Optimization Principles in Neural Coding and Computation

Crick-Jacobs Center Seminar
Salk Institute for Biological Studies
Monday, 20 December, 2004

William Bialek

Joseph Henry Laboratories of Physics, and
Lewis-Sigler Institute for Integrative Genomics
Princeton University

http://www.princeton.edu/~wbialek/wbialek.html

The brain is not a general purpose computer ...
so it does some things "better" than others

What do we mean by "better"?

- more reliable decisions
- actions that increase our rewards
- more accurate estimates
- use fewer resources
- faster reaction times
- learn the rules from fewer examples
- simpler processes for developing the circuitry

Saying what we mean by "better" forces us to be very specific about what problem we think the brain is solving ... note that we could be wrong.

If we understand "better" we may also find that there is a "best": the optimal performance given sensible constraints

The classic example: Photon counting in vision

pre-history 19th century measurements of minimum energy for a visible flash; Lorentz' suggestion (1911)

classical psychophysical experiments Variability of human responses to dim flashes follows Poisson statistics of photon arrivals; consistent with "threshold" of 5-7 photons spread over ~500 rod cells (1940s)

dark noise Detection as discrimination of signal vs noise, trade sensitivity vs reliability, can reach K=1; effective dark noise ~ 1 event/minute/rod, lifetime of rhodopsin ~ 1000 years! (1950s)

direct observation of rod responses to single photons single photon current ~ 10 x (continuous background noise), highly reproducible from photon to photon, dark noise at level predicted from behavior (1970s); "toad cooling" experiments (1980s)

These observations pose challenges at many levels:

Dynamics of the rhodopsin molecule itself

Dynamics of the biochemical network for amplification of single molecular events

Filtering and nonlinearity in the synaptic network of the retina

Learning

In each case there are notions of optimization ...

rod cells produce a current I(t) ...

although this is a model, you can measure everything (!)

$$I(t) = \underbrace{\sum_i I_0(t-t_i)}_{\text{responses to single photons at times } t_i} + \text{noise}$$

responses to single photons at times t_i

responses to thermal isomerization events
+
continuous (~ Gaussian) background noise

what does the brain want to do with these currents?

"noise" obviously is irrelevant and should be suppressed
more subtly, even photon arrival times are not relevant in themselves

because statistics of photon arrivals are Poisson, the only things of possible relevance to behavior are functions of the photon rate r(t)

tempting to formulate statistical inference problem:
rod currents $I(t)$ -> photon arrival rate $r(t)$

but this isn't right ... presumably brain isn't interested in $r(t)$ = light intensity, but only in some features.
which features?

at low light intensities it doesn't matter because we have sufficient statistics:

$P[r(t)|I(t)]$ depends only on a filtered version of the rod current
filter $F(t)$ = matched filter for $I_0(t)$ against spectrum $N(\omega)$ of "noise"

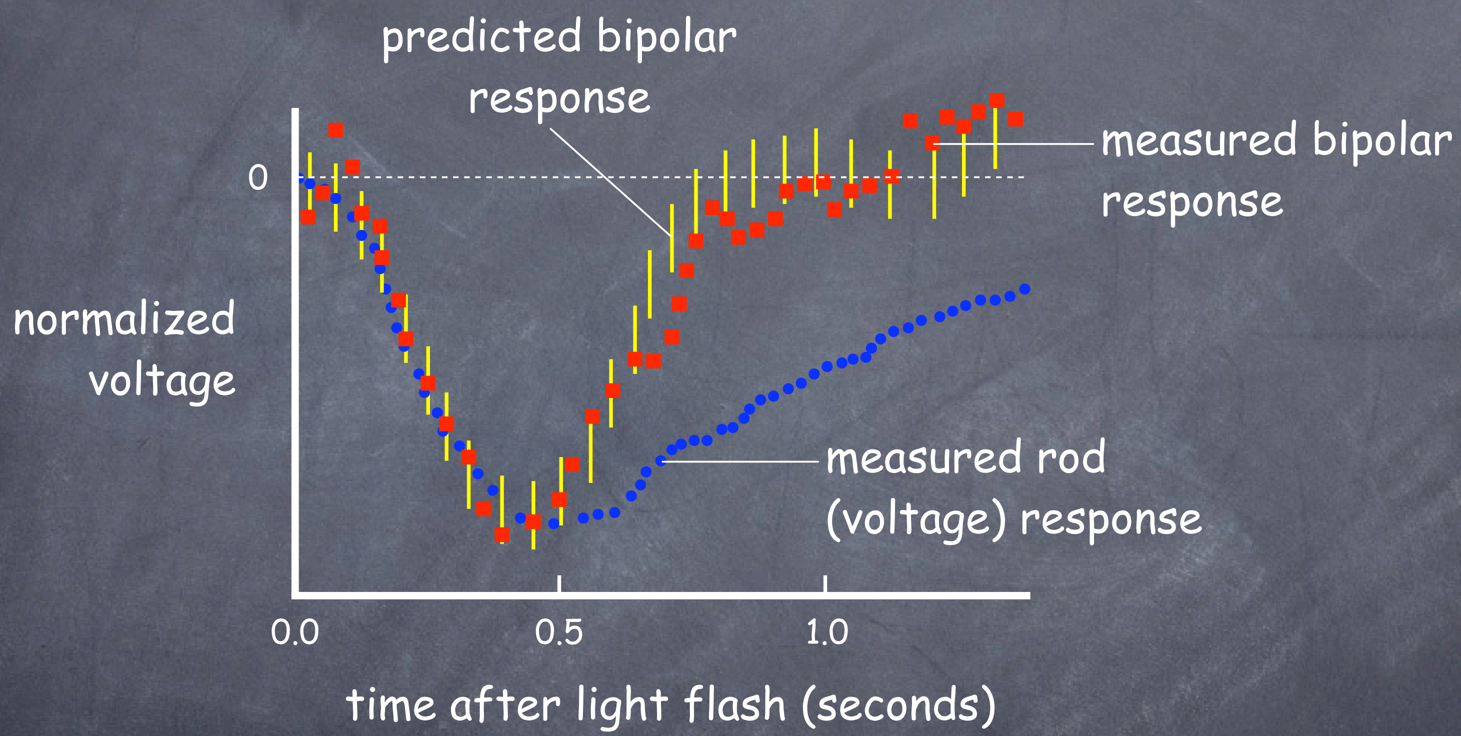
$$\tilde{F}(\omega) = \frac{\tilde{I}_0(\omega)}{N(\omega)}$$

again, these quantities are <u>measured</u>

this filter should be implemented at the first stage of visual processing

Temporal filtering in retinal bipolar cells:
Elements of an optimal computation?
W Bialek & WG Owen,
Biophys J <u>58</u>, 1227-1233 (1990).

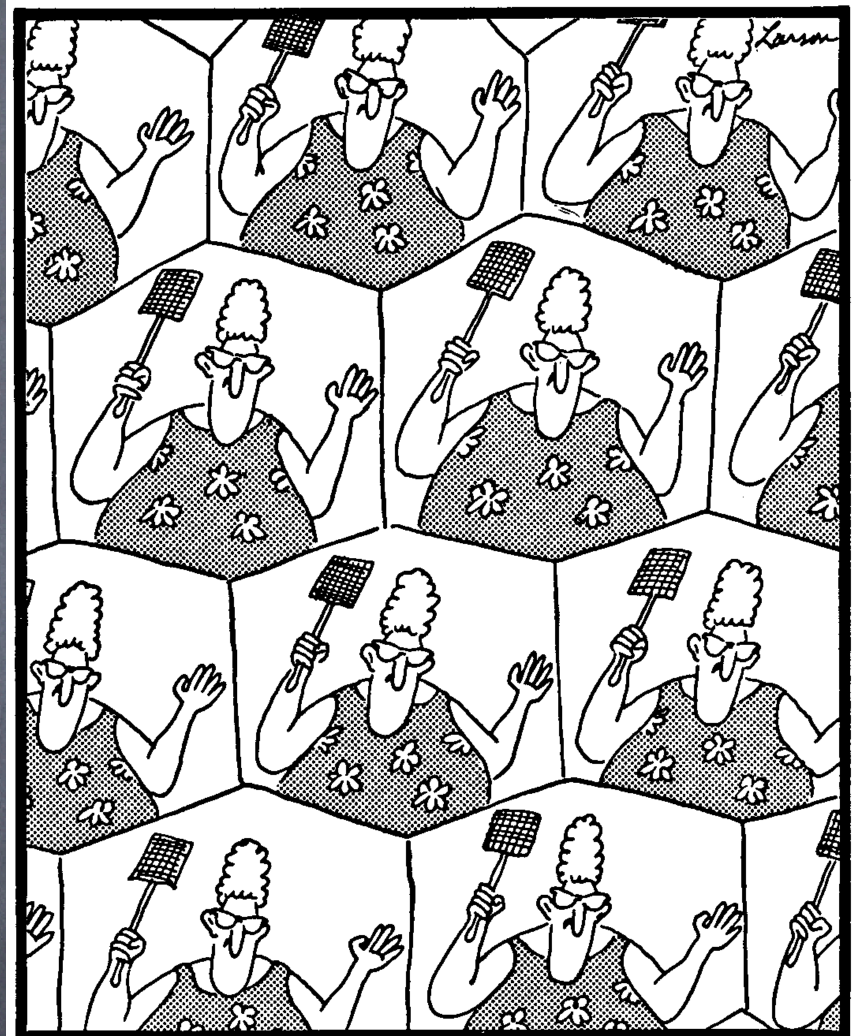
Testing the theory... <u>no free parameters</u>



Optimal filtering in the salamander retina.
F Rieke, WG Owen & W Bialek,
in Advances in Neural Information Processing 3,
R Lippman, J Moody & D Touretzky, eds, pp 377-383
(Morgan Kaufmann, San Mateo CA, 1991).

more to say about this synapse ...
Nonlinear signal transfer from mouse rods to bipolar cells
and implications for visual sensitivity.
GD Field & F Rieke, Neuron 34, 773-785 (2002).

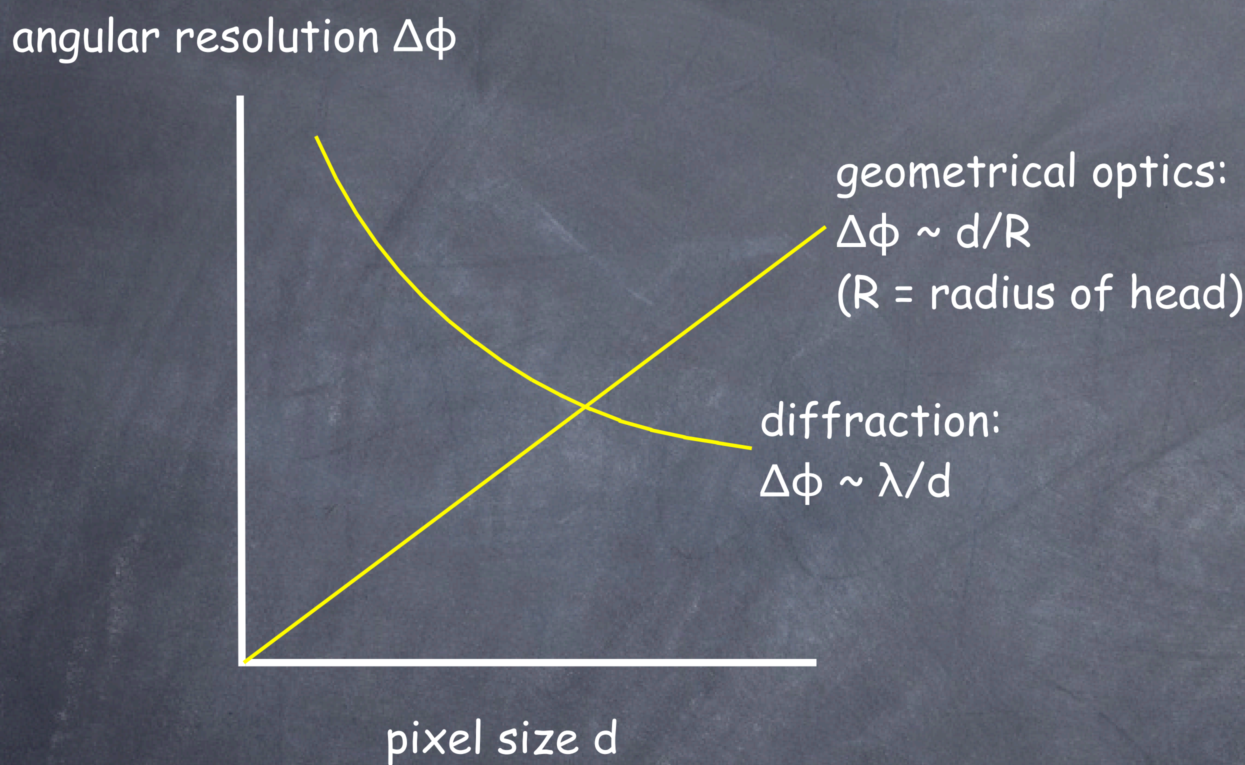
Vision, but in a different animal ...



The last thing a fly ever sees

not as different as Mr Larson thinks

how big should we make the pixels of the compound eye?

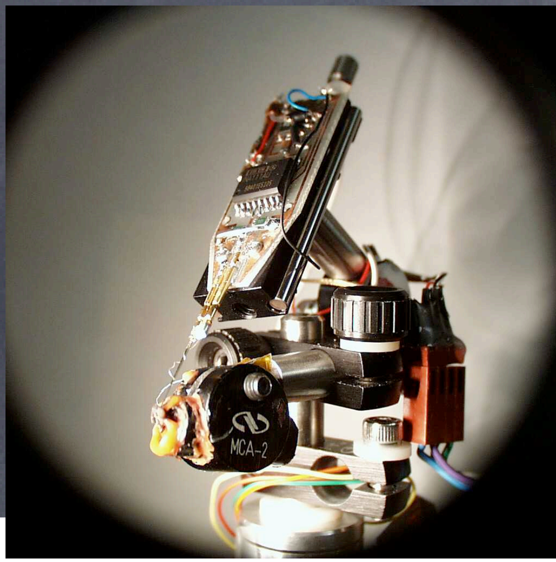


optimum: min $\Delta\phi$ at $d \sim (\lambda R)^{1/2}$

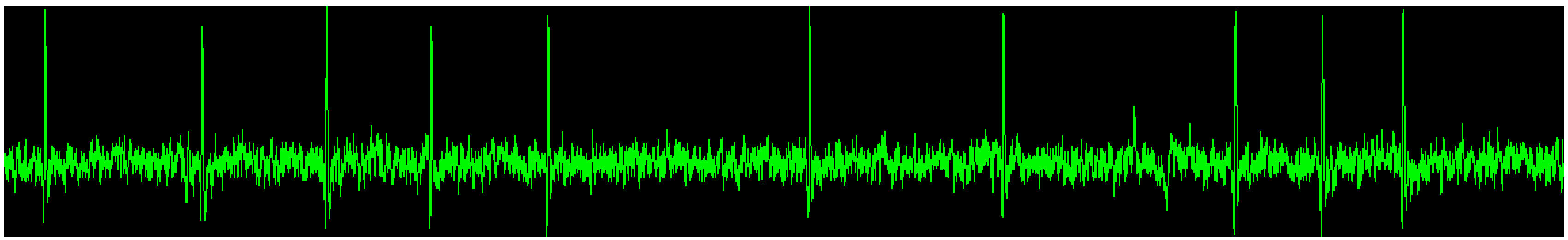
agrees with exp't on many insects with different R
not right when SNR is low ...
need to optimize information not resolution

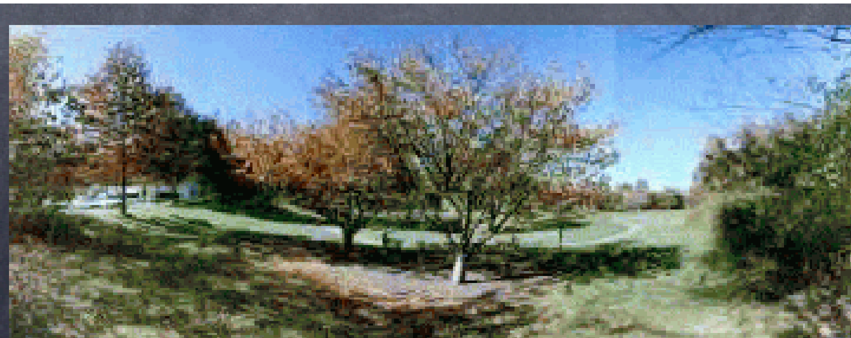
The size of ommatidia in compound eyes.
HB Barlow, J Exp Biol 29, 667-674 (1952).

see also the Feynman lectures!



place a small wire in the back of the fly's head
to "listen in" on the electrical signals from nerve cells
that respond to movement





Experiments from R de Ruyter van Steveninck & GD Lewen

The fly solves (at least) two problems:

computing motion from signals in the retina, and
coding the trajectory of motion in sequences of spikes

optimization in computation:

1. do motion sensitive neurons allow discrimination and estimation with precision close to limits set by noise at the sensory input?
2. what function(al) of the inputs provides the best estimate?
3. can we analyze the response of these neurons to 'dissect' the computation and test the predictions from (2)?

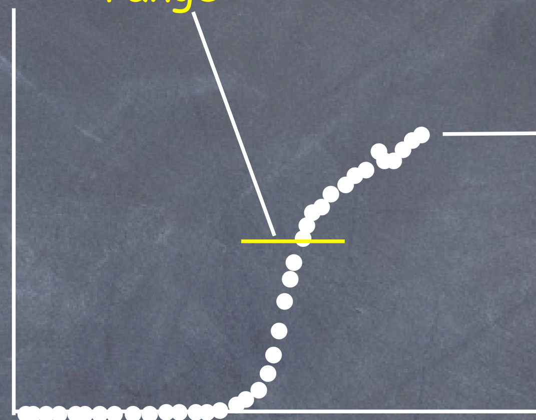
optimization in coding:

1. do these neurons use most of their dynamic range for real signals, or irrelevant variability?
2. what are the symbols in the code?
3. what events do these symbols signify in the motion signal?
4. is there an optimal choice of this mapping?

(to learn more about the other problems see http://www.princeton.edu/~wbialek/flypapers_links.html)

What determines the structure of input/output relations?

(not just a question about neurons!)



In particular, what sets the scale along the input axis?

Suppose we chose the input/output relation
to maximize the information I(input; output) ...

Because mutual information is context dependent, the optimal input/output relation is matched to P(input)

if P(input) is mostly in this range, then there is no built in scale ... $\Rightarrow$ the only way to get a scale on the input axis is from P(input) itself!

noise level

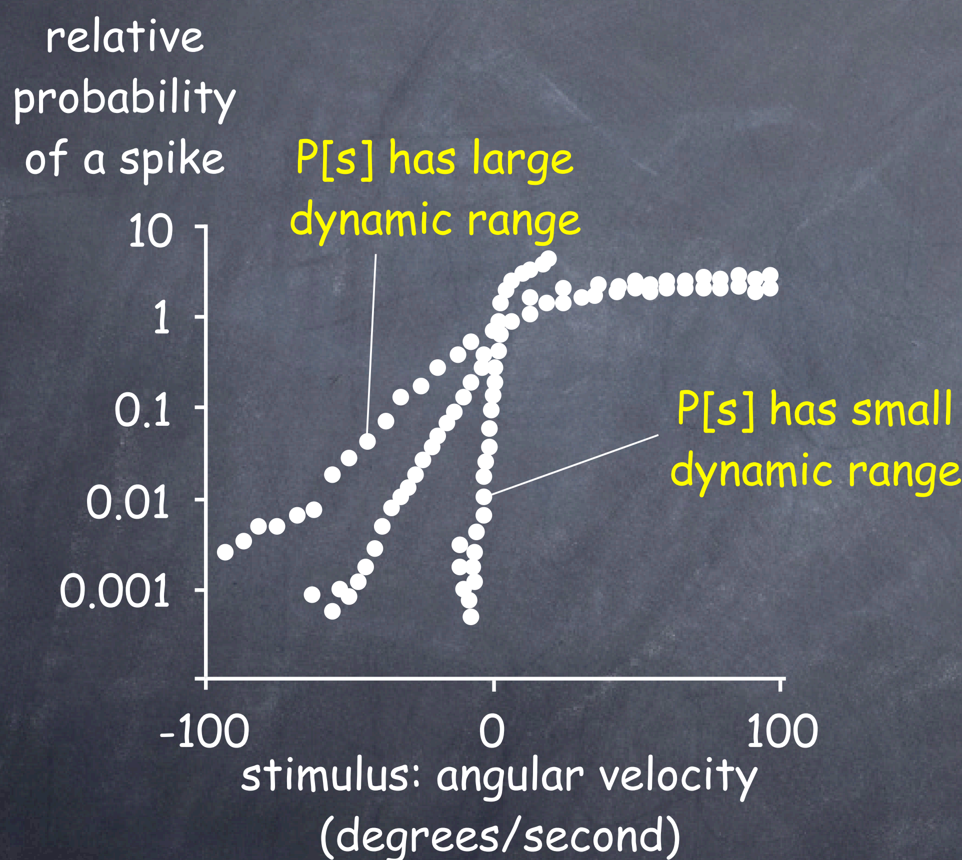
maximum transducable signal

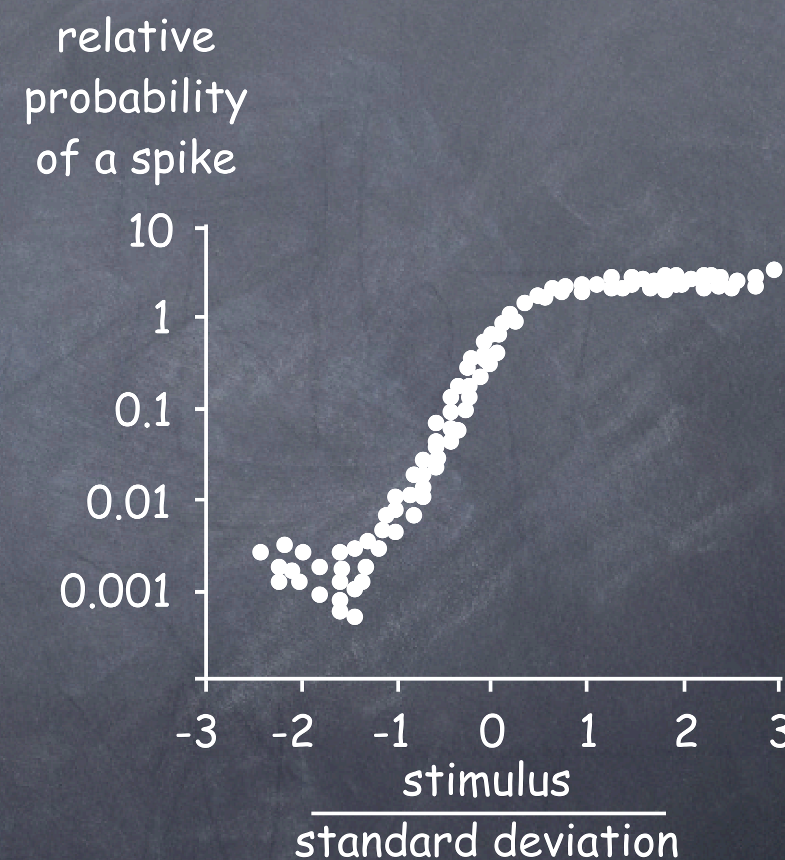
these are scales “built in” to the system itself

(somewhat embarrassingly, equations don’t add much to this picture)

Measure input/output relations when inputs are drawn from different distributions P[s]

(important technical question of how to do this!)



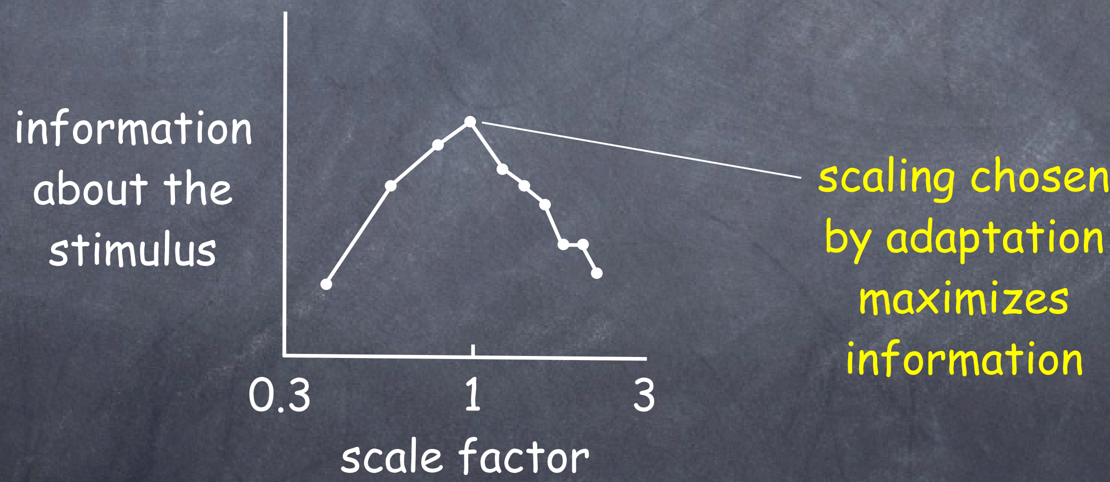


Adaptive rescaling optimizes information transmission.
N Brenner, W Bialek & RR de Ruyter van Steveninck,
Neuron 26, 695-702 (2000).

the code adapts to the distribution of inputs, and

the form of adaptation is consistent with the an optimization principle, but ...

how do we know that information is optimized?



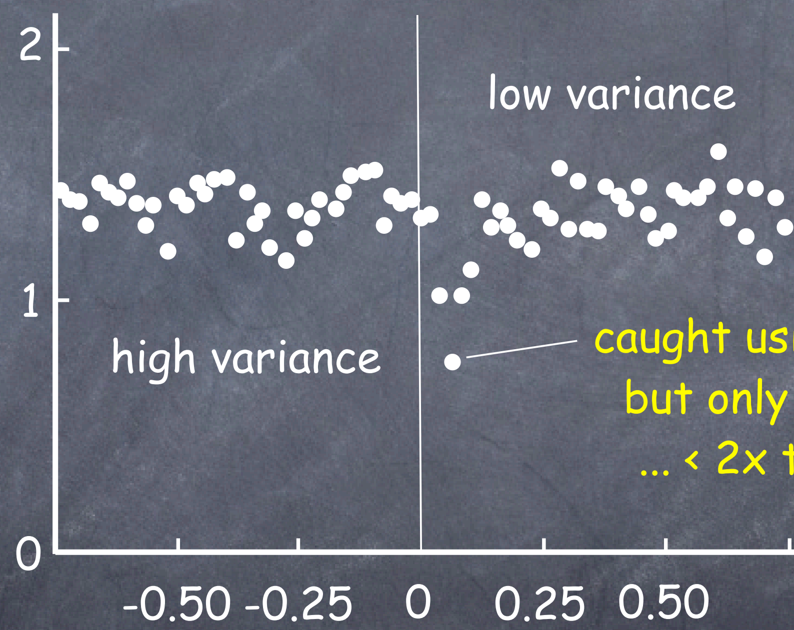
calculate the information that <u>would</u> be transmitted if P[s] is fixed and the cell chose different rescalings of the input/output relation ...

How quickly can the system adjust?

(interestingly difficult to measure)

how long does it take to be sure
that we are seeing a new distributions vs.
outliers in the old distribution?

information about
the rapidly varying
sensory input
(bits/spike)



caught using the wrong code,
but only for < 100 msec (!)
... < 2x the sampling limit

time relative to switch
(seconds)

Efficiency and ambiguity in an adaptive neural code.
AL Fairhall, GD Lewen, W Bialek & RR de Ruyter van Steveninck,
Nature 412, 787-792 (2001).

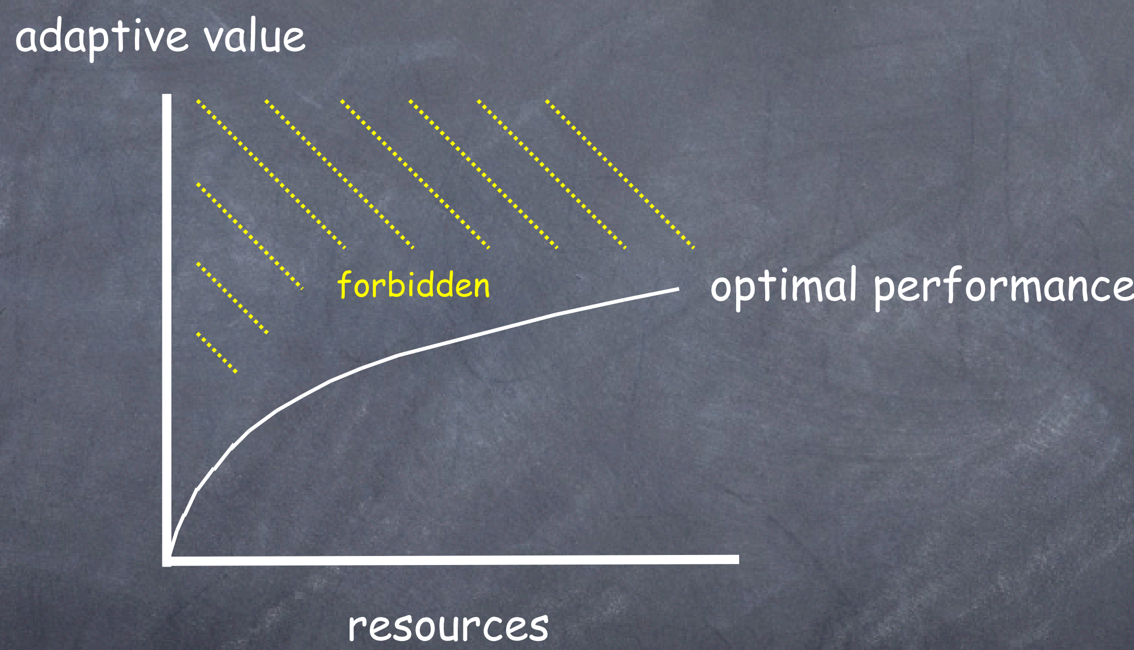
do we have one principle or many?

optimal estimation sounds unified, but comes out differently for each thing we want to estimate. how do we choose?

information theoretic principles have even more generality. but information always is about something. information about what?

(these are polite versions)

Can we find a general, but still biologically meaningful notion of optimal performance?

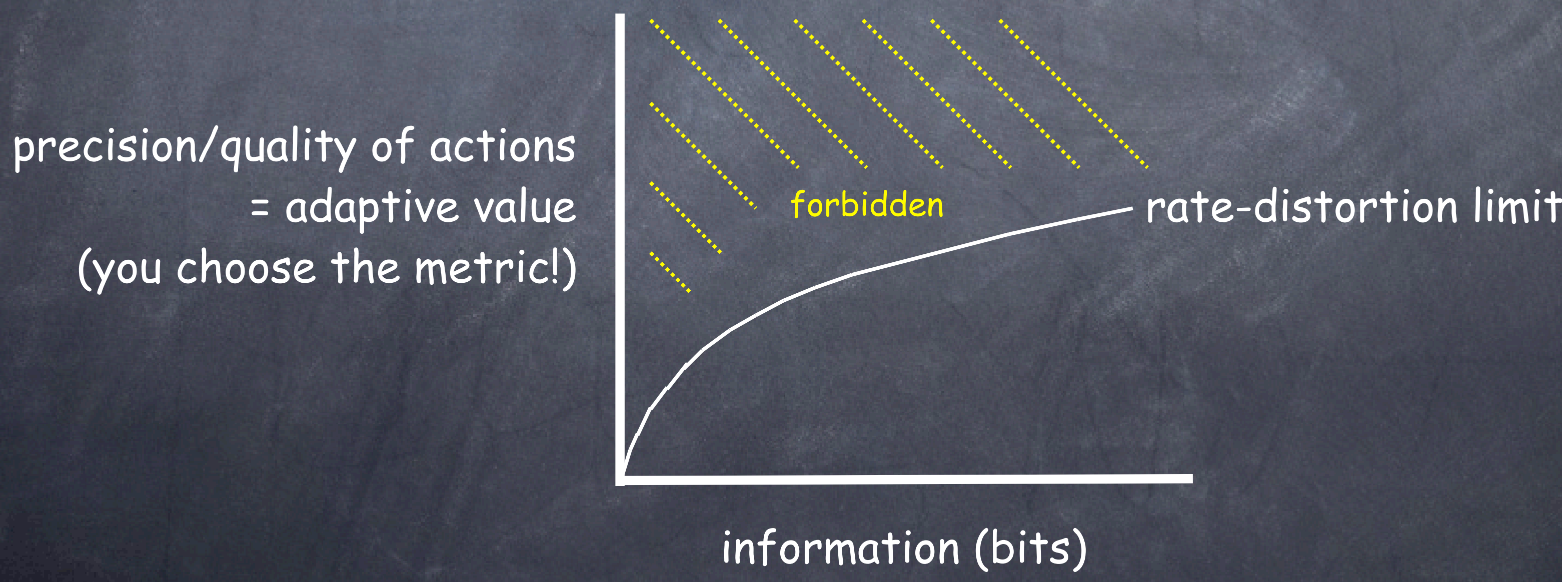


(note that we would still will have a family of solutions, but this is a good thing!)

these arguments based on unpublished work with N Tishby

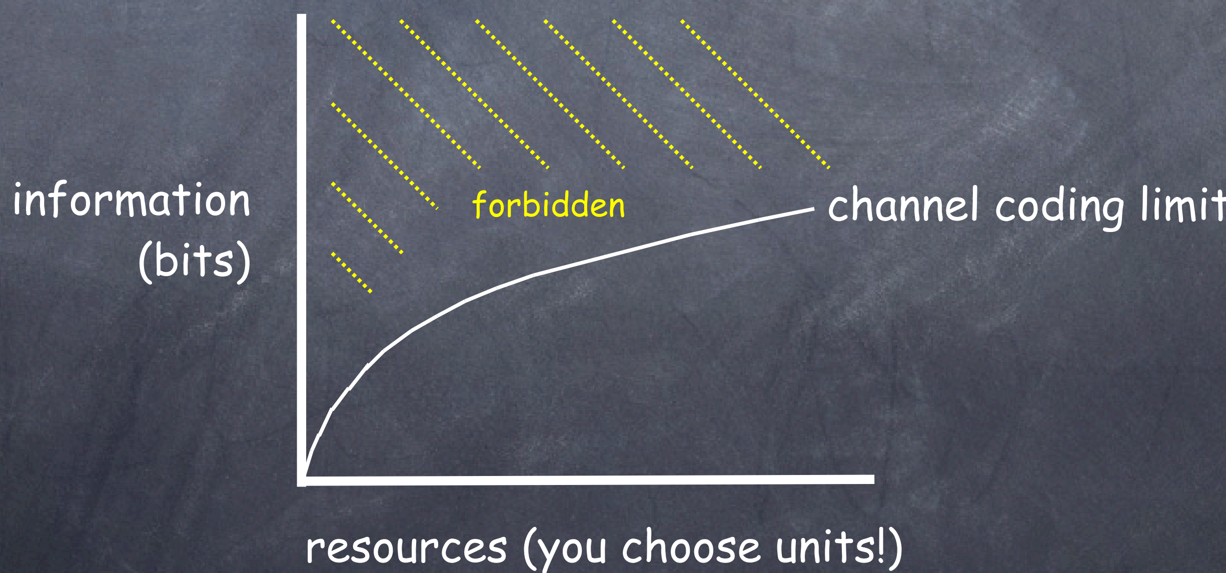
Organisms are not rewarded for maximizing information; they are rewarded for appropriate actions.

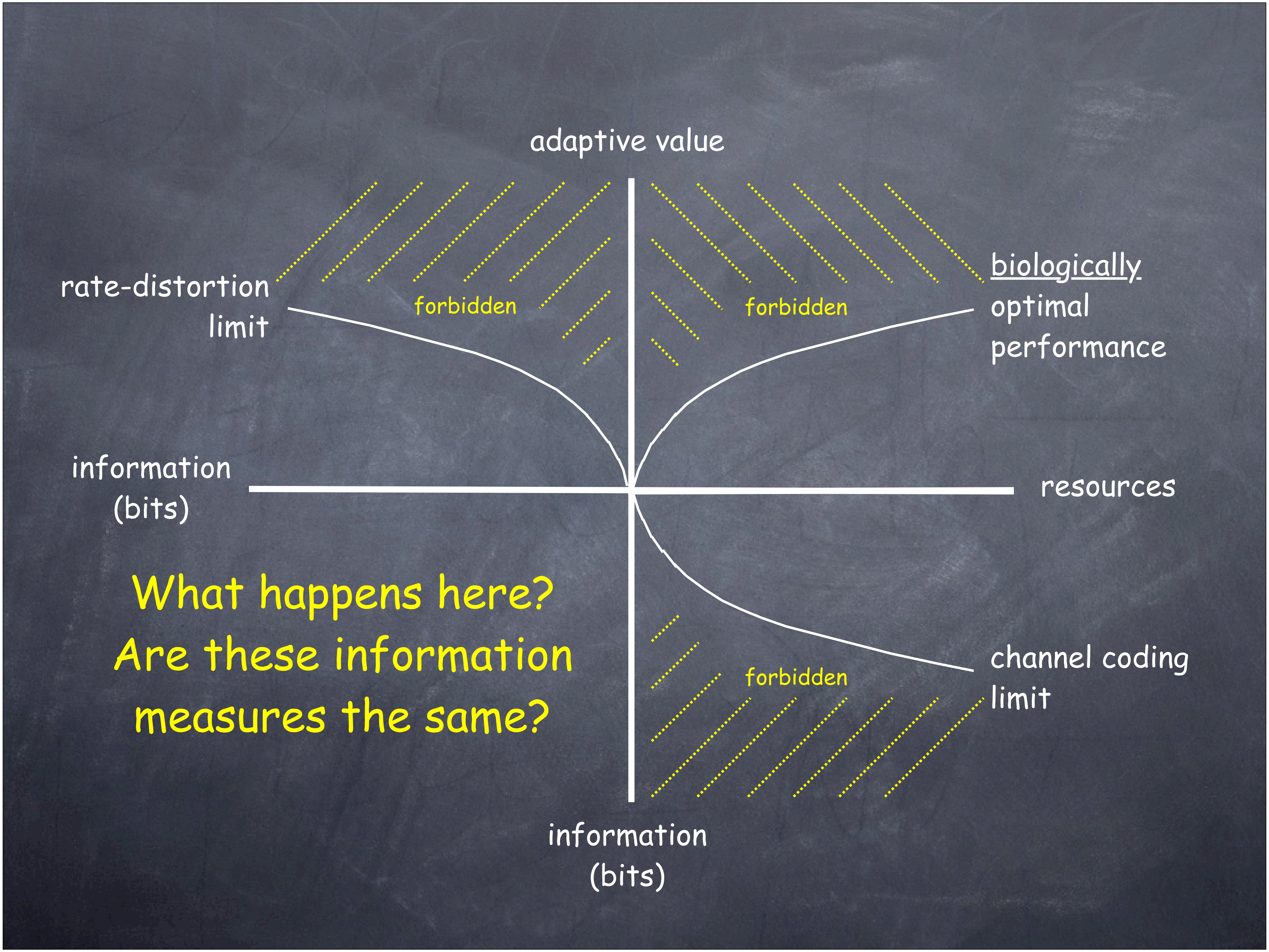
But to act with some level of precision or effectiveness requires a minimum amount of information ... this is the content of rate-distortion theory



Organisms are not just under pressure to act, they have to do so with limited resources.

But given fixed resources (e.g., energy), there is a limit to how much information we can represent ... this the content of channel coding theory



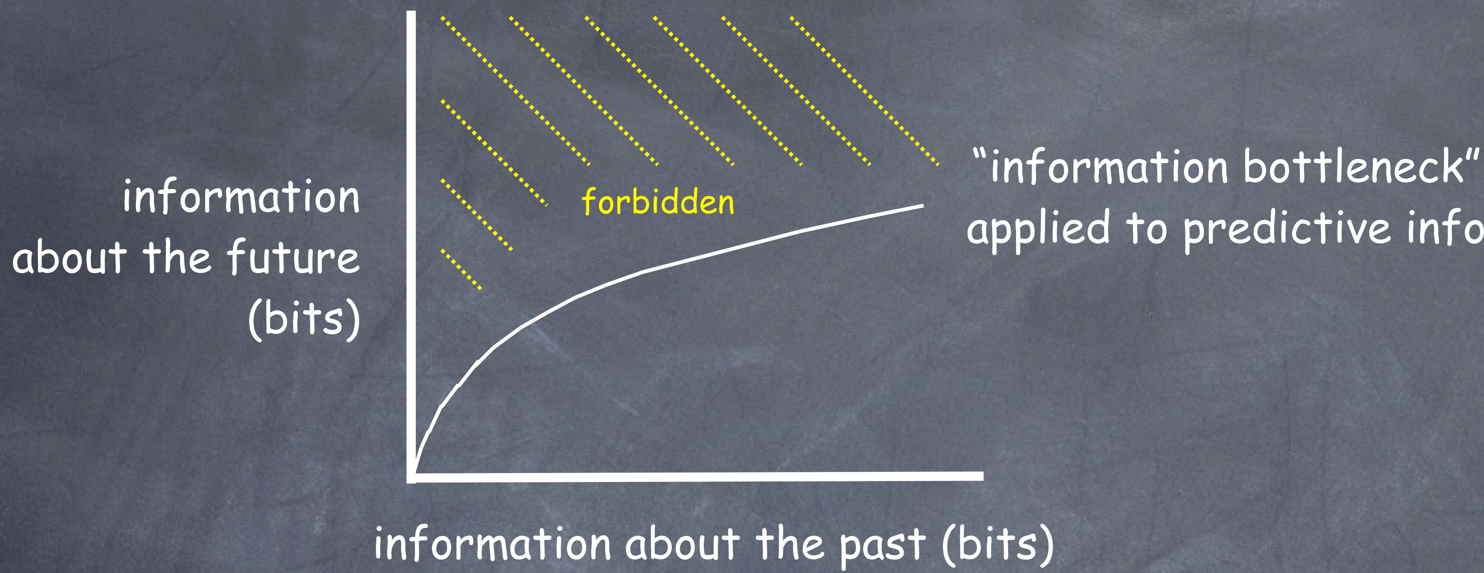

adaptive value
rate-distortion limit
forbidden
forbidden
biologically optimal performance
information (bits)
resources
What happens here?
Are these information measures the same?
forbidden
channel coding limit
information (bits)

Information we extract from sensory inputs must be information about the past (causality)

Information that has a chance of having adaptive value must be predictive: information about the future.

> Predictive information, perhaps surprisingly, provides a compelling measure of what is "complex" about a data stream, and connects to measures of complexity that arise in many different fields ...

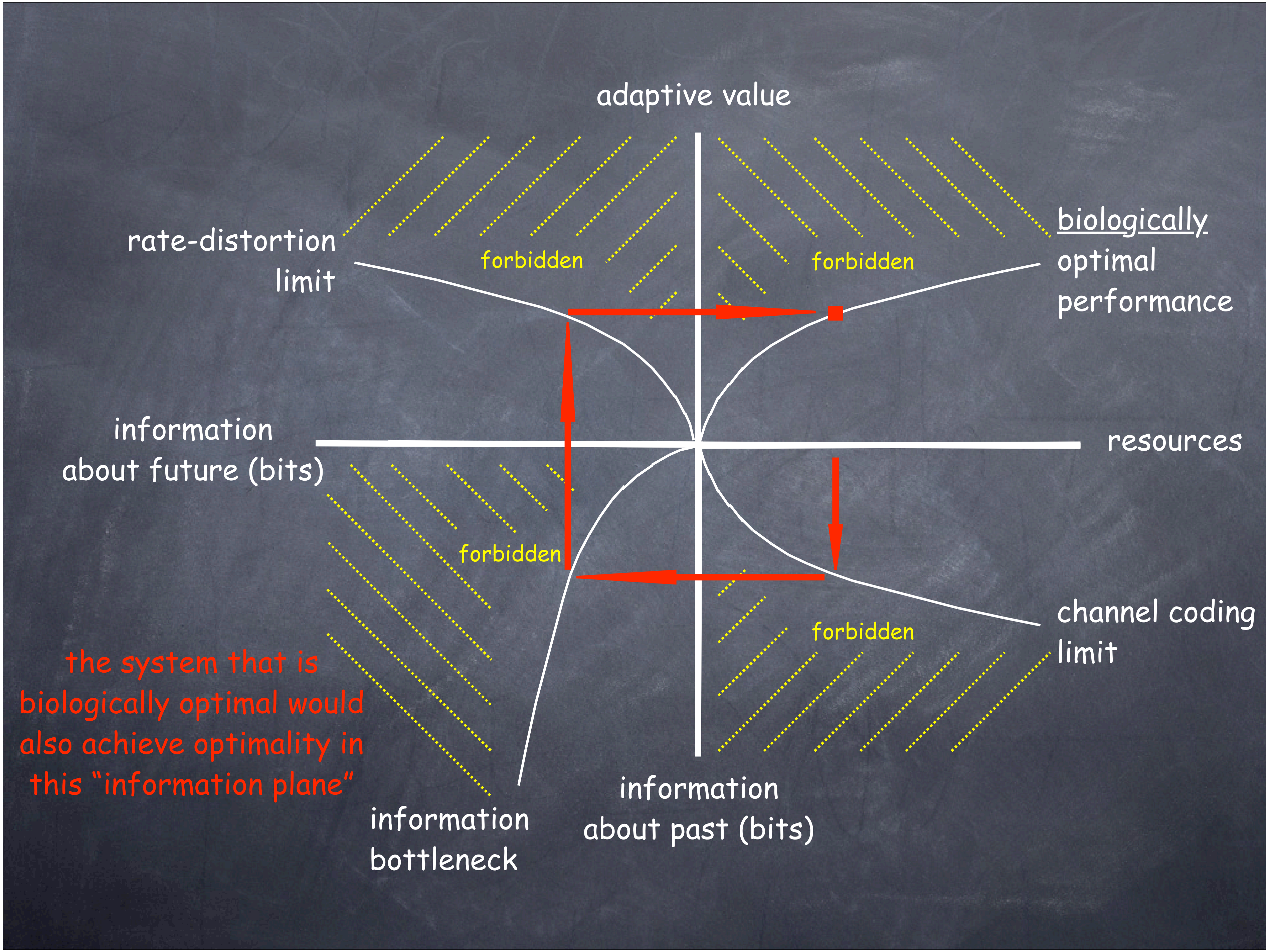
Given the statistical structure of the world, a certain amount of information about the past provides only a limited amount of information about the future.



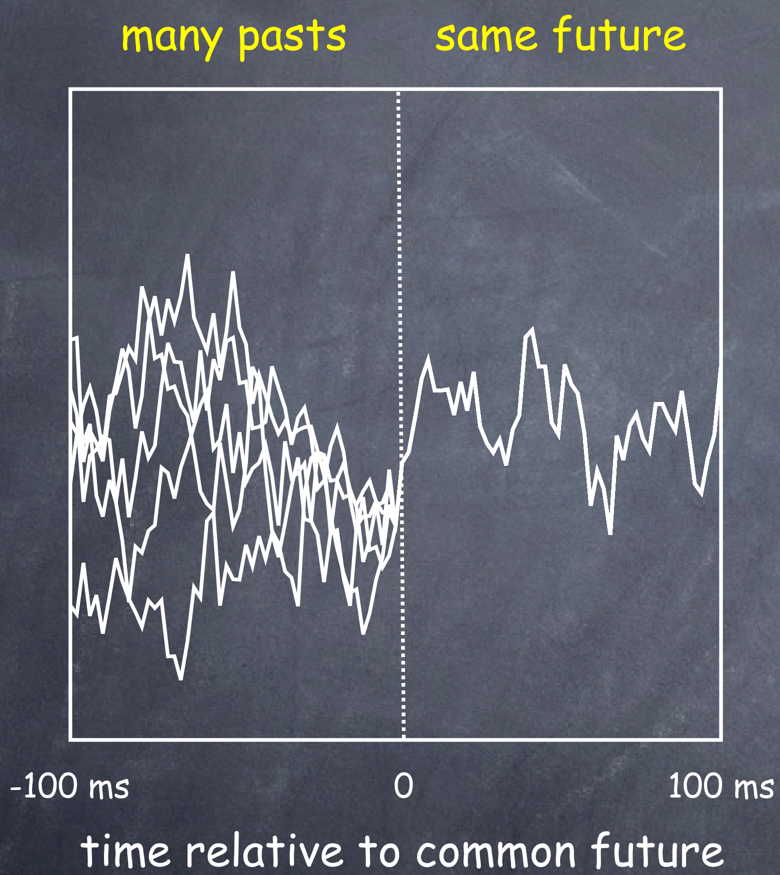
efficient representation of predictive information contains many other problems:
filtering to separate signal from noise
estimating parameters of a model to be learned from the data
...

Predictability, complexity and learning.
W Bialek, I Nemenman & N Tishby,
Neural Comp 13, 2409-2463 (2001).

The information bottleneck method.
N Tishby, FC Pereira & W Bialek,
in Proceedings of the 37th Allerton Conference, B Hajek & RS Sreenivas,
eds, pp 368-377 (University of Illinois, 1999).

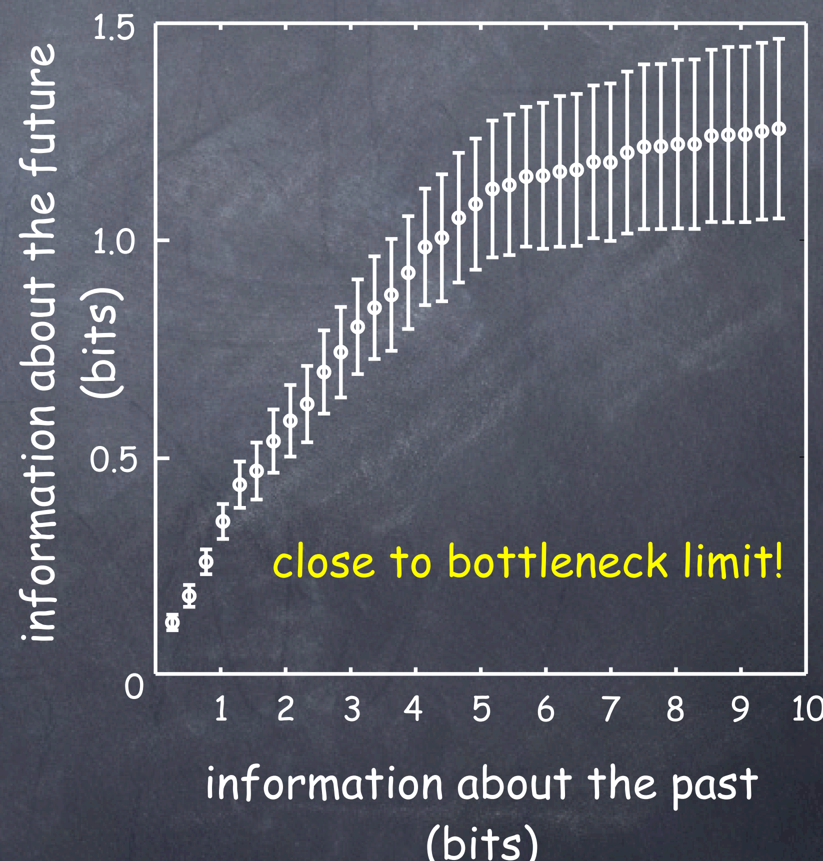

adaptive value
rate-distortion limit
forbidden
forbidden
biologically optimal performance
information about future (bits)
resources
forbidden
channel coding limit
forbidden
the system that is biologically optimal would also achieve optimality in this "information plane"
information bottleneck
information about past (bits)

can we test this picture directly?



generate stochastic trajectories that converge on a common future ...

reproducible responses before convergence = information about the future



Experiments in fly motion sensitive neurons, again unpublished work with RR de Ruyter van Steveninck