Learning on the Time Scale of Seconds

Each sensory and motor experience affects our subsequent behavior.

These effects on cognition are significant and long lasting.

Many psychological paradigms explore sequential effects.

- cue validity (e.g., Bodner & Masson, 2001; Posner, 1980)
- list composition (e.g., Taylor & Lupker, 2001)
- contingent cueing (e.g., Chun & Jiang, 1998, 1999)
- task switching (e.g., Rogers & Monsell, 1995)
- perceptual priming (e.g., Ratcliff & McKoon, 1997)
- response priming (e.g., Jentzsch & Sommer, 2002)

Continual adaptation to the ongoing stream of experience is an essential aspect of human learning.
Long-Term Repetition Priming

Refers to an increase in the efficiency of processing a stimulus with repeated exposure

e.g., Bar and Biederman (1998)
Long-Term Repetition Priming

Refers to an increase in the efficiency of processing a stimulus with repeated exposure

e.g., Bar and Biederman (1998)
Long-Term Repetition Priming

Refers to an increase in the efficiency of processing a stimulus with repeated exposure

e.g., Bar and Biederman (1998)
Repetition Suppression

Decreased neural response with each exposure of a stimulus

e.g., Rainer and Miller (2000)

e.g., Poldrack and Gabrieli (2001)
Repetition Suppression: Key Findings

- neural representations are sharpened
Repetition Suppression: Key Findings

• neural representations are sharpened

• item specific
Repetition Suppression: Key Findings

- neural representations are sharpened
- item specific
- long lasting
Repetition Suppression: Key Findings

- neural representations are sharpened
- item specific
- long lasting
- graded reduction in activity
Repetition Suppression: Key Findings

• neural representations are sharpened
• item specific
• long lasting
• graded reduction in activity
• widespread in neocortex
Repetition Suppression: Key Findings

- neural representations are sharpened
- item specific
- long lasting
- graded reduction in activity
- widespread in neocortex
- depends on mere repetition, not behavioral significance
Existing Accounts


Cells unnecessary for identifying a stimulus are suppressed.

Ringo (1996)

Suppression of familiar stimuli may contribute to automatic orientation to novel stimuli.
Our Account: Basic Assumptions

1. A subset of cortical neurons encode binary hypotheses.
Our Account: Basic Assumptions

1. A subset of cortical neurons encode binary hypotheses.

2. A neuron’s response is roughly sigmoidal.
1. A subset of cortical neurons encode binary hypotheses.

2. A neuron’s response is roughly sigmoidal.

- **Diagram:**
  - **xlabel:** net input strength
  - **ylabel:** firing rate
  - **Legend:**
    - true
    - false
Our Account: Basic Assumptions

1. A subset of cortical neurons encode binary hypotheses.

2. A neuron’s response is roughly sigmoidal.


Noise may be intrinsic to neural dynamics, due to integration of conflicting cues, or due to the failure of attention to suppress irrelevant input.
Equalization

In digital communication systems that operate on underlying analog representations, effect of noise is often mitigated by an equalization process.
Equalization

In digital communication systems that operate on underlying analog representations, effect of noise is often mitigated by an equalization process.

Adaptive equalization involves training to produce the correct response in the presence of noise.
Equalization In A Neural Network

input -> neural net communication & equalization -> corrected output

target output
Equalization In A Neural Network

If target response is known, adapt weights with LMS.

If target response not known but noise corruption is small, target can be inferred.
Blind Equalization

Unsupervised error-correction procedure based on inferred target

Complements supervised and reinforcement learning mechanisms
Neuron Model

Model a biological neural net via connectionist abstraction.

Activation in $[0, 1]$ interpreted as mean firing rate relative to the maximum firing rate.

$$net = \sum_{i} w_i x_i + b$$

$$y = f(net)$$

input firing rate  
weights and bias  
output firing rate
Simulation Details

Two layer architecture

- 1000 input and output neurons

Weights initialized randomly.

Bias set so for each output neuron so that it typically produces a “true” response for a fraction $\alpha$ of inputs.

Because neural codes are sparse, we assume $\alpha = .20$.

Two sets of randomly-generated input patterns

- Repeated
- Nonrepeated

Ten passes through data during which all patterns are tested, and repeated patterns are adapted via blind equalization.
Simulation Result

Graded, item specific reduction in activity

Long lasting

Depends on mere repetition
Sharpening of stimulus representation

- response of most neurons decreases
- response of a small fraction of neurons increases
Interpretation: Gain Control of Responses

- f(firing rate) vs. net input strength
- g(firing rate) vs. net input strength
Previous models have exploited gain control (e.g., Cohen & Servan-Schreiber, 1992; Kello & Plaut, 2003).

Here, gain adaptation is item specific.
Practice and Noise Robustness (Rainer & Miller, 2000)

Delayed match-to-sample paradigm

Familiar and novel stimuli

Recording from neurons in lateral prefrontal cortex
**Practice and Noise Robustness (Rainer & Miller, 2000)**

Delayed match-to-sample paradigm

Familiar and novel stimuli

Recording from neurons in lateral prefrontal cortex

**Manipulated stimulus noise**
Results (Rainer & Miller, 2000)

For familiar stimuli versus novel stimuli

• repetition-suppression-like sharpening of representations
Results (Rainer & Miller, 2000)

For familiar stimuli versus novel stimuli

- repetition-suppression-like sharpening of representations
- greater selectivity of neural responses
Results (Rainer & Miller, 2000)

For familiar stimuli versus novel stimuli

- repetition-suppression-like sharpening of representations
- greater selectivity of neural responses
- noise-robust neural responses
Results (Rainer & Miller, 2000)

For familiar stimuli versus novel stimuli

- repetition-suppression-like sharpening of representations
- greater selectivity of neural responses
- noise-robust neural responses

Can these effects of practice be explained by the same mechanism as that which produces repetition suppression?
Sigmoid squashing leads to less response variability when neurons are firing near min and max levels.
Does Noise Robustness Depend on Discriminative Training?

Rainer and Miller task involved discriminative training, associated with reward.

Model claims that observed noise robustness with practice is not related to reward or task relevance.
Does Noise Robustness Depend on Discriminative Training?

Rainer and Miller task involved discriminative training, associated with reward.

Model claims that observed noise robustness with practice is not related to reward or task relevance.

Model prediction

Rainer and Miller effects should be observed in passive viewing.

Supporting behavioral data (Goldstone, Macintosh)
Selectivity of Response

Neural activity for familiar objects is more narrowly tuned than for novel objects.

Rainer and Miller (2000)

\[ S_{\text{familiar}} = 0.30, \quad S_{\text{novel}} = 0.24 \]

\[ S_i = \frac{n - \sum y_i^j}{\max(y_i^j)} \]

where \( j \) is index over alternative objects

Simulation

\[ S_{\text{familiar}} = 0.66, \quad S_{\text{novel}} = 0.40 \]
Proportion of Active Neurons

Prediction

In representations with a high fraction of the population active, no repetition suppression, and possibly repetition enhancement.

Supporting data

No repetition suppression in V1, and amount of suppression increases with level in visual hierarchy

Slight repetition enhancement for pseudowords in left posterior fusiform gyrus
Repetition Suppression and Repetition Priming

Repetition suppression is thought to mediate priming.
Repetition Suppression and Repetition Priming

Repetition suppression is thought to mediate priming.

Model provides a concrete link.

Add temporal dynamics by turning neurons into leaky integrators (e.g., McClelland, 1979):

$$y_i(t) = \tau y_i(t - 1) + (1 - \tau)f(\text{net}_i(t))$$

Response rule: Initiate response when all binarized outputs are at their asymptotic value,

$$\Psi(y_i(t)) = \Psi(y_i(\infty))$$
Repetition Suppression and Repetition Priming

Repetition suppression is thought to mediate priming.

Model provides a concrete link.

Add temporal dynamics by turning neurons into leaky integrators (e.g., McClelland, 1979):

\[ y_i(t) = \tau y_i(t-1) + (1-\tau)f(n_{et_i}(t)) \]

Response rule: Initiate response when all binarized outputs are at their asymptotic value, \( \Psi(y_i(t)) = \Psi(y_i(\infty)) \)
Repetition Suppression and Repetition Priming

Repetition suppression is thought to mediate priming.

Model provides a concrete link.

Add temporal dynamics by turning neurons into leaky integrators (e.g., McClelland, 1979):

\[ y_i(t) = \tau y_i(t-1) + (1 - \tau) f(\text{net}_i(t)) \]

Response rule: Initiate response when all binarized outputs are at their asymptotic value, \( \Psi(y_i(t)) = \Psi(y_i(\infty)) \)
How To Interpret Learning Rule When Neurons Have Temporal Dynamics

Blind-Equalization Learning Rule

\[ \text{error}_i(t) = y_i(t) - \Psi(y_i(t)) \]
How To Interpret Learning Rule When Neurons Have Temporal Dynamics

Blind-Equalization Learning Rule
\[ \text{error}_i(t) = y_i(t) - \Psi(y_i(t)) \]

Temporal-Difference Learning Rule, TD(1)
\[ \text{error}_i(t) = y_i(t) - y_i(\infty) \]

Temporal-Difference Learning Rule, TD(0)
\[ \text{error}_i(t) = y_i(t) - y_i(t + 1) \]

All three rules have the same effect!

Learning rule can be viewed as compressing the time course of processing.
Recap

Individual second-to-second perceptual experiences have long-term effects on cognition

Proposed an unsupervised learning mechanism that

• explains changes to neural representations with passive viewing of stimulus
• explains noise robustness in perceptual discrimination
• links neuroscientific phenomena (repetition suppression) and psychological phenomena (repetition priming)

Can interpret the mechanism as

• mitigating the effects of noise on intrinsically binary representations
• improving processing efficiency
Levels of Abstraction

neurobiological  psychological  computational
Levels of Abstraction

blind-equalization model

neurobiological psychological computational
Levels of Abstraction

- Neurobiological
- Psychological
- Computational

- Blind-equalization model
- Probabilistic information transmission (PIT) model
Key Properties Shared Between Models

Transmission of information in a processing pathway is gradual.
Key Properties Shared Between Models

Transmission of information in a processing pathway is gradual.

With each experience, pathway tends to produce its response more rapidly.
Key Properties Shared Between Models

Transmission of information in a processing **pathway** is gradual.

With each experience, pathway tends to produce its response more rapidly.

This adaptation is **stimulus specific**.
Pathway as a Dynamic Belief Net

pathway output

pathway input

\[ y_0 \rightarrow y_1 \rightarrow y_2 \rightarrow y_3 \rightarrow \cdots \]

\[ x_1 \rightarrow x_2 \rightarrow x_3 \]
**Informal description**

Input at time $t$ combines with output at time $t-1$ to determine output at time $t$.

**Formal description**

Each node is a random variable.

Each arrow indicates a conditional dependency.
Illustration of Pathway Behavior

concept
visual pattern

"one" "two" "six" "nine"

y0 → y1 → y2 → ... → yt

x1 → x2 → xt
Simple Illustration of Pathway Behavior

![Diagram showing pathway behavior](image)

Graph showing change in p(y) with increase in # iterations.
Inference can be performed from x to y regardless of direction of arrows.
Knowledge in Model

Transmission probability distribution: $p(x_t = i \mid y_t = j)$

Strength of association between input and output

Transition probability distribution: $p(y_t = i \mid y_{t-1} = j)$

Output persistence
Transmission Probability Distribution

\( \alpha_{ij} \): count of past experiences with \( x = i \) leading to \( y = j \)

\[
p(x_t = i \mid y_t = j) \sim 1 + \alpha_{ij}
\]
Transmission Probability Distribution

\[ \alpha_{ij} : \text{count of past experiences with } x = i \text{ leading to } y = j \]

\[ p(x_t = i | y_t = j) \sim 1 + \alpha_{ij} \]

\[ \gamma_{ik} : \text{similarity between inputs } x = i \text{ and } x = k \]

\[ p(x_t = i | y_t = j) \sim 1 + \sum_{k} \gamma_{ik} \alpha_{ik} \]
Inference in a Dynamic Belief Network

Given
- transmission probability distribution, $p(x_t | y_t)$
- transition probability distribution, $p(y_t | y_{t-1})$
- prior distribution, $p(y_0)$

can compute $y_t$ based on $x_1 ... x_t$:

$$p(y_t = j | x_1, x_2, x_3, ..., x_{t-1}, x_t) \sim \left[ \sum_{k=1}^{N_s} p(y_{t-1} = k)p(y_t = j | y_{t-1} = k) \right] \left[ \sum_{i=1}^{N_x} p(x_t = i)p(x_t = i | y_t = j) \right]$$
Full Model

Action Pathway

Perceptual Pathway

Action pathway reset
Action pathway output
Action pathway input
Perceptual pathway reset
Perceptual pathway output
Perceptual pathway input
World state
Priming

Logically, what learning mechanisms could shift the response curve with experience?
Priming

Logically, what learning mechanisms could shift the response curve with experience?

Experience with input $i$ leading to output $j$ could cause:

- more efficient signal transmission, $\alpha_{ij}$

![Graph showing the probability of correct response over iterations](image-url)
Priming

Logically, what learning mechanisms could shift the response curve with experience?

Experience with input $i$ leading to output $j$ could cause:

- more efficient signal transmission, $\alpha_{ij}$
- revising priors, $y_j(0)$

(model of environment)
Primeing

Logically, what learning mechanisms could shift the response curve with experience?

Experience with input i leading to output j could cause:

- more efficient signal transmission, $\alpha_{ij}$
- revising priors, $y_j(0)$ (model of environment)
- both

![Graphs showing the effect of iterations on the probability of correct response for different scenarios.](image-url)
Priming

Logically, what learning mechanisms could shift the response curve with experience?

Experience with input i leading to output j could cause:

- more efficient signal transmission, $\alpha_{ij}$
- revising priors, $y_j(0)$ (model of environment)
- both

Both mechanisms are required to explain behavioral data.
Bowers (1999); McKoon and Ratcliff (2001)

Study phase

Perceptual identification task

both studied  neither studied
Study phase

Perceptual identification task

both studied  neither studied  both studied  neither studied
high frequency  low frequency
Discriminability of low-frequency words increases.  
Discriminability of high-frequency words does not.
Discriminability of low-frequency words increases.
Discriminability of high-frequency words does not.
Results

*Human data*

- **p(correct response)** vs. **word frequency**
  - Low
  - High
  - *p < .05*
  - n.s.
  - ○ both
  - × neither

*Model*

- **p(correct response)** vs. **word frequency**
  - Low
  - High
  - ○ both
  - × neither

Explanation

- **time to criterion** vs. **alpha**
  - 0
  - 0.5
  - 1
  - 50
  - 100
  - 150

- **log(time to criterion)** vs. **log(alpha)**
  - -3
  - -2
  - -1
  - 0
Data Explained By The PIT Model

Priming results in a bias to report primed words (Ratcliff & McKoon, 1997, Experiment 3).

With low frequency words, priming also results in increased sensitivity to primed words (Bowers, 1999; McKoon & Ratcliff, 2001).

Bias effects are negligible if response alternatives are dissimilar (Ratcliff & McKoon, 1997, Experiment 4).

Greater effect of target presentation duration on accuracy if response alternatives are dissimilar (Ratcliff & McKoon, 1997, Experiment 4).

Decay of bias effect when time between study and test is increased (Ratcliff & McKoon, 1997, Experiments 6 & 7).

Interaction between bias effect and response paradigm—naming, forced choice, and yes/no matching (Ratcliff & McKoon, 1997, Experiments 6 & 8; Wagenmakers et al., 2000)

Interaction between bias effect and lexicality (Steyvers et al., 2001)

Sensitivity effects in stem completion (Zeelenberg, et al., 2002)

Effects of trial sequence and response conflict in speeded discrimination (Jones et al., 2002)

Power law of practice

Hick-Hyman law of choice reaction time

Subliminal priming phenomena (Abrams & Greenwald, 2000; Bar & Biederman, 1998)
Do We Need Yet Another Modeling Framework?

Several other models can explain all or most of these data.

REM (Schooler, Shiffrin, & Raaijmakers, 2001)
Counter model (McKoon & Ratcliff, 2001; Ratcliff & McKoon, 1997)

Strengths of PIT

• Crisp, parsimonious theoretical account
• Many cognitive phenomena well characterized at level of information flow
• Incorporates what’s right about connectionist models
  treating inference as probabilistic (explicitly!)
  treating low-level learning as statistical
  allows for generalization by virtue of similarity of representations
• Avoids what’s wrong
  currency of “activation”
  arbitrary dynamics
  uninterpretable parameters
How Do Individual Second-to-Second Experiences Impact Cognition?

Described two complementary models incorporating unsupervised learning:

- pseudo-neural model
- probabilistic information transmission (PIT) model

**Claims of Both Models**

- learning makes subsequent processing more noise robust and efficient
- power law of practice
- item specific
- long lasting
How Do Individual Second-to-Second Experiences Impact Cognition?

Described two complementary models incorporating unsupervised learning

- pseudo-neural model
- probabilistic information transmission (PIT) model

<table>
<thead>
<tr>
<th>Claims of Both Models</th>
<th>Facts about Skill Learning (Refinement, Practice, Tuning, Rehearsal, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>learning makes subsequent processing more noise robust and efficient</td>
<td>improvement in fluency, flexibility, and performance with practice</td>
</tr>
<tr>
<td>power law of practice</td>
<td>power law of practice</td>
</tr>
<tr>
<td>item specific</td>
<td>specificity (Healy &amp; Bourne, 1995)</td>
</tr>
<tr>
<td>long lasting</td>
<td>durability (Healy &amp; Bourne, 1995)</td>
</tr>
</tbody>
</table>