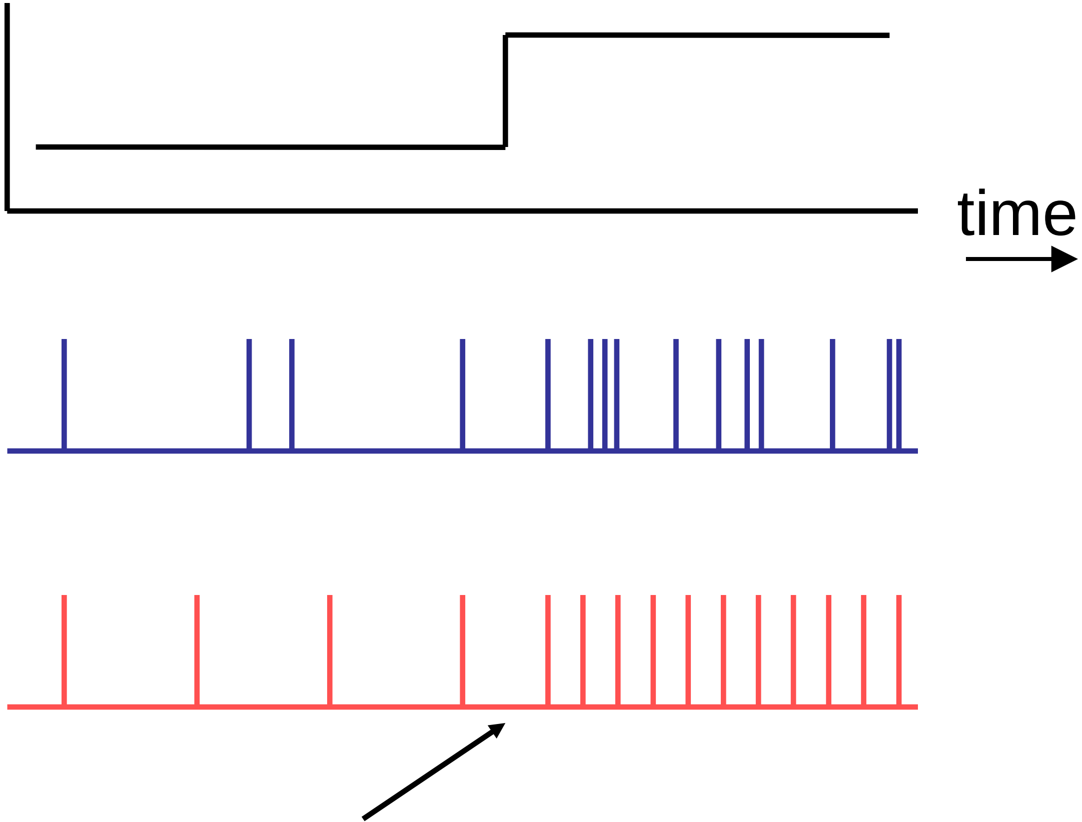
Rate codes are "easier" to create and read out, but codes that use spikes in an intrinsic way can be more "efficient"

Some coding hypotheses

- At the level of individual neurons
  - Spike count
  - Firing rate envelope
  - Interspike interval pattern, e.g., bursts

- At the level of neural populations
  - Total population activity
  - Labeled lines
  - Patterns across neurons, e.g., synchrony
  - Oscillations

Coding by intervals can be faster than coding by count



Signaling a step change in a sensory input

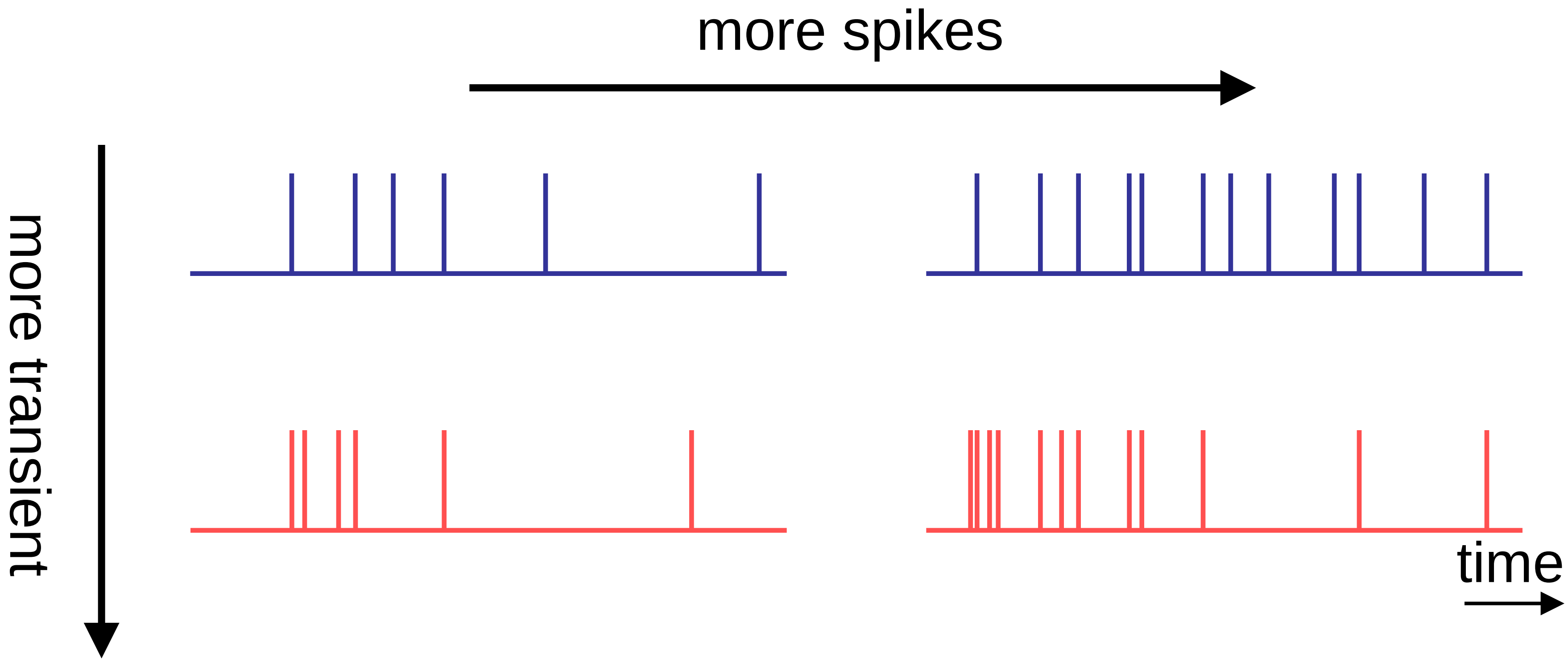
time

coding by count (rate averaged over time)

coding by interval pattern

One short interval indicates a change!

Coding by rate envelope supports signaling of multiple attributes



more spikes

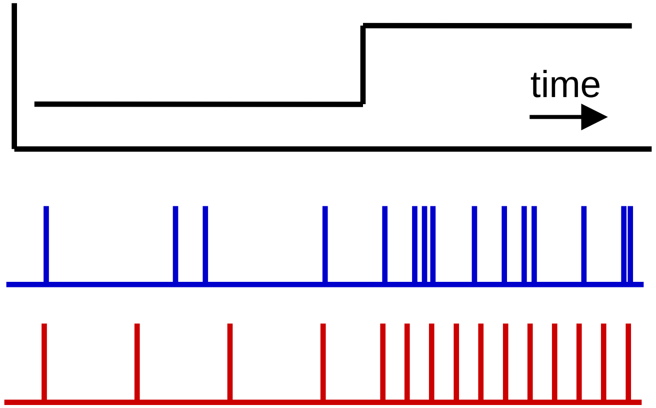
more transient

time

Codes based on spike patterns can also support signaling of multiple attributes.

Codes that use temporal pattern can be *faster* and *richer*

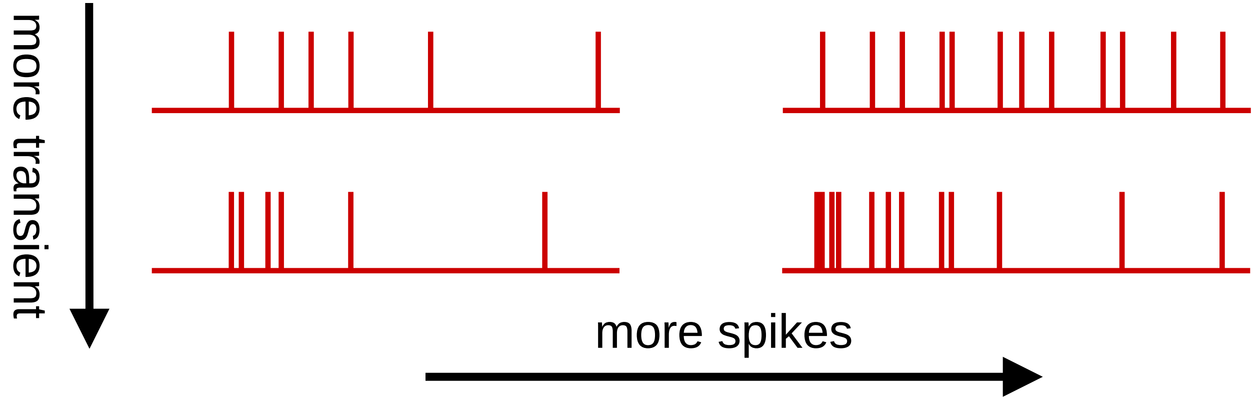
faster: detecting a step change in a sensory input



by count: response must be averaged over time

by intervals: a single interval signals a change

richer: signaling more than one attribute



Firing Patterns are a Nuisance (and rebuttals)

Neurons must fire irregularly.

No, sensory neurons have low variability, and may be clock-like.

Cortical neurons are intrinsically variable.

No, if their inputs are precisely controlled, they work like machines.

Controlling firing patterns is not worth the effort.

No, specific channels lead to particular firing patterns (e.g., thalamocortical neurons).

Reading firing patterns is not worth the effort or not biologically plausible.

No, postsynaptic mechanisms are highly specialized.

- coincidence detection
- “facilitating” and “depressing” synapses (sensitivity to intervals)
- dendritic processing (not just global summation of inputs)

Firing Patterns are Important (and rebuttals)

EEG patterns (oscillations) reflect state of arousal.

But that’s clinical. (Or that’s an epiphenomenon).

Burst and tonic modes in the thalamus reflect state of arousal.

That’s clinical too.

What about spike time dependent plasticity!

That’s just learning.

Direct evidence: locust olfactory system (Laurent)

But that’s a locust. What about Newsome’s cortical microstimulation experiments?

Microstimulation doesn’t only change rate, but also pattern. And who knows what the animal experiences?

Firing Patterns are Important (and rebuttals)

EEG patterns, such as oscillations reflect state of arousal.

But that's an epiphenomenon. (Or that's clinical).

Burst and tonic modes in the thalamus reflect state of arousal.

That's clinical too.

What about spike time dependent plasticity!

That's just learning.

Direct evidence: locust olfactory system

But that's a locust. What about Newsome's cortical microstimulation experiments?

Microstimulation doesn't only change rate, but also pattern. And who knows what the animal experiences?

A direct experimental test of a neural coding hypothesis is difficult

- Count, rate, and pattern are interdependent

  “Time is that great gift of nature which keeps everything from happening at once.” (C.J. Overbeck, 1978)

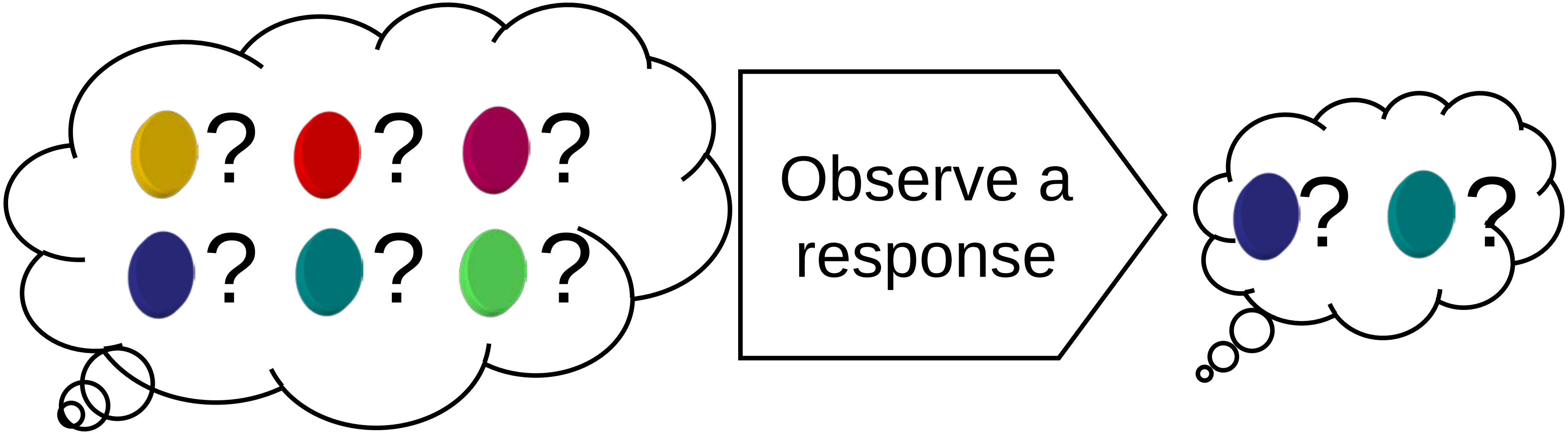
- We’d have to manipulate count, rate, and pattern selectively *AND observe an effect on behavior*
- So, we need some guidance from theory

A Principled Approach to Neural Coding

- Goal: account for the behavior
- *Information* determines the limits of behavior
  - We can measure it from behavior (in principle)
  - We can measure it from neural activity (in principle)
- Two properties of information
  - Data Processing Inequality
  - Independent channels combine additively
- Good news and bad news
  - These properties imply a unique definition of information
  - But they also imply a fundamental problem in implementing the definition
- Addressing this problem refocuses us on biology

Information = Reduction in Uncertainty

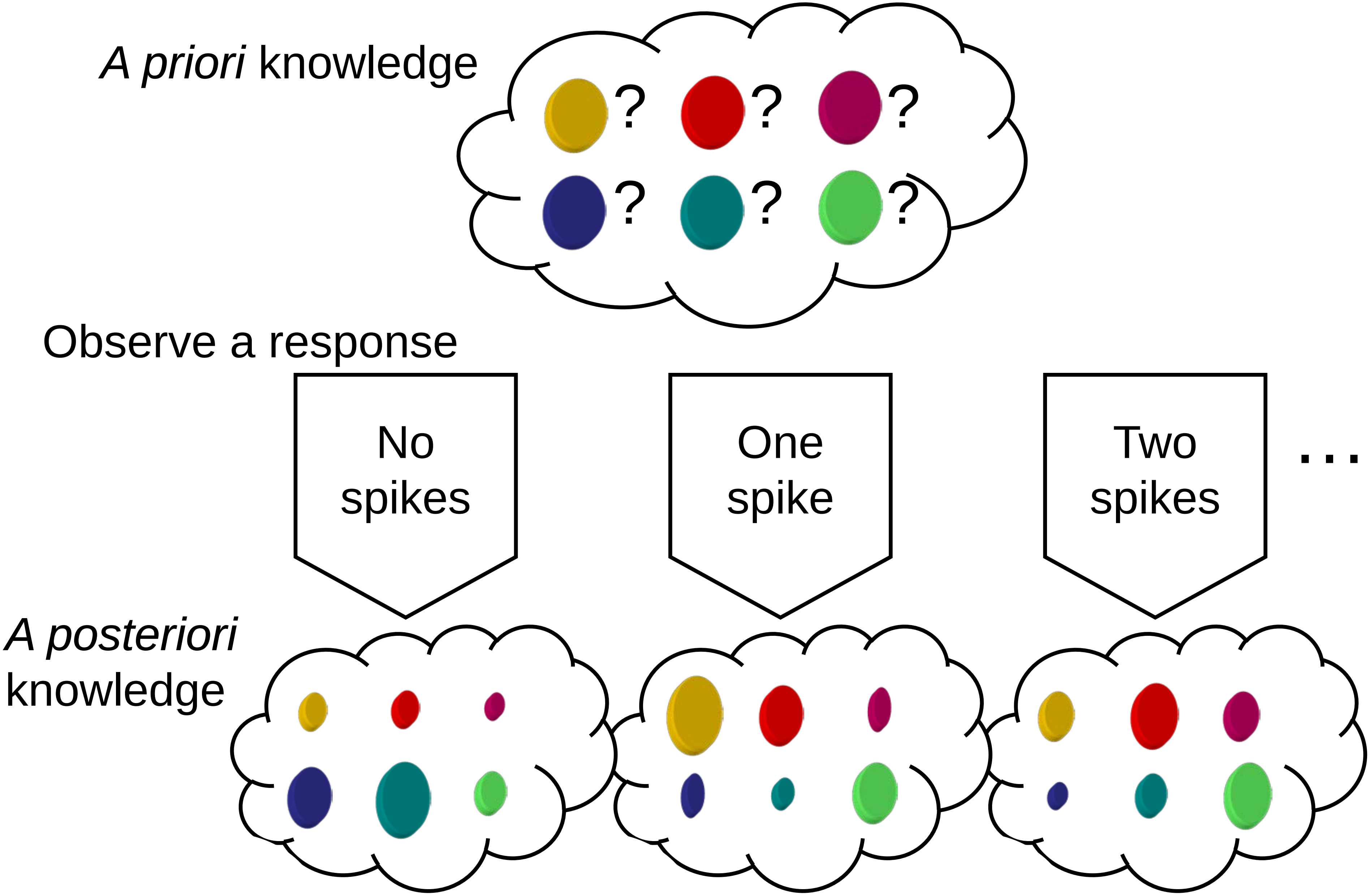
(Claude Shannon, 1948)



Observe a response

- Reduction in uncertainty from 6 possibilities to 2
- Information = log(6/2)

In a bit more detail:



A priori knowledge

Observe a response

No spikes

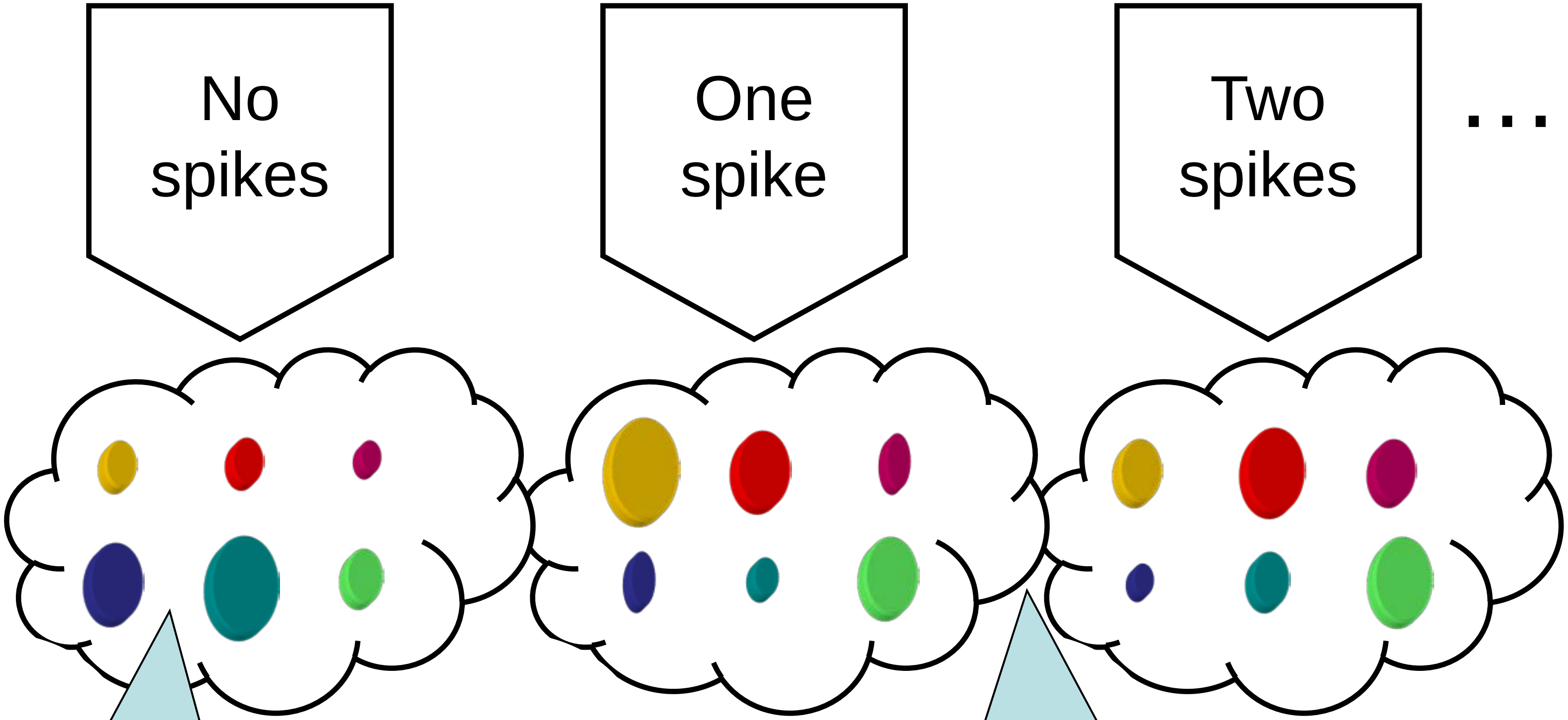
One spike

Two spikes

…

A posteriori knowledge

Second-guessing shouldn’t help



No spikes

One spike

Two spikes

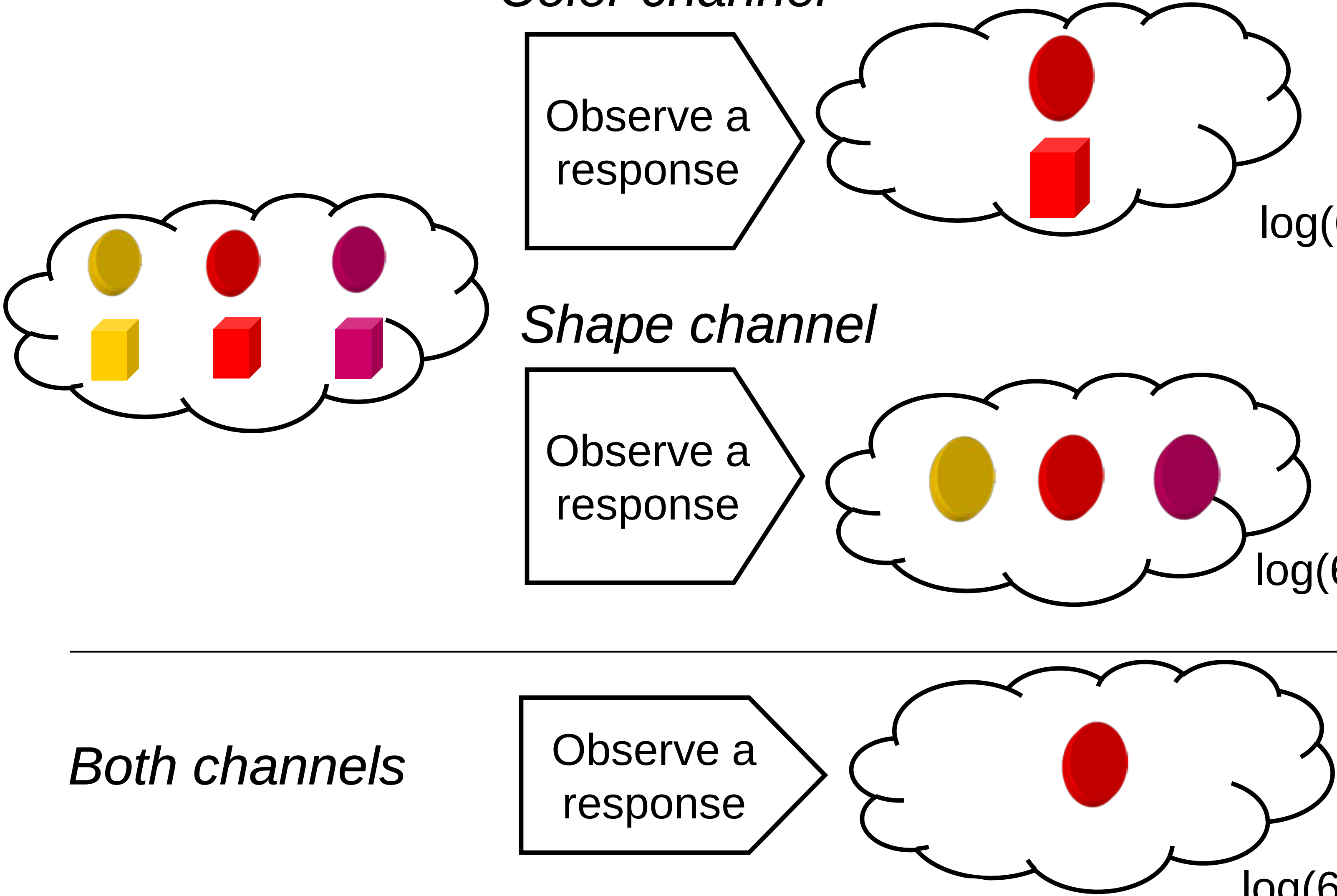
…

Maybe there really should have been a spike?

Maybe these two kinds of responses should be pooled?

The “Data Processing Inequality”: information cannot be increased by re-analysis.

Information on independent channels should add



Color channel

Observe a response

log(6/2) = log(3)

Shape channel

Observe a response

log(6/3) = log(2)

Both channels

Observe a response

log(6/1) = log(6)

$$\log(6) = \log(3 \times 2) = \log(3) + \log(2)$$

Surprising Consequence

- Two ingredients
  - Independent channels combine additively
  - Data Processing Inequality
- Stir well (and add continuity)
- Unique definition of information:

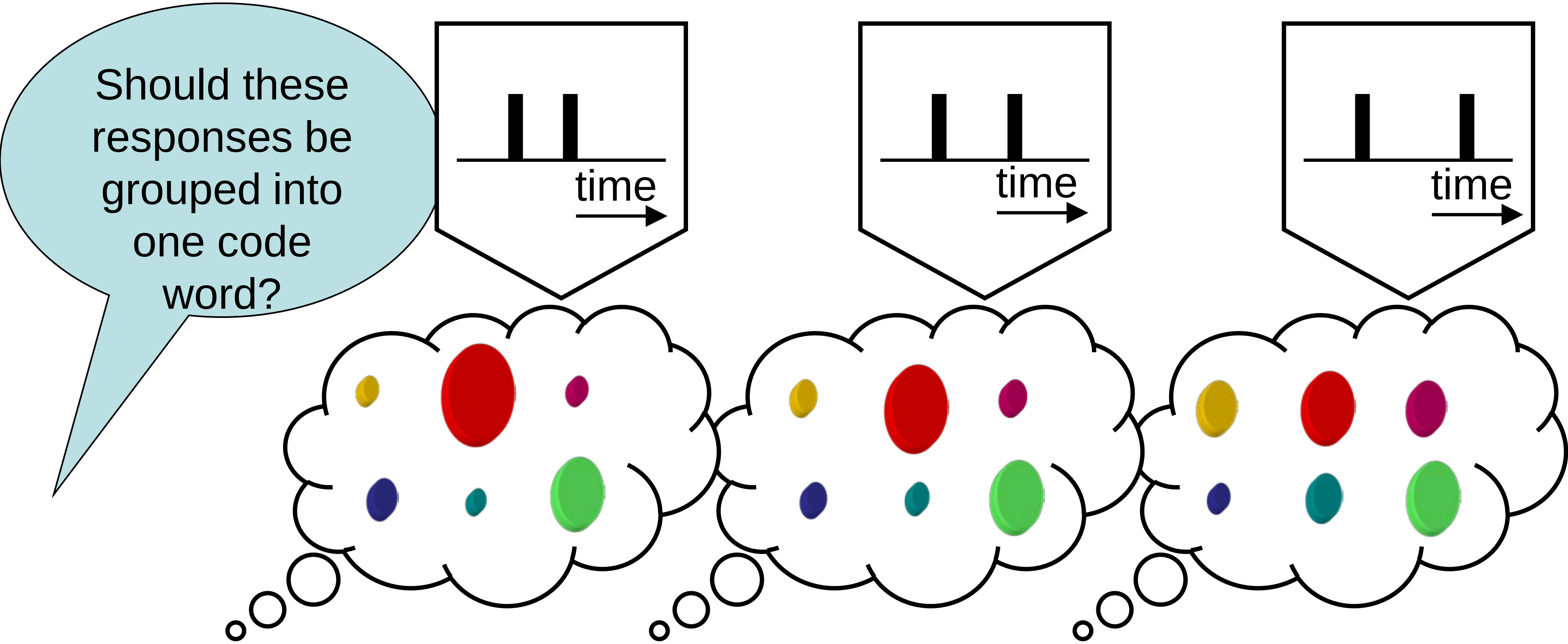
{entropy of stimulus} + {entropy of response} –
{entropy of stimulus, response pairs}

where

entropy= - sum(p_j log p_j), p_j is the probability of each kind of event

- But what are the different kinds of events?
  - Different kinds of stimuli – usually up to the experimenter (but even so, a potential pitfall)
  - Different kinds of responses – that is what we are trying to discover

Revenge of the Data Processing Inequality



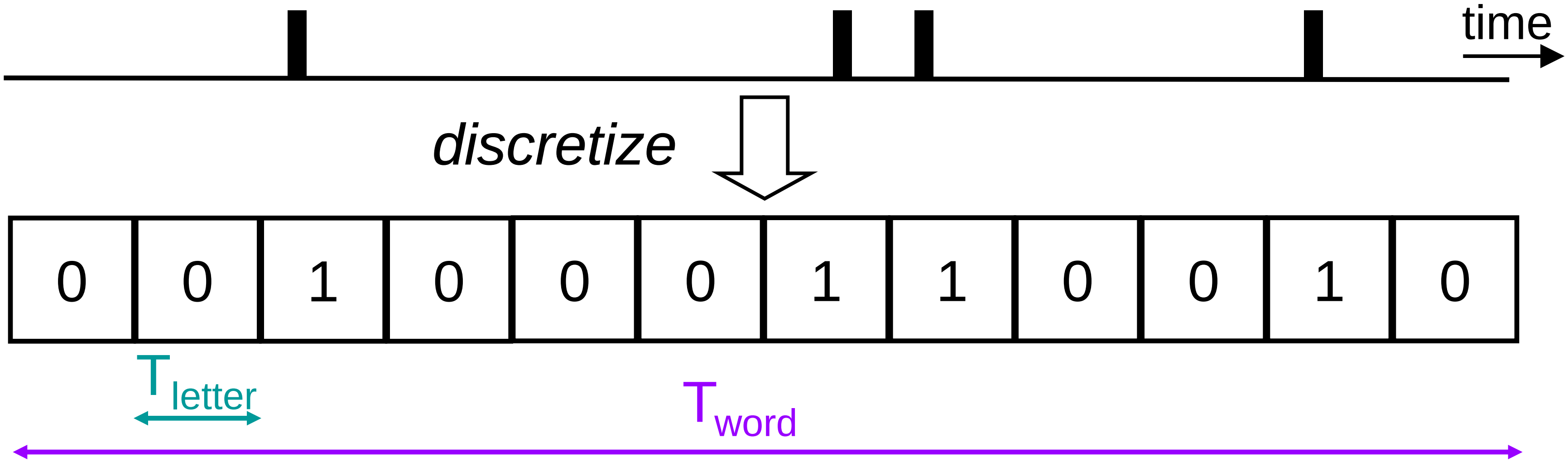
Should these responses be grouped into one code word?

Data Processing Inequality says ***NO***: If you group, you underestimate information

Revenge of the Data Processing Inequality – Consequences

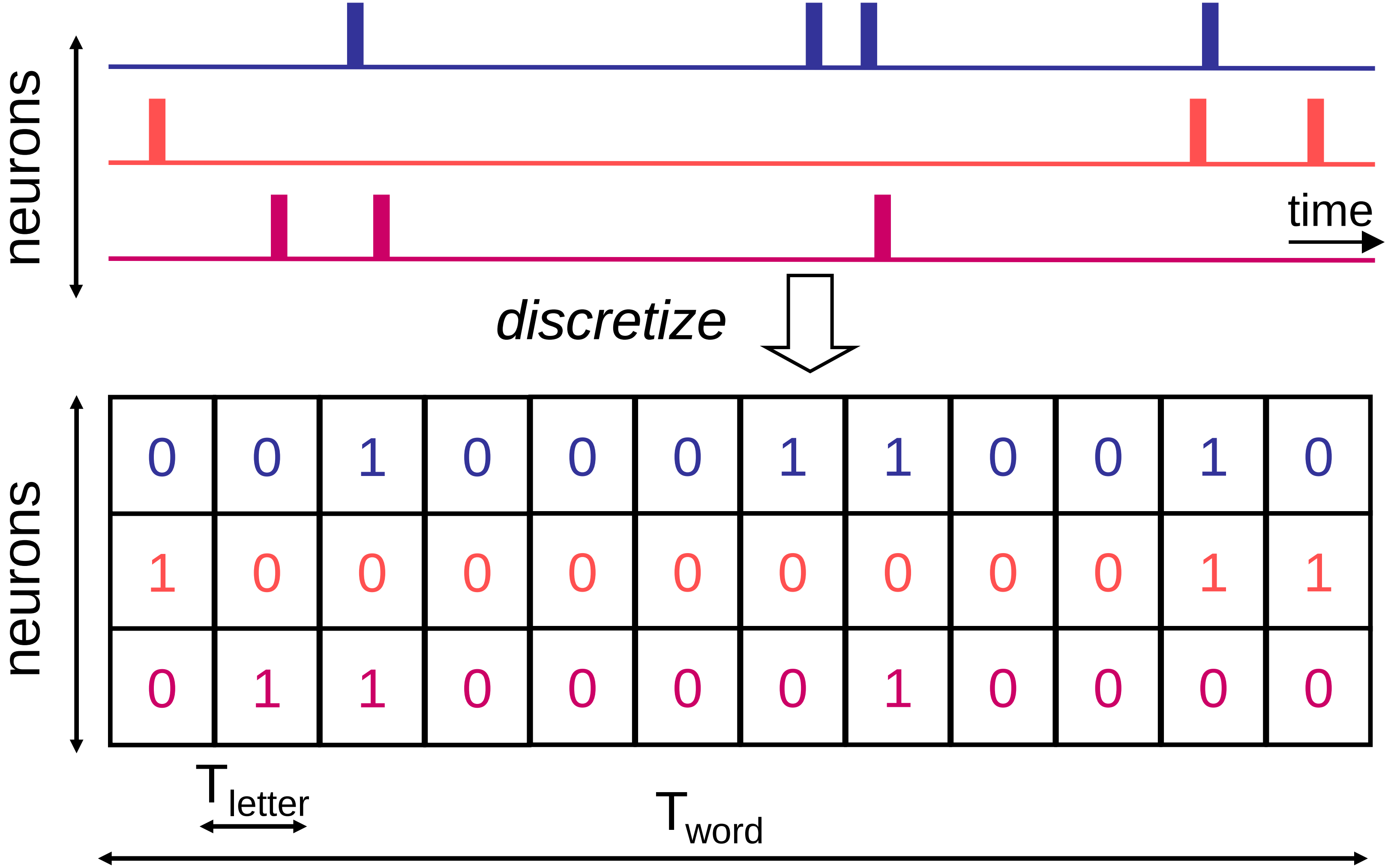
- To avoid underestimating information, if two responses are measurably different, they must be considered separate code words
- Recipe, revised: information is the average of the log of a ***large*** number of ***small*** probabilities
- Why is this a problem?

How many code words?



- T_{letter} must be small (< 1 msec) to resolve the precise biophysics of spiking
- T_{word} must be large (>100 msec) to encompass
  - – behaviorally relevant periods
  - – “slow” biophysical processes (inhibitory potentials)
  - – macroscopic brain rhythms
- T_{word}/T_{letter} > 100. Up to $2^{(T_{word}/T_{letter})}$ code words whose probabilities must be estimated.

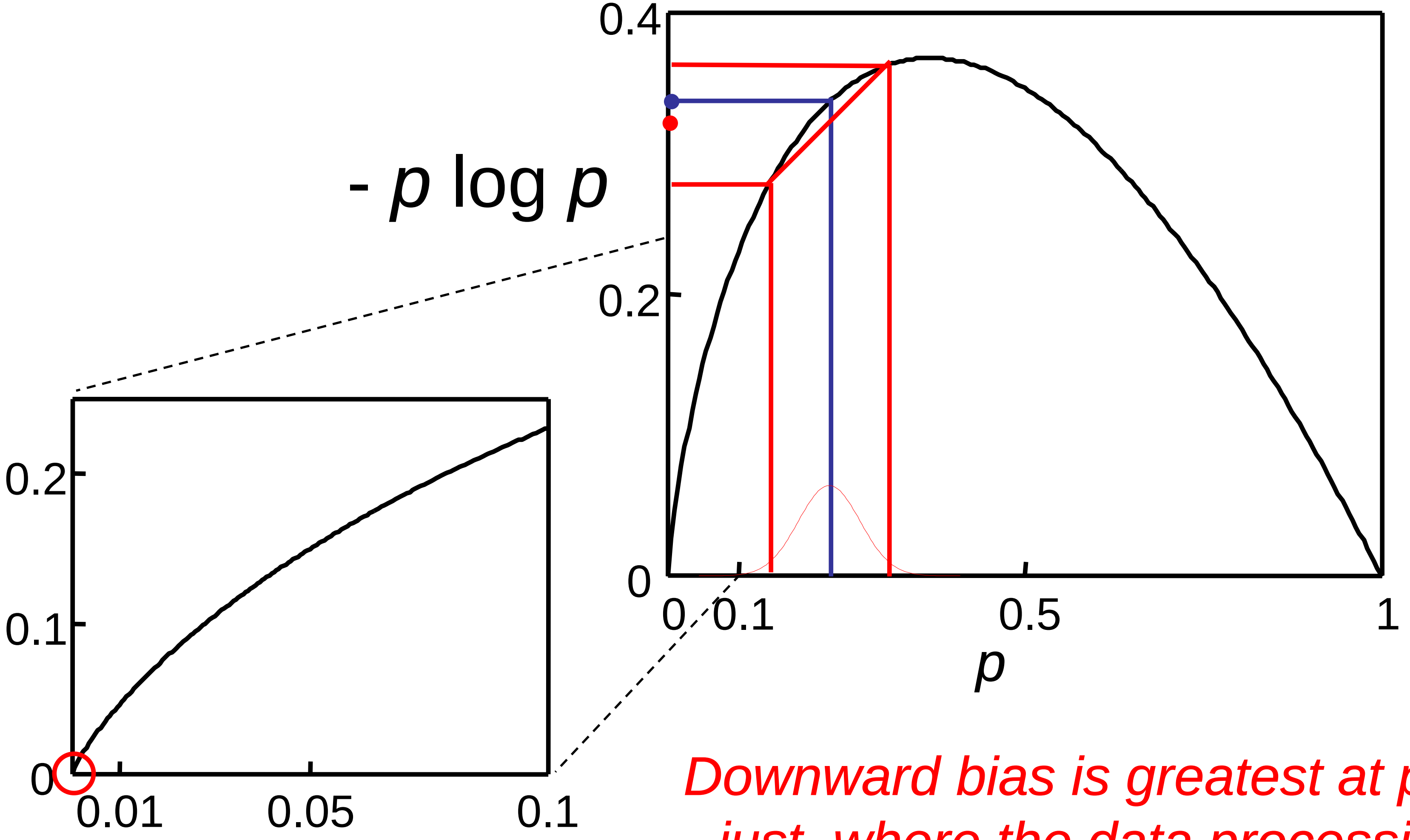
Multiple neurons



0	0	1	0	0	0	1	1	0	0	1	0
1	0	0	0	0	0	0	0	0	0	1	1
0	1	1	0	0	0	0	1	0	0	0	0

- $2^{(T_{word}/T_{letter})*(\text{Number of neurons})}$ probabilities must be estimated. 2^300 is larger than the number of particles in the universe.

It is difficult to estimate expected values of logs of small probabilities



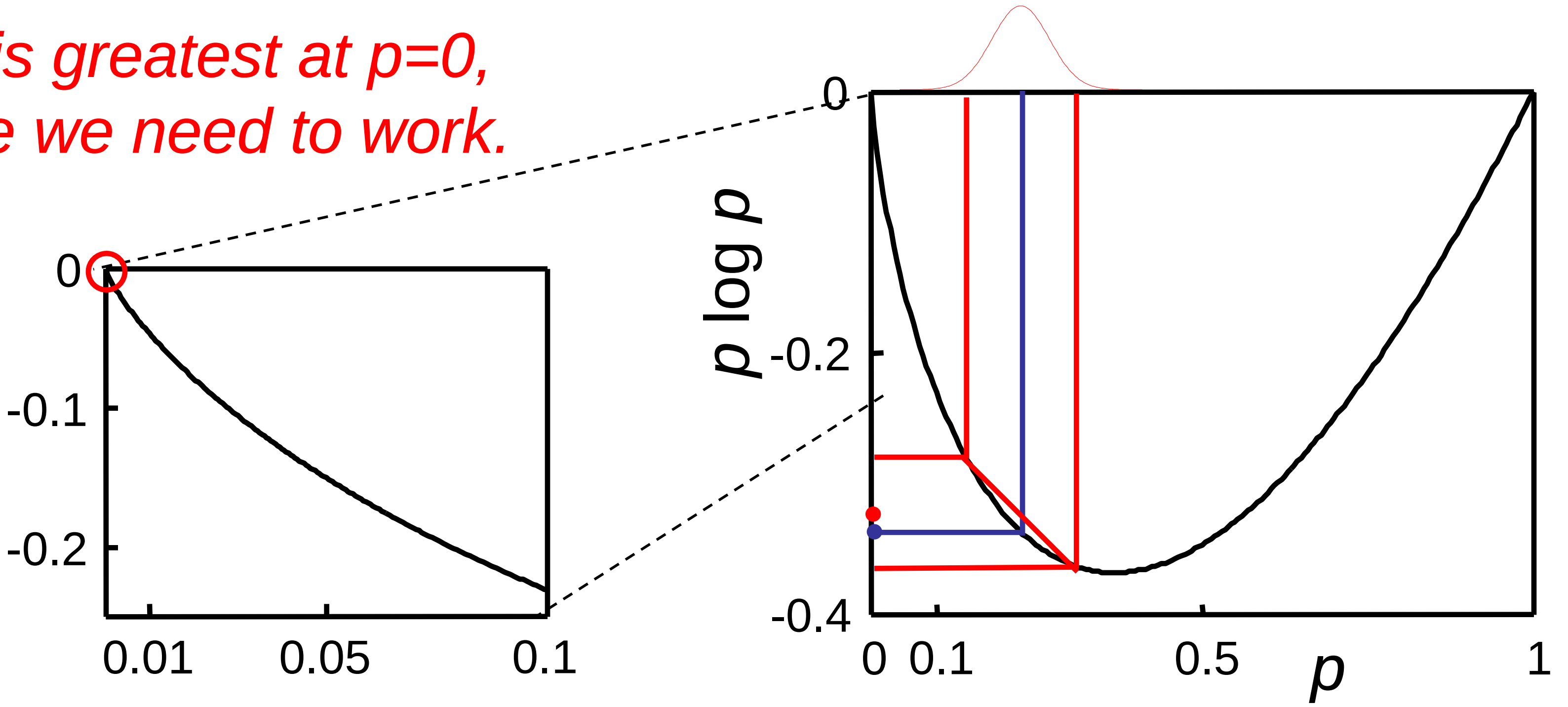
Downward bias is greatest at p=0, just where the data processing inequality wants us to work.

But now there’s another problem

To avoid underestimating information, we need to partition the responses into tiny pieces, each of which has a low probability,p_j.

Our goal is to determine $<p_j \log p_j>$, but we only have an estimate of p_j, not its exact value. This incurs a bias.

The bias is greatest at p=0, just where we need to work.



Avoiding a downward bias due to lumping responses together incurs an upward bias due to the nonlinearity of the log.

Good News/ Bad News

We don’t have to debias every *p* log *p* term, just the sum.

The good news:

The entropy estimate has a bias proportional to **(*k*-1)/*N*,** where *N* is the number of samples and *k* is the number of different symbols (Miller, Carlton, Treves, Panzeri).

The bad news:

Unless *N*>>*k*, the asymptotic correction may be worse than none at all.

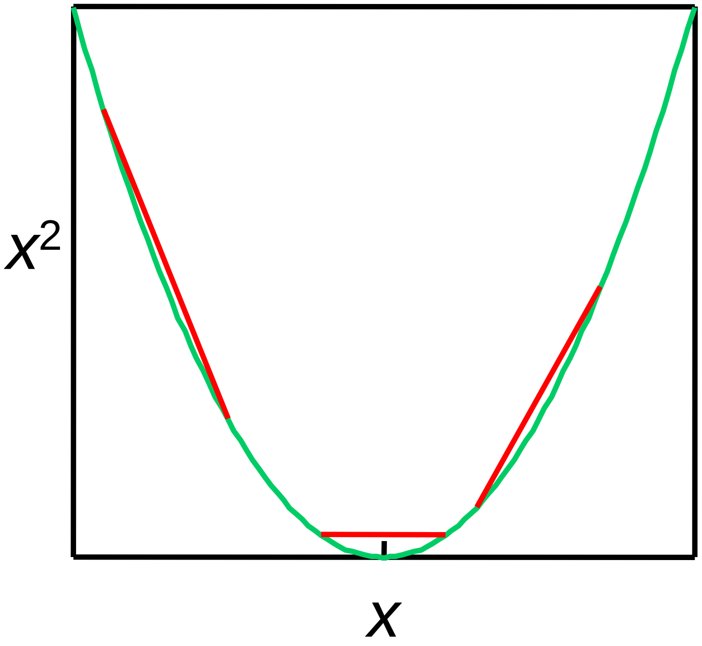
More bad news:

We don’t know what *k* is.

Another debiasing strategy

Toy problem:

$<x^2> \neq <x>^2$



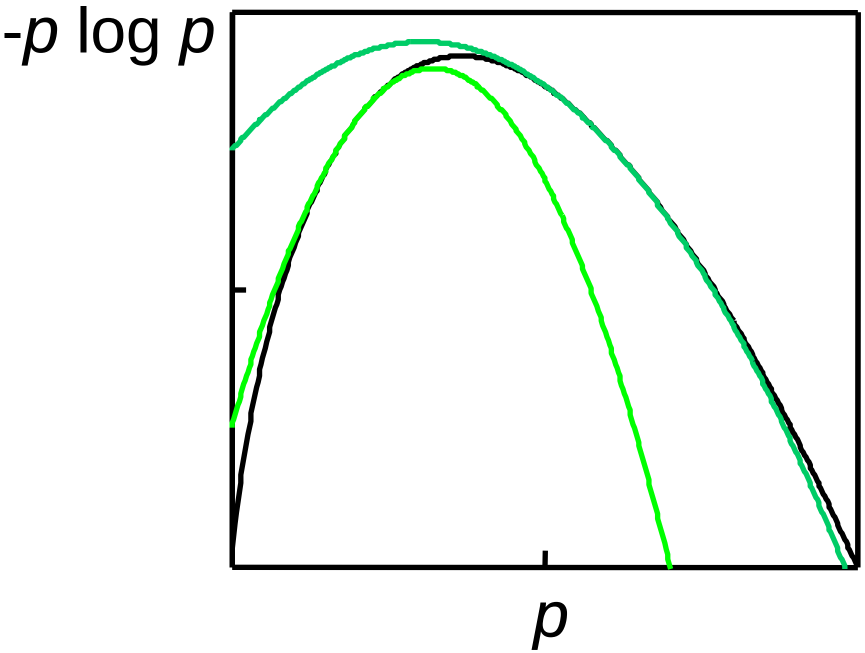
For a parabola, bias is constant.

This is why the naïve estimator for variance can be simply debiased:

$\sigma^2_{est}=<(x-<x>)^2>/(N-1)$

Our problem:

$<- p \log p> \neq -<p> log <p>$



Bias depends on the best local parabolic approximation. This leads to a polynomial debiaser. (Paninski)

Better than classical debiaser, but p=0 is still worst case. And it still fails in the extreme undersampled regime.

“Birthday Paradox” Methods

a.k.a. NSB (Nemenman, Shafee, Bialek 2002, 2004)

Basic idea:

- Don’t need to estimate individual log(p_j)’s, just their average
- Parametric (Dirichlet) model for distribution: probability of p_j is proportional to $p_j^{\beta-1}$
- If β known, then Bayesian estimate of entropy has a closed form (Wolpert and Wolf 1995):

$$H_\beta = -\frac{1}{\ln 2}\left(-\Psi(N+k\beta+1) + \sum \frac{N_j+\beta}{N+k\beta}\Psi\left(N_j+\beta+1\right)\right) \quad \text{where} \quad \Psi(u) = \Gamma'(u)/\Gamma(u)$$

- Bayesian estimate of β from the data
- Clever choice of *a priori* distribution of β *:* flatten the prior for *H*

Summary

- To determine β, you need a good estimate of number of kinds of responses that occur at least twice
- That is: $k \sim N^2$ (“Birthday paradox”), much better than $k \sim N$
- But $k \sim 2\hat{}(T_{word}/T_{letter})$
- Net result: temporal resolution is doubled (halves T_{letter})

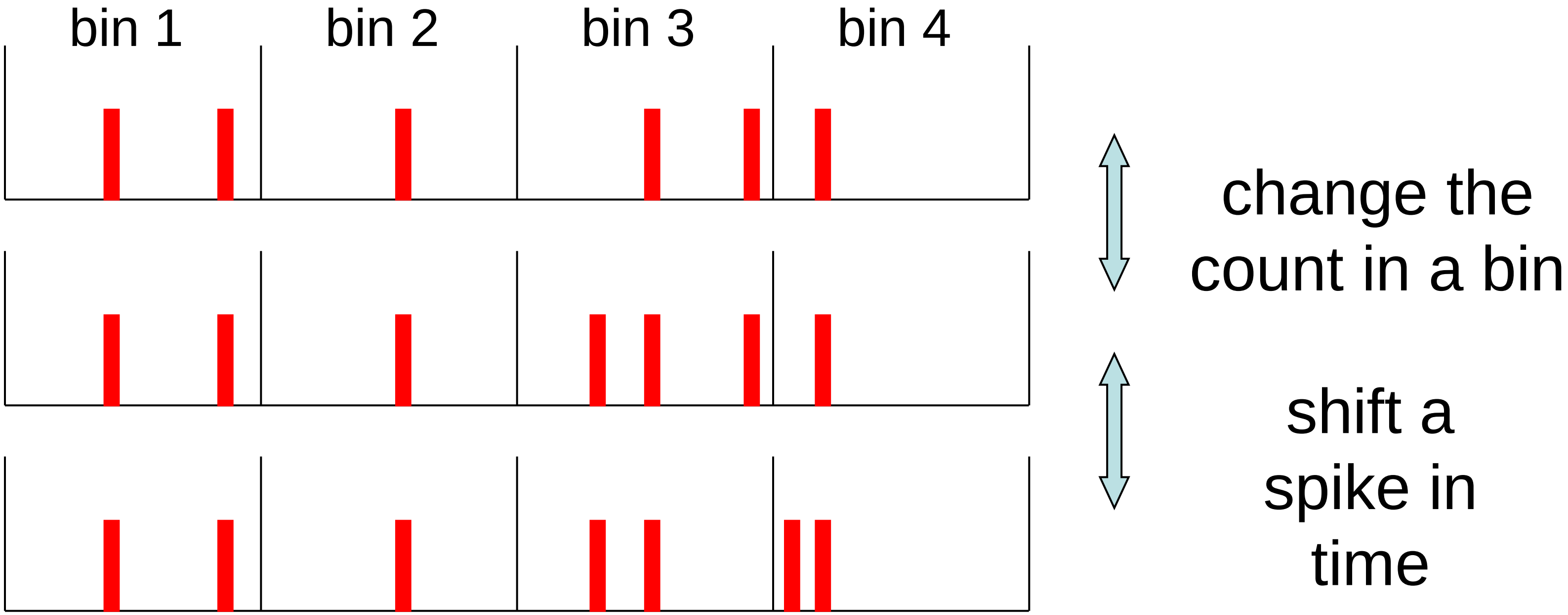
None of these strategies make any use of the fact that the data are spike trains.

Good?

Bad?

What are we missing?

- Each response is a sequence of events
- There are natural notions of “similar” responses
- Similar responses typically have similar probabilities

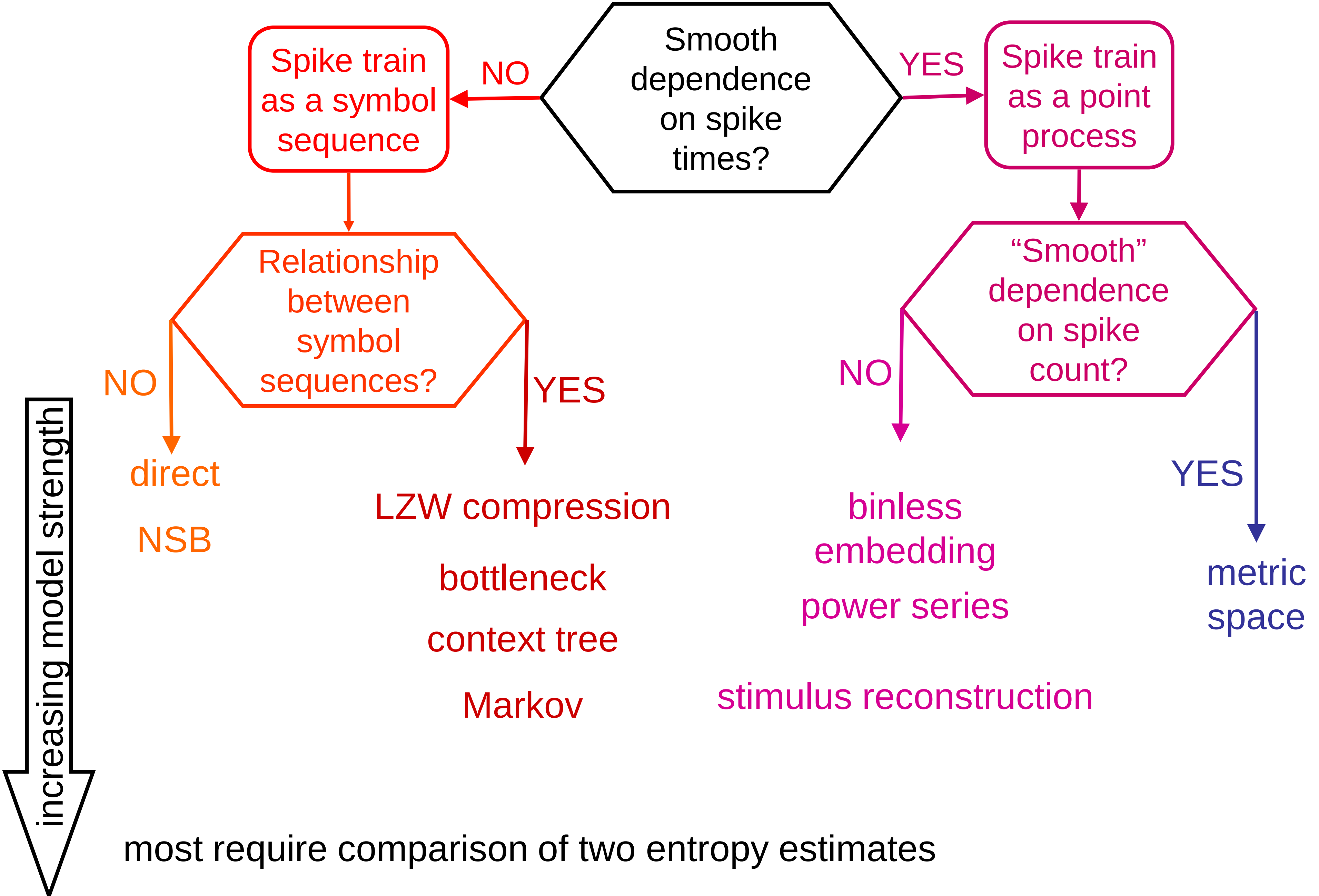


change the count in a bin

shift a spike in time

- But there are many notions of “similarity”
- Entropy can depend on topology, and entropy *estimates* can depend on metrical structure

Strategies for Estimating Entropy and Information



Smooth dependence on spike times?

NO

Spike train as a symbol sequence

YES

Spike train as a point process

Relationship between symbol sequences?

NO

direct

NSB

YES

LZW compression

bottleneck

context tree

Markov

"Smooth" dependence on spike count?

NO

binless embedding

power series

stimulus reconstruction

YES

metric space

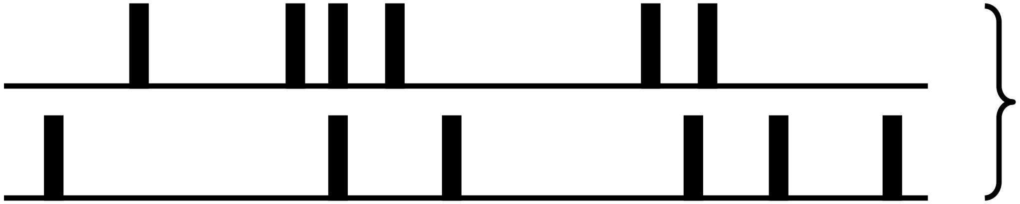
increasing model strength

most require comparison of two entropy estimates

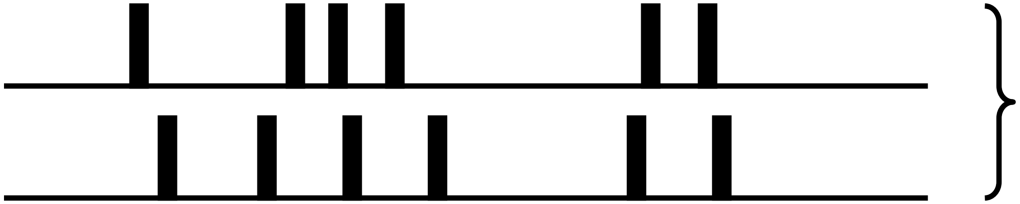
What to do? (Back to Biology)

- Hypothesize the nature of the relationship between the code words.
- This leads to indirect ways of estimating stimulus-response probabilities
- Two strategies
  - If the hypothesis does not yield sufficient information to account for behavior, one can rule out codes (Sheila Nirenberg)
  - More commonly, the calculated information is much more than is necessary to account for behavior. But accounting for behavior is not enough: *do the relationships between the code words account for the perceptual relationships?*

Coding hypotheses: in what ways can spike trains be considered similar?



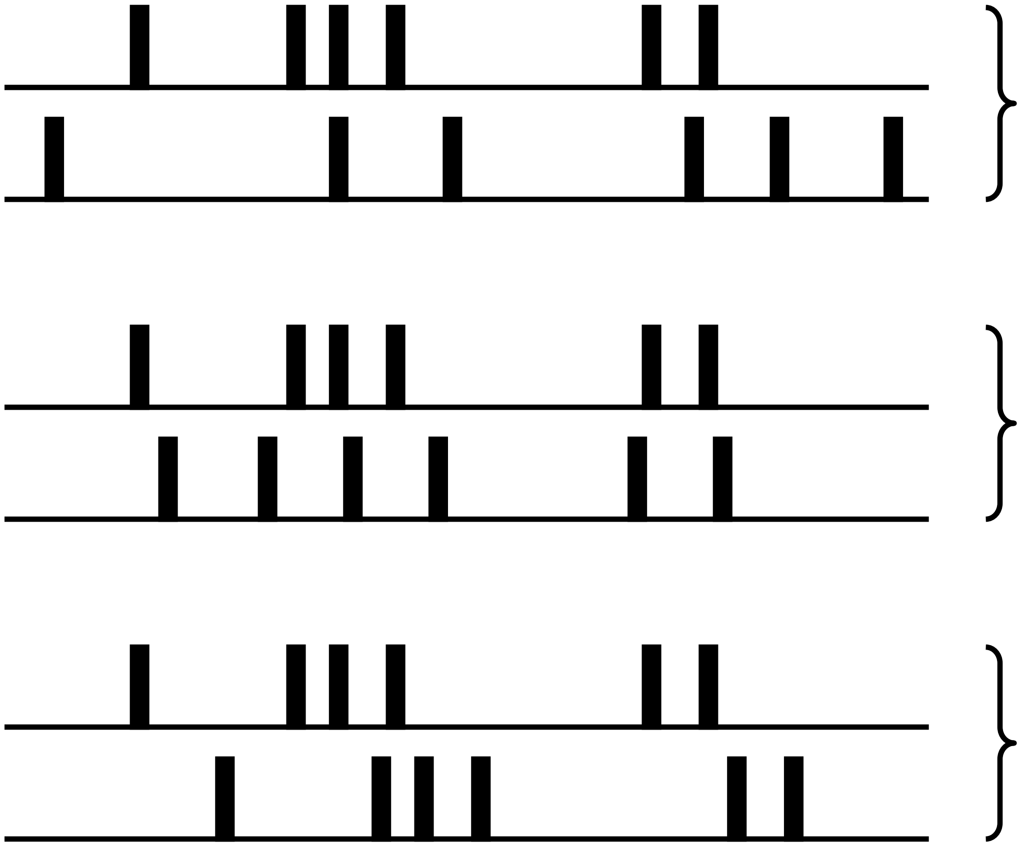
Similar *spike counts*



Similar *spike times*

Similar *interspike intervals*

Coding hypotheses: what notions of similarity are applicable to spike trains?



Similar *spike counts*

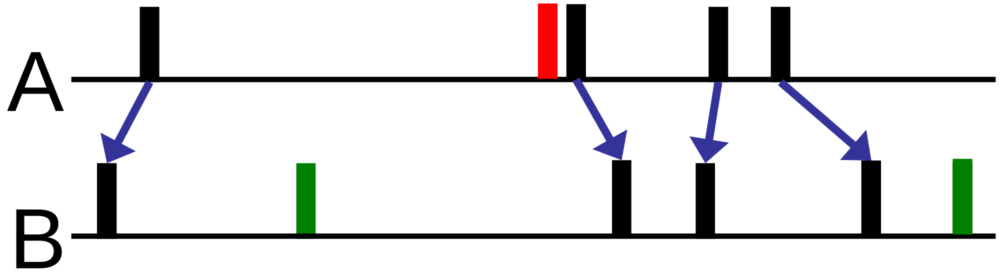
Similar *spike times*

Similar *interspike intervals*

How do we formalize and exploit these notions?

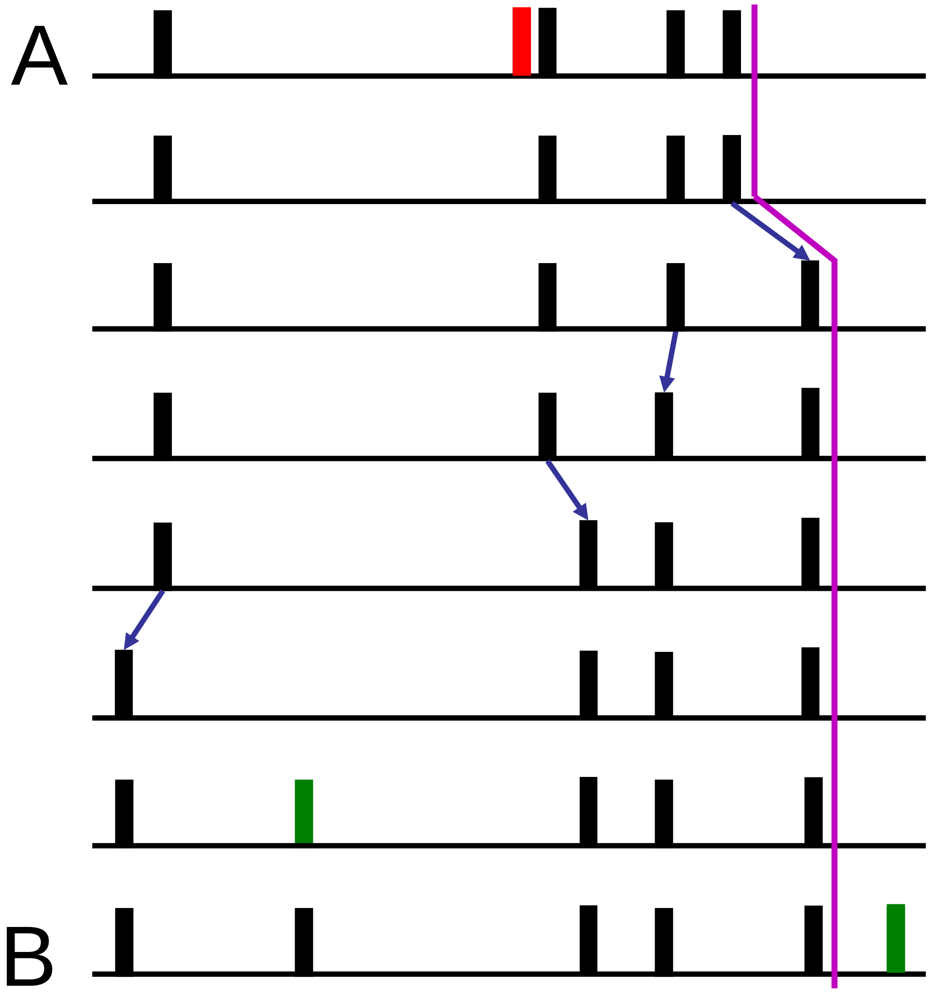
Measuring similarity based on spike times

- Define the “distance” between two spike trains as the simplest morphing of one spike train into the other by inserting, deleting, and moving spikes



- Unit cost to insert or delete a spike
- We don’t know the relative importance of spike timing, so we make it a parameter, q: shift a spike in time by ΔT incurs a cost of q ΔT
- Spike trains are similar only if spikes occur at similar times (i.e., within 1/q sec), so q measures the informative precision of spike timing

Identification of Minimal-Cost Paths



“World lines” cannot cross.

So, either

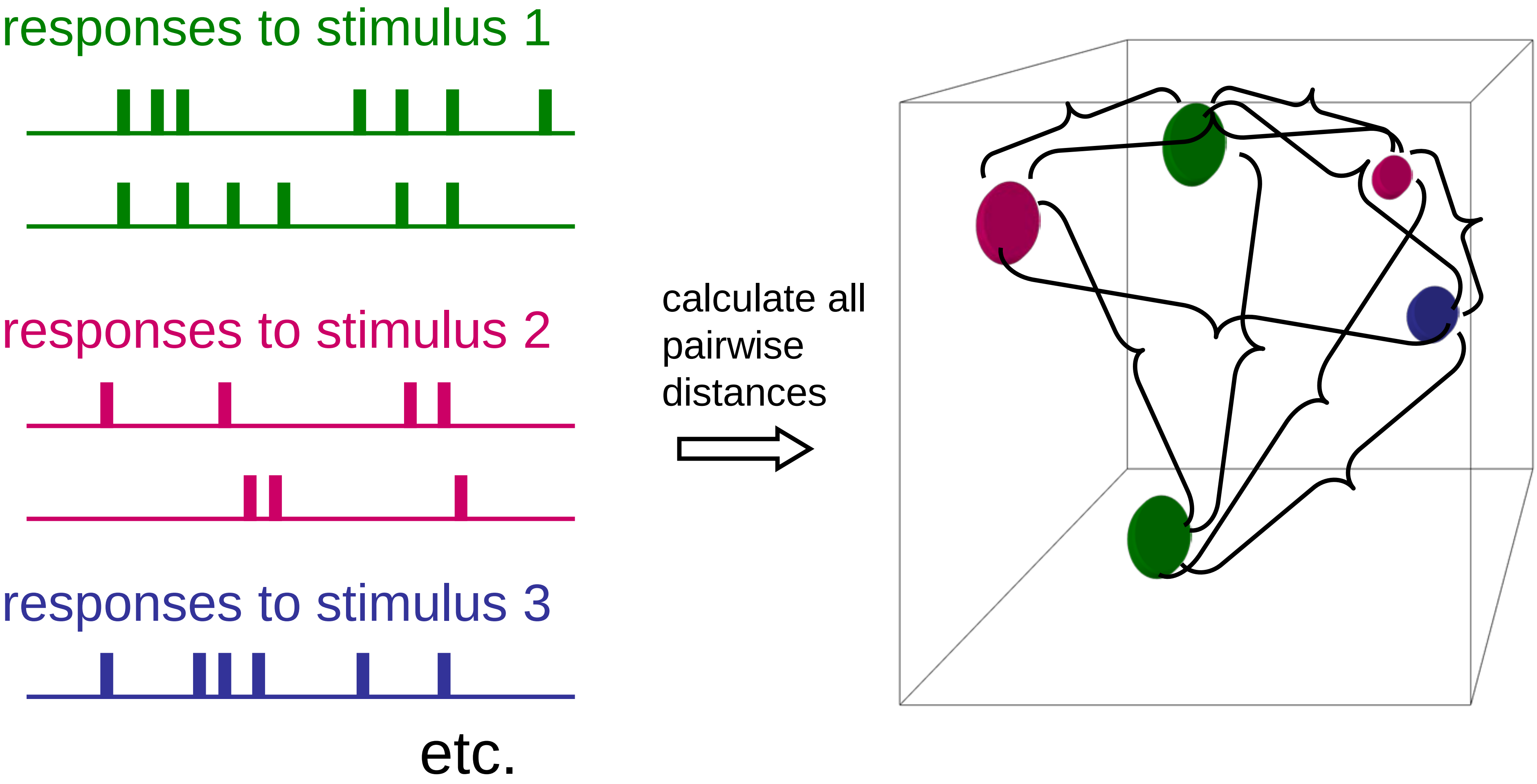
(i) The last spike in A is deleted,

(ii) The last spike in B is inserted

(iii) The last spike in A and the last spike in B must correspond via a shift

The algorithm is closely analogous to the Needleman-Wunsch & Sellers (1970) dynamic programming algorithms for genetic sequence comparisons.

Distances between all pairs of responses determine a response space



responses to stimulus 1

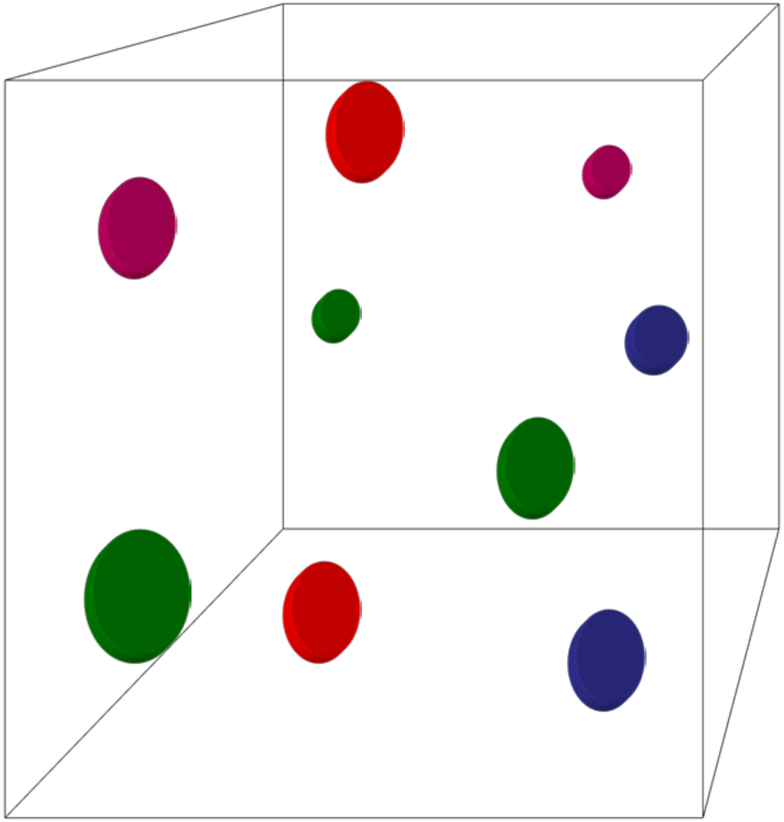
responses to stimulus 2

responses to stimulus 3

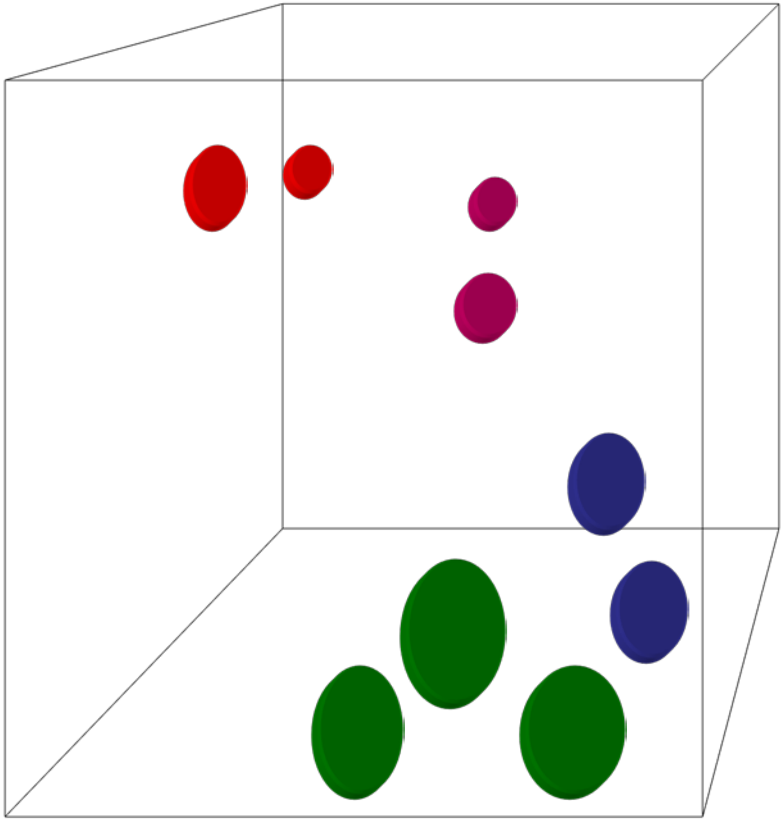
etc.

calculate all pairwise distances

Configuration of the response space tests whether a hypothesized distance is viable



Random: responses to the four stimuli are interspersed



Systematic clustering:

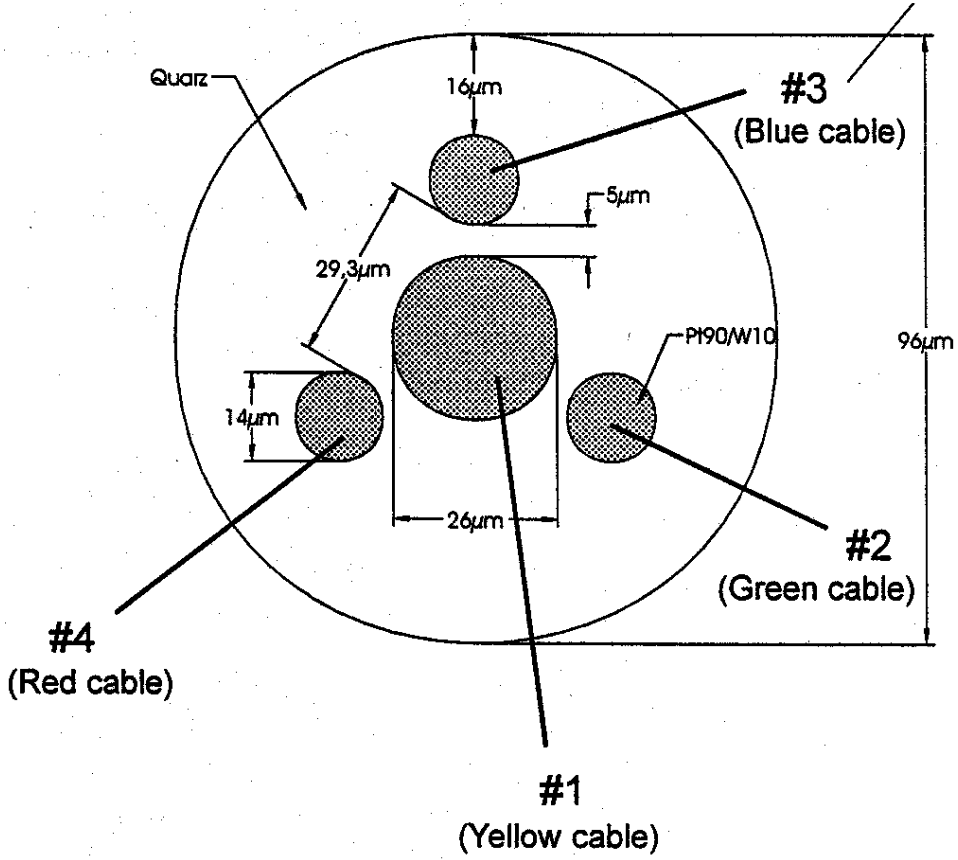
responses to the stimuli are grouped

and

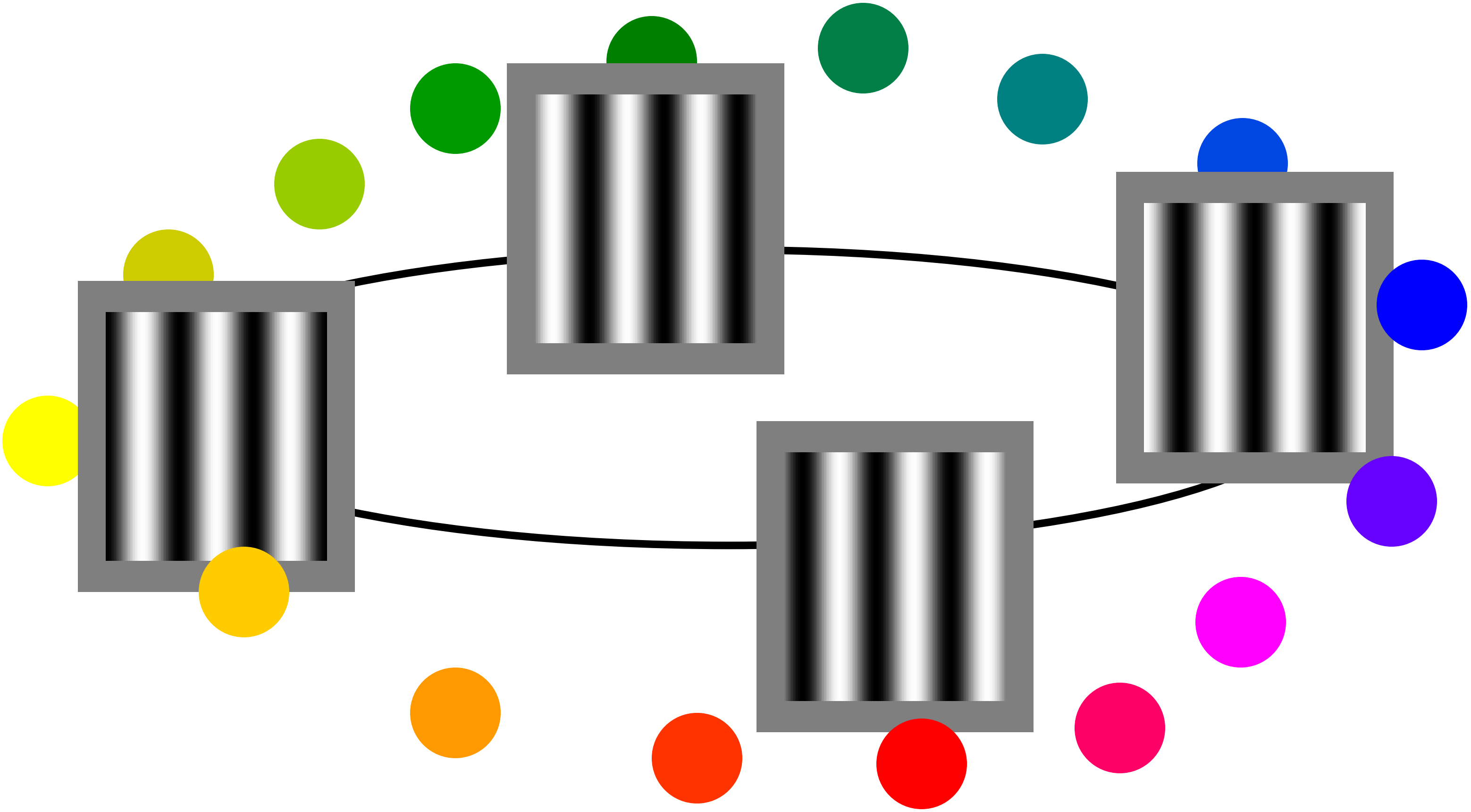
nearby groups correspond to similar stimuli

Preparation

- Recordings from primary visual cortex (V1) of macaque monkey
- Multineuronal recording via tetrodes
  - ensures neurons are neighbors (ca. 100 microns)



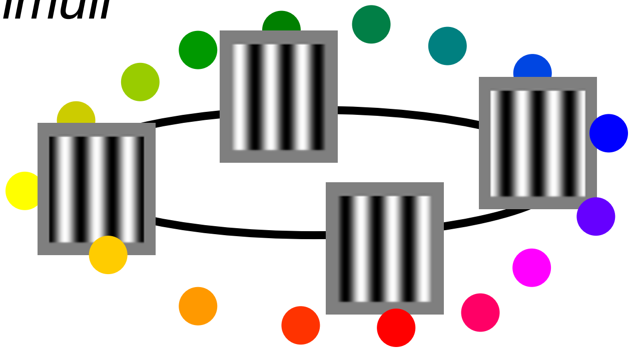
The stimulus set: a cyclic domain



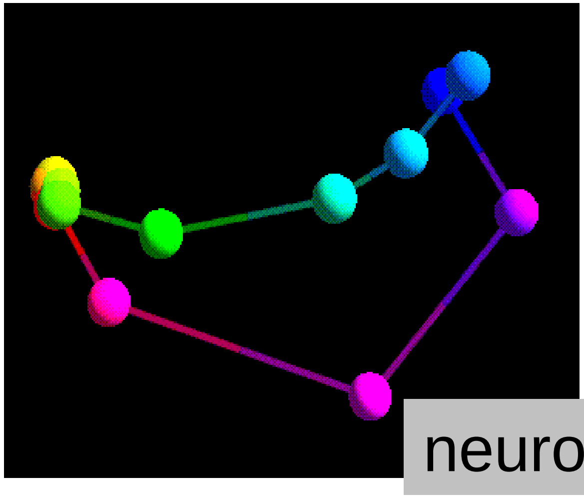
16 kinds of stimuli in the full stimulus set

Representation of phase by single neurons

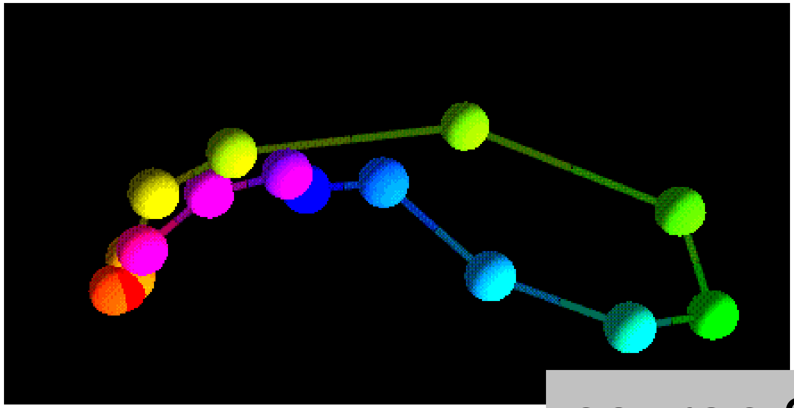
stimuli



reconstructed response space
informative precision q ~ 30 msec



neuron 1

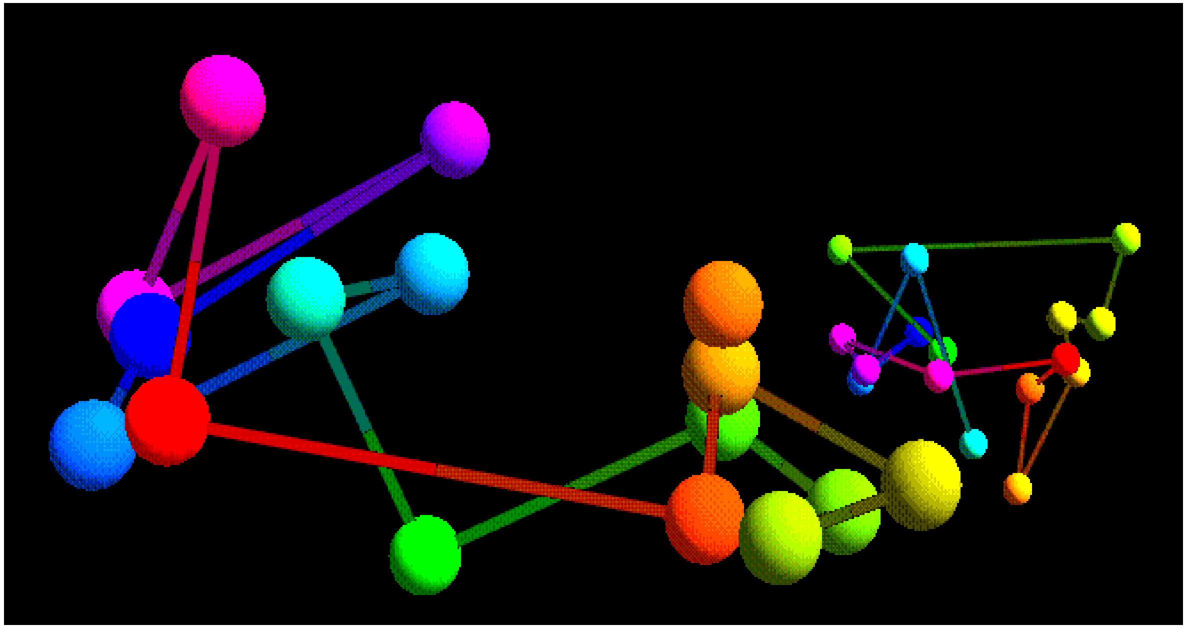


neuron 2

- Spike timing supports representation of a stimulus space

- Nearby neurons represent (somewhat) different portions of the stimulus space

Joint representation of contrast and spatial phase



Distances calculated by spike time metric, with q=64 (informative precision ~15 ms)

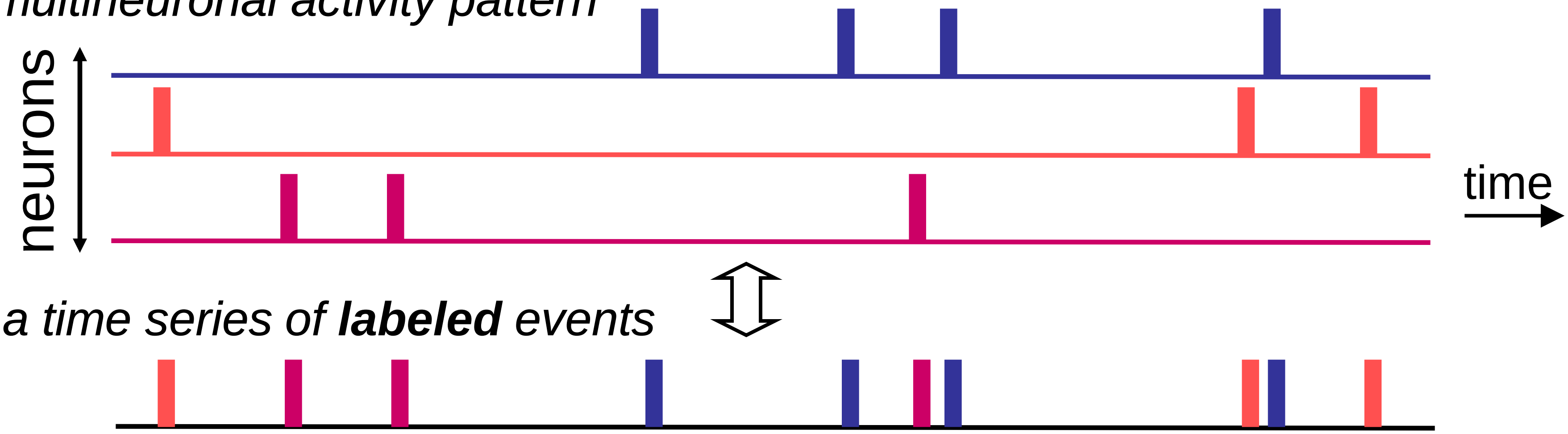
Colors indicate spatial phase (22.5 deg intervals)

Sphere sizes indicate contrast (0.25 and 1.0)

Analyzing coding across multiple neurons

A multineuronal activity pattern

is a time series of ***labeled*** *events*



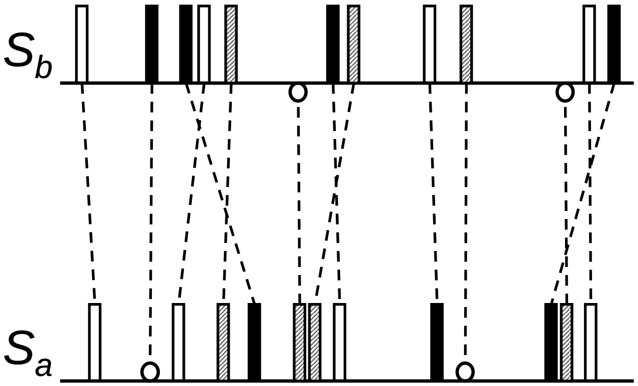
Distances between labeled time series can also be defined as the minimal cost to morph one into another, with one new parameter:

- Cost to insert or delete a spike: 1
- Cost to move a spike by an amount ΔT: q ΔT
- Cost to change the label of a spike: k

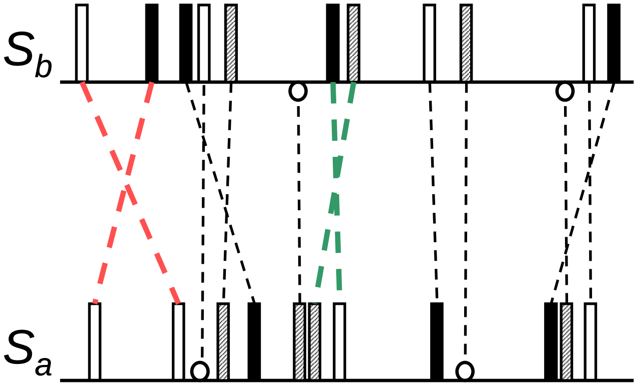
k determines the importance of the neuron of origin of a spike.

- k=0: summed population code
- k large: labeled line code

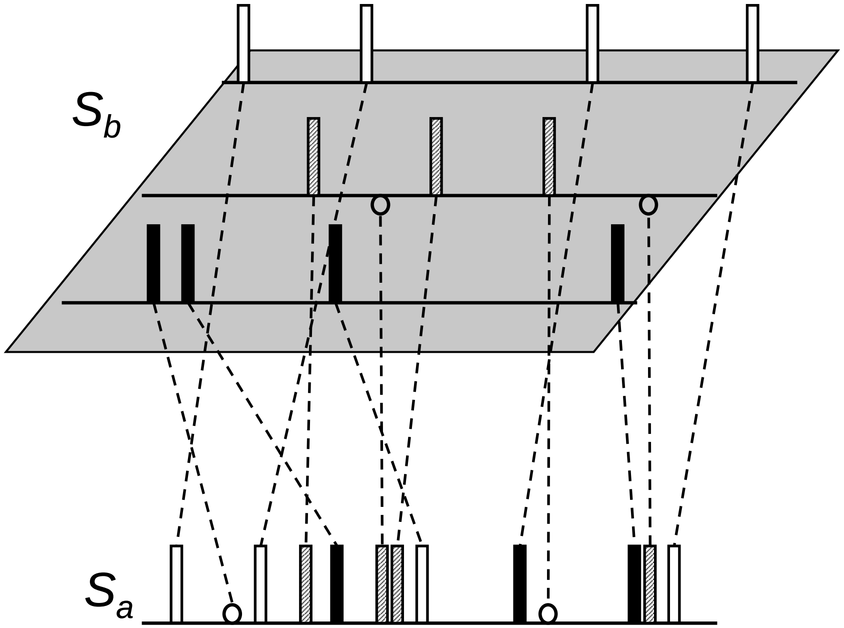
The key to the multineuronal algorithm



a possible alignment



forbidden and allowed crossings

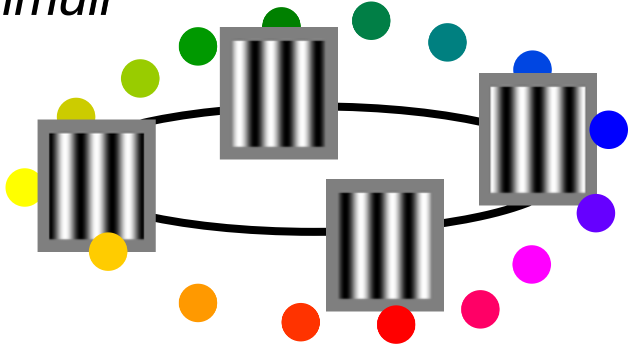


summarized by a 3-d geometry

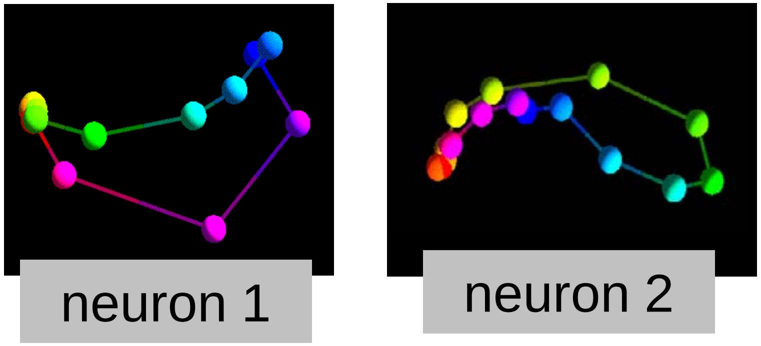
Aronov, 2003

Representation of phase by a neuron pair

stimuli



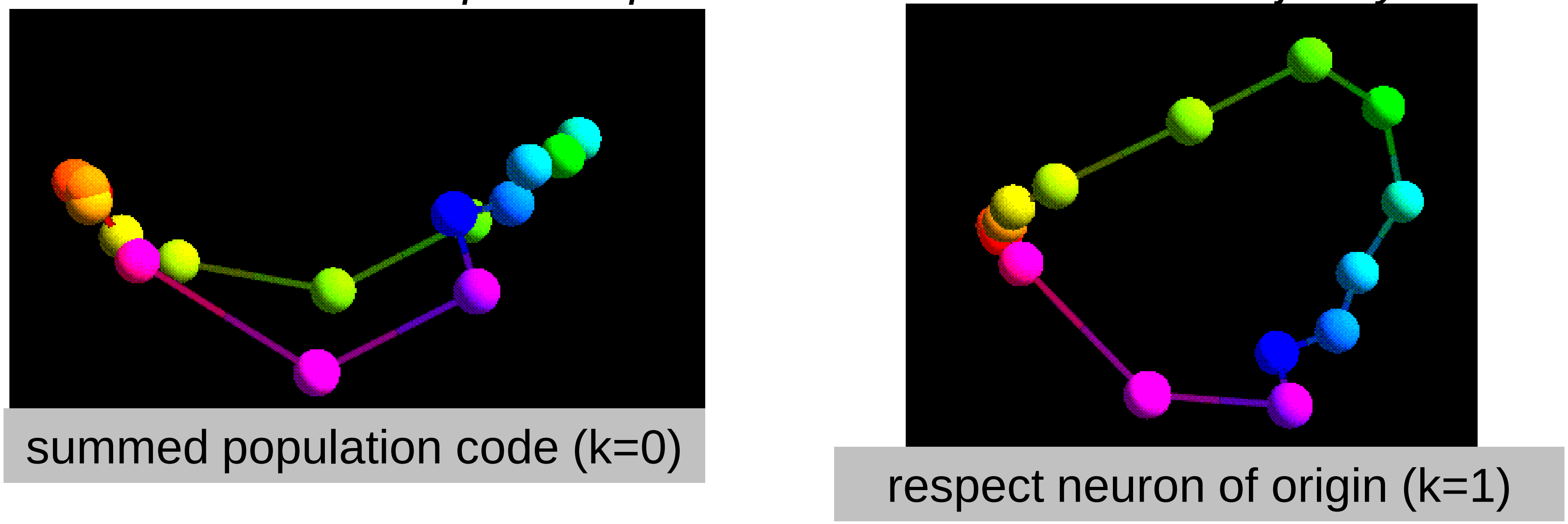
reconstructed response space: each neuron considered separately



neuron 1

neuron 2

reconstructed response space: two neurons considered jointly

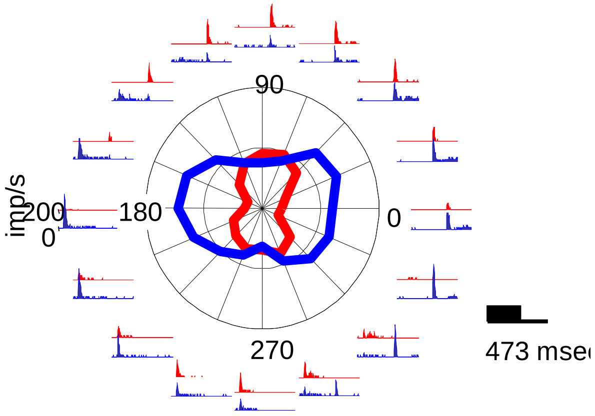


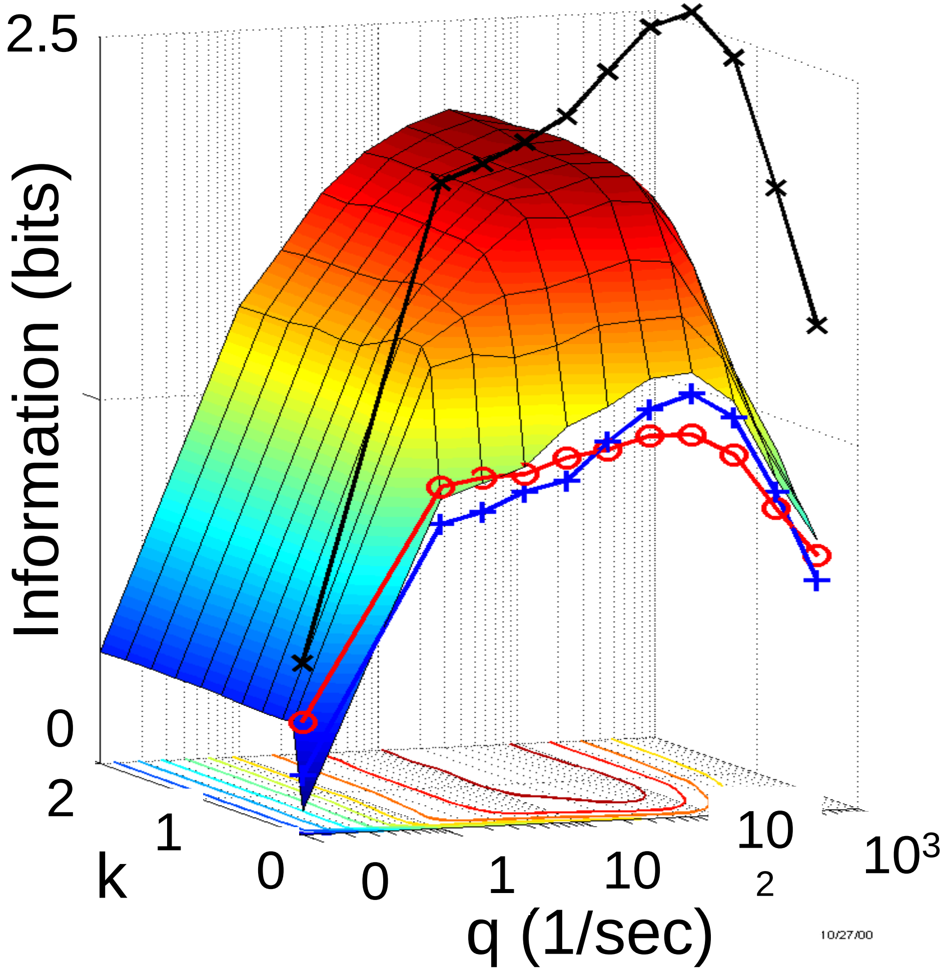
summed population code (k=0)

respect neuron of origin (k=1)

Representation of stimulus space is more faithful when neuron-of-origin of each spike is respected.

Spatial phase coding in two simple cells

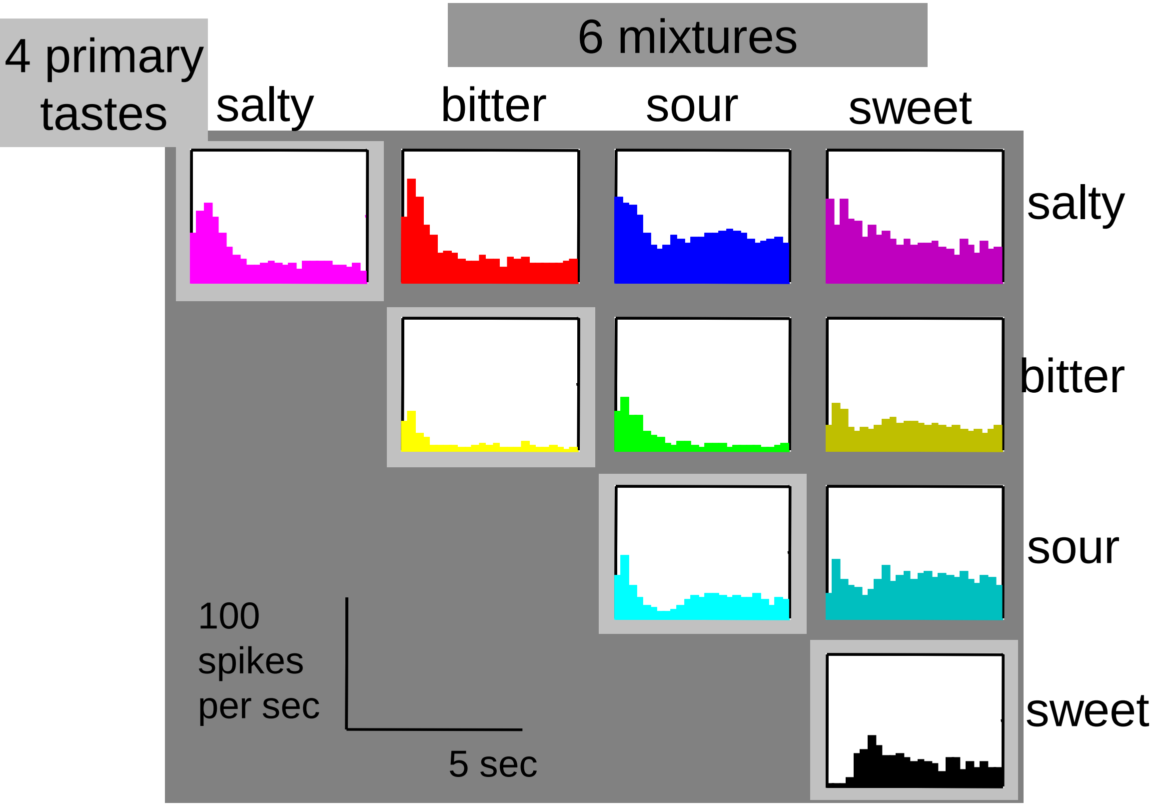




Temporal Representation of Taste

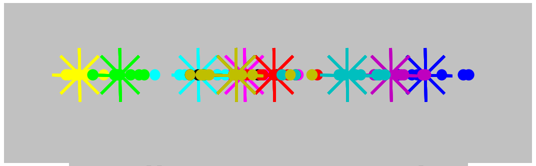
responses of a rat solitary tract neuron

P. DiLorenzo et al., JNeurosci 2009

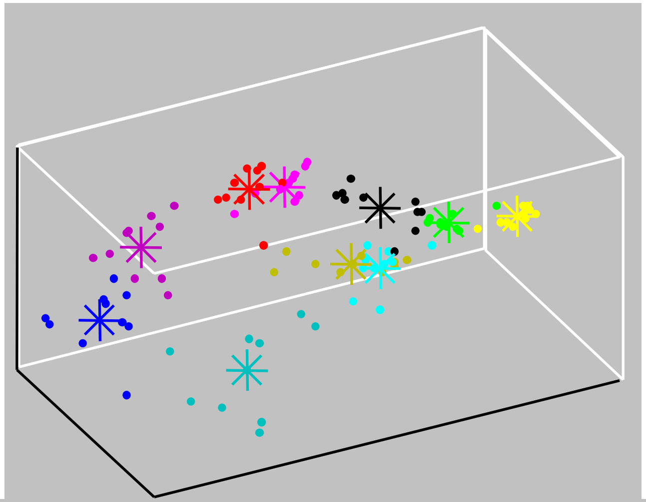


Temporal pattern supports full representation of the 4 primary tastes and their 6 mixtures

reconstructed response spaces



spike count code



spike timing code,
informative precision q ~ 200 msec

Conclusions

- Understanding how neurons represent information is intrinsically both a mathematical and experimental question
- The geometry of a perceptual space can be recovered from temporal features of neuronal responses
  - Single neurons can represent a multidimensional perceptual space
  - In a cluster of neurons with similar properties, the neuron of origin of each spike augments the faithfulness of the representation
  - Firing pattern, as well as overall time course, contributes to this representation
- Neuronal activity is not “intended” to be averaged over time, or averaged over a local population, but to be decoded spike by spike