Lecture 5: Learning II

Contents

Major Phenomenon of Classical Conditioning

Acquisition of Conditioned Responses

Pavlov noticed that the tendency for the CS (bell) to elicit a conditioned response (salivation) increased with the number of with the number of times the two are presented together. Thus, acquisition (i.e., learning a new, conditioned response) in classical conditioning is gradual.

A pairing that increases the strength of the CR (e.g., amount of salivation) is said to reinforce the connection. When the US (e.g., meat) is presented, it is a reinforcement trial; when the US is omitted, it is a unreinforced trial.

One can measure strength of the CR in terms of amplitude, probability of occurrence, or latency.

The time between the CS and US is important. If there is too much time between the two stimuli, the animal is unlikely to associate them.

The temporal relation is also important. Maximal conditioning usually occurs when the onset of the CS precedes the US. This is known as **forward conditioning**. **Simultaneous conditioning** is usually less effective, and **backward conditioning** (where the US is presented before the CS) is still less effective.

Second-order conditioning. Reflexes can be conditioned directly or indirectly. For example, the eye blink reflex in response to a puff of air (US) can be associated with a tone (CS). The tone can then be presented with a light (without the air puff) and the light will be come a second-order CS. Second-order conditioning is usual rather fragile and weak. (Pavlov and his colleagues even tried third-order conditioning.)

Extinction

The adaptive value of conditioning is self-evident. (Zebra fearing a location where a lion pounced before.) However, we might not want to keep all associating forever as this would be inefficient and paralyzing.

Extinction refers to the process whereby the conditioned reflex becomes weaker when the US and the CS stop being associated.

Pavlov showed that the CR (salivation) will gradually disappear if the CS (bell) is repeatedly presented without being reinforced by (i.e., paired with) the US (meat).

Reconditioning (after extinction) typically takes less time (i.e., fewer trials). This suggests that the CR is not really abolished but is somehow masked.

Spontaneous Recovery

Spontaneous recovery refers to a sudden reappearance of a CR (salivation in response to a bell) some time after extinction. (This happens in many instances.) This adds credence to the idea that the CR is not abolished in extinction.

Generalization

Generalization occurs when stimuli similar to original CS also evoke CR. For example, a change in the tone (frequency) of voice.

Generalization results in a "generalization gradient" where the the greater the difference between the original CS and the new CS, the weaker the CR. (Example: tone of voice.)

For example, Baby Albert was conditioned to fear rats but Albert was also fearful of the dog, the fur coat, etc. (i.e., other furry objects).

Discrimination

Discrimination is the "flip-side" of generalization. It refers to the ability to discriminate the CS from similar but unimportant stimuli. Discrimination can also be adaptive - we would not want to respond to a tiger and a kitten in the same way!

Example:

- Dog first conditioned to respond to a black square (CS+)
- Then reinforcement trials (with CS+ and US) are randomly interspersed with non-reinforced trials with another stimulus, say a gray square (CS-).
- This continues until the dog always salivates in response to CS+ and never to CS-.
- Initially, during learning, the dog will sometimes salivate in response to CS- (generalization) and will also sometimes not salivate in response to CS+.

Discrimination gets worse the more stimuli there are and the finer the required discrimination.

Applied Examples of Classical Conditioning

Borrowed from Larry Symon's overheads.

1. Food Advertisements

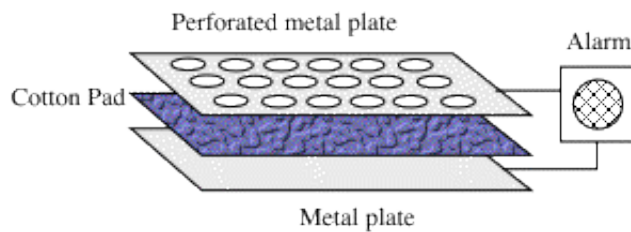- CS=Pizza Hut Logo
- US=Ooey Gooey Pizza Scenes
- UR=Salivation

- CR=Salivation to Pizza Hut Logo

2. Allergic Reactions

- CS=Flower
- US=Pollen
- UR=Sneezing, Wheezing
- CR=Sneezing to Flower

3. Treatment for Enuresis (bed wetting)

- CS=Full Bladder
- US=Alarm
- UR=Awakening
- CR=Awakening to Full Bladder

Other Types of Learning

This lecture focuses on types of learning that are different than classical conditioning. Classical conditioning looks at how we react to stimuli, both unconditioned and conditioned. Thus:

- Pavlov looked how dogs reacted to meat (US) and a bell (CS)
- Watson and Rayner looked at how Baby Albert reacted to a loud sound (US) and a white rat (CS)
- Garcia and colleagues looked at how a wolf reacted to lithium (US) and a sheep (CS)
- Advertisers are interested in how consumers react to their product (US) and their logo (CS)

However, humans and animals do not simply react to stimuli or events in the environment, they interact with the environment. In other words, not only is our behaviour influenced by events, events are often determined by our behaviour.

We often act to produce or obtain specific events or stimuli by behaving in specific ways. For example, I know that by pressing the red button on the overhead projector, I can turn the light on or off. Another example is the seal who learns to do a somersault in order to get a fish from the zoo attendant.

Such actions are called **instrumental responses**. "Instrumental" because the response acts like an instrument or tool to achieve a desired effect. "Response" because the action is produced in a particular context or situation. (Not a reflexive reaction.)

Thorndike and the Law of Effect

Edward Lee Thorndike (1874-1949)

While Pavlov was developing a general model of learning involving "reflexes" and classical conditioning (an approach that was becoming popular in Europe), Thorndike was also carrying out experiments on animal learning. Thorndike was interested in how animals learn to solve problems. His approach was fundamentally different than Pavlov's. While Pavlov was interested in how animals react to various stimuli, Thorndike was interested in how the animal responds to a situation in the environment in an effort to achieve some result.

If Thorndike had been in Pavlov's lab he would have wondered how dogs learn to produce specific behaviour in order to get food. (For example, some dog owners insist that their dog sit before being given food. Thorndike would have been interested in how the animal learns this behaviour.)

Note that people had been interested in instrumental learning for a number of years before Pavlov and Thorndike started their experiments on learning. In particular, they were interested in showing that animals were capable of intelligent behaviour as a way of defending Darwin's theory of evolution. This was considered important because people who attacked the Theory of Natural Selection argues that humans were fundamentally different than other animals in terms of there ability to reason. What set Thorndike apart from his predecessors was that he was the first to investigate instrumental learning systematically using sound experimental methods.

Thorndike's Puzzle Box Procedure

Thorndike placed a hungry cat inside a "puzzle box" with food outside. Initially, the cat would become agitated and produce many different "random" behaviours in an attempt to get out of the cage. Eventually, the cat would press the paddle by chance, the door would open and the cat could escape and get the food. The cat would then be placed inside the box again and would again take a long time (on average) to escape after exhibiting many different behaviours.

Puzzle Box
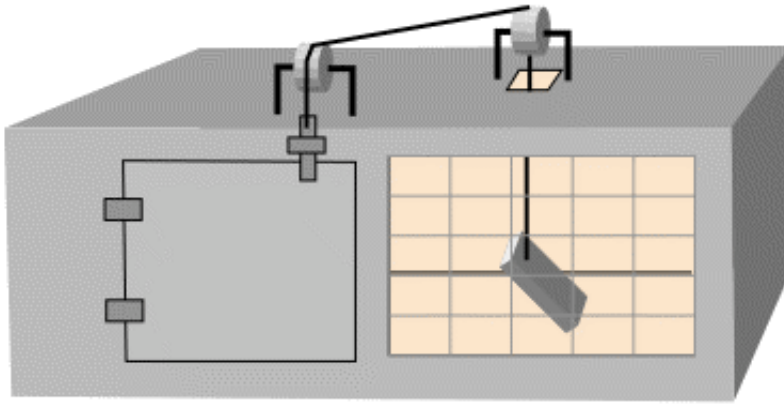
Thorndike examined the time to escape (his operational definition of learning) as a function of trials. The learning curve was gradual and uneven (see below). There was little evidence of sudden insight. Nevertheless, after about thirty trials, the cats would press the paddle almost as soon as they were placed in the cage. Thorndike concluded that the animals learned by "trial and error".

Successive trials in the puzzle box
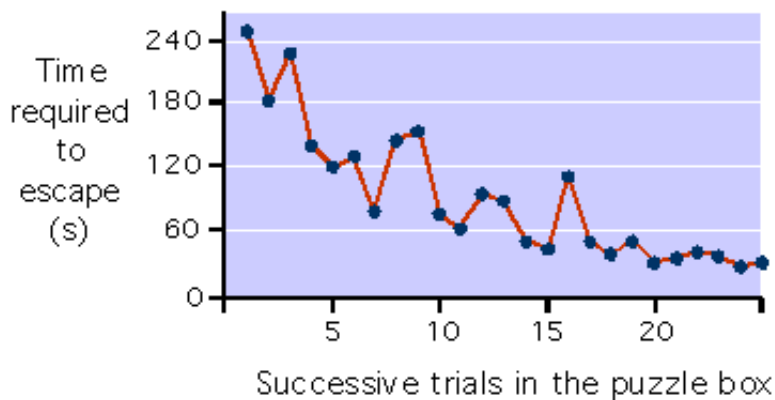
Based on observation such as these, Thorndike proposed a general theory of learning which is called the Law of Effect. This law of effect states that:

> The consequences of a response determine whether the tendency to perform it is strengthened or weakened. If the response is followed by a satisfying event (e.g., access to food), it will be strengthened; of the response is not followed by a satisfying event, it will be weakened."

The Law of Effect starts with the assumption that when an animal encounters a new environment, it will initially produce largely random behaviours (e.g., scratching, digging, etc.). Over repeated trials, the animal will gradually associate some of these behaviours with good things (e.g., access to food) and these behaviours will be more likely to occur again. In Thorndike's terms, these behaviours are "stamped in". Other behaviours that have no useful consequences are "stamped out" (see below).
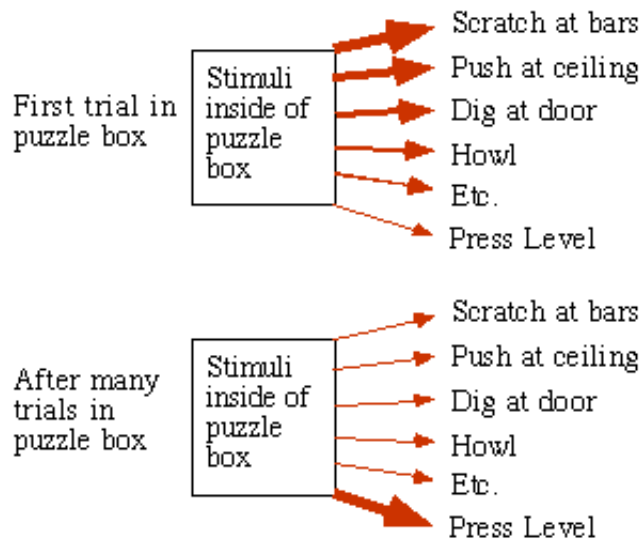
Because, the more useful behaviours are more and more likely to be performed, the animal is more and more likely to complete the task quickly. Thus, in the cat in the box example, the time to escape will tend to decrease.

Note that according to Thorndike's view of learning, there is no need to postulate any further intelligent processes in the animal. There is no need to assume that the animal notices the causal connection between the act and its consequence and not need to believe that the animal was trying to attain some goal. The animal simply learns to associate certain behaviours with satisfaction such that these behaviours become more likely to occur.

Thorndike called this type of learning **instrumental learning**. The animal learns to produce an instrumental response that will lead to satisfaction.

Skinner and Operant Learning

Burrhus Fredric Skinner (1904-1990)

Most of the early research on instrumental learning was performed by B. F. Skinner. Skinner proposed that instrumental learning and classical conditioning were fundamentally different processes.

In classical conditioning:

- a biologically significant event (US - meat) is associated with a neutral stimulus (CS - bell),
- a neutral stimulus becomes association with part of a reflex.

In instrumental learning:

- a biologically significant event is followed by a response, not a stimulus,
- a satisfying or non-satifying event alters the strength of association between a neutral stimulus (e.g., the cage) and a quite arbitrary response (e.g., pressing the paddle). The response is not any part of a reflex.

Skinner called instrumental responses **operant responses** or simply **operants** because they operate on the world to produce a reward. He also referred to instrumental learning as **operant conditioning**. Thus, operant conditioning is:

> The learning process through which the consequence of an operant response affects the likelihood that the response will be produced again in the future.

Unlike reflexes, operant responses can be accomplished in a number of ways (compare an eyeblink to pressing a paddle) and are what we normally think of as voluntary actions. In operant learning, the emphasis is on the consequences of a motor act rather than the act in and of itself.

Skinner, like Thorndike, believed in the Law of Effect. He believed that the tendency to emit an operant response is strengthened or weakened by the consequences of the response. However, he avoided mentalistic terms and interpretations. Thus for example he used the term **reinforcer**, instead of reward, to refer to the stimulus change that occurs after a response and tends to make that response more likely to occur in the future. (The term "satisfaction" was distasteful to Skinner. After all, how do we know if a cat is satisfied by the food is gets when it escapes from the cage? All we really know is that the response leading to the food will be more likely to occur again in a similar situation.)

Skinner Box

Skinner developed a new method for studying operant learning using what is commonly called a "Skinner box". Skinner boxes are also called operant chambers.

Operant Chambers
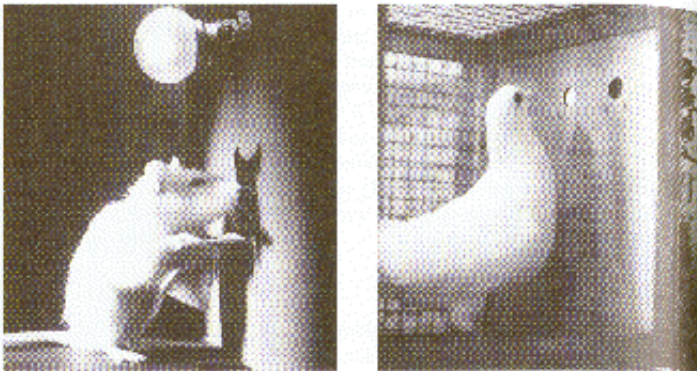
A Skinner box is a cage with a lever or some other mechanism that the animal can operate to produce some effect, such as the delivery of a small amount of juice. The advantage of the Skinner box over Thorndike's puzzle box is that the animal does not have to be replaced into the cage on each trial. With the Skinner box, the animal is left in the box for the experimental session and is free to respond whenever it wishes. The standard measurement used by Skinner to assess operant learning was the rate of responses. (This was Skinner's operational definition of learning.)

Skinner and his followers argued that virtually everything we do can be understood as operant or instrumental responses that occur because of their past reinforcement and that this is independent of whether or not we are aware of the consequences of our behaviour.

For example, if the students in a class all smiled when the Professor walks to the right side of the room but put on blank expressions when he or she walks to the left, there is a good chance that the Professor will end up spending most of the lecture on the right - even thought he or she is not aware of what is happening.

A more simple effect you can have on your Professor is to simply be alert and enthusiastic. This will tend to make him or her more enthusiastic and you will get a better lecture.

Three Consequences of Behaviour

As mentioned above, Skinner believed that operant behaviour (i.e., operant responses) is determined by its consequences. He identified three possible consequences of behaviour:

1) Positive Reinforcement

> Any stimulus that increases the probability of a behaviour (e.g., access to fish is a positive reinforcer for a cat).

> Familiar examples of positive reinforcement: studying and gambling.

2) Negative Reinforcement

> Any stimulus whose removal increases the probability of a behaviour. For example, bar pressing that turns off a shock.

3) Punishment

> Any stimulus whose presence (as opposed to absence in -ve reinforcement) decreases the probability of behaviour. For example, bar press that leads to a shock.

Skinner thought that punishment was the least effective of the 3 possible consequences for learning.

Processes Associated with Operant Conditioning

As with classical conditioning, there are a number of processes in operant conditioning.

Shaping

Imagine a rat in Skinner box where a pellet of food will be delivered whenever the animal presses a lever. What happens if the rat in a Skinner box never presses the lever?

To deal with this problem, one can use a procedure know as "shaping".

One might start by providing the reinforcement when the rat gets close to the lever. This increases the chance that the rat will touch the lever by accident. Then you provide reinforcement when the animal touches the lever but not when the animal is near the lever. Now you hope the animal will eventually press the lever and, when it does, you only reinforce pressing.

Thus, shaping involves reinforcing behaviours that are increasingly similar to desired response. (Shaping is sometimes called the method of successive approximations.)

Highly "Shaped" Behaviours
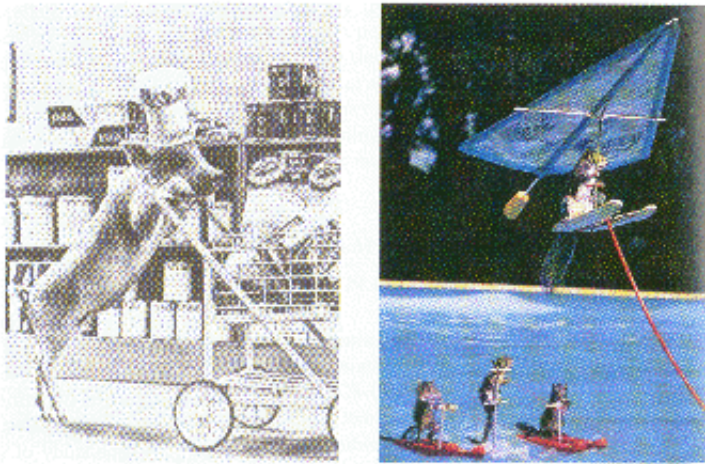
Extinction and Spontaneous Recovery

Extinction in operant conditioning is similar to extinction in classical conditioning. If the reinforcer is no longer paired with the response, the response decreases.

> e.g., people stop smiling if you do not smile back.

The response also exhibits spontaneous recovery some time after the extinction session.

Extinction of operant learning has two non-intuitive facts:

1. The larger the Reinforcer, the more rapid the extinction.
2. The greater the number of training trials, the more rapid the extinction.

This may reflect the fact that the onset of extinction is more "obvious".

Stimulus Control

The instrumental or operant response in operant condition is not elicited by an external stimuli but is, in Skinner's terms, emitted from within. But this does not mean that external stimuli have no effect. In fact they do exert considerable control over behaviour because they serve as **discriminative stimuli**.

Suppose a pigeon is trained to hop on a treadle to get some grain. When a green light comes on, hopping on the treadle will pay off, but when a red light comes on it will not. In this case, the green light becomes a positive discriminative stimuli (S+) and the red light becomes a negative discriminative stimuli (S-).

Note that the S+ does not signal food in the way that the CS+ might in a Pavlov's laboratory. (Recall the example with the black and gray squares where after training the animal salivates in response to a black square, the CS+, but not a gray square.) Instead, the S+ signals a particular relationship between the instrumental response and the reinforcer telling the pigeon "if you jump now, you will get food."

A variety of techniques have been used to study the role of discriminative stimuli in operant learning and many of the results mirror those of generalization of discrimination in classical conditioning.

For example, if a pigeon is trained to respond only when a yellow light appeared, after training, it will also respond to lights of a different colour. However, there is a response gradient - the response decreases with the size of the difference (measured in terms of wave frequency) between the test light and original yellow light (i.e., the original discriminative stimulus).

(After Reynolds, 1968)
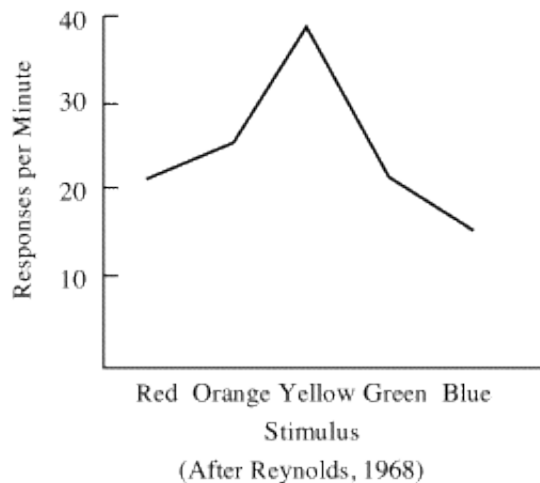
The following cartoon illustrates an experiment in which a rat learns to discriminate between a triangle and a square in order to get food.

Discriminative Stimuli (after Lashley, 1930)
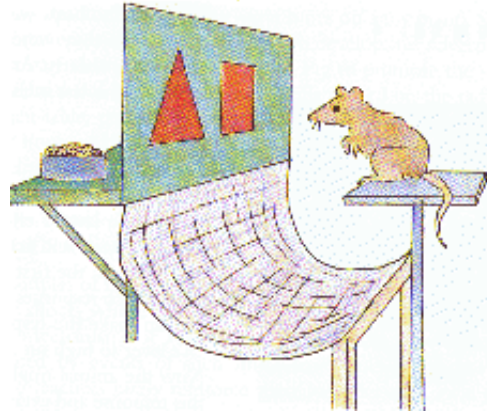
Reinforcement Schedules in Operant Conditioning

A major area of research in Operant Learning is on the effects of different reinforcement schedules. The first distinction is between partial and continuous reinforcement.

- Continuous Reinforcement: every response is reinforced
- Partial or Intermittent Reinforcement: only some responses are reinforced.

In initial training, continuous reinforcement is the most efficient but after a response is learned, the animal will continue to perform with partial reinforcement. Extinction is slower following partial reinforcement than following continuous reinforcement.

Skinner and others have described four basic schedules of partial reinforcement which have different effects on the rate and pattern of responding.

We have fixed and variable interval and ratio schedules.

- Ratio schedules: reinforcer given after some number of responses.
- Interval schedules: reinforcer given after some time period.
- Fixed: the number of responses or time period is held constant.
- Variable: the number of responses or the time period is varied.

Typical Behaviour with the 4 Schedules

- Fixed-Ratio: bursts of responses.
- Variable-Ratio: high, steady rate of responding. (Slot machines work on a V-R schedule).
- Fixed-Interval: pauses with accelerating responses as the time approaches.
- Variable-Interval: after training, a slow, steady pattern of responses is usually seen.

Response rate is generally higher with the ratio schedule.

Top of Page