19th AWCBR (Australian Winter Conference on Brain Research), 2001, Queenstown, AU
https://www.otago.ac.nz/awcbr/proceedings/otago394614.pdf

Do local modification rules allow efficient learning about distributed representations?

A. R. Gardner-Medwin
Dept. Physiology, University College London, London WC1E6BT, UK

Distributed representations permit very many distinguishable events to be coded on a set of cells, with each cell used in many events. Since synaptic modifications can only depend on local influences, there is a fundamental problem learning how often and under what conditions distributed patterns of activity may occur. Frequencies of use and the associations of individual active elements can be measured locally and pooled for active elements within an event, but overlap leads to interference that can only be compensated on an average basis, with inevitable added variance. This constrains the compactness of distributed representations if they are to operate efficiently. Gardner-Medwin & Barlow (2001, *Neural Computation* **13**: 477-504) have employed counting (a form of familiarity discrimination) to explore such constraints quantitatively. Counting underlies estimation of probabilities and association, and is fundamental to learning. Though precise counts are only possible if events have direct representations, high efficiency (i.e. effective use of available data samples) is only possible with distributed representations at the cost of high redundancy. With counts based on usage of individual cells, efficiencies >50% require a number of cells (Z) at least comparable to the number of events (N). Synapses counting activity in pairs of cells can reduce this to around $Z=(6N)^{0.5}$ cells. More sophisticated use of dendritic information (e.g. modification conditions requiring combined activity in 3 cells: adjacent presynaptic terminals and a post-synaptic cell) can improve performance, but nowhere near the information theoretic limit: $Z=\log_2(N)$.

DO LOCAL MODIFICATION RULES ALLOW EFFICIENT LEARNING ABOUT DISTRIBUTED REPRESENTATIONS ?

A. R. Gardner-Medwin

THE PRINCIPLE OF LOCAL COMPUTABILITY

Neural computations must satisfy a constraint - each modifiable parameter is a function only of the history of locally available information.

Realistic models must be capable of expression in this form - without functions of global patterns.

In synaptic terms:

modification can only depend on the history of pre- and post-synaptic membrane and cytoplasm states and local extracellular modulators.

DISTRIBUTED REPRESENTATIONS

Events (significant for learning & inference) correspond to spatial patterns - sets of active cells - and are not usually the sole trigger features for individual cells .

Distributed representations can be compact (representing many distinct events on few cells), but .....

THE OVERLAP PROBLEM

Each cell and synapse is involved in experience of many events, causing interference and inevitable errors in learning and statistics for individual events.

THESE CONSIDERATIONS SEEM TO ENFORCE THIS BROAD STRATEGY :-

Measure statistics for individual cells and synapses → **Combine for all cells or synapses active in an event** → **Correct for the average interference due to overlap** → **Tolerate the variance introduced**

Questions:

- **How can we quantify the variance that is tolerable? - the concept of *efficiency*.**
- **What are the constraints on the number of events that can be learned about, on a given number of cells?**
- **What are the implications for the nature of distributed representations and trigger features, for efficient learning?**

Counting Events - a challenge for distributed representations

With H. Barlow (2001, *Neural Computation* 13: 477-504) we examined the problem of counting events.

We argue that unless a representation can sustain efficient counting of events, it cannot sustain efficient learning.

But ... neural systems do not need to count precisely

Why Count?

Biological success is about prediction
Prediction is about conditional probabilities
Probabilities come from frequencies
Frequencies require counting

But NB typically:

Actual N = expected value $\mu \pm \sqrt{\mu}$

Note that what matters is the estimate of μ, not the value of N

This implies that there is little point in counting much more accurately than $\pm\sqrt{\mu}$ (roughly, $\pm\sqrt{N}$)

Statistical efficiency of an estimation (Fisher)

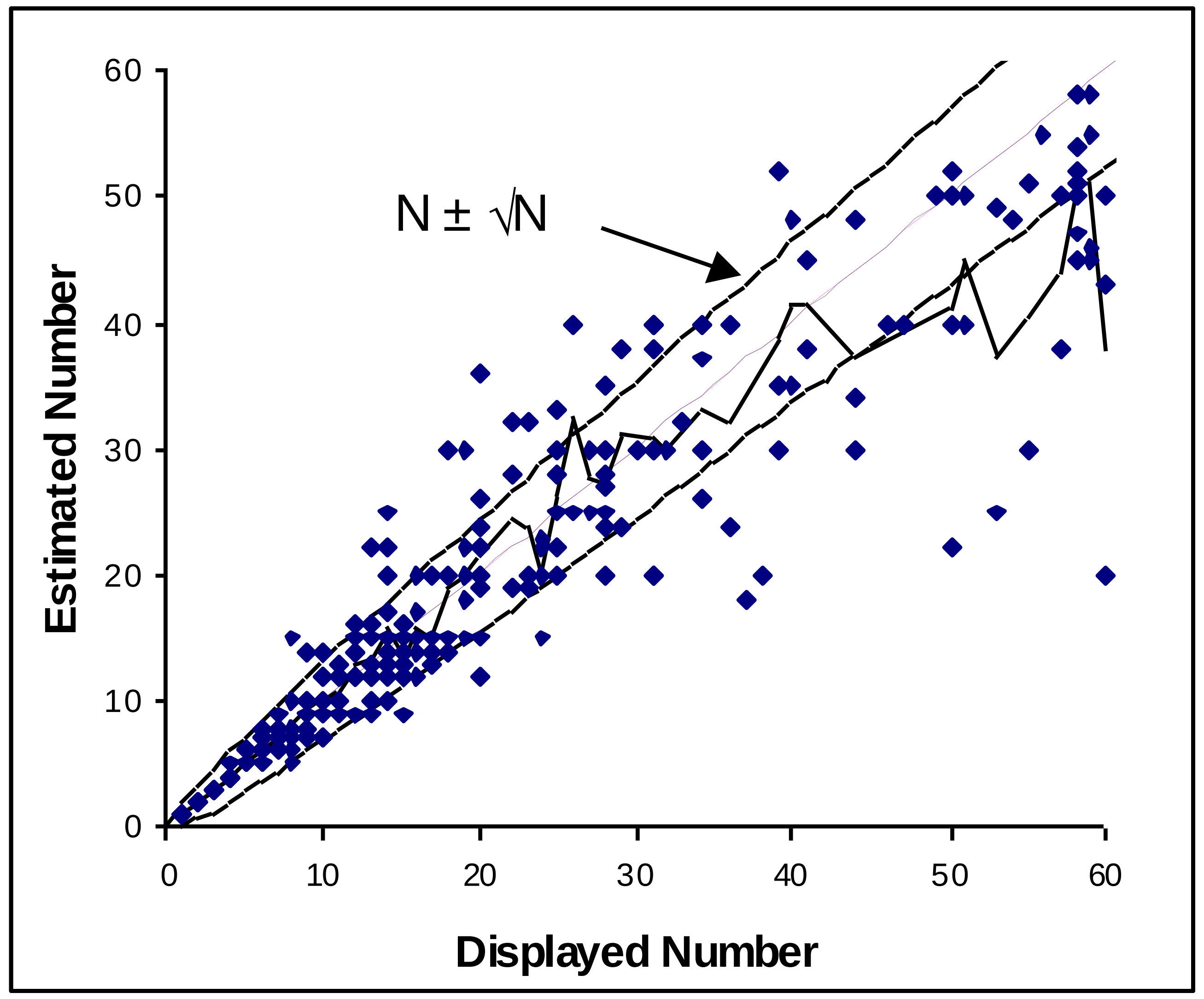
Efficiency = $\dfrac{\text{data needed for a given reliability, with a perfect (counting) algorithm}}{\text{data needed when there is added variance}}$

50% efficiency $\Rightarrow$ it takes twice as long to achieve reliable inferences

Counting a Poisson process with variance = V :- ***Efficiency = 100% x $\mu / (\mu + V)$***

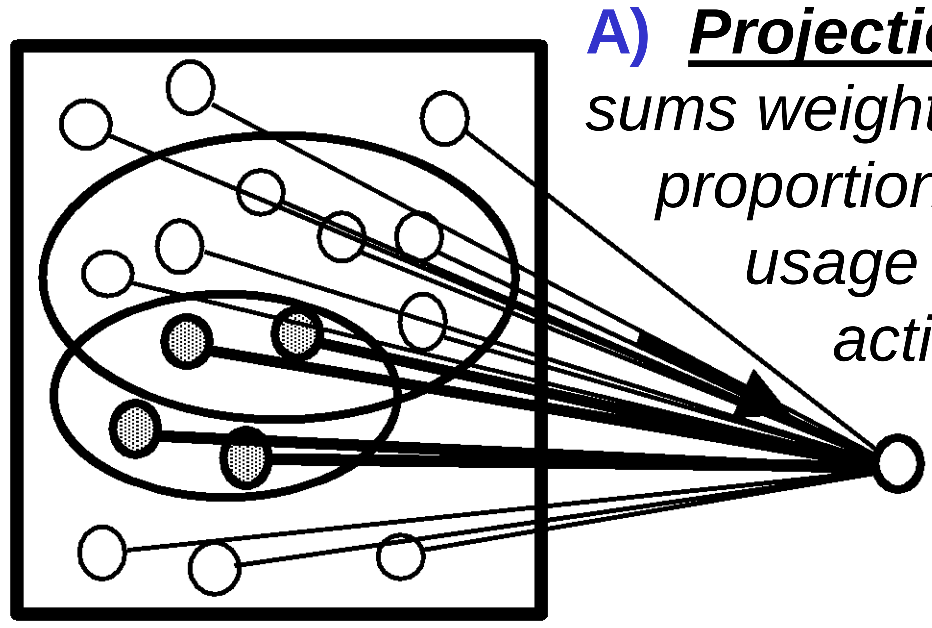
Estimating the count of randomly scattered X's

When forced to subitise (i.e. to estimate a count without sequential matching to linguistic symbols, fingers, etc.) human counting efficiency falls with large numbers, but is typically >50% for counts in the spatial domain.

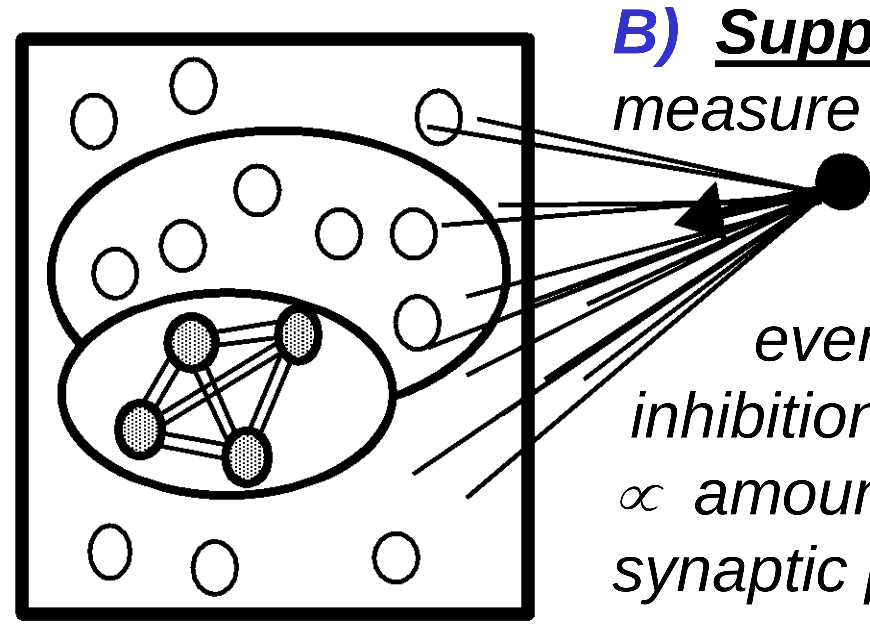


Dept. of Physiology, University College London, London WC1E 6BT, UK

Models & simulation examples of performance based on usage of (A) cells, (B) pre- & post-synaptic cell pairings and (C) triplets of cells

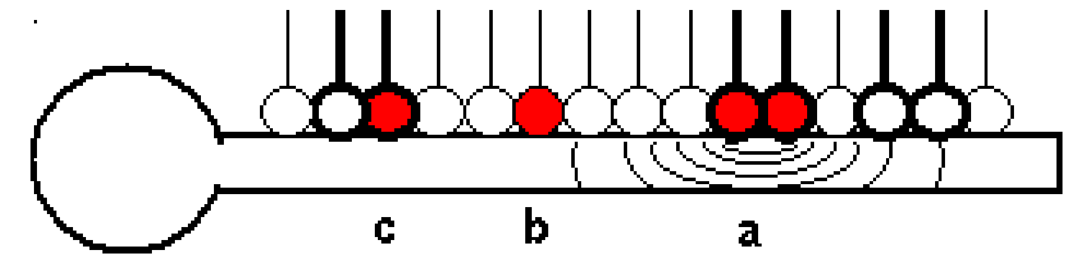


A) ***Projection model :***
sums weights proportional to usage of cells active in a test event.

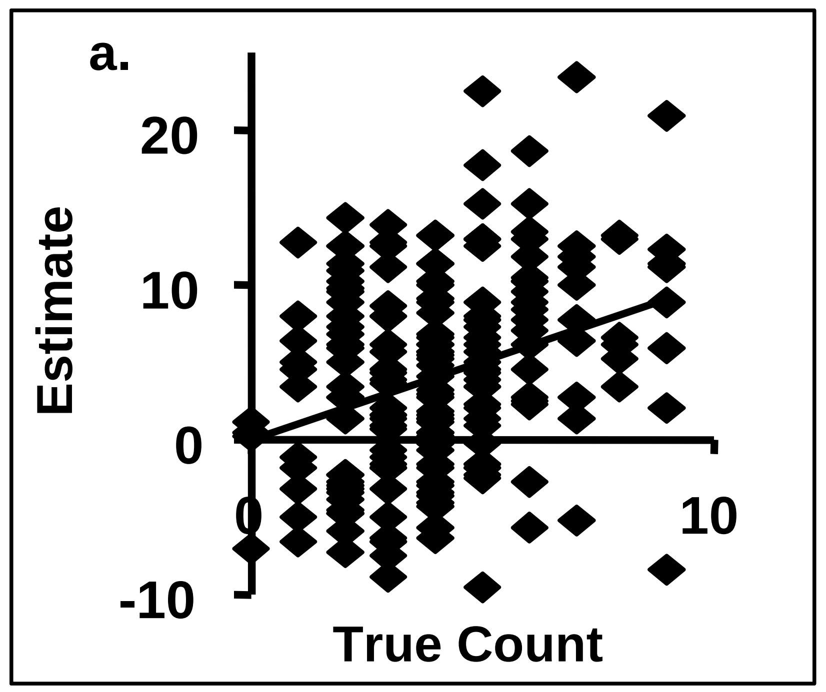


B) ***Support model :***
measure auto-associative support for activity in a test event using imposed inhibition. Synapse weights $\propto$ amount of pre- and post-synaptic paired activity.

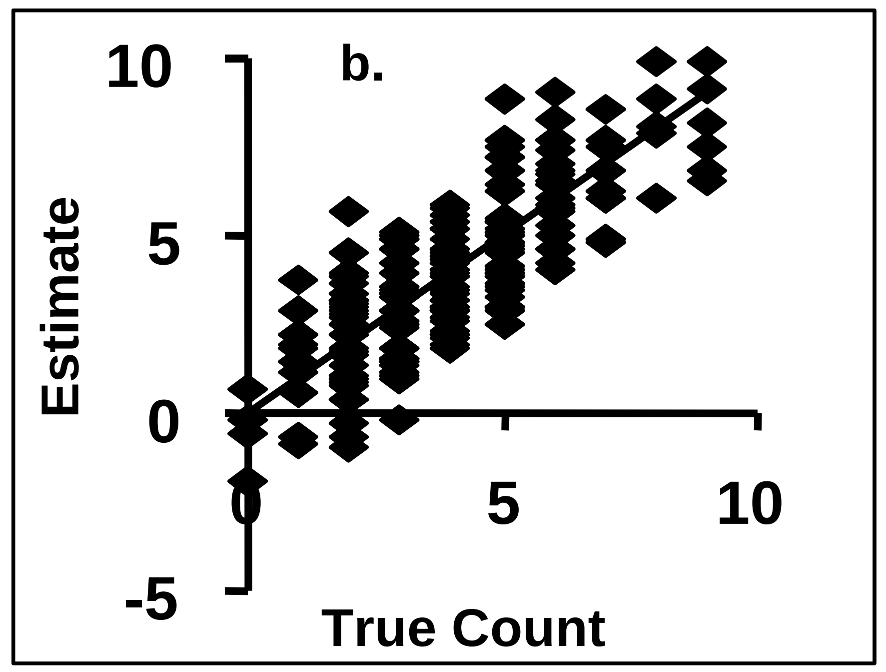
C) **Triplet support**



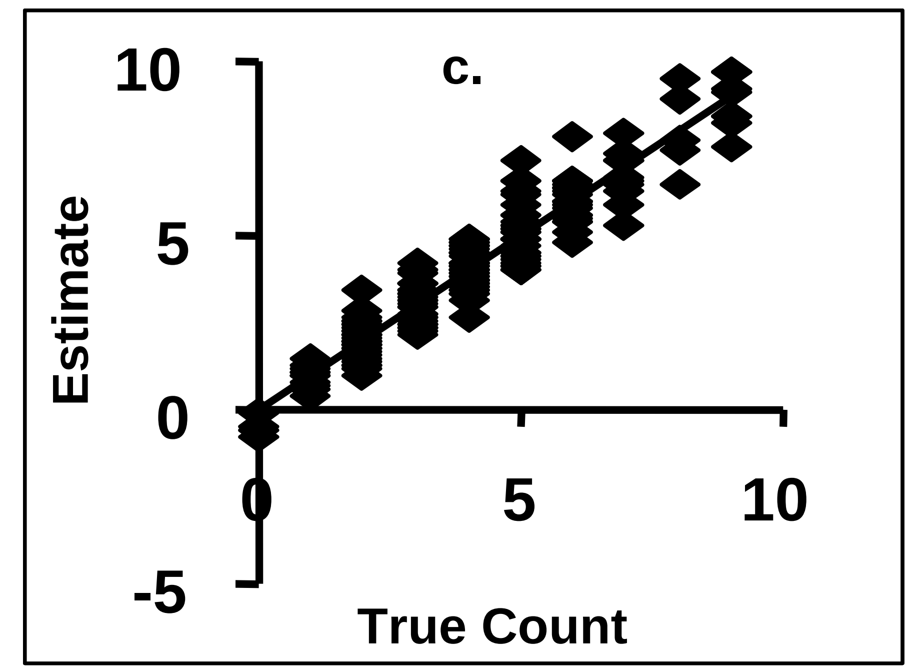
As B, but with synapses that strengthen with co-occurrences of activity in 3 cells: pre- and post- plus an adjacent pre- axon. Under test conditions (recall) active synapses (red) sum their strengths, but only when adjacent pairs are active (a), not single synapses (b,c).



A) Counting efficiency =10%



B) Counting efficiency =65%

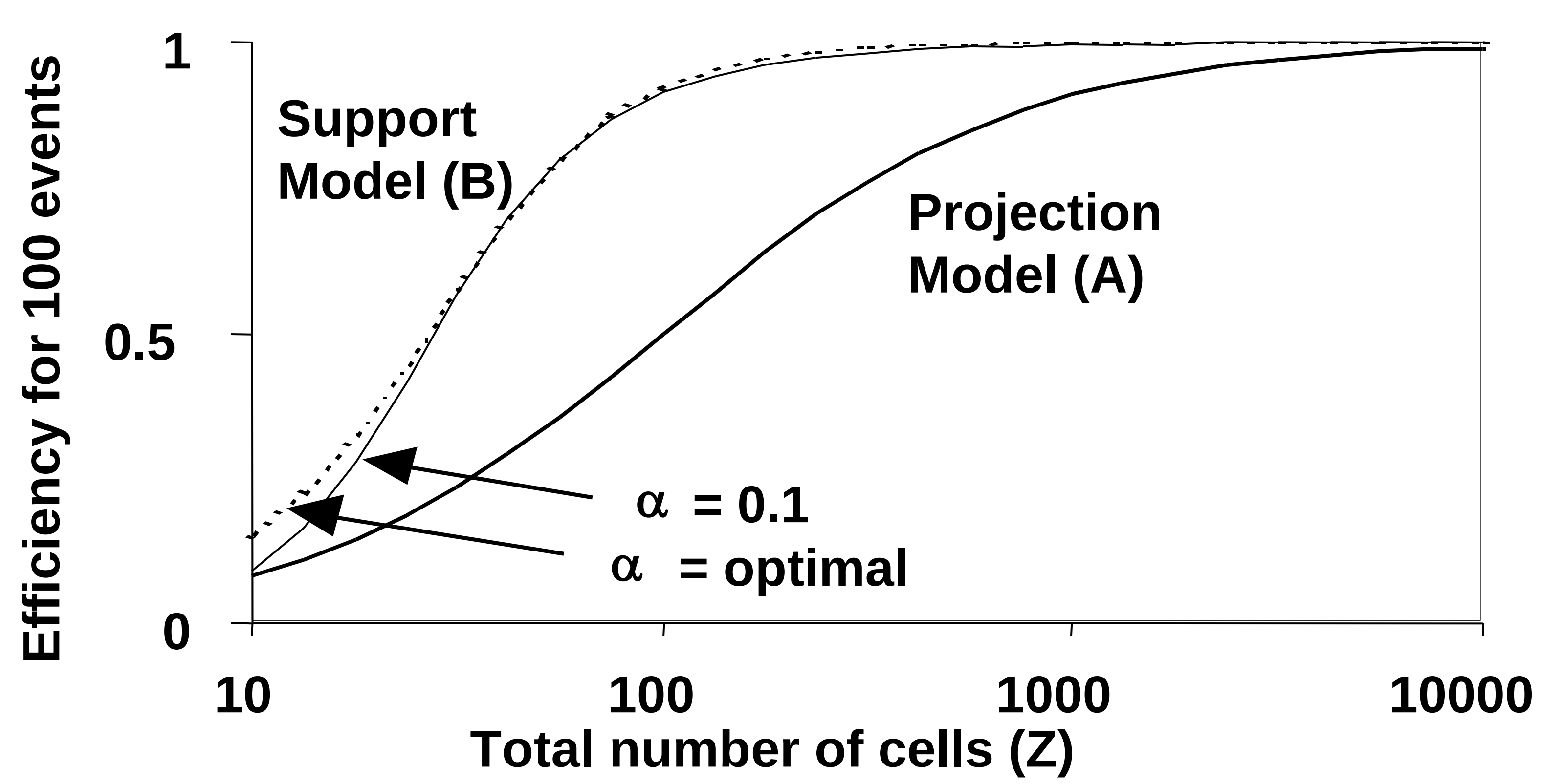


C) Counting efficiency =91%

Simulation results (shown above) with the 3 models:-

- 200 distinct events were represented by random selections of 10 binary active cells on a network of 100 cells
- Events occurred differing numbers of times shown on the horizontal axes (Poisson distributed, mean=4)
- Simulated estimates of the number of occurrences are plotted on the vertical axes, from activation on test presentation of each event.
- For (C) cells had 20 afferent synapses from each other cell, shuffled so that neighbours were never identical, but otherwise random.

Distributed representations can support high efficiency, but they must employ very many cells compared with a compact representation for the counted events.



- NB 100 cells would allow each event to have a direct representation, giving 100% counting efficiency.
- 7 cells suffice for a compact representation distinguishing 100 events.

α (the activity ratio) is the fraction of cells active in each event.

How can better representations improve efficiency?

1. Use more cells & trigger features than there are events of interest

E.g. counting 100 equiprobable events:-

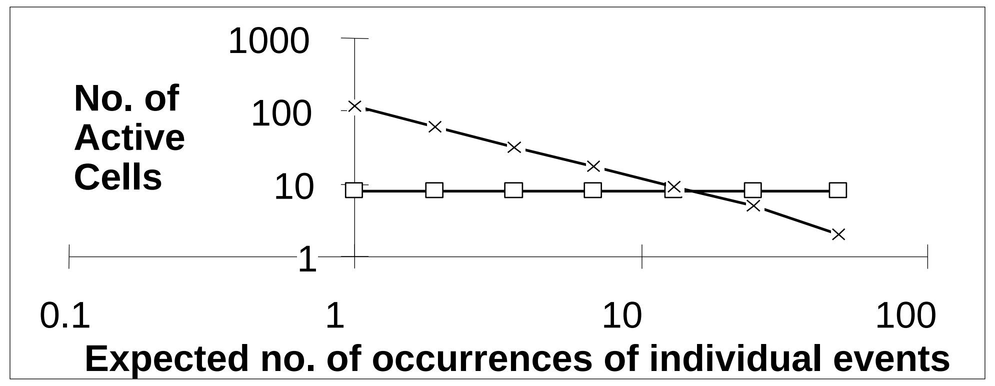
active cells	total cells	Efficiency (projection model - A)	Efficiency (support model -B)	number of distinct reps
12	30	22%	45%	10^8
4	30	22%	63%	10^4
5	100	50%	93%	10^8
4	300	75%	98%	10^8
Direct Reps:				
1	100	100%	-	10^2
1	10^8	100%	-	10^8

- NB distributed reps handle almost any 100 of the many distinct events.
- Direct reps must be set up in advance ($\Rightarrow$ restricted to familiar events)

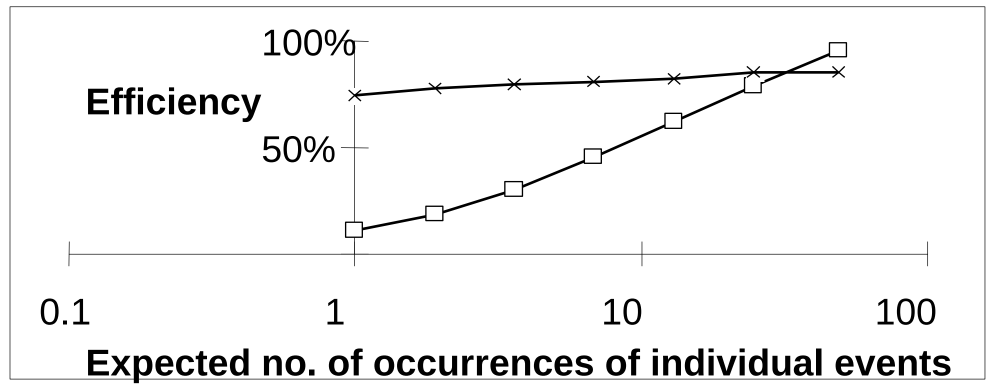
2. Vary the number of active cells between events

Rare events tend to have poor counting efficiency because of overwhelming interference from overlap with common events.

- Increase the number of cells active in rare (and important) events, e.g. through alerting & orienting reactions and selective attention.



- Reduce the number of cells active in common events through habituation, adaptation.



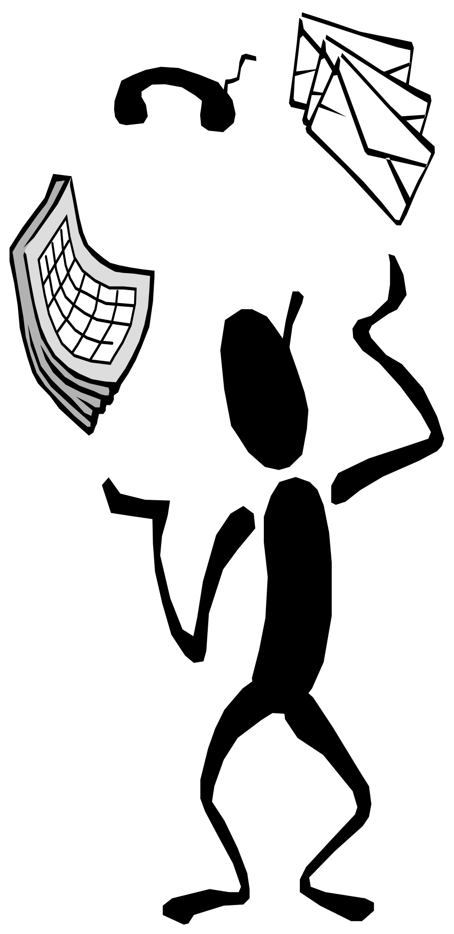
3. Choose trigger features appropriately

Trigger features (i.e. the conditions that lead to activity of individual cells) should ideally be such that :-

- there is minimum overlap between representations of different events, especially ones that have different significance in relation to learning.
- the subset of events in which a feature is active tend to have similar associations, and significance in relation to learning.

Conclusions - Distributed vs. Direct representations

- Distributed representations can be compact (requiring few cells), but do not permit 100% counting efficiency.
- Distributed reps do permit high efficiencies (>>50%) but only if they are highly redundant (comparable to or greater than direct reps).
- **Distributed reps allow novel events to be counted without prior provision for special encoding, even when the interest in a type of event emerges only after the experience. In this context distributed reps have a huge advantage over direct reps.**
- Distributed reps permit the boosting of counting efficiencies for rare or significant events by variation of activity ratios through mechanisms such as attention and habituation.
- Direct reps allow the processing of simultaneous events in parallel. Distributed reps require that simultaneous events be processed serially, as in a stream of conscious awareness.



Serial processing of simultaneous events