# The Role of the Environment in Synaptic Plasticity: Towards an Understanding of Learning and Memory


Brian Scott Blais

Professor Leon N Cooper

Dr. Harel Shouval

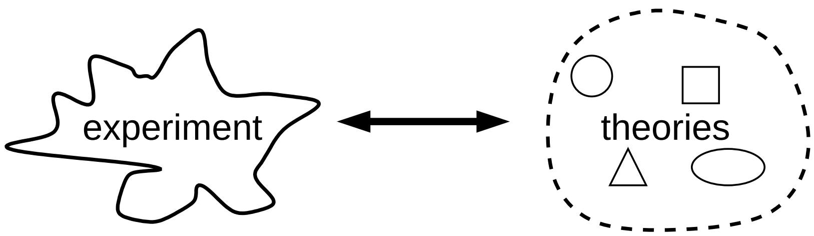
Dr. Nathan Intrator

## Outline of Thesis

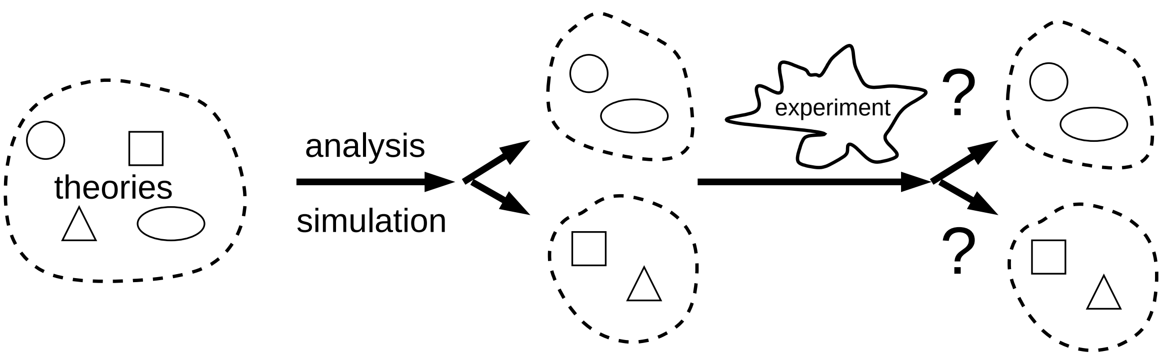
- Oscillations in 1D
  - differences between learning rules, BCM and PCA
  - possible measurements of model parameters
- Visual Deprivation
  - valid parameter regime for BCM
  - bounds on BCM parameters, in real-time units
  - experimental predictions for BCM and PCA
- Projection Pursuit
  - classes of learning rules, and low-D environment
  - analysis of dynamics of synaptic modification
- Extensions
  - direction selectivity
  - structure removal in natural scene environment

# OUTLINE

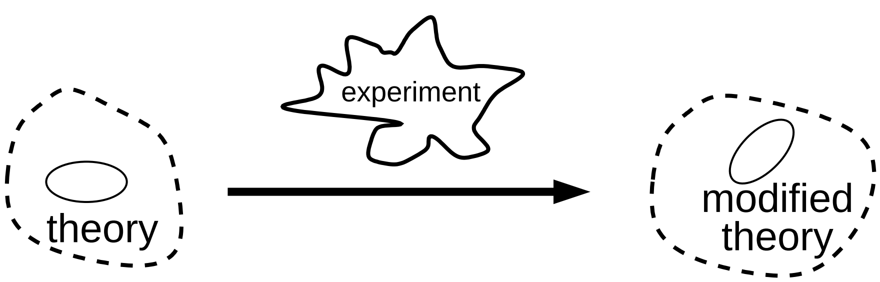
- initial match of theories to experiment



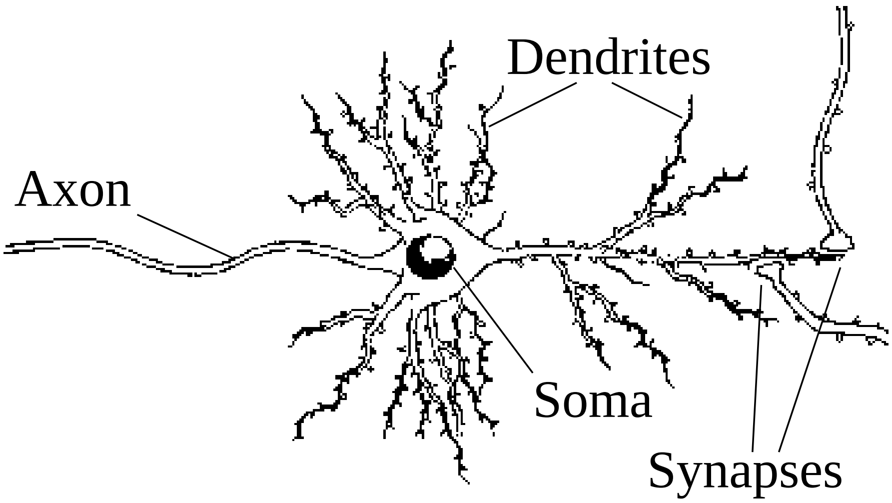

- analysis/simulation to distinguish between theories




- may need to modify theory for new experiments

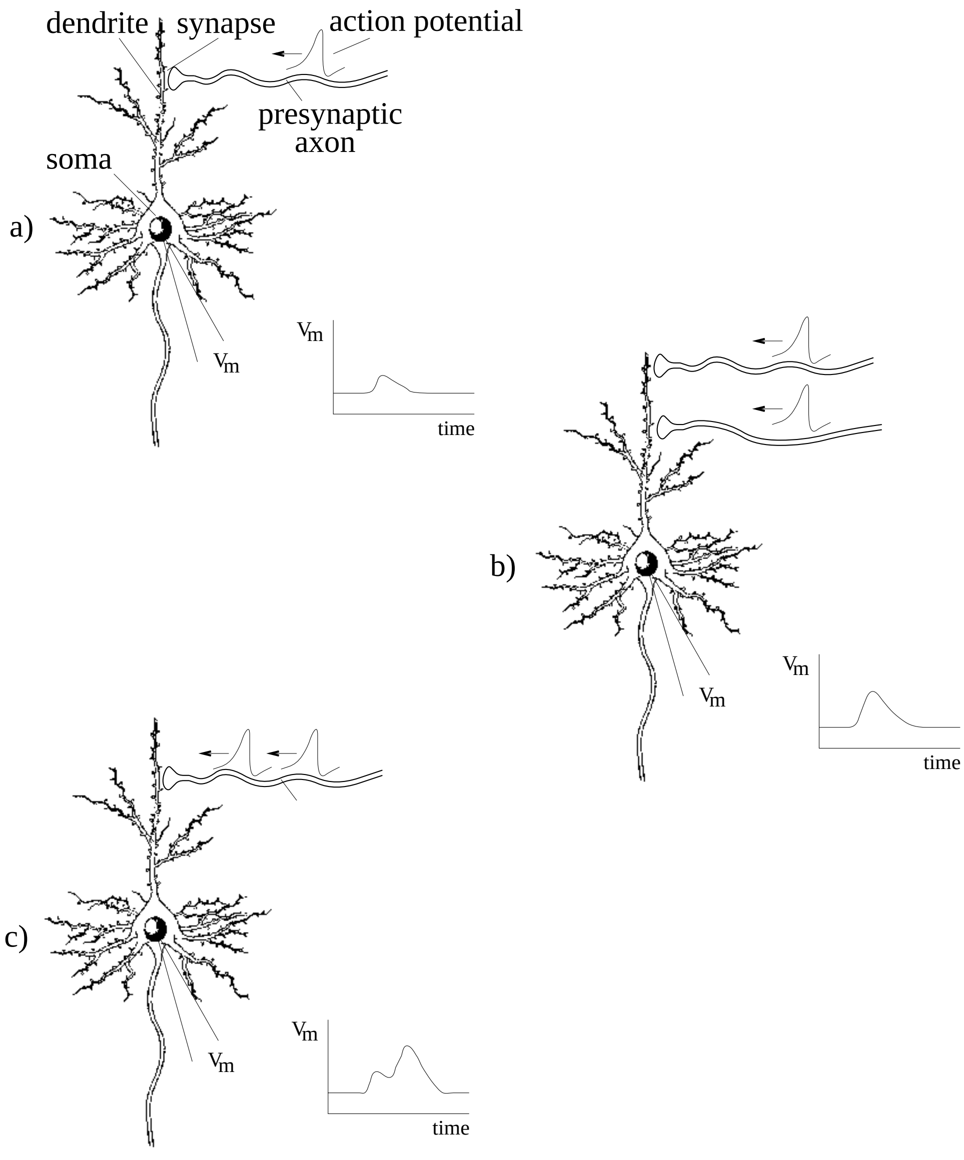


# THE NEURON

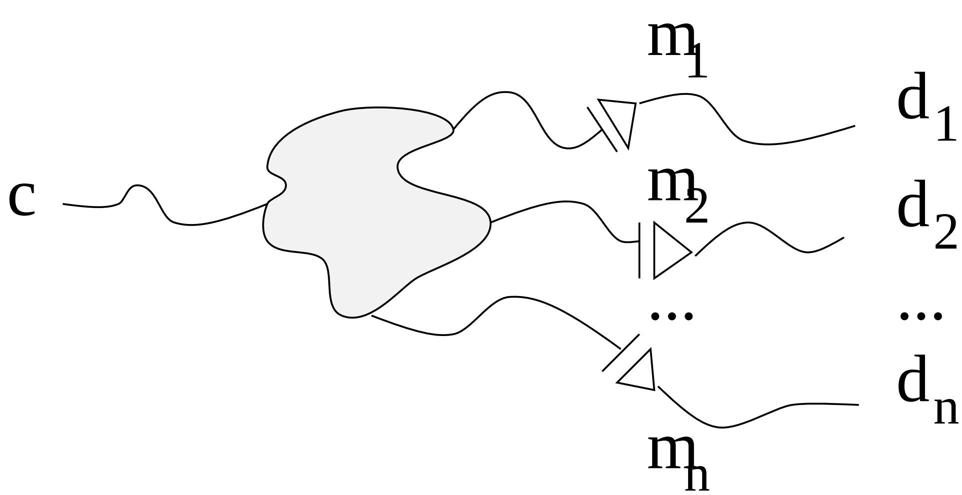



- electrical signals travel along axon
- at synapse: electrical $\longrightarrow$ chemical $\longrightarrow$ electrical
- efficacy of the synapse can change
- synaptic changes depend on history of the inputs

# SPATIAL AND TEMPORAL INTEGRATION OF THE INPUT SIGNALS

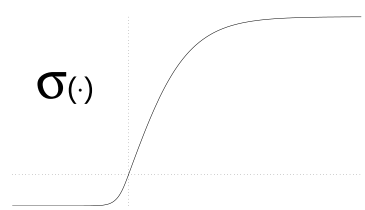


# MODEL NEURON




- $c$ and $\mathbf{d}$ are time-averaged activities over a short time window
- response of cell

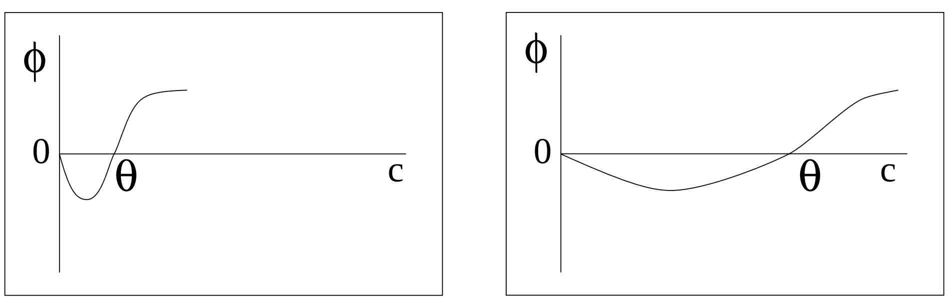
$$c = \sigma(\mathbf{m} \cdot \mathbf{d})\text{, where } \sigma(\cdot) \text{ looks like}$$



# BCM (Bienenstock, Cooper, Munro) Learning Rules

$$\dot{\mathbf{m}} = \phi(c,\theta)\mathbf{d} - \epsilon(c)\mathbf{m}$$

- $\phi(c,\theta)$ must satisfy
  - $\phi(0,\theta) = 0$
  - $\phi(c,\theta) > 0$ for $c > \theta$
  - $\phi(c,\theta) < 0$ for $0 < c < \theta$,
  - $\theta$ averaged superlinear function of $c$




- weight decay, $\epsilon(c)$, is often not necessary

# Example Learning Rules

- quadratic form (QBCM)

$$\begin{aligned} \dot{\mathbf{m}} &= c(c-\theta)\mathbf{d} \\ \theta &= E[c^2] \end{aligned}$$

  - stable with sliding threshold

- kurtosis 2 ($K_2$)

$$\begin{aligned} \frac{d\mathbf{m}}{dt} &= c(c^2-\theta)\mathbf{d} \\ \theta &= 3E[c^2] \end{aligned}$$

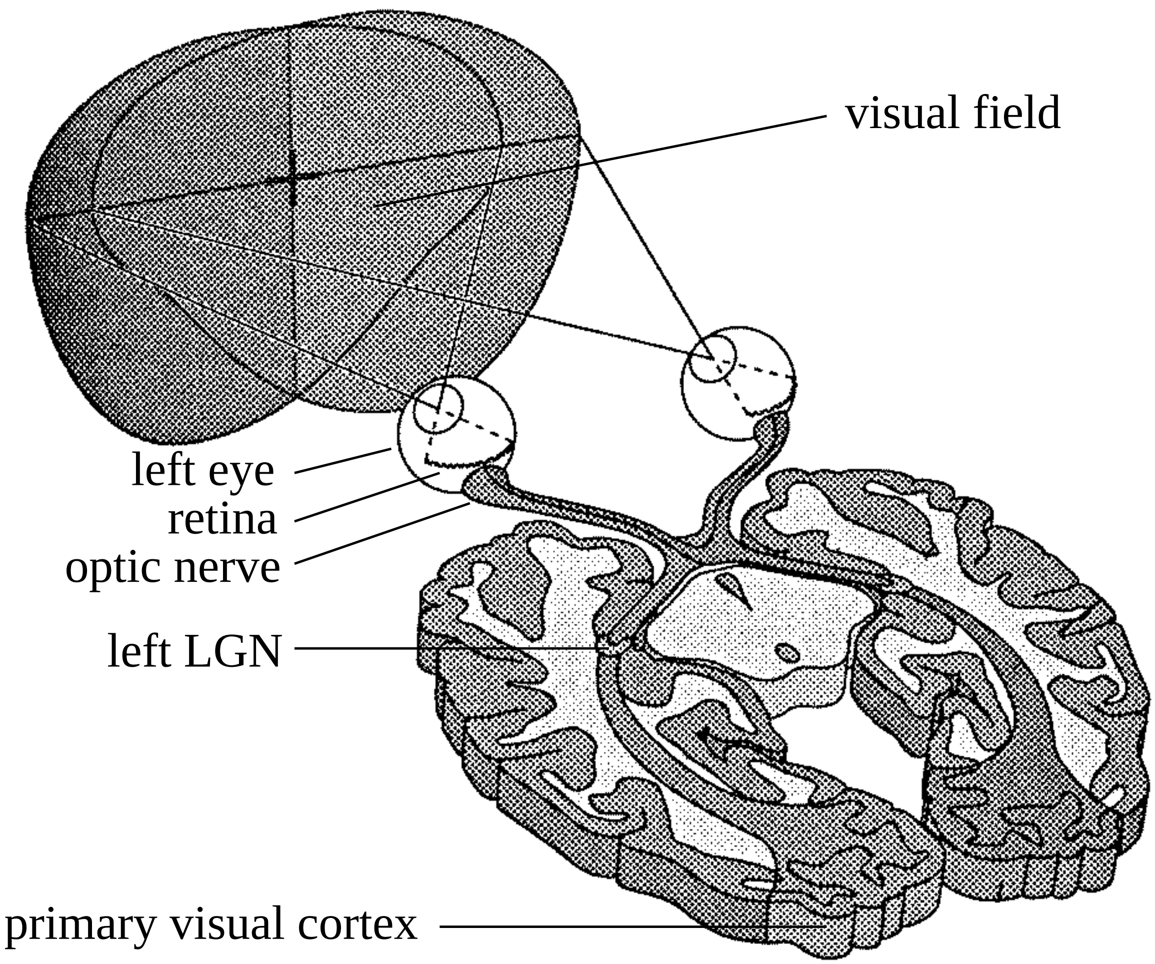
  - not stable $\longrightarrow$ introduce weight decay to normalize weight

$$\begin{aligned} \frac{d\mathbf{m}}{dt} &= \phi(c)\mathbf{d} - \phi(c)c\mathbf{m} \\ \mathbf{m}_{n+1} &= \frac{\mathbf{m}_n + \eta\phi(c_n)\mathbf{d}_n}{\sqrt{(\mathbf{m}_n + \eta\phi(c_n)\mathbf{d}_n)^2}} \\ &\approx \mathbf{m}_n + \eta\phi(c_n)(\mathbf{d}_n - c_n\mathbf{m}_n) + O(\eta^2) \end{aligned}$$

## How do We Compare with Experiment?

- spin out the consequences of the theory, and make predictions
- directly measure some of the basic postulates of the theory

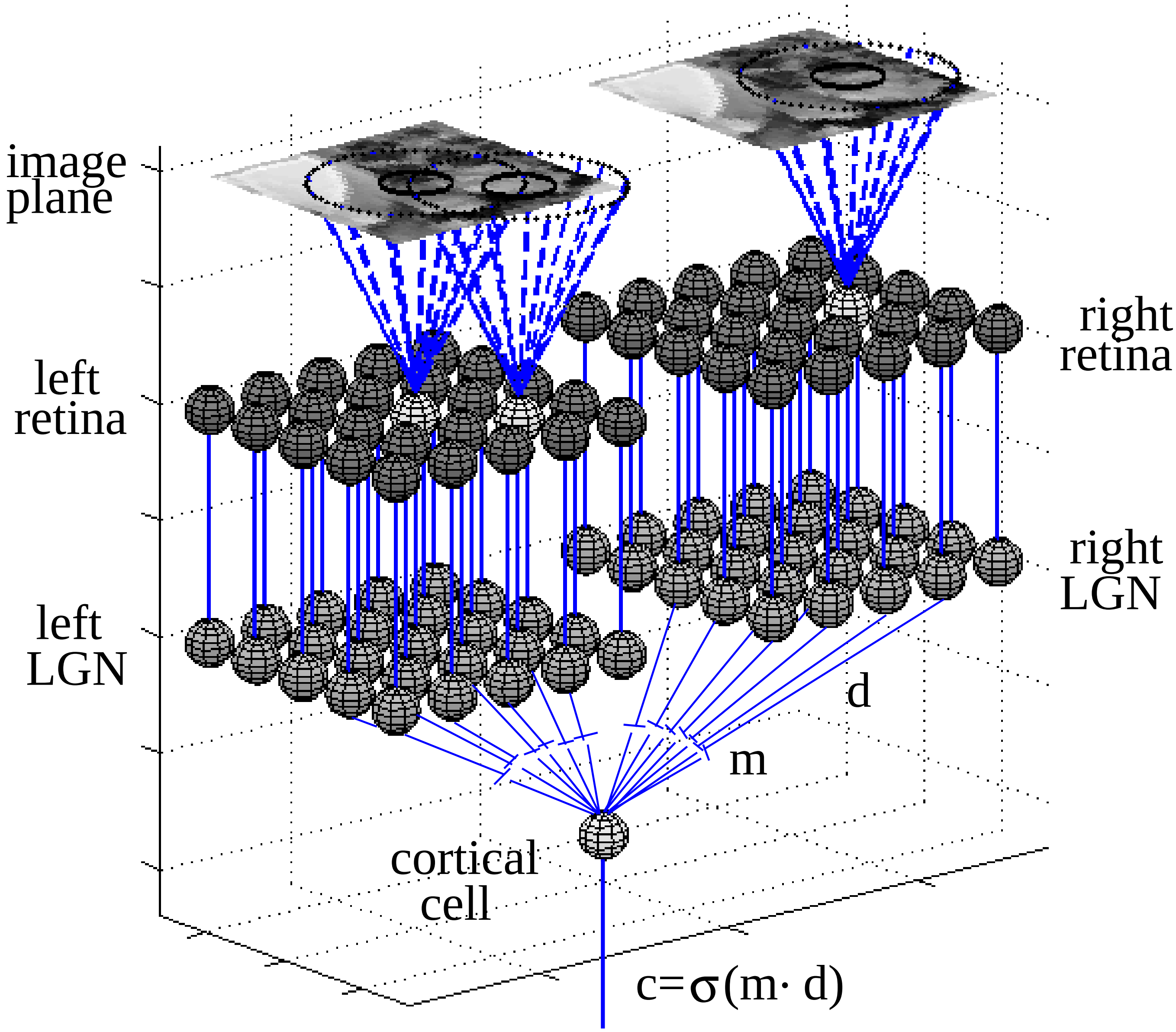
# VISUAL PATHWAY



# EXPERIMENTS (1)

- cortical cells are orientation selective and binocular(Hubel and Wiesel, 1959)
- orientation selectivity requires visual experience(Frégnac and Imbert, 1978) during a critical period
- brief periods of monocular deprivation (MD) cause shift in ocular dominance from binocular to almost totally monocular(Mioche and Singer, 1989; Wiesel and Hubel, 1963)
- loss of patterned input, not reduced variance, required for monocular deprivation(Blakemore, 1976; Wiesel and Hubel, 1965)

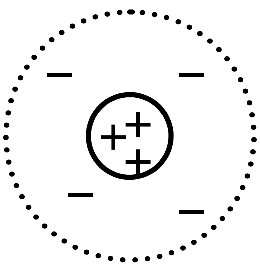
# Model of Visual Pathway



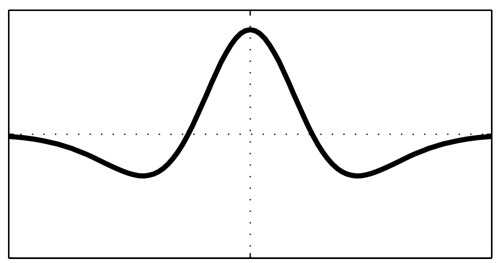
# INPUT ENVIRONMENT

Original Images

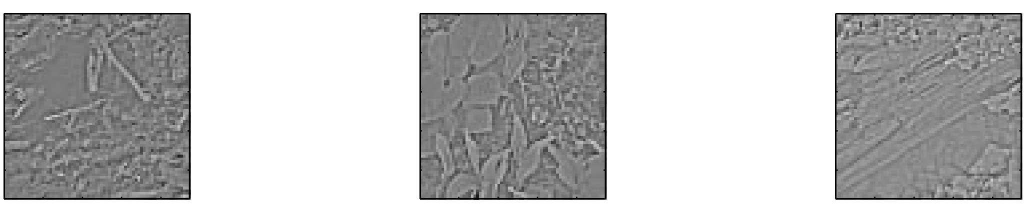
Retina RF



Difference of Gaussians

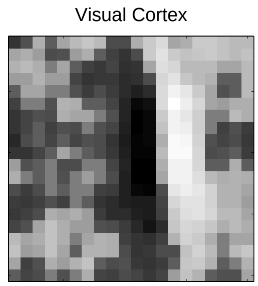


Retinally Processed Images

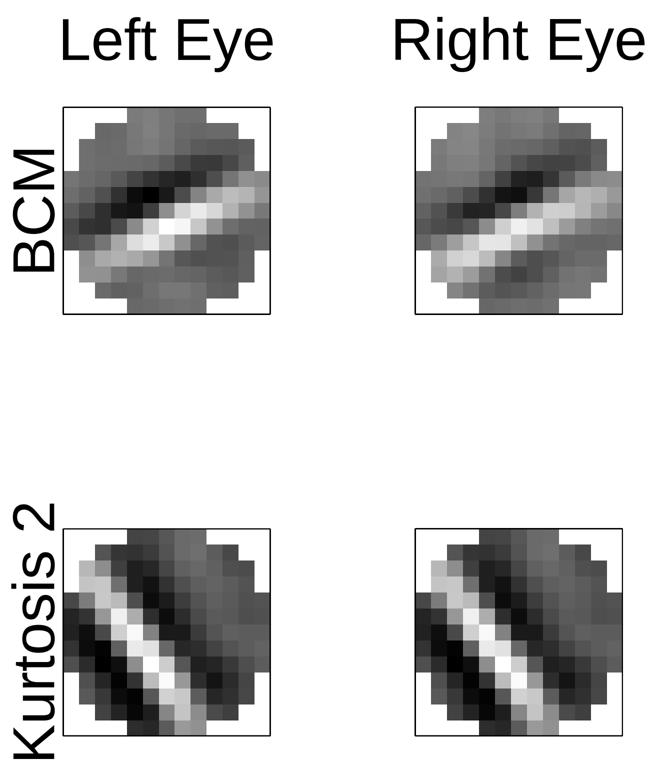


# ORIENTED RECEPTIVE FIELDS

- Experiment (Izumi Ohzawa, 1995. Reverse Correlation)




- Simulation

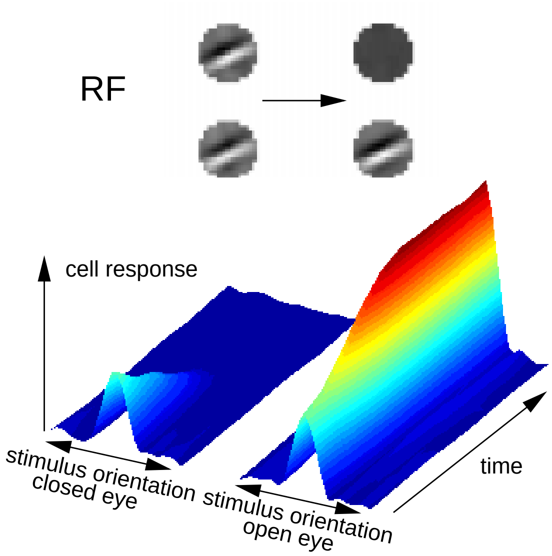


# MODELING DEPRIVATION

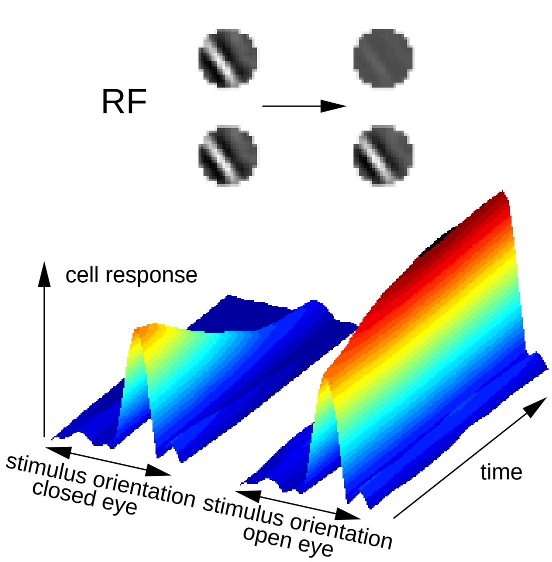
- neurons in the retina and LGN fire spontaneously with no input
- Assumption 1: spontaneous firing carries no structure: noise
- Assumption 2: variance of noise would be larger for more light intensity into closed eye: diffuse lens vs. opaque eye patch, lid suture vs. TTX

# MODELING DEPRIVATION

- QBCM



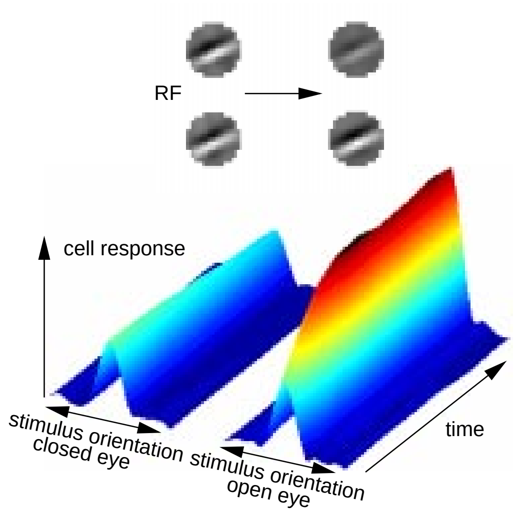

- Kurtosis 2



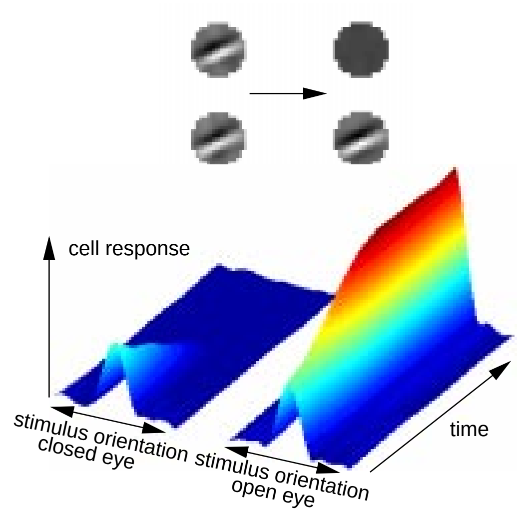
# Noise Dependence of Deprivation

- QBCM

Low Noise

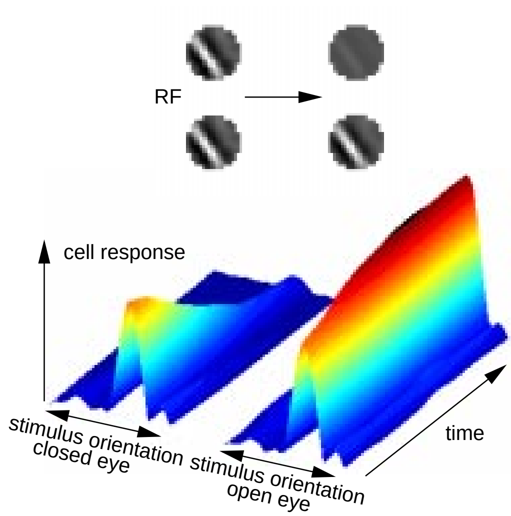



High Noise

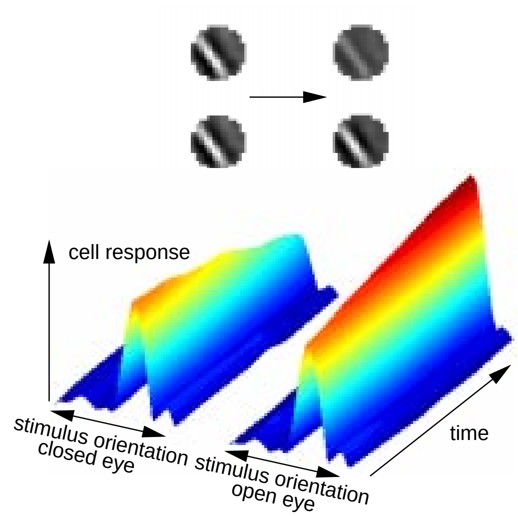



- Kurtosis 2

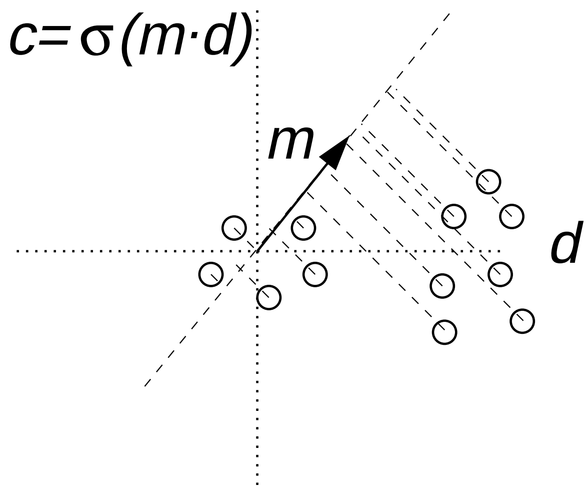
Low Noise




High Noise

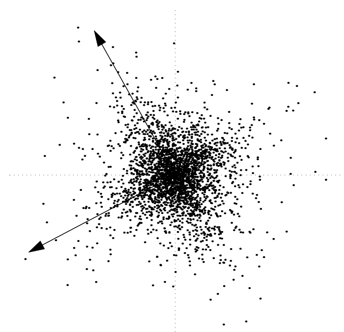


# Projection Pursuit

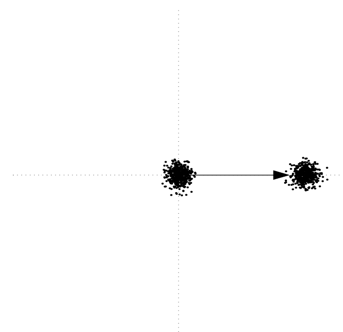



- a cost function is chosen which measures deviation from Gaussian distribution, usually in terms of polynomial moments
- gradient descent/ascent of the cost function w.r.t. weights

Maximizing Kurtosis



Multi-Modality



## Two Examples

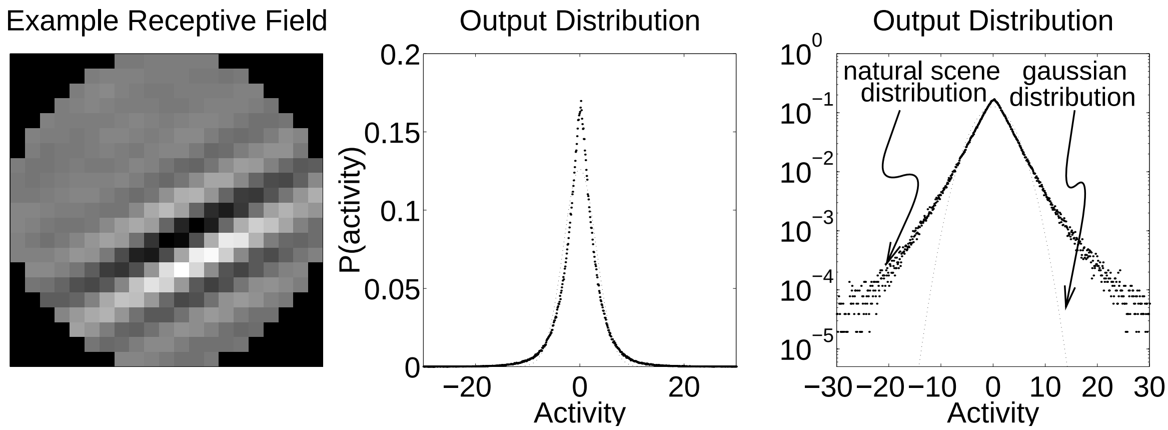
- Kurtosis 2: maximizes kurtosis

$$
\begin{aligned}
K_2(\mathbf{m}) &= E[c^4] - 3E^2[c^2] \\
\frac{d\mathbf{m}}{dt} &= \frac{\partial}{\partial \mathbf{m}} K_2(\mathbf{m}) \\
&= 4E\left[c^3\mathbf{d}\right] - 4 \cdot \underbrace{3E\left[c^2\right]}_{\theta} E\left[c\mathbf{d}\right] \\
&= 4E[c(c^2 - \theta)\mathbf{d}]
\end{aligned}
$$

- QBCM: maximizes multimodality(Intrator and Cooper, 1992)

$$
\begin{aligned}
R(\mathbf{m}) &= \frac{1}{3}E\left[c^3\right] - \frac{1}{4}E^2\left[c^2\right] \\
\frac{d\mathbf{m}}{dt} &= \frac{\partial}{\partial \mathbf{m}} R(\mathbf{m}) \\
&= E\left[c^2\mathbf{d}\right] - \underbrace{E\left[c^2\right]}_{\theta} E\left[c\mathbf{d}\right] \\
&= E[c(c - \theta)\mathbf{d}]
\end{aligned}
$$

# Distribution from Natural Scenes




- distribution is very nearly double-exponential

$$P(\text{activity}) = \frac{1}{2\lambda} e^{-|\text{activity}|/\lambda}$$

# Cost Functions for Learning Rules

- Quadratic BCM

$$R_{\mathrm{QBCM}} = \frac{1}{3}E\left[c^3\right] - \frac{1}{4}E^2\left[c^2\right]$$

- Skewness 1

$$S_1 = E[c^3]/E^{1.5}[c^2]$$

- Skewness 2

$$S_2 = E[c^3] - E^{1.5}[c^2]$$

- Kurtosis 1

$$K_1 = E[c^4]/E^2[c^2] - 3$$

- Kurtosis 2

$$K_2 = E[c^4] - 3E^2[c^2]$$

## LEARNING RULES

- Quadratic BCM

$$\frac{d\mathbf{m}}{dt} = c(c - E[c^2])\mathbf{d}$$

- Skewness 1

$$\frac{d\mathbf{m}}{dt} = c(c - E[c^3]/E[c^2])\mathbf{d}/E^{1.5}[c^2]$$

- Skewness 2

$$\frac{d\mathbf{m}}{dt} = c(c - E^{0.5}[c^2])\mathbf{d}$$

- Kurtosis 1

$$\frac{d\mathbf{m}}{dt} = c(c^2 - E[c^4]/E[c^2])\mathbf{d}/E^2[c^2]$$

- Kurtosis 2

$$\frac{d\mathbf{m}}{dt} = c(c^2 - 3E[c^2])\mathbf{d}$$

# Two Classes of Learning Rules

- Class 1

$$\frac{d\mathbf{m}}{dt} = \phi(c)\mathbf{d}$$

  - BCM
  - Skewness 1
  - Kurtosis 1

- Class 2 (unstable without decay term)

$$\frac{d\mathbf{m}}{dt} = \phi(c)(\mathbf{d} - c\mathbf{m})$$

  - Hebb
  - Skewness 2
  - Kurtosis 2

# Low Dimensional Environment

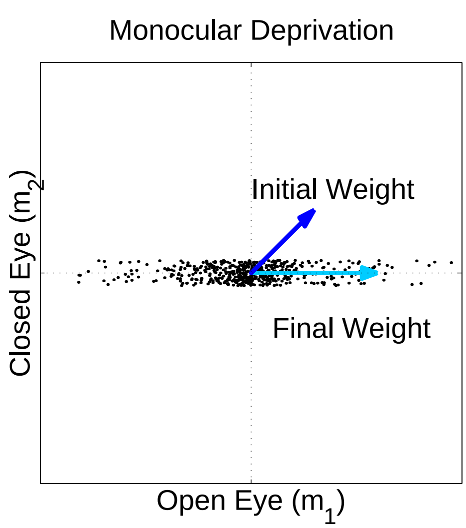
- structure has double exponential distribution (positive kurtosis)

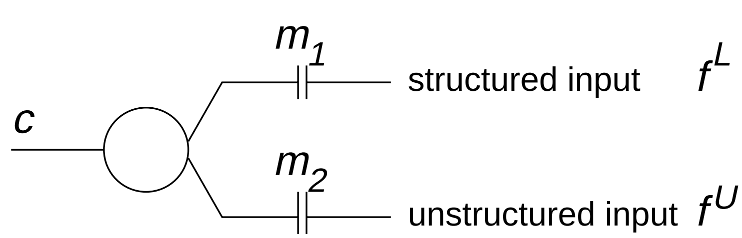
$$f^L(d) = \frac{1}{2\lambda} e^{-|d|/\lambda}$$

- noise has uniform distribution (negative kurtosis)

$$f^U(d) = \frac{1}{2a} \text{ in the range } [-a..a]$$

- output of the cell $c = \sigma(\mathbf{d} \cdot \mathbf{m})$
- calculate cost function, maximize w.r.t. weights



# Calculating Cost Functions

- input distributions

$$
\begin{aligned}
f_{d_1}(d_1) &= \frac{1}{2\lambda} e^{-|d_1|/\lambda} \\
f_{d_2}(d_2) &= \frac{1}{2a} \text{ in the range } [-a..a]
\end{aligned}
$$

- output distributions

$$
\begin{aligned}
c &= d_1 m_1 + d_2 m_2 \equiv c_1 + c_2 \text{ (note: no sigmoid)} \\
f_{c_1}(c_1) &= \frac{1}{2\lambda m_1} e^{-|c_1|/m_1 \lambda} \\
f_{c_2}(c_2) &= \frac{1}{2m_2 a} \text{ in the range } [-m_2 a..m_2 a] \\
f_c(c) &= \int_{-\infty}^{\infty} f_{c_1}(c - c_2) f_{c_2}(c_2) dc_2 \\
\tilde{c} \equiv \sigma(c) &\Rightarrow f_{\tilde{c}}(\tilde{c}) \Rightarrow E\left[\tilde{c}^n\right]
\end{aligned}
$$

# Monocular Deprivation: QBCM

- QBCM

$$
\begin{aligned}
R_{\mathrm{QBCM}} &= \frac{1}{3}E\left[c^3\right] - \frac{1}{4}E^2\left[c^2\right] \\
&= \frac{1}{2}\lambda^2 m_1^2 a m_2 + \frac{1}{24}a^3 m_2^3 \\
&\quad +\lambda^4 m_1^4\left(1-e^{-am_2/\lambda m_1}\right)/am_2 \\
&\quad -\frac{1}{144}a^4 m_2^4 - \frac{1}{12}a^2 m_2^2 \lambda^2 m_1^2 - \frac{1}{4}\lambda^4 m_1^4
\end{aligned}
$$

  – low noise approximation $a \ll \lambda$

$$
\frac{dm_2}{dt} \equiv \frac{\partial R_{\mathrm{QBCM}}}{\partial m_2} \quad \approx \quad -\frac{3}{2}a^2 m_2
$$

- more noise into closed eye $\longrightarrow$ faster loss of response

# MONOCULAR DEPRIVATION: KURTOSIS 2

- Kurtosis 2

$$
\begin{aligned}
K_2 &= E[c^4] - 3E^2[c^2] \\
&= \frac{1}{60} a^4 m_2^4 + 9\lambda^4 m_1^4 + \lambda^2 a^2 m_1^2 m_2^2
\end{aligned}
$$

  – convert to polar, enforce constraint $|m|^2 = 1$ with $R = 1$

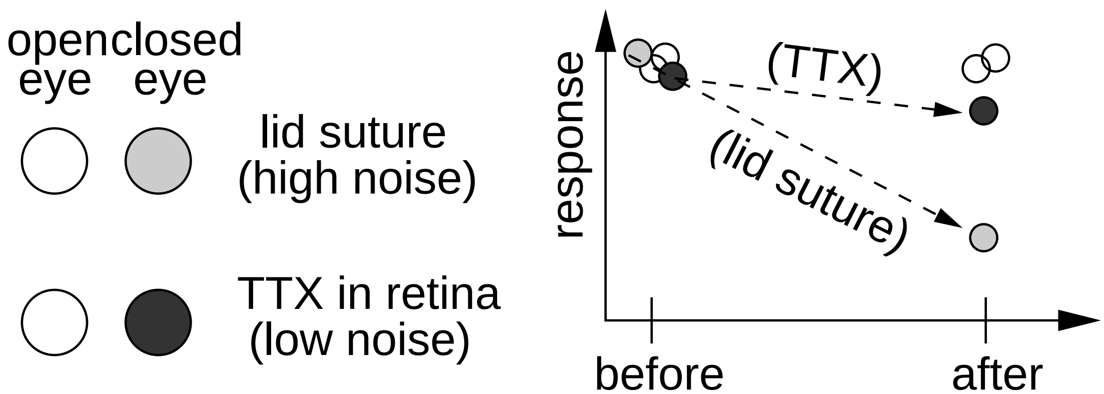
$$
\begin{aligned}
\frac{\partial K_2}{\partial \theta} &\overset{a \ll \lambda}{\approx} \left(-36\lambda^4 + 4\lambda^2 a^2\right) \sin(\theta) \cos^3(\theta) \\
&\quad -2\lambda^2 a^2 \sin(\theta) \cos(\theta)
\end{aligned}
$$

$$
\begin{aligned}
\theta &\equiv \pi/4 + x \\
\frac{dx}{dt} &\approx \left(18\lambda^4 - 2\lambda^2 a^2\right) x - 9\lambda^4 \\
x(t) &= -\frac{9\lambda^2}{18\lambda^2 - 2a^2} \left(e^{\left(18\lambda^4 - 2\lambda^2 a^2\right)t} - 1\right)
\end{aligned}
$$

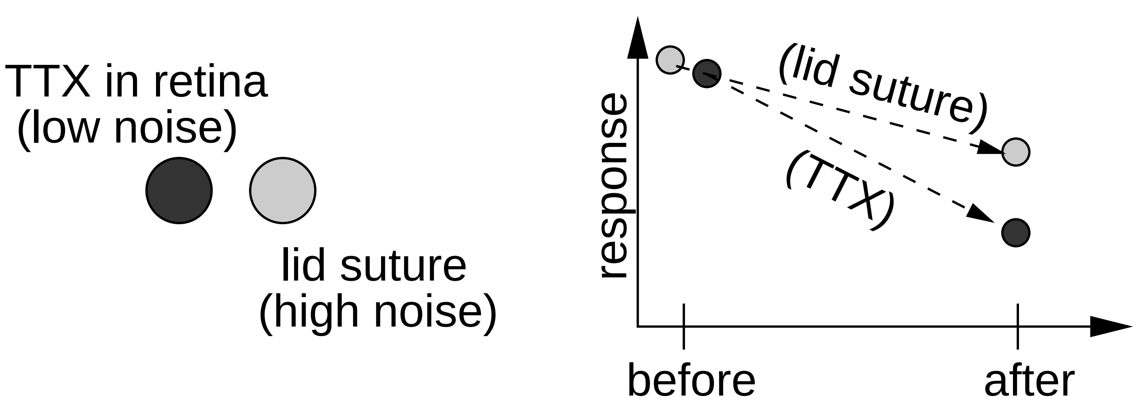
- more noise into closed eye $\longrightarrow$ slower loss of response

# EXPERIMENTS (2)

- MD using lid suture produces a stronger OD shift then MD using TTX in the retina(Rittenhouse et al., 1999)




- TTX in one eye **and** lid suture on the other, produces an OD shift towards the lid suture eye(Chapman et al., 1986)



## CONCLUSIONS

- experimental observations show synaptic modification dependent on structure in the environment
- model can be made consistent with many experiments
  - assumptions of environment
  - learning rules
- simulation results in natural scene environment, and analysis in simplified environment, show dynamics of synaptic changes can be used to distinguish learning rules
- distinguish learning rules $\longrightarrow$ distinguish mechanisms
- future work
  - structure in the noise
  - statistics of natural scenes

# References

Blakemore, C. (1976). The conditions required for the maintenence of binocularity in the kitten's visual cortex. *Journal of Physiology*, 261(2):423–44.

Chapman, B., Jacobson, M. D., Reiter, H. O., and Stryker, M. P. (1986). Ocular dominance shift in kitten visual cortex caused by imbalance in retinal electrical activity. *Nature*, 324:154–156.

Frégnac, Y. and Imbert, M. (1978). Early development of visual cortical cells in normal and dark reared kittens: relationship between orientation selectivity and ocular dominance. *J. Physiol., Lond.*, 278:27–44.

Hubel, D. H. and Wiesel, T. N. (1959). Integrative action in the cat's lateral geniculate body. *J. Physiol*, 148:574–591.

Intrator, N. and Cooper, L. N. (1992). Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5:3–17.

Mioche, L. and Singer, W. (1989). Chronic recording from single sites of kitten striate cortex during experience-dependent modification of synaptic receptive-field properties. *J. Neurophysiol.*, 62:185–197.

Rittenhouse, C. D., Shouval, H. Z., Paradiso, M. A., and Bear, M. F. (1999). Evidence that monocular deprivation induces homosynaptic long-term depression in visual cortex. *Nature*, 397:347–350.

Wiesel, T. N. and Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, 26:1003–1017.

Wiesel, T. N. and Hubel, D. H. (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol.*, 28:1029–1040.