An Introduction to Learning

Lecture 5/14

Todd M. Gureckis
Department of Psychology
New York University
todd.gureckis@nyu.edu
Agenda for Today

- Classical Conditioning (Part II)
  - Attention, associability, and learning (picking up where we left off last time)
  - Neural substrates, neural dissociations in conditioning
  - Temporal-Difference Learning and the relationship to Rescorla-Wagner model
  - Modern computational theories - causal theories of conditioning
1 Attention, associability, and learning.
Modulations in CS Effectiveness: Selective Attention?

Learning and selective attention

Peter Dayan¹, Sham Kakade¹ and P. Read Montague²

¹ Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR, UK
² Division of Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA
Correspondence should be directed to P.D. (dayan@gatsby.ucl.ac.uk)

- Selective attention to cues is usually studied in a resource-limited fashion (we can’t attend to everything due to capacity limitations, so must choose)

- One way to think about the relationship between learning and CS effectiveness is as an **attentional processes that highlight information in the world deemed useful for learning**
Modulations in CS Effectiveness: Selective Attention?

**Fig. 1.** Uncertainty and unreliability. Top, rewards of various magnitudes (size or number of food pellets) are given. Bottom, lights of different colors represented by binary $x_i(t)$ potentially predict reward. Time $t_*$, the first time the reward associated with the blue stimulus changes; time $t_s$, when the red light is introduced; time $t_f$, the end of the plots in Fig. 2; time ‘?’ when the prediction associated with the combination of red and green lights is assessed.
The Kalman Filter

Optimal model for mean-tracking in a non-stationary task (developed for NASA to predict motion of flying objects... key point - system under internal dynamics and measurement uncertainty)

associability term just like in Pearce-Hall model!

\[ \hat{w}_i(t+1) = \hat{w}_i(t) + \alpha_i(t)\delta(t) \]
\[ \delta(t) = r(t) - x(t) \cdot \hat{w}(t) \]
\[ \alpha_i(t) = \frac{\sigma^2_i(t)x_i(t)}{\sum_j \sigma^2_j(t)x_j(t)+E} \]

However, it takes into account our uncertainty in the estimate of our prediction

Basic idea: More learning should come when we are uncertain, and when our predictions are off
**The Kalman Filter**

**Fig. 2.** Prediction with the Kalman filter up to time $t_F$. (a) Mean value of predictions associated with the blue and red lights. (b) The net prediction (brown) of the actual reward (black) that would be made according to the sum of predictions of the stimuli. Note how the prediction adapts to the change in the actual rewards. (c) Uncertainties associated with the blue and red lights. Larger uncertainties (as at the start of learning for each stimulus) allow larger changes in the predictions. Note that there is asymptotic uncertainty, that is, the estimates do not become arbitrarily accurate (because of the continual possibility of change in the world). Predictions and uncertainties for the green and yellow lights are not shown.
Modulations in CS Effectiveness: Selective Attention?

Fig. 1. Uncertainty and unreliability. Top, rewards of various magnitudes (size or number of food pellets) are given. Bottom, lights of different colors represented by binary $x_i(t)$ potentially predict reward. Time $t_*$, the first time the reward associated with the blue stimulus changes; time $t_S$, when the red light is introduced; time $t_F$, the end of the plots in Fig. 2; time ‘?’ when the prediction associated with the combination of red and green lights is assessed.
Modulations in CS Effectiveness: Selective Attention?

Since green has been a less perfect predictor, the combination R+G combines them in a normative way.

Fig. 4. Competitive combination of predictions. Red and green lines, example distributions of predictions of reward from the red and green lights at the time marked by ‘?’ in Fig. 1. The prediction of the red light is more reliable, as its distribution is more sharply peaked. Brown line, combined prediction after information from both stimuli has been integrated. As the red light is more reliable, the mean of the combined estimate is closer to $w_2$. 
Normative accounts of changes in CS effectiveness

- In contrast to somewhat arbitrary (but successful mechanistic approaches) it is possible to view learning as a process of statistical estimation of relevant properties in the world.

- Leveraging these ideas (and mathematical tools) can give new insight into old problem and frame them in a more transparent way.

- “Psychological” terms like associability are given new meaning in these approaches by casting them in clear, easy to understand, and well motivated terms.

- Emphasis on more sophisticated understandings of what it means to learn (why would the human or rat assume that the fixed contingencies in the experiment are all there is to learn?). The world is changing and different assumptions of that change influence what is or isn’t learnable.
2 Neural Substrates
Classical Conditioning, Kandel, and Aplysia

Aplysia californica from Elkhorn Slough, California

dorsal aspect, from life

Aplysia californica from Elkhorn Slough, California

Before conditioning

CS+ Sensory neuron

Motor neuron

Conditioned

CS+ Sensory neuron

Motor neuron

Retracted gill

Retracted retraction

Retracted

Sensory neuron

CS− Siphon

US Tail

Modulatory Interneuron

Motor neuron

Gill
Cellular Mechanisms of Classical Conditioning in Invertebrae
Classical Conditioning in Mammals

A. Eye Blink Preparation

B. Conditioned Responses

- Day 1: Eyeblink UR
- Day 3: Weak Eyeblink CR
- Day 5: Strong Eyeblink CR

CS (tone) on
US (airpuff) on

Time
Classical Conditioning in Mammals

(b)

Frontal lobe

Parietal lobe

Occipital lobe

Temporal lobe

Cerebellum

Brainstem
Cerebellum is a critical brain region involved in the control and timing of motor processes.

Cerebellar cortex & deeper nuclei - interpositus nuclei both receive inputs from sensory pathways relevant for eyeblink conditioning (CS tone and US - air puff).

Deactivation or damage to interpositus nucleus impairs acquisition of association, but inhibition of red nucleus/motor nuclei do not.
Cerebellar contributions to classical conditioning

Figure 7.16 Electrophysiological recordings in the rabbit cerebellum during classical conditioning
(a) Response of a trained rabbit to the CS. (b) Response of an untrained naive rabbit to the CS alone [top] and to the US alone (bottom). The blue line shows the eyeblink behavior (the extent of eyelid closure over time), while the lower graph shows the frequency of neuronal firing in the interpositus nucleus. Adapted from McCormick and Thompson, 1994.

Neural activity in the interpositus shows no or low response to the CS alone, US alone.

In contrast, in the trained animal, interpositus activity follows the CS and is a strong predictor of the CR.
Cerebellar contributions to classical conditioning

Error-correction process in the cerebellar circuit!

Inhibitory pathways from interpositus nucleus to trigeminal nucleus.

These effectively reduces the size of the activation coming from trigeminal nucleus (actual-expected), helping us compute the prediction error.

Indeed, possible to block “blocking” effects by stopping these inhibitory connections (Kim, Krupa, & Thompson, 1998)
Contributions of other systems

- Hippocampus activity often observed during classical conditioning
- One possible role is in modulating CS effectiveness (Mackintosh, Pearce-Hall, etc.)
- Another is for building representations between context, pairs of cues, etc...
Hippocampal contribution to sensory preconditioning

Percent CRs to Tone During Phase 3 Testing

- Compound Exposure Normal Rabbits
- Separate Exposure Normal Rabbits
- Compound Exposure Lesioned Rabbits
- Separate Exposure Lesioned Rabbits
Hippocampal contribution to trace conditioning

(A) Delay conditioning

US

Delay

CS

Time

CS on

US on

CS&US off

(B) Trace conditioning

US

Trace interval

CS

Time

CS on

CS off

US on

US off

(C) Trace conditioning in rabbits

Trials to Criterion

2000

1600

1200

800

400

0

Control Rabbits

Lesioned Rabbits
Hippocampal amnesics selectively impaired at trace conditioning

In healthy normals awareness of contingency lead to normal patterns
no difference between groups in the delay condition

selective impairment for unaware participants and for amnesics
The take-home

- Classical conditioning manifests itself in different ways across species.

- There is unlikely to be a single “classical conditioning” system but rather is a basic principal of learning.

- In mammals, a popular task is the eyeblink conditioning experiment and careful neuropsychological work highlights the role of the cerebellum. This circuit has properties much like Rescorla-Wagner learning process (error-corrective learning of CS-US relationship).

- However, other systems (e.g., the hippocampus) are likely involved in mediating processes like sensory preconditioning, latent inhibition, and the hippocampus (and awareness of the contingencies) appears critical for the acquisition of trace conditioning.
3 Temporal-difference (TD) learning
Rescorla & Wagner (1972)

<table>
<thead>
<tr>
<th>phenomena explained</th>
<th>RW</th>
<th>a better theory?</th>
</tr>
</thead>
<tbody>
<tr>
<td>acquisition curves</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>extinction curves</td>
<td></td>
<td>kinda</td>
</tr>
<tr>
<td>blocking</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>overshadowing</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>temporal relationships</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>overexpectation</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>second-order conditioning</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>latent inhibition</td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>
Latent Inhibition

Phase 1
Phase 2

Result: Unpaired presentations of light lead to shock
No learning/prediction error on unpaired trials? RW can’t learn anything then.
Second-order conditioning

Phase 1

Phase 2

Test

Why does the RW/PH theory have trouble with this?

Special status of the US stimulus... it must be the thing linked to the UR
Timing

RW/PH are silent on issues of timing.
Rescorla & Wagner (1972)

- Prediction error all the way!

\[ \delta = R_{US} - V_{s_1,s_2} \]

prediction error = actual US - expected US

- Update Equations

\[ \Delta V_{s_1} = \alpha_{s_1} \beta_1 (R_{US} - V_{s_1,s_2}) \]

\[ \Delta V_{s_2} = \alpha_{s_2} \beta_1 (R_{US} - V_{s_1,s_2}) \]
What are the weights ($V$s)? Associative weights between the CS and the US.

Another way to say it is that the representation in $V$ is the predictive “value” of the cue (in fact that is why we used the letter $V$). Under Rescorla-Wagner, when the value for the organism of the US is positive (money, sex, drugs, food) then the value of $V$ will end up mostly positive. If the value is low (shock, pain, airpuff) then the value of $V$ will be negative.

The values thus represent our expectation about the future. Indeed differences between the actual outcomes ($R_{\text{US}}$ and $V$) drive learning via the prediction error!}

Rescorla & Wagner (1972)
Maximizing Reward

- But how can we explain the above phenomena? One approach is to start over with a computational-level description of the problem and see where it gets us.

- Critically, we will take time (either within a trial or across a series of trials) seriously right from the start.

- In particular, we want our values ($V_t$) to reflect the expected reward that we will get in the future. Sounds reasonable enough, this is was RW was kind of already doing, although the values were linked to particular CSs rather than particular states in time.

\[
R_t = r_{t+1} + r_{t+2} + r_{t+3} + \ldots
\]

\[
V_t = E\left[ \sum_{i=t+1}^{\infty} r_i \right]
\]
Maximizing Reward

How do we learn that? Before we just made the error term the difference between our expectation and the outcome we were interested in predicting

\[ \delta_t = \sum_{i=t}^{T} r_i - V_t \]
Maximizing Reward

How do we learn that? Before we just made the error term the difference between our expectation and the outcome we were interested in predicting

$$\delta_t = \sum_{i=t}^{T} r_i - V_t$$
Maximizing Reward

Can basically solve this using the Bellman Equation

\[ V_t = E[r_{t+1} + r_{t+2} + ... + r_T] \]

\[ V_t = E[r_{t+1}] + E[r_{t+2} + ... + r_T] \]

\[ V_t = E[r_{t+1}] + V_{t+1} \]

Notice this “chaining” relationship, value at time \( t \) depends on expected reward on the next trial and the value going of being the state \( t+1 \) (defines a system of equations were some values can be computed based on others)
Maximizing Reward

The learning algorithm which approximates the full Bellman Solution (temporal difference):

\[ V_t = E[r_{t+1}] + V_{t+1} \]

\[ V_t \leftarrow V_t + \eta(r_{t+1} + V_{t+1} - V_t) \]

\[ V_t \leftarrow (1 - \eta)V_t + \eta(r_{t+1} + V_{t+1}) \]

Compare with Rescorla-Wagner

\[ V_t \leftarrow V_t + \eta(r_{t+1} + V_t) \]
Temporal-difference learning and “backing up”
Dopamine neurons do it too...
Dopamine neurons do it too...

Also correlates with magnitude of reward, degree of intermittent reinforcement, etc...
The role of dopamine in reward learning
Temporal Difference Learning

- Temporal difference learning is in many ways a “better” version of Rescorla-Wagner learning.

- Derived from first principles (from definition of the problem as estimating future rewards).

- Explains everything that R-W does, and more (even 2nd order conditioning).

- Basically a generalization of R-W to real time. This is the key difference, RW was meant to be a trial level model, while TD models the continuous flow of time.

- TD provides a basic mechanism for solving the **credit assignment problem** (how to credit earlier actions or effects for later rewards). This is why it is so interesting to both neuroscientists and computer scientists. ==> It can actually be used to DO stuff like fly airplanes, play backgammon, etc...
<table>
<thead>
<tr>
<th>phenomena explained</th>
<th>RW</th>
<th>a better theory?</th>
</tr>
</thead>
<tbody>
<tr>
<td>acquisition curves</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>extinction curves</td>
<td>kinda</td>
<td>kinda</td>
</tr>
<tr>
<td>blocking</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>overshadowing</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>temporal relationships</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>overexpectation</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>second-order conditioning</td>
<td>×</td>
<td>√</td>
</tr>
</tbody>
</table>
4 Modern theories - causal approaches to conditioning
Causal Interpretations of Conditioning  

(Michael Waldmann, Ralph Miller, Nathaniel Daw, Yael Niv, and others...)

- Classical models of conditioning (e.g., Rescorla-Wagner) assume that the organism is hard-wired with some mechanisms for assessing the associative relationship between events in the environment (CS - US)

- However, the number of possible pairings is quite large (CSxUS where CS is all possible stimuli and US are all possible outcomes/rewards) and so if our knowledge about the world was based strictly on this we would need a lot of experience

- An alternative approach is to assume the organisms attempt to build a **generative model** of their environment that includes the possibilities of latent (unobservable) causes... note that this draws the learning closer to the unsupervised structure learning we discussed in lecture 3
The role of latent variables

Assuming latent structures/causes can reduce the number of parameters that need to be estimated to structure a domain
Generative vs. Discriminative Learning

- Discriminative models use cues (x) in the environment and attempt to learn the mapping to outcomes (c). All learning is conditional on the initial stimulus (x).

- Generative models assume that all events come from (are generated by) some latent cause or reason and the goal of learning is to estimate the properties of this generating distribution.

- Generative models notably include a model of all variables of interest, not just the rewards.
Conditioning as inference of a Latent Causal Model

Critically, recasts learning not as the estimate of a particular set of parameters within a particular structure (i.e., structure 1) but inferences of the appropriate model itself (e.g., Courville, et al. reading)

Learning via Bayes Rule

\[ P(S|D) = \frac{P(D|S)P(S)}{\sum_{S'} P(D|S')P(S')} \]

Occam’s Factor

All else being equal, a more complex structure/model (blue) will be able to explain a wider range of observable patterns. As a result these model assign lower probability to any single observation (compared with a simpler, more limited model)

This gives an automatic preference for simple models/explanations with limited data

What is the relationship between that irritating tone and the shock?
What is the experimenter trying to get me to learn this time about that irritating tone and the shock?

experimenter/context/cause

CS  US
Model Uncertainty in Classical Conditioning

A. C. Courville*1,3, N. D. Daw2,3, G. J. Gordon4, and D. S. Touretzky2,3
1Robotics Institute, 2Computer Science Department,
3Center for the Neural Basis of Cognition,
4Center for Automated Learning and Discovery
Carnegie Mellon University, Pittsburgh, PA 15213
{aaronc,daw,ggordon,dst}@cs.cmu.edu

(a) Sigmoid belief network

(b) Marginal likelihood
Model Uncertainty in Classical Conditioning

A. C. Courville\textsuperscript{1,3}, N. D. Daw\textsuperscript{2,3}, G. J. Gordon\textsuperscript{4}, and D. S. Touretzky\textsuperscript{2,3}

\textsuperscript{1}Robotic Institute, \textsuperscript{2}Computer Science Department, \textsuperscript{3}Center for the Neural Basis of Cognition, \textsuperscript{4}Center for Automated Learning and Discovery
Carnegie Mellon University, Pittsburgh, PA 15213
\{aaronc, daw, ggordon, dst\}@cs.cmu.edu

Two key learning questions:

\begin{itemize}
  \item What is the structure?
  \item What is the parameters of that structure (the weights between latent causes and observable variables)?
\end{itemize}
Prediction of any outcome (US) is made by averaging over all possible settings of the parameters/weights in the system

\[
P(US \mid CS, m, D) = \int P(US \mid CS, w_m, m, D)p(w_m \mid m, D) \, dw_m
\]
Select models using Bayes rule (conveying the automatic Occam’s razor, as well as a in this specific instance a prior that favors a smaller number of nodes in the representation)

\[
P(US \mid CS, \mathcal{D}) = \sum_m P(US \mid CS, m, \mathcal{D})P(m \mid \mathcal{D})
\]

\[
= \sum_m \int P(US \mid CS, w_m, m, \mathcal{D})p(w_m \mid m, \mathcal{D})P(m \mid \mathcal{D})d w_m
\]

The posterior over models, \(p(m \mid \mathcal{D})\), is given by:

\[
P(m \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid m)P(m)}{\sum_{m'} P(\mathcal{D} \mid m')P(m')}, \quad P(\mathcal{D} \mid m) = \int P(\mathcal{D} \mid w_m, m)p(w_m \mid m)d w_m
\]
Shifting between second-order and conditioned inhibition

<table>
<thead>
<tr>
<th>Group</th>
<th>A-US</th>
<th>A-X</th>
<th>B-US</th>
<th>Test (\sim)Result</th>
<th>Test (\sim)Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-X</td>
<td>96</td>
<td>0</td>
<td>8</td>
<td>(X \sim)–</td>
<td>(XB \sim)CR</td>
</tr>
<tr>
<td>Few-X</td>
<td>96</td>
<td>4</td>
<td>8</td>
<td>(X \sim)CR</td>
<td>(XB \sim)CR</td>
</tr>
<tr>
<td>Many-X</td>
<td>96</td>
<td>48</td>
<td>8</td>
<td>(X \sim)–</td>
<td>(XB \sim)–</td>
</tr>
</tbody>
</table>

Table 1: A summary of some of the experiments of Yin et al. [11]. The US was a footshock; \(A\) = white noise or buzzer sound; \(X\) = tone; \(B\) = click train.

Few A-X pairing mean X predicts CS (second order)
Shifting between second-order and conditioned inhibition

Table 1: A summary of some of the experiments of Yin et al. [11]. The US was a footshock; A = white noise or buzzer sound; X = tone; B = click train.

<table>
<thead>
<tr>
<th>Group</th>
<th>A-US</th>
<th>A-X</th>
<th>B-US</th>
<th>Test → Result</th>
<th>Test → Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-X</td>
<td>96</td>
<td>0</td>
<td>8</td>
<td>X →—</td>
<td>XB → CR</td>
</tr>
<tr>
<td>Few-X</td>
<td>96</td>
<td>4</td>
<td>8</td>
<td>X → CR</td>
<td>XB → CR</td>
</tr>
<tr>
<td>Many-X</td>
<td>96</td>
<td>48</td>
<td>8</td>
<td>X →—</td>
<td>XB →—</td>
</tr>
</tbody>
</table>

With more, X becomes an inhibitor.
Small amount of data, X is integrated into common cause
Later, X is separately represented as an inhibitor
Complexity of model grows with more training
Context effects, extinction, and the creation of new latent causes

What is the experimenter trying to get me to learn this time about that irritating tone and the shock?

- CS
- US
- light
- shock

Blue room
Context effects, extinction, and the creation of new latent causes

What is the experimenter trying to get me to learn this time about that irritating tone and the shock?

CS

blue room

light

US

shock

CS

red room

light
Context effects, extinction, and the creation of new latent causes

Maybe there are two contexts here? Red room is safe, blue room isn’t? How did I learn that?
Basic idea is that contextual effect represent adding a new latent cause into the model.

What is the procedure for doing this? How many latent causes should there be?

Highly related to models of categorization that try to model the process of how people discover the latent structure of categories in the world (in fact all the math is borrowed from classic work in categorization, Anderson, 1991)

Argument is that this can be done in a normative fashion using a model that places no a-priori restriction on the number of states (called an infinite-capacity model)
Context, Learning, and Extinction

Samuel J. Gershman, David M. Blei, and Yael Niv
Princeton University

Generative model

1. Choose a latent state or context

2. Choose a set of rewards/stimuli conditioned on that context
Prior over latent causes/clusters:

\[
P(c_{t+1} = k|c_{1:t}) = \begin{cases} 
\frac{N_k}{t + \alpha} & \text{if } k \leq K_t \text{ (i.e., } k \text{ is an old cause)} \\
\frac{\alpha}{t + \alpha} & \text{if } k = K_t + 1 \text{ (i.e., } k \text{ is a new cause),}
\end{cases}
\]

\[a=0 \rightarrow 1 \text{ latent cluster, } a=\infty \text{ new latent cluster on every trial}\]
same basic story applies to the context-dependence of latent inhibition
Hippocampal pathologies negatively impact contextual conditioning!

Figure 4. Effect of hippocampal lesions on ABA renewal. a. Experimental conditioned responding to a cue during the test phase in control rats (CON) and those that received pretraining electrolytic lesions of the dorsal hippocampus (HPC). Data replotted from Ji and Maren (2005). b. Simulated conditioned responding following restriction of the model’s capacity to infer new clusters prior to training.

- Reduce the tendency to create new clusters/causes in the hippocampal damage simulations
- Shows extinction effects that generalize across contexts, since a single latent state is being inferred.
The take-home for today

- Modern theories of conditioning draw from normative, statistical models as inspiration.

- Rather than assuming arbitrary mechanistic-level descriptions, attempt to cast key ideas (associability, error-driven learning, context, renewal, cue combination, etc...) as ways that organisms learn about the structure of their environment in the face of considerable uncertainty.

- As in the classic work, heavily inspired by the technology and tools of the time (particular advances in Bayesian statistics, graphical and generative models, etc...).
Next time
From observation to action!
Instrumental Conditioning
Readings

Coming soon. Check the website Friday afternoon.
References for Slides


Lecture notes from Yael Niv (http://www.princeton.edu/~yael/PSY338/index.html) and Peter Dayan

The interweb.