3 Conditioning Principles and Theories

CONTENTS

Classical conditioning is not just a neat way to get dogs to salivate. As we shall see in Chapter 4, it affects crucial aspects of our behavior including sexual attraction, fear, and drug addiction. Our goal in this chapter, therefore, will be to gain a deeper understanding of conditioning. We will begin by looking at what determines how strongly a response will be conditioned. We will then look at theories of conditioning: What are the underlying processes that eventually lead to the drops of saliva or the surge in fear?

Principles

The British Associationists, sitting in their armchairs several centuries ago, identified a number of laws of association, of which the most important were *contiguity*, *frequency*, and *intensity*. We will begin our survey of the principles of conditioning by considering the extent to which these laws have been supported by experiments.

Contiguity, Frequency, and Intensity

Contiguity

The most important principle of association was thought to be **contiguity**. The very concept of an association – a bond between two events that occur

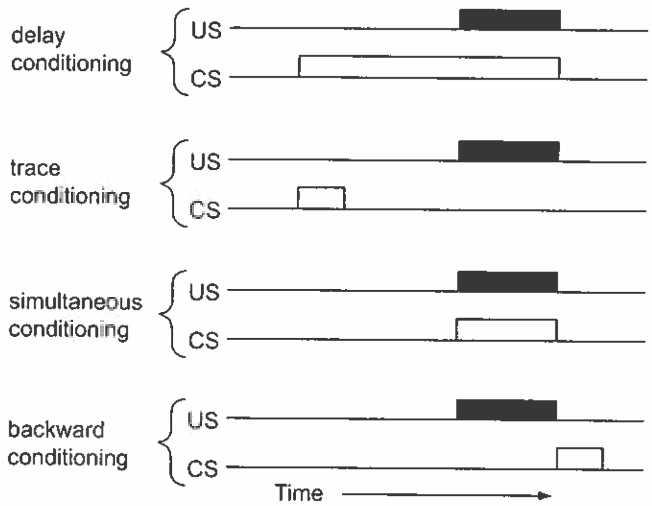

Figure 3.1 Paradigms for four varieties of classical conditioning. The bars on the time line indicate periods during which a stimulus is presented. In simultaneous conditioning, for example, the US occurs at the same time as the CS and for the same duration.

together – implicitly assumes that contiguity is necessary, and considerable effort has been devoted to exploring the role of contiguity in classical conditioning.

As with most other aspects of conditioning, Pavlov was the first to explore it. He experimented with four different temporal arrangements between the CS and the US (see Figure 3.1). In **delay conditioning**, once the CS comes on, it remains on until the US is presented. In **trace conditioning** the CS ends before the US begins. As the British Associationists would have predicted, Pavlov found that conditioning was much stronger in the delay conditioning paradigm, where the CS and the US were on at the same time.

Subsequent research confirmed the importance of contiguity. In a typical study, Moeller (1954) looked at the effects of the CS–US interval on GSR conditioning. He used a trace conditioning paradigm in which a brief burst of white noise (100 ms) was followed after a delay by a weak electric shock, with the interval between the onset of the CS and the onset of the US (the interstimulus interval, or ISI) set at either 250, 450, 1,000, or 2,500 milliseconds (ms). Moeller's results are illustrated in Figure 3.2, which shows that the strength of the conditioned response was greatest in the group with a 450 ms gap, conditioning was weaker with a delay of one second, and virtually no conditioning occurred when the delay was increased to 2.5 seconds. Both the optimum interval and the maximum interval that will sustain conditioning vary somewhat for different responses (see Cooper, 1991, for a discussion of why this might be), but as a general rule the shorter the interval between the CS and the US, the better the conditioning.

In Pavlov's third basic procedure, **simultaneous conditioning**, the CS and the US come on at the same time. According to the British Associationists, simultaneous

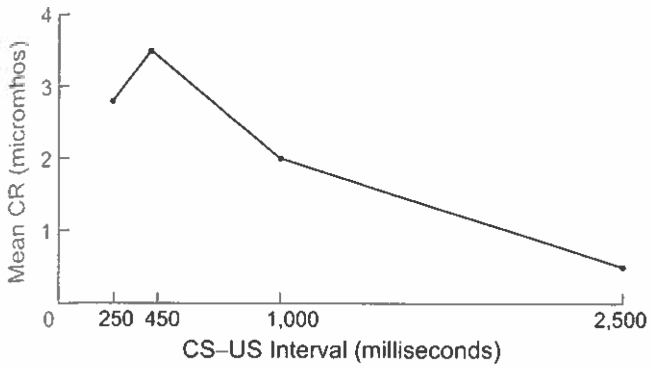

Figure 3.2 GSR conditioning as a function of the CS–US interval during training. (Adapted from Moeller, 1954.)

conditioning should be the optimal procedure for producing conditioning – two stimuli can't be any closer than being on simultaneously. That, however, was not what Pavlov found: Simultaneous presentations of the CS and US produced virtually no conditioning, and this result has generally been confirmed in subsequent experiments (e.g., Heth, 1976).

The results for Pavlov's fourth procedure, **backward conditioning**, produced a similar puzzle. The CS typically came on soon after the US, but again Pavlov found little or no conditioning (Hall, 1984).

Why not? In simultaneous and backward conditioning the CS and US either come at the same time or, at any rate, occur in close proximity. Why, then, aren't they associated?

The perhaps surprising answer is that an association *is* formed, it is just that this association does not lead to a conditioned response (see, for example, Rescorla, 1980b; Matzel, Held, and Miller, 1988). Consider a dog in a delay conditioning experiment in which a tone is followed by food. The tone has *predictive value* – it warns the dog that food is about to appear. The dog can then prepare for the arrival of this food by salivating, which will allow it to digest the food more efficiently. If the tone and food come on at the same time, however, then the tone has no predictive value – the dog doesn't need the tone to tell it that food is coming, the food has already appeared! In other words, it appears that a CS will elicit a response only if the CS has predictive value – that is, it will allow the dog (or person) to prepare for the forthcoming US.

As we shall see shortly, conditioning is an adaptive process whose purpose is to allow organisms to prepare for forthcoming events. When responding would serve no purpose, as in simultaneous and backward conditioning, it would be pointless to respond, and the processes involved in conditioning have evolved so that we don't.

Further evidence for the importance of contiguity comes from a phenomenon called **sensory preconditioning**. Brogden (1939) gave dogs 200 pairings of a bell with a light; they were on together for 2 seconds. He then presented the bell on its own, but now followed by a mild shock:

Sensory preconditioning:	{Light + bell}
Conditioning:	Bell → shock

Finally, he presented the light on its own:

Test: Light → ?

His perhaps surprising finding was that the light now also elicited a conditioned response, even though it had never been followed by shock. The mere fact of the bell and light occurring together in the first phase had been enough for them to become associated, and thus for the response conditioned to the bell to be transferred to the light.[1] (For appropriate controls, see Rescorla and Cunningham, 1978.) Sensory preconditioning is thus further evidence of the importance of contiguity: If two salient stimuli occur together, there is a good chance we will associate them, so that thereafter one will remind us of the other.

There are important aspects of contiguity that are still not well understood (see Molet and Miller, 2014; Kirkpatrick and Balsam, 2016; Delamater, Derman, and Harris, 2017 for discussions) but as a general rule the formation of an association does depend on the contiguity of the CS and the US. Whether this association then leads to a response, though, depends on whether the CS is also a good predictor of the US. To understand conditioning, we need to separate the processes that lead to the formation of an association, and the quite separate processes that determine whether the CS will then elicit a response.[2]

Frequency

A second variable that the British Associationists thought determined the strength of an association between two events was the frequency of their pairing. Pavlov's

---

[1] There is intriguing evidence that it isn't even necessary for two stimuli to be present together for a response conditioned to one to transfer to the other. If we even *think* of two stimuli at the same time, conditioning can transfer from one to the other (Ward-Robinson and Hall, 1998; Holland and Sherwood, 2008).

[2] In the learning literature, this idea is often expressed in terms of a distinction between *learning* and *performance*. The fact that we have learned something does not necessarily mean that we will act on this knowledge. Suppose you learned to ride a bicycle as a child but now you don't even own one. Your childhood learning might currently be invisible, but it's there, available to be tapped. Similarly in simultaneous conditioning, when a CS and a US occur together we do seem to learn about this relationship, but we don't act on it because there is no point in preparing for a US that has already occurred.

research on salivary conditioning strongly supported this view, and so has subsequent research. In general, the strength of the conditioned response seems to increase most during the early trials of conditioning, with the rate of increase gradually declining as training continues, until performance eventually reaches a stable plateau, or **asymptote**.

Intensity

The third major principle proposed by the British Associationists was that the strength of any association depends on the vividness or intensity of the stimuli involved. Associations involving emotional or traumatic events, for example, were thought to be better remembered. Again, research on conditioning strongly supports this principle. Annau and Kamin (1961), for example, found that the amount of fear conditioned to a tone depends on the intensity of the shock that follows the tone (see Figure 2.11). There is also evidence that the intensity of the CS is of some importance, although this effect appears weaker (Grice, 1968).

On the whole, then, the armchair speculations of the British Associationists have been impressively confirmed by research under controlled conditions. Stimuli that occur together do tend to be associated, and the strength of this association depends on how often they are paired. Whether this association then leads to a preparatory response, though, depends on whether the CS predicts that a US is coming, and on the intensity of that US.

Challenges to Contiguity

Until 1966, all the available evidence converged on a coherent and satisfying picture of conditioning in which the foundation stone was contiguity: If two events are contiguous, then an association will be formed between them. The strength of this association might be modulated by other factors such as the intensity of the stimuli involved. Fundamentally, though, conditioning appeared to be a simple process in which associations were automatically formed between contiguous events. In 1966, however, two landmark papers were published in *Psychonomic Science*, ironically a relatively obscure journal with a reputation for publishing competent but minor studies. These two papers posed a fundamental challenge to traditional views of the role of contiguity and unleashed an intellectual ferment – revolution would not be too strong a word – that is still continuing.

Contingency

The first of these papers was the work of Robert Rescorla, then a graduate student at the University of Pennsylvania. In his paper, Rescorla suggested that contiguity between two events was not sufficient for conditioning; something more was needed. Specifically, he suggested that a CS must not only be contiguous with a US but must also be an accurate *predictor* of the US. To understand what he meant

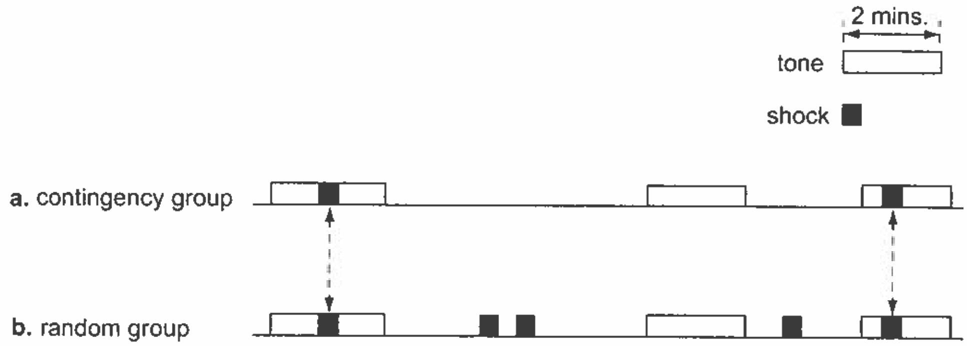

Figure 3.3 An illustration of contingency. The period during which a tone is present is indicated by a shaded bar, and the presentation of a shock is indicated by a solid bar. In situation **a**, the contingency group, a shock is sometimes presented while the tone is on but never when the tone is absent. In **b**, shocks are also sometimes presented during the tone but now shocks are also presented in the absence of the tone. (As indicated by the broken lines, both groups received the same number of shocks in the presence of the CS; they differ only in what happens in the absence of the tone.) In situation **a** there is a contingency between the tone and the shock – the onset of the tone tells you that there is now a greater danger of shock. In situation **b**, there is no contingency between the tone and the shock – there is no relationship between them, and the tone's onset does not signal any greater probability of shock.

by this, consider the situations outlined in Figure 3.3, in which a series of tones and shocks are presented. In situation a, the shock sometimes occurs while the tone is on, but not when it is off. The tone is thus a useful predictor of the shock; shock is more likely when the tone is on. But now consider situation b. Again, shocks sometimes occur while the tone is present, but shocks now also occur in the tone's absence. In fact, shocks occur just as frequently when the tone isn't there as when it is. In this situation the tone would have no predictive value, as it would not help you to estimate when you would receive a shock.

What these examples tell us is that the predictive value of a CS can vary quite widely. At one extreme, a tone might be a perfect predictor of shock, with shock occurring whenever the tone was presented but never at its absence. At the other extreme, the tone might have no predictive value, with shock just as likely to occur in its presence and absence. It would be quite useful, therefore, if we had some way of measuring predictive value. In fact, there are several such measures, but one of the most useful is a mathematical statistic called a **contingency**. Because contingencies are defined in terms of probabilities, however, we need to start by quickly reviewing what we mean by a probability.

You may already know that a probability is essentially just a mathematical expression of the likelihood that an event will occur. If there is no chance of an event occurring, the probability is said to be 0; if the event is certain to occur, the probability is said to be 1.0. Suppose that a tone was presented 100 times, and that every one of these presentations was followed by a shock. In that case, the probability of a shock following the tone would be 1.0. This can also be expressed as the probability of a shock occurring "given that" the tone has already occurred, or, in probability notation, as

$$p(\text{shock} \mid \text{tone}) = 1.0$$

Let us further suppose that the shock never occurs in the absence of the tone. The probability of a shock in the absence of the tone would then be 0,

$$p(\text{shock} \mid \text{no tone}) = 0$$

This would be similar to the situation previously outlined in Figure 3.3a – shock would be more likely when the tone was on than when it was off.

Now consider a situation more like that in Figure 3.3b, in which shocks still occurred in the presence of the CS but now also occurred in its absence. Specifically, suppose that the probability of a shock in the absence of the tone was exactly the same as in its presence:

$$p(\text{shock} \mid \text{tone}) = p(\text{shock} \mid \text{no tone}) = 1.0$$

Because the shock would occur just as often in the absence of the tone as in its presence, the tone would no longer help you to predict when shock would occur.

So, even though tone is followed by shock equally often in both of our examples, its predictive value would be very different. One way of capturing this idea is to say that the predictive value of a CS depends on the extent to which the probability of the US *changes* when the CS is present. And that is what a contingency statistic measures. The contingency between a CS and a US is defined as the difference between the probability of the US when the CS is present and the probability when it is absent (Allan, 1980). Or, in probability notation:

$$\text{contingency} = p(\text{US} \mid \text{CS}) - p(\text{US} \mid \text{no CS})$$

If the two probabilities are different, then there is a contingency between the two stimuli, and the CS can help us to predict when the US will occur.

In a typical conditioning experiment, the CS is closely followed by the US, and the US is never presented in the absence of the CS. The CS and the US are thus contiguous in time, and, because they always occur together, there is also a strong contingency between them. Until Rescorla, everyone had simply taken it for granted that it was the contiguity between the two stimuli that determined the outcome – if a CS and a US occur together, then an association will be formed. Rescorla, however, wondered whether the predictive value of the CS might also be important. What would happen, he asked, if a tone and a shock were presented contiguously, as in most fear-conditioning experiments, but the shock was also presented in the *absence* of the tone, so that the tone would have no predictive value?

To find out, Rescorla ran several experiments using procedures similar to the ones outlined in Figure 3.3. In one of these experiments (Rescorla, 1968), rats were exposed to a series of tones and shocks. In the random group, there was no relationship between the stimuli; they were presented at totally random intervals. In the contingency group, on the other hand, shocks occurred only when the tone was on. The conditions were arranged so that the probability of a shock during a tone was exactly the same in both groups; they differed only in that the random group also received shocks in the absence of the tone.

What should we expect to happen? If conditioning depends simply on contiguity, then conditioning should be equal, since both groups had the same number of tone-shock pairings. The groups differed, however, in terms of contingency: There was a strong contingency between the tone and shock in the contingency group but none in the random group. Insofar as contingency is important, therefore, we should expect stronger conditioning in the contingency group. And that is what Rescorla found. To measure fear, he used a CER test in which he presented the tone while the rats were pressing a bar to obtain food. The rats in the contingency group immediately stopped responding when the tone was presented, indicating strong fear. When the tone was presented to the rats in the random group, however, they carried on as if nothing had happened. Despite repeated pairings of the tone with shock, these rats showed no sign of fear when the tone was presented.

These findings do not mean that contiguity is not important; in GSR conditioning, for example, we saw that a delay of even 2 seconds could largely prevent conditioning. Rescorla's results, however, suggest that contiguity is not enough; for conditioning to occur, a CS must also be a good predictor of the US. In one sense, this is hardly surprising. If a tone and a shock occur at random intervals, the tone will not signal an increase in the likelihood of shock. There is no reason, therefore, why the tone should elicit fear. When viewed from the traditional perspective of contiguity, however, these results are deeply puzzling. According to Pavlov, an association is formed whenever CS and US centers in the brain are active at the same time. But in the random group, the tone and shock occurred together many times. Why was no association formed? How could presentations of the shock by itself have prevented conditioning?

We will return to this question soon; for now, we simply note that Rescorla's work provided one of the first indications that conditioning is not just a simple process based on contiguity; something rather more complex seemed to be going on. For the first time, the simple picture painted by Pavlov was beginning to unravel.

Preparedness

The second seminal paper of 1966 was by Garcia and Koelling, and they also challenged the assumption that any two events that were contiguous would be associated. In particular, these researchers challenged the idea that it did not matter what stimulus was chosen as a CS. Pavlov had claimed, "Any natural phenomenon chosen at will may be converted into a conditioned stimulus . . . any visual stimulus, any desired sound, any odor, and the stimulation of any part of the skin" (1928, p. 86). Subsequent research almost universally supported Pavlov's position – until, that is, the publication of Garcia and Koelling's paper.

As we saw in Chapter 2, Garcia and Koelling's experiment had its origins in the phenomenon of bait shyness, the fact that rats might eat a poisoned bait once, but that if they survived, they would rarely if ever touch that food again. Classical conditioning provided a possible explanation: Ingestion of the poisoned bait produces nausea, and this reaction becomes conditioned to the gustatory and olfactory cues that precede the nausea. On future occasions, the rats avoid the bait because its odor or taste makes them ill.

As plausible as this explanation is, it cannot account for one aspect of the rats' behavior. Although the poisoned rats later avoided the bait, they showed no reluctance to return to the *place* where they had been poisoned and consume other foods there. If associations form between any contiguous events, then we should expect place cues to be associated with illness as readily as taste and odor cues, but this did not appear to be happening. Was it possible that the rats could associate nausea with tastes but not visual cues?

To test this hypothesis under controlled laboratory conditions, Garcia and Koelling allowed rats to taste distinctly flavored water from a drinking tube that was wired so that every lick produced not only water but a brief noise and a flash of light. Following exposure to this taste–noise–light compound, they received a dose of radiation to make them ill. Then, on a test trial, the rats were exposed to each of the compound stimuli separately, to determine which ones had become aversive. A lick produced either the flavored water or plain water plus the noise–light compound. As shown in Figure 3.4a, the rats were now very reluctant to drink the flavored water, but they had no such compunctions about the bright–noisy water. As suggested by the naturalistic observations, it looked as if nausea could be conditioned to gustatory cues but not visual ones.

An alternative explanation, however, was possible: Perhaps the noise and light used in the experiment were simply too faint to be detected, so conditioning would not have occurred with any US. To test this hypothesis, Garcia and Koelling repeated their experiment with the same compound CS, but with electric shock as the US instead of X-rays. The results for the suppression test are shown in Figure 3.4b, which illustrates that the audiovisual stimulus produced suppression of drinking while the taste stimulus had no effect. We thus face this strange situation in which nausea cannot be conditioned to a noise, nor fear to a taste, even though each of these conditioned stimuli is easily associated with the other US.

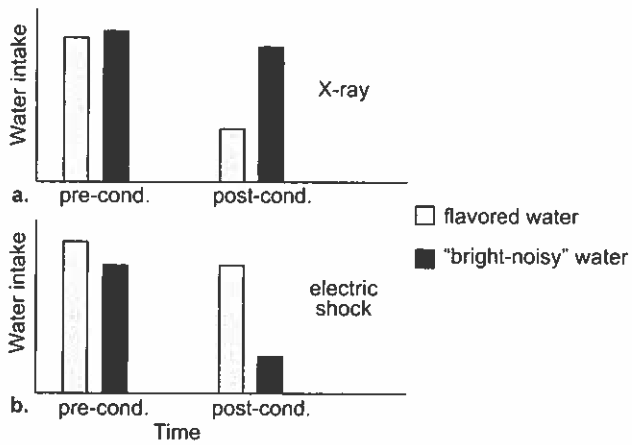

Figure 3.4 Water intake before (pre) and after (post) conditioning: (a) when X-rays were used as the US; (b) when shock was used as the US. The gray bars represent intake of the flavored water; the dark bars represent intake of plain water when licking produced a noise and a light. (Based on Garcia and Koelling, 1966.)

Subsequent research has established that it is possible to associate taste with shock and noise with illness, but it is much more difficult, requiring many more trials (for example, Best, Best, and Henggeler, 1977). Seligman (1970) has coined the term **preparedness** to refer to the fact that we seem prepared to associate some CS–US combinations more readily than others.

The evidence for preparedness provided another important challenge to the principle of contiguity. In Garcia and Koelling's experiment, noise and taste were both contiguous with illness, and yet the rats did not develop any aversion to the noise. Contiguity, in other words, is not sufficient for learning to take place. Moreover, subsequent research on taste-aversion learning established that contiguity is not even necessary. In one memorable experiment by Etscorn and Stephens (1973), conditioning occurred despite a delay of 24 hours between eating a food and becoming ill. The irresistible conclusion is that conditioning is not just a simple process of hooking together any events that occur together; some more complex process or processes must be involved.

We will return to the question of what these other processes might be shortly. Before leaving taste-aversion learning, though, we will briefly comment on one other question raised by their results, namely *why* it should be easier to condition nausea to a taste than to a noise. To answer this question, it might be helpful to begin by standing back a bit and considering the broader question of why classical conditioning occurs in the first place.

In discussing Pavlov's research, we referred to his view that the process of conditioning evolved because it helps animals survive in their natural environments. Conditioning, in this view, evolved as a means of identifying stimuli that cause or predict important events. If an animal knows where food is available, for example, or which of the other animals in its vicinity is likely to attack it, then it can use this information to guide its behavior. Culler (1938) expressed this view with some eloquence:

> [Without a signal] the animal would still be forced to wait in every case for the stimulus to arrive before beginning to meet it. The veil of the future would hang just before his eyes. Nature began long ago to push back the veil. Foresight proved to possess high survival-value, and conditioning is the means by which foresight is achieved. (Culler, 1938, p. 136)

Salivary conditioning provides one example of the advantages of foresight: If a dog knows when food is coming, it can begin to salivate beforehand, and this will allow it to consume the food more quickly – not a small advantage when predators or other hungry dogs are around. (See also Hoffman, 2011; Domjan and Akins, 2011). In the real world, though, some cues are much more likely to be helpful than others. Consider a rat that became ill after eating rancid meat. If it developed an aversion to all the stimuli that were present when it began to feel ill, it would be as likely to develop an aversion to a bird that happened to be singing as to the rancid meat that it had eaten earlier. And if it thereafter scurried to its burrow whenever it heard a singing bird, it would have been more likely to die of hunger and exhaustion than to prosper. The pressures of natural selection would thus favor rats (and people) that associated illness with preceding tastes, rather than with irrelevant lights or sounds.

Have you ever developed an aversion to a food that you ate prior to becoming ill?

Blocking

The 1960s were a difficult time for the principle of contiguity. First, Rescorla showed that temporal contiguity between a CS and a US is not sufficient to ensure conditioning; the CS must also be a good predictor of the US. Then Garcia and Koelling showed that even valid predictors are not always conditioned. In 1969, a third event undermined still further the traditional view of contiguity and suggested an alternative analysis to replace it. That event was the publication of a paper by Leo Kamin.

Kamin (1969) gave rats fear-conditioning trials in which two stimuli, a noise (N) and a light (L), were paired with an electric shock. The noise and the light came on together, remained on for 3 minutes, and were immediately followed by the shock. To assess conditioning to the light, Kamin used a CER test in which the light was presented while the rats pressed a bar to obtain food. The suppression ratio for the light was 0.05, indicating substantial fear conditioning. (Recall that a suppression ratio of 0.50 indicates no fear and zero indicates maximal fear.)

Kamin was interested primarily in a second group, though. The subjects in this second group received identical pairings of the noise–light compound with shock, but these compound trials were preceded by trials in which the noise by itself was paired with shock.

	Pretraining	*Conditioning*
blocking group:	N → shock	NL → shock
control group:		NL → shock

For subjects in the blocking group, therefore, the noise already elicited fear when the compound trials began. What effect should we expect this to have on conditioning to the light?

According to a contiguity analysis, fear should be conditioned to the light in both groups, because, in both, the light was repeatedly and contiguously paired with the shock. The results for the two groups, however, proved to be very different. The suppression ratio in the control group was 0.05; the ratio for subjects given preliminary conditioning to the noise was 0.45, a statistic only barely distinguishable from the 0.50 level representing no fear. In other words, prior conditioning to the noise had blocked conditioning to the light. Kamin thus called this phenomenon, in which prior conditioning to one element of a compound prevents conditioning to the other element, **blocking**.

Blocking thus provided yet another case in which pairing a CS with a US did not result in conditioning. The message was becoming overwhelming: Conditioning was more than just a simple process of associating any two events that occurred together; something more must be involved.

Surprise!

To account for blocking, Kamin proposed an intriguing explanation. When an important event such as shock occurs, he said, animals search their memories to identify cues that could help to predict the event in the future. Imagine that a rat foraging for food in a forest is suddenly attacked by an owl. If the rat survives the attack, it will search its memory to identify cues that preceded the attack and thus help it avoid such an event in the future. If the rat had seen the owl in a tree just before the attack, for example, then the next time it saw an owl it would dive for cover.

Kamin's first assumption, then, was that unconditioned stimuli trigger memory searches for predictive cues. His second assumption was that such searches require effort. In taste-aversion conditioning, for example, we have seen that animals may develop an aversion to foods consumed as much as 24 hours earlier, indicating that any memory search must cover events spread over at least this time period. Such a search would require considerable time and effort, and Kamin speculated that, to save energy, subjects would scan their memories only if the US were unexpected or surprising. If the US were expected, then by definition some cue predicting its occurrence must already have been available, so that no further search would be needed.

To see how this analysis can account for blocking, consider first the control group that received only the compound trials. The first shock would have been unexpected and would have triggered a memory search for the cause. The rats would remember the preceding noise and light, and thus both cues would be associated with the shock.

Similarly in the blocking group, presentation of the shock during the preliminary phase would have surprised the rats and thus triggered a memory search in which the rats recalled the noise and associated it with the shock. When the noise was then presented as part of the noise–light compound, the rats would have expected the shock to follow and hence would not have been surprised. As a result, they would not have searched their memories and thus would not have learned about the relationship between the light and the shock.

According to Kamin, then, the reason that blocking occurs is that the US is expected. To test this analysis, he used an ingenious design in which he changed the US used during the compound trials so that its presentation would come as a surprise. As before, a noise was paired with shock during preconditioning, but during conditioning the shock presented at the end of each noise–light compound was unexpectedly followed by a second shock 5 seconds later:

Noise — Shock   Noise – Light — Shock . . . Shock

During pretraining, the rats would have learned that the noise was followed by a shock, and thus on the compound trials they would have expected the first shock. The second shock, however, would have been a surprise. The rats should therefore search their memories for possible causes, notice the light, and associate it with the shock. And that is what Kamin found: When the light was later presented on its own, it produced powerful fear in the group that had received two shocks. This result, and those of similar experiments (e.g., Dickinson, Hall, and Mackintosh, 1976), suggests that surprise is one of the key factors in conditioning: Conditioning will occur if and only if the US surprises us.

We suggested earlier that conditioning cannot just be a matter of associating brain centers that are active at the same time; one or more other processes must be involved. And Kamin's analysis now provides us with one possible model of what these other processes might be. According to this analysis, when we experience an unexpected and important event, it triggers an active search through memory for possible causes, a search that can (as in taste-aversion learning) extend to events that occurred many hours earlier. Then, once possible causes have been identified, there appears to be some sort of selection process to choose which is the likeliest cause, or, if not the cause, the best predictor. If you became ill, for example, your brain would be much more likely to identify an earlier meal as the cause, rather than a song you had heard on the radio. This model of conditioning is only provisional, but it does provide us with an intriguingly different framework for thinking about conditioning. Conditioning was beginning to look as if it might involve rather more sophisticated processes than suggested by the mindless image of salivating dogs.

contiguity Literally, proximity or closeness. In learning, the principle of contiguity says that the formation of an association between two events depends on their closeness in time.

delay conditioning A procedure in which a CS is presented and remains on until a US is presented.

trace conditioning A procedure in which a CS is presented but terminates before the US is presented.

simultaneous conditioning A procedure in which the CS and the US are presented at the same time.

backward conditioning A procedure in which a US is presented and then followed by a CS.

sensory preconditioning A procedure in which two stimuli are presented together and then one of them is paired with a US. The other stimulus will now also elicit the conditioned response, even though it was never paired with the US.

contingency A measure of the extent to which two events occur together, or covary, over time. A contingency coefficient is a mathematical statistic determined by two probabilities: the probability that a US will occur in the presence of a CS, and the probability that is will occur in the absence of the CS.

preparedness The tendency to associate some CS-US combinations more readily than others.

blocking A phenomenon in which prior conditioning to one element of a compound reduces conditioning to the other elements.

Suppose you were a farmer, and a salesman for a commercial weather service said they had developed an extraordinarily accurate system for forecasting the weather – for every 100 days they predicted rain in your area, it rained on 99. Assuming the salesman's figures were accurate, should you buy his service?

Theories

Kamin's findings posed a serious challenge to some earlier views of conditioning as an essentially simple process in which contiguous stimuli are automatically associated. The clash between these two views has a long history, and in this section we'll look more closely at how it started and how it evolved.

Conditioning in Animals

According to one of the earliest cognitive theorists of associative learning, Edward Tolman (1932), the pairing of a CS and a US leads to the formation of an expectation. If a tone is followed by food, for example, then a dog will form an expectation that future tones will also be followed by food. Tolman was not very specific about how this expectation would then be translated into a conditioned response, but the general notion was that the dog would take whatever action was appropriate to prepare for the expected food. Thus, a dog would salivate when it expected food because such anticipatory salivation would help it to digest the food more quickly and efficiently (see Hollis, 1982); a rabbit would blink when it expected a puff of air to its eye because this blink would protect the eye.

Stimulus Substitution

Pavlov's interpretation was very different. As we have seen, he believed that the CS and the US centers became linked so that activation of the CS center would lead to activation of the US center. The CS would therefore elicit the same behaviors as the US did; in effect, it was as if the CS had *become* the US – hence the term **stimulus substitution**. For Tolman, the CS became a signal that food was coming. (Imagine the dog thinking, "Oh boy, I'm about to get food.") For Pavlov, the CS effectively became a substitute for that food, in that it elicited the same responses. (You can imagine the dog thinking, "Oh boy, what lovely food this is." Note, though, that Pavlov did not actually speculate about what dogs were thinking. His view was simply that the CS would elicit the same response as the food.)

At first, Pavlov's substitution theory might seem silly. A dog may not be a brilliant scholar, but surely it has enough sense to be able to distinguish a tone or a light from food. On closer examination, however, this claim is perhaps not as outrageous as it sounds. The assumption that a dog knows that a light is not food begs the important question of *how* it knows. We tend to think that identifying food is trivially simple; everyone knows, for example, that apples are edible but pebbles are not. Babies, however, do not know this and will often try to ingest objects that are emphatically not edible. This is also true of many other species, which have to learn which objects in their environment are edible. If a visual cue is repeatedly followed by food in the mouth, therefore, it is not inconceivable that a conditioning process might lead us to respond to this cue as food in the future.

But if Pavlov's dogs viewed the light as food, you may be asking yourself, why didn't they try to eat it? The answer is, they did! Pavlov found that when a dog is released from its harness after pairings of a light bulb with food, it eagerly runs over to the light bulb and licks it. The dogs didn't actually chew or swallow the bulb, but this might have been because the bulb's hardness inhibited the dog's swallowing reflexes.

In other studies in which the physical characteristics of the CS have been more appropriate, animals *have* tried to ingest the CS. One example comes from a phenomenon known as **autoshaping**. It was discovered by Brown and Jenkins (1968), who trained pigeons in a cage that had a circular plastic disk mounted on one wall, with an opening beneath it through which the experimenters could deliver food. Roughly once a minute they illuminated the key for 8 seconds, at the end of which they gave the birds access to food for 4 seconds. There was no need for the birds to peck the key; they received the food regardless of their behavior. Nevertheless, 36 out of 36 birds began to peck the key every time it was illuminated, and continued to do so for session after session.

The birds' behavior was puzzling. They didn't need to peck the key; all they had to do was wait a few seconds and then collect the food when it arrived. Why, then, were they pecking the key? The answer, according to stimulus substitution, is simple: Because the lit key had been paired with food, the key now elicited the same behavior as food. The birds were pecking the key because they were trying to eat it!

To test this interpretation, Jenkins and Moore (1973) ran an autoshaping experiment with two groups. As in previous experiments, one of the groups received food following the key's illumination, but the other received water. The authors knew that birds have very different behaviors for trying to ingest food and water. When given food, a pigeon pecks with its beak open, in order to ingest the food; when given water, it pecks with its beak almost closed, and uses its tongue to pump the water into its mouth. Also, a pigeon pecks water with its eyes open, but it pecks food with its eyes closed. (Food pecking is much more forceful, and the pigeon probably closes its eyes to protect them from ricocheting pebbles.) According to the substitution hypothesis, therefore, pigeons exposed to light-food pairings should try to eat the key with an open beak and closed eyes, whereas those exposed to light-water pairings should peck with a closed beak and open eyes. As Figure 3.5 shows, that is exactly what happened. As strange as it may sound, the birds in the food group seemed to trying to eat the key, and the birds in the water group seemed to be trying to drink it. (For some striking cases where animals actually *have* eaten stimuli paired with food, see Breland and Breland, 1961.)

Evidence of the power of the conditioning mechanism involved, and of its irrationality, comes from another study by Jenkins (reported in Hearst and Jenkins, 1974), in which the key light was located along one wall of a 6-foot-long box and the food source along another (Figure 3.6). The key light was occasionally presented for 5 seconds, followed by the raising of a grain magazine so that the grain was accessible for 4 seconds. Because of the layout of the box, if the pigeons approached the key and pecked it when the light came on, they could not return to the food dispenser in time to eat all the food. Nevertheless, Jenkins found that his birds would run over to the light as soon as it came on, peck it, and then quickly hurry back to the food magazine. Because of the length of the box, they missed most or all of the food on the trials in which they pecked. Despite this, they continued to peck the key in session after session (see also Williams and Williams, 1969). Pecking the key seemed more important to the birds than eating.

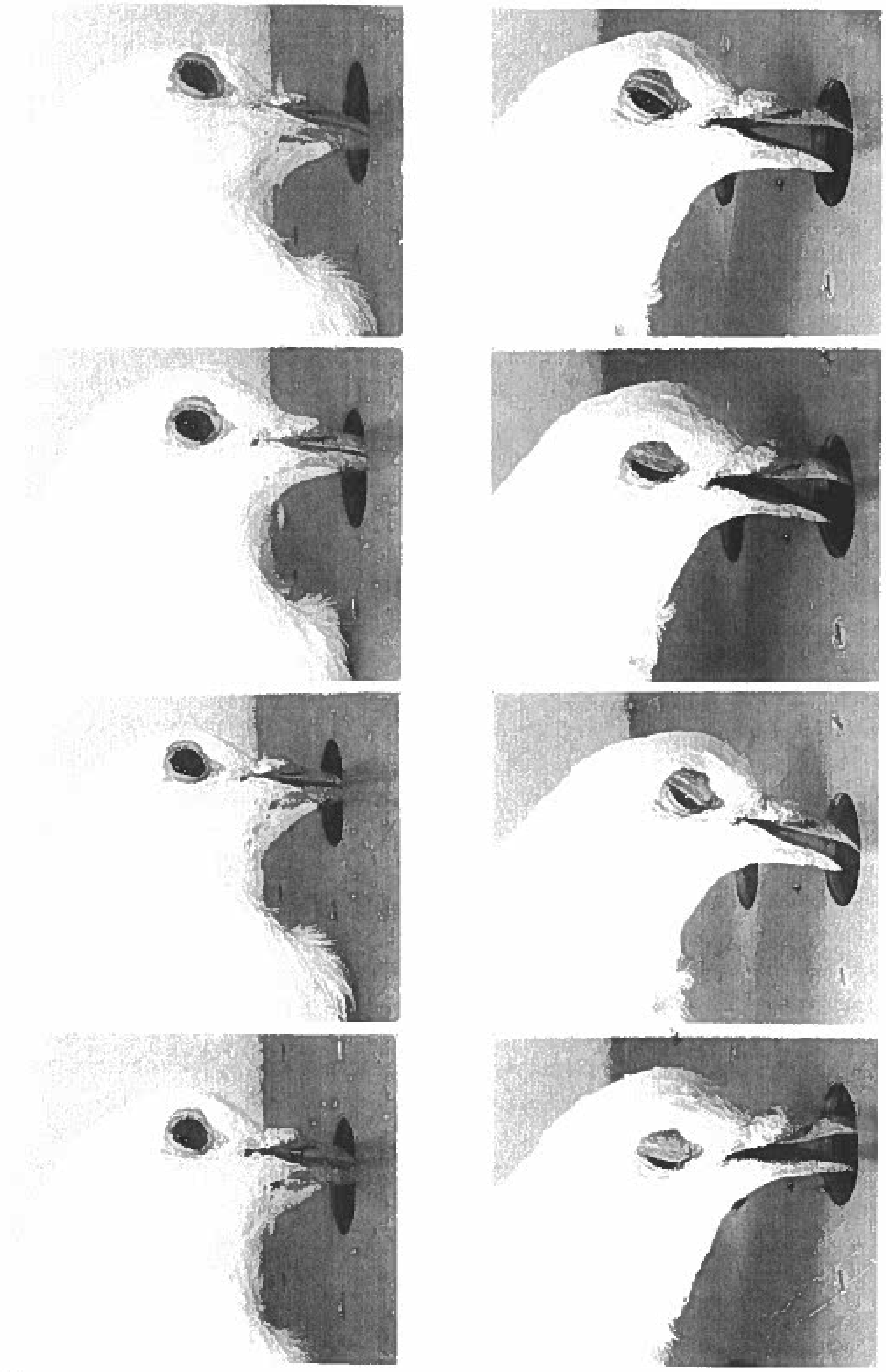

Figure 3.5 Typical responses of key-pecking during autoshaping trials. The photographs on the left show typical pecking when the US was water; the photographs on the right show typical pecking when the US was food. (Jenkins and Moore, 1973; the photographs were provided by Bruce Moore.)

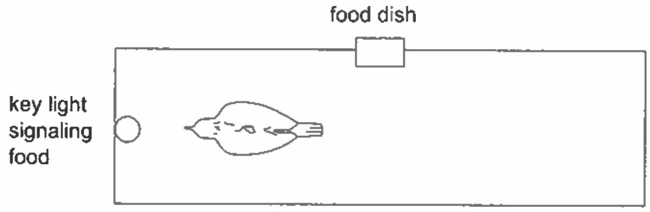

Figure 3.6 Top view of the apparatus used by Jenkins to study autoshaping in pigeons. (Adapted from Domjan and Burkhard, 1986.)

Tolman's Expectations

Pavlov's substitution theory thus needs to be taken seriously: In many situations, animals do behave as if a CS paired with food really is food, and they will persist in trying to eat the CS even if it costs them real food. On the other hand, evidence also supports Tolman's view that a CS acts as a signal that the US is coming. One source of support comes from observations of dogs' behavior during salivary conditioning experiments. They do not just salivate; they will also turn toward the food tray and, if released from the harness, approach the tray (Zener, 1937). Their behavior strongly suggests that they expect to find food there, and similar results have been obtained in other experiments. In Jenkins's long-box experiment, for example, the pigeons usually approached the key light when it was illuminated, but in some cases they moved toward the food dispenser instead. (See also Colwill and Motzkin, 1994.)

Moreover, there is evidence that these expectations can be quite specific, containing detailed information about the forthcoming US. The first indication of this came in Pavlov's work on delay conditioning, in which a CS comes on and then remains on until the US appears. Pavlov found that at first his dogs would salivate as soon as the CS came on, but as training progressed they began to salivate closer to the time when the food would arrive. If there was a 10 second gap between the CS and US, for example, the dogs might initially salivate during the first second, but as trials continued they might not salivate until 7 or 8 seconds had passed. The dogs did not simply learn that food was coming, they learned *when* it was coming, and their salivation moved to the point where it would be most useful in preparing for that food. (See also Williams, Todd, Chubala, and Ludvig, 2017.)[3]

Subsequent research suggested that animals also know what that food will be. Holland (1990), for example, gave rats conditioning trials in which two tones served as conditioned stimuli. One of the tones was followed by access to a peppermint

[3] Matzel, Held, and Miller (1988) called the idea that animal expectations about a US include information about when it will arrive the *temporal coding hypothesis*. See also Molet and Miller (2014).

flavored sucrose solution, delivered in a food cup on one of the cage walls. The second tone was also followed by sucrose, but this time flavored with wintergreen:

Tone 1 → peppermint sucrose
Tone 2 → wintergreen sucrose

As in other experiments using procedures like this, as conditioning progressed the rats began to approach the food cup when they heard either tone; they acted as if they knew that sucrose was coming. Holland, however, now took this a step further: He gave the rats training designed to reduce the attractiveness of one of the sucrose solutions. He allowed the rats to drink this sucrose, and then gave them an injection of a toxin to make them ill. For example, one group was allowed to drink the peppermint sucrose and then made ill:

peppermint sucrose → illness

(This procedure is known as **postconditioning devaluation**, because after conditioning the value or attractiveness of the US is devalued by pairing it with an aversive event.) Finally, several days later Holland presented the tones again.

Holland found that when the tone that signaled the devalued version of sucrose was presented, the rats were far less likely to approach the food cup. (In a comparable experiment by Robinson and Berridge [2013], some rats actually turned away when the CS was presented and pressed themselves against the opposite wall.) They had not simply learned that a tone signaled food, or even that it specifically signaled sucrose; they knew which *flavor* of sucrose was coming. And if it was one they no longer liked, they wouldn't approach the cup.

A Two-System Hypothesis

We now have what seems a distinctly confusing situation in which animals sometimes behave as if a CS paired with food actually is food and try to eat it, but at other times behave as if the CS is simply a signal that food is coming and initiate appropriate action to obtain it (see also Jenkins, Barrera, Ireland, and Woodside, 1978; Timberlake, Wahl, and King, 1982). One way to resolve this conflict is to assume that both views are correct, and that the reason why the outcome varies is that classical conditioning actually involves two distinct learning systems – an *associative* system in which the CS elicits responses automatically, and a *cognitive* system in which expectations guide responding.

Perhaps the first learning system to evolve was a relatively primitive one in which the CS was simply associated with the US, and when presented would automatically elicit responses to prepare for that US. In the course of time, a more sophisticated system developed that involved active anticipation of the US; this allowed subjects to select flexibly from a much wider range of preparatory responses, taking into account other information available at the time. Insofar as both systems still coexist in vertebrates, this would explain why animals act sometimes as if the CS is a signal for food and at other times as if it actually is food.

The idea that the brain contains two distinct learning systems has been proposed by a number of theorists over the years (for example, Konorski, 1967; Razran, 1971; Squire, 1992; Öhman and Mineka, 2001; LeDoux and Pine, 2016). Each theorist has attributed somewhat different properties to the two systems, but a common theme has been that one system is essentially simple and automatic, whereas the other involves some sort of expectation about the properties of the forthcoming US. We will call the assumption of two systems, one associative and one cognitive, the **two-system hypothesis** (also sometimes called the *dual-process* or *dual-system* hypothesis).

The Brain's Evolution

Indirect evidence for two learning systems comes from what is known about the evolution of the vertebrate brain. Studies of fossil records and of the brains of living species suggest that the vertebrate brain has changed enormously in the course of evolution. These changes, however, have consisted not so much in the disappearance of old structures – most of the primitive structures of an alligator or rat brain can still be easily recognized, almost unchanged, in that of a human – as in the elaboration of new structures. In particular, there has been a massive increase in the outer covering of the brain known as the *neocortex*.[4] In humans, for example, the proportion of the brain devoted to neocortex is 150 times greater than in the tree-shrewlike mammals from which we are thought to have descended, even after adjusting for differences in body weight. The functions of this vastly expanded neocortex include cognitive processes such as thinking, language, and consciousness.

In the course of evolution, then, the central core of the brain has remained unchanged to a remarkable extent, with emerging cognitive functions concentrated in a massively expanded outer region. Insofar as the older core has retained its old functions as well as structure, a relatively primitive associative system might still be present in vertebrates along with a more advanced cognitive one.

Epstein has proposed a similar hypothesis to account for human emotions, suggesting that a largely preconscious system evolved first, and that this system remained when a more rational and analytical system emerged. In his words,

> It is inconceivable that, with the advent of language and the capacity for analytical thought, the hard-won gains of millions of years of evolution were summarily abandoned.

---

[4] The outer covering of the brain is called the *cortex*. Some form of cortex is present in most vertebrates, but it is expanded considerably in mammals, and the larger, newer section of the cortex is called the *neocortex*. In humans, most of the cortex consists of neocortex.

> It can more reasonably be assumed that the same principles . . . that apply to nonhuman animal cognitions apply as well to human cognitions, wherein they influence and are in turn influenced by a newly acquired verbal-analytical rational system. (Epstein, 1994, p. 714)

Two Routes to Fear

More direct evidence for the existence of two learning systems has come from research on the physiological mechanisms underlying fear conditioning. The area of the brain primarily responsible for fear conditioning is a structure called the **amygdala** (LeDoux, 2014; LeDoux and Pine, 2016). LeDoux found that in rats there are two pathways leading from the senses to the amygdala: a direct path that can trigger a fear response very quickly, and an indirect path that goes first to the cortex and only then to the amygdala (Figure 3.7). He suggested that the direct path allows a rapid, automatic response to signals of possible danger, whereas the cortical path, although slower, allows subjects to evaluate the signal more carefully and decide what response would be most appropriate. If you glimpsed a fast-moving object out of the corner of your eye, for example, the direct path might trigger immediate arousal, while the cortical path would allow you to weigh the situation and decide whether the moving object was a dangerous predator or just a harmless bird.

An Unconscious System?

What about in humans? Do we also possess separate associative and cognitive systems for conditioning?

No one doubts that participants in typical human conditioning experiments are aware of the relationship between the CS and the US – if you are sitting in a chair and an experimenter periodically turns on a light and then blows air into your eye, it would be hard not to notice. In humans, in other words, our ability to form conscious expectations about forthcoming events is not in doubt; the question is whether we also possess an unconscious system that can detect relationships between events without our realizing it. It is to this issue that we now turn.

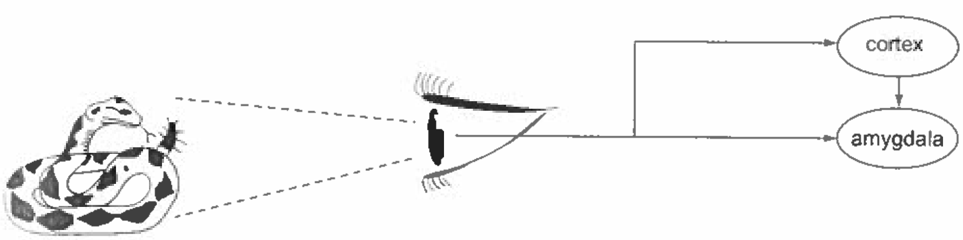

Figure 3.7 Two cortical routes to fear.

Irrational Fears

One argument in favor of a non-cognitive system is what might be called the "common sense" argument, that there *must* be such a system, because otherwise it would be impossible to explain some behavior. In our discussion of taste-aversions, for example, we noted that some people who develop an aversion to a food they ate before becoming ill know that the food was not the actual cause – perhaps they just happened to come down with the flu that day – but this conscious knowledge could not temper their aversion.

Similar arguments have been put forward about fear. Öhman and Mineka (2001) put the case eloquently:

> According to the DSM-IV (APA, 1994), one of the defining features of phobias is that the victim recognizes the fear as excessive and unreasonable. Thus, at the heart of phobia, there is a dissociation between fear and cognitive understanding that is consistent with the automaticity . . . of fear. If phobias result from a defense system of ancient origin in the class of mammals, this system must have evolved to serve organisms with much more primitive brains than those of contemporary phobics. This happened hundreds of millions of years before the emergence of language and thought in the recently evolving hominids. From this perspective, it is not surprising that strong fears and phobias may not be amenable to cognitive control. (Öhman and Mineka, 2001, p. 501)

In both taste-aversions and fear, our emotions seem to be controlling our behavior; we may rationally know that our behavior doesn't make sense but be unable to alter it. Our conscious beliefs are being overridden by . . . something.

Subliminal Presentations

More direct evidence for an unconscious system has come from studies in which the conditioned stimulus is presented subliminally, so that participants are not aware of it. In one experiment exemplifying this approach, Öhman and Soares (1998) presented participants with pictures of either a snake or a spider, followed on some trials by a mild electric shock. A discriminative conditioning procedure was used in which one of the stimuli was followed by the shock while the other was not:

CS+ — shock

CS– — ____

To see if conditioning could occur without awareness, the authors presented the stimuli only very briefly, for less than 1/50th of a second, and each presentation was followed by a *masking stimulus*, a meaningless jumble of dark and light shapes. Previous research had shown that masking stimuli presented under these conditions effectively erase preceding stimuli before subjects can become consciously aware of them. The procedure is thus sometimes referred to as *subliminal* presentation – *limen* is the Latin word for threshold, and the stimulus remains below the threshold of consciousness.

To check that subjects were genuinely not aware of the CS, the experimenters ran an additional group in which there was a 4 second gap between the CS and the US, and during this gap participants were asked to report whether the picture had been of a snake or a spider. The percentage of correct responses was almost exactly at chance (50.5 percent), confirming that participants had no idea what stimulus had been presented. And yet, despite this lack of awareness, conditioning occurred normally, as presentations of CS+ eventually elicited substantially higher GSRs than CS–. The CS+ was thus eliciting fear even though participants did not know that it had been presented. (See also Bechara *et al.*, 1995; Diano, Celeghin, Bagnis, and Tamietto, 2017; Greenwald and De Houwer, 2017.)

Evaluation

At this point you might feel the evidence for conditioning without awareness is compelling. But this evidence has not gone unchallenged; the problem is that awareness can be difficult to assess. Suppose that at the end of a conditioning experiment participants are asked if they had noticed any relationship between the CS and the US, and they say no. It is possible that they actually did notice a relationship at the time, but had forgotten about it by the end of the experiment, or perhaps they suspected there had been a relationship but weren't sure, and so said no. It turns out that alternative explanations are often possible, and it has proven difficult to rule them out conclusively (Lovibond and Shanks, 2002; Lähteenmäki, Hyönä, Koivisto, and Nummenmaa, 2015; Weidemann, Satkunarajah, and Lovibond, 2016). At least for now, this is an area where it has not yet been possible to reach a definitive conclusion. I think the balance of the evidence does strongly favor the assumption of two systems, but the jury (10 to 2? 11 to 1?) has not yet reached a unanimous verdict.

If there truly are two separate systems, conscious and unconscious, this could help to explain some otherwise puzzling aspects of our behavior, including why we sometimes feel anxious but cannot say why. In one clinical case, a woman who had been raped had no conscious memory of the incident, but she nevertheless became extremely upset when she returned to the scene of the crime (Christianson and Nilsson, 1989).

Another case involved a French woman who suffered from a condition called Korsakoff's syndrome, one consequence of which was that she lost the ability to form new memories. Each time her physician, Edouard Claparède, came to see her, she failed to recognize him, even if their last encounter had been only minutes previously. (See the discussion of retrograde amnesia in Chapter 10.) One day, as a test, he concealed a pin within his hand when he greeted her and shook her hand.

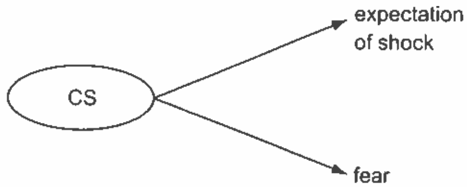

Figure 3.8 Following conditioning, a CS can independently elicit an expectation of shock and fear; one can occur without the other.

When he next met her, he again offered to shake her hand but this time she refused, even though she could not say why (Claparède, 1911).

Figure 3.8 provides one way to understand what is happening in situations like this. When a stimulus is followed by an unpleasant event such as shock, that stimulus will normally come to elicit a conscious expectation that the shock is imminent, and also an emotional state of fear. The perhaps surprising suggestion is that these two forms of learning occur independently, so that it is possible for one to occur without the other. In particular, a stimulus can come to elicit fear even though we have no conscious awareness that an unpleasant event is about to occur. We can suddenly find ourselves frightened, for no obvious reason.

One final note. In previous sections we have traced a gradual shift in psychologists' understanding of classical conditioning. In Pavlov's stimulus substitution view, it was all very simple: activation of the CS center activates a US center, and so the CS elicits the same response as the US. The research reviewed in this chapter, however, has revealed a far more complex process:

- Conditioning is not simply about contiguity. We do not just associate a US with whatever stimuli happen to precede it; conditioning focuses on which of these stimuli is the best *predictor* of the US.
- In order to find the best predictor, we may search our memories for events that occurred minutes or even hours earlier.
- Once we identify a useful predictor, there is some sort of decision process to determine how we respond. In simultaneous conditioning, we know that no US is coming and so we don't respond; in taste-aversion learning we come to regard the CS as itself disgusting and so avoid it like the plague (if we're a coyote, even urinating on it and trying to bury it); in fear conditioning, if we're a rat we may freeze to hide from a predator or else run to a hiding place, if that option is available (e.g., Bolles, 1970; Timberlake, 1994). Our response is not just determined by the CS; it depends on the circumstances.

In sum, learning psychologists no longer see conditioning as a primitive process for connecting brain centers that happen to be active at the same time; it is a far more sophisticated system for detecting relationships, allowing us to anticipate when important events are going to occur and then deciding how best to prepare for them (see also Dickinson, 1980).

This shift in perspective was neatly captured in the title of an article by Robert Rescorla (1988): "Pavlovian conditioning: It's not what you think it is." He went on to write:

> Pavlovian conditioning is not a stupid process by which the organism willy-nilly forms associations between any two stimuli that happen to co-occur. Rather, the organism is better seen as an information seeker using logical and perceptual relations among events . . . to form a sophisticated representation of its world. (Rescorla, 1988, p. 154)

stimulus substitution Pavlov believed that activation of the CS center in the brain would be transferred to the US center, and the CS would therefore elicit the same behaviors as the US. In effect, the CS becomes the US.

autoshaping A procedure originally developed to train (shape) pigeons to peck a plastic disk called a key. The key is illuminated for several seconds, and this is followed by access to food. Within a few trials, the pigeons begin to peck the key when it is illuminated. Traditionally, training pigeons to peck a key was a time-consuming process requiring an experimenter to watch the bird and reinforce it for progressively closer movements toward the key. Autoshaping was much faster, as the equipment could be programmed to automatically turn on the key light and then deliver food.

postconditioning devaluation Following conditioning, the US is paired with an aversive event – typically, a toxin to produce illness – in order to make the US less attractive.

two-system hypothesis The proposal that two different learning systems can be involved when we learn about the relationship between successive events. One is a relatively primitive system that forms an association between the events; the other is a more sophisticated cognitive system that creates an expectation – in humans, conscious – that the first event will be followed by the second.

amygdala An almond-shaped organ – the term comes from the Greek word for almond – located deep within the center of the brain. It plays an important role in emotion, and, in particular, fear conditioning.

The Rescorla–Wagner Model

We're going to conclude our discussion of conditioning theories by looking at one proposed by Robert Rescorla and Allan Wagner. Their theory set out to account for almost every important aspect of conditioning – conditioning, extinction, blocking, contingency, and so on – and to do so using only a single, simple equation.

Their model has proved to be one of the most remarkable and influential theories in psychology, and we therefore will examine it in some detail. Before we begin, it might be worth noting that some parts of the exposition are difficult and may require careful rereading. This might seem to contradict the claim that the model is simple, but once you understand the model, it really is simple. The catch is that it is stated in mathematical form, so you will have to master some unfamiliar symbols before the model begins to make sense. Mastering this new terminology is not easy, but the potential reward is an insight into how complex behavior can sometimes be explained by remarkably simple processes.

The Model

The main impetus for the Rescorla–Wagner model came from Kamin's work on blocking. As we saw earlier, Kamin used the concept of surprise to account for blocking. If a powerful event surprises us, he said, we search for an explanation, and it is this search that produces learning.

Rescorla and Wagner took this fundamental insight of Kamin's and modified it in several important respects. In essence, they introduced three major changes:

- They extended the model. Where Kamin had assumed that surprise determines *whether* conditioning occurs, Rescorla and Wagner assumed that the amount of surprise on any trial would also determine *how much* conditioning occurred on that trial.
- In order to be able to predict the amount of conditioning, they stated their model in mathematical form.
- They changed the terminology. Kamin couched his explanation in terms of cognitive concepts such as expectations and surprise; Rescorla and Wagner were reluctant to speculate about mental states and so adopted more neutral terminology.

Translating Surprise

Rescorla and Wagner assumed that how much conditioning occurred on any trial would depend on how surprising the US was. Put in intuitive terms, if you're very surprised by the US, you obviously weren't expecting it, and thus you need to modify your expectations substantially. If you're only a little surprised, on the other hand, your expectations must have been just about right, and so there's less need to change them. So, how much you adjust your expectations after a US will depend on how much it surprised you:

$$\text{amount of conditioning} \approx \text{amount of surprise}$$

(As used here, the symbol $\approx$ means "approximately equals," or "depends on.")

Rescorla and Wagner, however, didn't want to use mental terms like expectations and surprise – how do we know what a dog or a rat is thinking, and whether or not

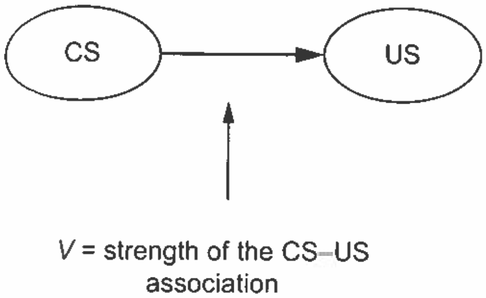

Figure 3.9 In Rescorla and Wagner's model, conditioning involves the formation of an association between a center in the brain representing the CS and another center representing the US. V measures the strength of this association.

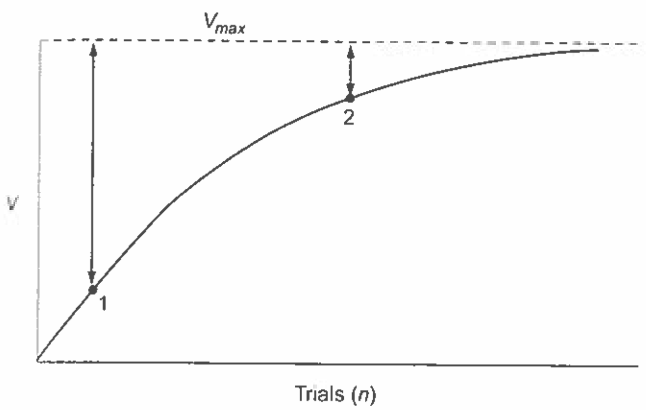

Figure 3.10 The relationship between V and V_{max} early and late in conditioning. Early in conditioning (point 1), the difference between V and V_{max} is great. Later (point 2), the difference is much smaller.

they are expecting food? So instead they formulated their model using the more neutral language of associations. Instead of talking about an *expectation* of a US, they talked about the *association* between the CS and the US. And they used the symbol ***V*** to represent the strength of this association.

Figure 3.9 illustrates this idea. When a CS is paired with a US, an association is assumed to be formed between a center in the brain representing the CS and a center representing the US. V represents the strength of this association. The stronger the association, the more likely that the CS will activate the US center, and thus that a conditioned response will occur.

So, using this associative terminology, how can we capture the concept of surprise?

To understand Rescorla and Wagner's solution, take a look at Figure 3.10, which shows the results of a typical conditioning experiment, perhaps a salivary conditioning experiment involving dogs. At first conditioning would increase rapidly, but over time associative strength would level off, approaching a maximum value. In mathematics, a maximum like this is called an **asymptote**, and we will use the symbol V_{max} to represent this maximum value.

We've marked two points on the curve. At point 1, early in conditioning, associative strength is far below its maximum value – in Kamin's terms, the dogs weren't really expecting food, and thus its appearance was still surprising. At point 2, on the other hand, associative strength was near its maximum, and so the appearance of food was less surprising. The relationship between V and V_{max} thus seems to closely parallel surprise: When the value of V is far from its maximum (point 1), there is a lot of surprise; when V's value is close to V_{max} (point 2), there is little surprise. The difference between V and V_{max} thus gives us a way to predict surprise:

$$\text{amount of surprise} \approx (V_{max} - V)$$

The greater the difference, the greater the surprise.

Summarizing our discussion to this point, we first suggested that conditioning is determined by surprise, and now we've suggested that surprise can be predicted by how far V is from V_{max}. Combining these assumptions, the implication is that how much conditioning occurs on any trial will depend on how far the current value of V is from its maximum value:

$$\text{amount of conditioning} \approx (V_{max} - V)$$

Rescorla and Wagner used the symbol ΔV to represent the amount of conditioning. Δ is the mathematical symbol for change; ΔV is the *change* in the strength of the association produced by a trial. To use a concrete example, suppose we knew that the strength of association at the beginning of a trial was 20, and that the pairing of tone and food on this trial increased the tone's associative strength from 20 to 30. If so, the change in the strength of the association produced by this trial (ΔV) would be 10.

So, Kamin's idea that conditioning depends on surprise becomes, in Rescorla and Wagner's version:

$$\Delta V \approx (V_{max} - V)$$

Our statement of the model is now almost complete, we just need to add one final assumption. Rescorla and Wagner knew that the precise amount of conditioning on any trial depends on more than just surprise – for example, we know that fear conditioning is faster when a more intense shock is used. To take account of some of these other factors, they added a constant to their equation, c. Their final formula, then, was that the amount of conditioning on any trial would be:

$$\Delta V = c(V_{max} - V)$$

For example, suppose we knew that the maximum value of V was 100, and that the value of c was .5. On the first trial, V's strength would initially be 0, and so the change in associative strength on that trial would be:

$$\begin{aligned}\Delta V &= c(V_{max} - V)\\ &= .5(100 - 0 = .5(100) = 50\end{aligned}$$

The model is letting us predict precisely how much conditioning will occur on any trial.[5]

Parameters

Ah, but there's a catch. In our example, in order to predict what would happen we had to know what values to use for the two constants in the equation, c and V_{max}; if we had used different values, the model would have predicted a different outcome. In mathematics, constants in an equation are called *parameters*; in order to use the equation we need to know what values to use for our parameters.

There are procedures for determining the values of parameters, but they turn out to be very demanding. In the entire history of learning theory there has been only one sustained effort to estimate a theory's parameters, by Clark Hull and his colleagues (Hull, 1943). When this failed, after more than a decade of effort, it convinced most learning theorists that the field simply wasn't ready for mathematical models.

Given this history, Rescorla and Wagner decided not even to try to determine the correct values for c and V_{max}; instead, they just used arbitrary values! This use of arbitrary values may seem pointless, to put it mildly, but although this strategy precludes *quantitative* predictions, it turns out that the model can still make some very interesting *qualitative* predictions. That is, although the exact points on the learning curve depend on the values of c and V_{max} used in the equation, the general shape of the curve turns out to be the same no matter what values are used. Thus although the model does not let us predict the exact number of drops of saliva, it always makes the same qualitative predictions about whether salivation will increase or decrease. This might still seem a waste of time – we hardly need a sophisticated mathematical model to tell us that conditioning will increase over

---

[5] Rescorla and Wagner actually used two different constants, α and β, rather than the single parameter c, and they also used the symbol λ to represent asymptotic conditioning rather than V_{max}. We have altered the symbols to make the exposition of the model easier to follow. Should you read the model in the original, you will find the equation stated as

$$\Delta V_n = \alpha\beta(\lambda - V_n)$$

The subscript n refers to the trial number. V_1, for example, would be the associative strength at the beginning of trial 1.

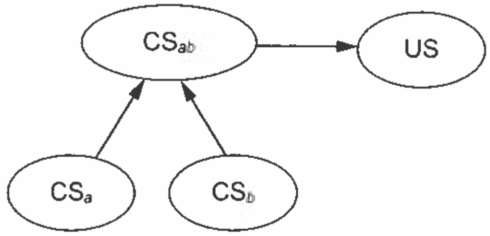

Figure 3.11 When two stimuli, A and B, are presented together, their centers in the brain both activate a center representing the AB compound. The level of activation of the AB compound, V_{ab}, would then be the sum of the activation it receives from the A center (V_a) and the B center (V_b).

trials – but Rescorla and Wagner were able to show that even simple statements of this kind can lead to interesting and unexpected predictions.

The Model's Successes

The model can account for most of the basic characteristics of conditioning, such as conditioning itself, extinction, and contingency, but it would take us too far afield to trace how it does it. (If you're interested, you can find accounts in Lieberman, 2004 and 2012.) To illustrate how the model works, though, we'll look at its account of two phenomena, blocking and overexpectation.

Blocking

In Kamin's blocking experiment, he first gave rats training with a noise followed by shock. He then paired a noise–light compound with shock, and subsequent test trials showed that no fear had been conditioned to the light, despite the fact that it had been paired with a powerful electric shock.

How would the Rescorla–Wagner model explain this? To understand the model's account, we need to first consider how conditioning is affected if two stimuli are present on a trial instead of one. We said earlier that conditioning on any trial depends on surprise, which in turn depends on how strongly the subject expected the US. Rescorla and Wagner assumed that if two conditioned stimuli, *a* and *b*, were presented together, the subject would take both stimuli into account in estimating the likelihood of the US. Specifically, they proposed that the associative strength at the beginning of a trial would be the sum of the strengths of each of the stimuli present:

$$V_{ab} = V_a + V_b$$

Figure 3.11 provides a graphic representation of this idea.

We are now in a position to explain blocking. When the noise was repeatedly paired with shock in the first phase, its associative strength would have reached its asymptote. If, for example, the value of V_{max} in this experiment had been 100, then at the beginning of the compound trials the associative strength of the noise would have been:

$$V_{noise} = V_{max} = 100$$

Because the light had never been paired with shock, however, its associative strength would have been 0:

$$V_{light} = 0$$

Their combined associative strength would thus have been

$$V_{noise+light} = V_{noise} + V_{light} = 100 + 0 = 100$$

At the beginning of the compound trials, therefore, the associative strength of the compound would already have been at its maximum value of V_{max}. (In Kamin's terms, the rats would now be strongly expecting shock.) The amount of conditioning to the light on this trial would thus be

$$\Delta V = c(V_{max} - V_{noise+light}) = c(100 - 100) = c(0) = 0$$

In other words, no conditioning would occur, which is exactly what Kamin found. And note that this would be the predicted result no matter what values we used for the parameters c and V_{max}. Because V on the compound trials would have already reached its asymptotic value of V_{max}, $(V_{max} - V)$ would always be zero, and c multiplied by 0 would also be 0. The model always predicts blocking, no matter what values we choose for the parameters.

Overexpectation

There are three main criteria used in evaluating any scientific theory. The first is whether it can account for known phenomena, and we have seen that the Rescorla–Wagner model has had considerable success in this regard, accounting not only for conditioning itself but also for phenomena such as extinction, blocking, and contingency.[6]

A second criterion is the scientific principle of **parsimony**. In essence, this says that where competing theories can account for the same data, preference should be given to whichever of the theories is the simplest. If one theory requires 5 assumptions to explain a phenomenon and another can do it with just 2, then we should prefer the theory with the fewer and simpler assumptions. Note that this principle is not, strictly speaking, logically necessary: It is entirely possible that the more complex theory could be the correct one. Scientists, however, have a deep love of simple explanations – there is something beautiful, something deeply satisfying, about being able to explain a wide range of complex phenomena in terms of just a few simple assumptions. And, as it turns out, simpler theories often have proved right. Newton's theory of gravity is one impressive example, as Newton was able to

[6] The model's account of contingency is available in Rescorla and Wagner (1972).

account for an astonishing range of phenomena – the orbits of the planets, the movements of the tides, the fall of an apple – in terms of just one fundamental force. In any case, for better or worse, scientists prefer simpler accounts wherever possible, and the Rescorla–Wagner model is just about the simplest account imaginable – it basically says that almost every aspect of conditioning can be explained by the difference between two quantities, V and V_{max}.

There is, however, a third, even more demanding criterion used to judge scientific theories, and that is its ability to predict new, previously unknown phenomena. So, how well has the Rescorla–Wagner model fared by this more daunting criterion? To assess the model's success in this crucial respect, we will focus on one of its strangest and most counterintuitive predictions, that in some circumstances pairing a CS with a US will result not in conditioning but in extinction!

Suppose that we exposed rats to conditioning trials in which a tone and a light were separately paired with an intense shock:

$$\text{tone} \rightarrow \text{shock}$$

$$\text{light} \rightarrow \text{shock}$$

Then suppose that we presented the tone and light together on the next conditioning trial:

$$\text{tone} + \text{light} \rightarrow \text{shock}$$

What effect should this additional pairing have on fear of the tone? Because the tone is again followed by an unpleasant shock, you might expect a further increase in fear, but, according to the Rescorla–Wagner model, the situation is not that simple. As we've already seen, the amount of conditioning in any situation depends not simply on the US but also on the associative strength at the beginning of the trial. Suppose, for example, that only a few trials were given before the compound trial, so fear levels to the two stimuli were only moderate:

$$V_a = V_b = 0.20$$

On the compound trial, V_{ab} would be 0.40. If we assign c the arbitrary value of 0.5, then the change in associative strength on that trial would be

$$\Delta V_a = \Delta V_b = c(V_{max} - V_{ab}) = 0.5(1.0 - 0.4) = 0.30$$

In accordance with common sense, in other words, the model predicts an increase in fear conditioning on this trial.

Now suppose that more extensive conditioning to the tone and light took place before the first compound trial, with this result:

$$V_a = V_b = 0.9$$

In this case, the associative strength of the compound would be 1.8, so that on the compound trial:

$$\Delta V_a = \Delta V_b = c(V_{max} - V_{ab}) = 0.5(1.0 - 1.8) = -0.40$$

Even though the compound is still being followed by a powerful electric shock, the model now predicts a *decrease* in fear levels!

Rescorla (1970) tested this prediction. In the first phase, rats were given extensive pairings of both a tone and a light with shock, so that fear conditioning to each would be essentially at asymptotic levels. An experimental group was then given 12 compound trials in which the tone and the light were presented together and followed by the same shock as in training; a control group received no further training. Finally, fear conditioning to the 2 stimuli was assessed by presenting them separately in a CER (conditioned emotional response) test. (See Chapter 2 for a discussion of the CER test.)

Figure 3.12 shows the results of this experiment. Let us look first at the results for the light: Note that responding was suppressed much more in the control group (suppression ratio of 0.03) than in the experimental group (suppression ratio of 0.17). The initial pairing of the light with shock, in other words, had resulted in strong fear conditioning, but the additional pairings in the experimental group actually reduced that fear. The effect on the tone was, if anything, even more

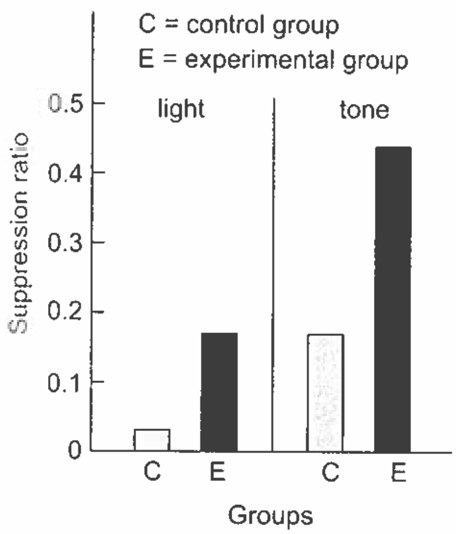

Figure 3.12 Fear elicited by a light and a tone following just conditioning (control group) or following additional trials in which the tone and light were presented jointly, followed by shock (experimental group). Fear was measured with a suppression ratio in which 0.5 represents no fear and 0.0 represents strong fear. The data thus show that additional conditioning trials actually reduced fear. (Based on data from Rescorla, 1970.)

dramatic, with the extra compound trials resulting in an even greater decrease in fear. Indeed, the tone no longer appeared to elicit any fear at all; the observed suppression ratio of 0.44 was virtually indistinguishable from the neutral point of 0.50. Extra pairings of the tone and light with shock not only did not increase fear, as common sense might predict, but actually reduced or even eliminated it! It is a bizarre result, but precisely what the model predicts. This phenomenon is often referred to as the **overexpectation effect**, because it is the result of cues predicting the US more strongly than is justified. (See also Arico and McNally, 2014.)

The Model's Limitations

We have examined in some detail one prediction of the model; other predictions have also been tested and many have been confirmed (for example, Blough, 1975). In some important respects, however, the model's predictions have proved incorrect. We will consider one example here, involving *latent inhibition*. In the first demonstration of this phenomenon, Lubow and Moore (1959) gave goats and sheep conditioning trials in which a flashing light was followed by a mild shock.[7] Before beginning conditioning, though, they presented the light on its own 10 times, for 10 seconds each time:

$$\text{Preexposure : light}$$

$$\text{Conditioning : light} \rightarrow \text{shock}$$

A control group received the identical conditioning trials but without preexposure to the light.

How should preexposure to the light affect subsequent conditioning? Because no US was presented, the model says that there should be no conditioning. At the beginning of the conditioning phase, the light should have had no associative strength in either group, and learning in the two groups should have proceeded identically. Contrary to this prediction, Lubow and Moore found that conditioning was significantly slower in the group preexposed to the light, a phenomenon they termed **latent inhibition**, because they believed that the CS became inhibitory during the preexposure phase.[8]

An obvious explanation for latent inhibition is that because the light in the first phase was not followed by any significant event, the animals learned to ignore it. As simple as this explanation is, the Rescorla–Wagner model cannot accommodate it,

---

[7] It is clearly not true that learning psychologists only study rats and pigeons.

[8] Subsequent evidence made it clear that the CS is neither excitatory nor inhibitory following preexposure; it is simply difficult to condition (for example, Reiss and Wagner, 1972). To prevent confusion, therefore, many researchers now prefer to use the term *CS-preexposure effect* rather than latent inhibition.

because the model does not include any mechanism for changing how much attention is paid to a stimulus. (See Escobar, Arcediano, and Miller, 2002, for a review of other possible explanations.)

To accommodate attention, theorists have proposed several ways in which the model could be modified. In essence, they propose that we cannot attend to all features of our environment, and that we learn to preferentially attend to stimuli that reliably signal important events. You can find discussion of some of the issues and theories in Mackintosh (1975), Pearce and Hall (1980), Pearce and Mackintosh (2010), Hogarth, Dickinson, and Duka (2010), Le Pelley *et al.* (2016), and Uengoer, Pearce, Lachnit, and Koenig (2018).

Evaluation

We have seen that there are three main criteria for evaluating scientific theories: the capacity to explain known phenomena, to predict new ones, and to do all this using the simplest possible set of assumptions. By all three criteria, the Rescorla–Wagner model has been impressive. In addition to explaining a wide range of phenomena using only a single, simple equation, the model has also generated a variety of counterintuitive predictions – for example, that a conditioning trial can reduce associative strength – and many of these predictions have been supported. As researchers have continued to test the model's predictions, however, it has become clear that there are also many phenomena that the model cannot explain. (For a review of the model's weaknesses as well as its strengths, see Miller, Barnet, and Grahame, 1995.)

Other theories have been proposed to try to address some of the model's deficiencies (for example, Mackintosh, 1975; Pearce and Hall, 1980; Schmajuk, Lam, and Gray, 1996; Denniston, Savastano, and Miller, 2001; Wagner and Brandon, 2001). One striking feature of many of these theories, however, is that they incorporate a version of the Rescorla–Wagner model's basic formula; they add other features to address processes such as attention, but the basic formula remains. Thus, although the Rescorla–Wagner model undoubtedly requires modification and extension, many psychologists believe that the insight at its heart – that conditioning depends on the discrepancy between current associative strength and asymptotic strength – captures a fundamental truth about conditioning.

Theories inevitably evolve, and perhaps this insight will eventually be seen as mistaken and supplanted by a deeper one, as Newton's ideas about gravity were eventually overtaken by Einstein's. Even were that to happen, the model has contributed to a rebirth of interest in mathematical models of learning with its remarkable demonstration of the power of a few simple assumptions to explain a wide range of seemingly complex phenomena. Its impact has been profound: Miller *et al.* (1995) called it "the most influential theory of associative learning to emerge . . . over the last 25 years," and Siegel and Allan (1996) wrote that "there have been few models in experimental psychology as influential as the Rescorla-Wagner model." Whatever its ultimate fate, the Rescorla-Wagner model is likely to prove a historic landmark in the evolution of our understanding of learning.

V In the Rescorla-Wagner model, the strength of an association between a CS and a US.

ΔV The change in the strength of a CS-US association produced by a trial.

asymptote In mathematics, a stable value that a curve approaches but never quite reaches. In learning, it describes the level of performance at which improvement ceases, so that further training would produce no additional improvement.

parameter A constant in a mathematical formula. The value of a parameter remains the same in successive applications of the formula in the same situation, but different values can be used in different situations. In the Rescorla-Wagner model, the parameters c and V_{max} determine the speed and asymptotic level of conditioning.

parsimony A criterion for evaluating theories which states that where several explanations are possible, we should prefer whichever is the simplest. A theory might be simpler because it involves fewer assumptions, or because those assumptions postulate simpler processes.

overexpectation effect A decrease in the strength of conditioning on compound trials in which the elements of the compound have previously been strongly conditioned separately.

latent inhibition Slower conditioning to a CS because of previous presentations of the CS by itself. Also known as the *CS preexposure effect*.

SUMMARY

- Psychologists initially believed that conditioning was a fundamentally simple process in which any stimulus that preceded a US would become associated with it. The strength of this association would depend on their contiguity, frequency, and intensity.
- This belief was challenged by research on preparedness, contingency, and blocking. Together, they suggested that conditioning depended not simply on whether a CS preceded a US but on whether it was a good predictor of that US.
- To explain blocking, Kamin suggested that conditioning occurred only when a US was surprising. When an unexpected event occurs, we search our memory for possible causes.
- Kamin's theory revived a longstanding dispute about what is learned during conditioning. According to Pavlov, conditioning is essentially a simple process

in which an association is formed between the CS and US centers in the brain. According to Tolman, on the other hand, conditioning is a sophisticated process involving the formation of an expectation that the CS will be followed by the US.

- There is evidence to support both accounts, and this has led some theorists to the view that in the course of evolution two fundamentally different systems emerged, first a relatively simple associative system and then a more sophisticated cognitive system. Evidence that conditioning can affect us without our awareness supports the idea that we still have a primitive system within our brains, as well as the more sophisticated cognitive one.
- Rescorla and Wagner developed a mathematical model of conditioning that was based on Kamin's work, though it avoided speculation on mental states. The model said conditioning causes the formation of an association between the CS and the US, and that the change in the strength of this association on any trial would depend on the difference between its existing strength and the maximum possible strength.
- Their model was able not only to account for most known phenomena in conditioning but also to successfully predict many new ones. Its success was particularly impressive because of its parsimony – its predictions were based on a small number of simple assumptions.
- There were also phenomena that the model had difficulty explaining, such as latent inhibition. Whatever its ultimate success, it is unquestionably the most successful – and influential – model in the history of learning.

Review Questions

1 Why did simultaneous and backward conditioning seem to pose problems for the principle of contiguity? How can these apparent anomalies be explained?

2 The principle of contiguity suggests that contiguity is both *necessary* for conditioning (conditioning will occur only if the CS and the US occur closely together in time) and *sufficient* (if a CS and a US occur together, then conditioning will occur). How did the research of Rescorla, Garcia, and Kamin pose problems for this principle?

3 How did Rescorla disentangle the roles of contiguity and contingency in conditioning?

4 How did Garcia and Koelling show that the conditioning of a stronger aversion to a taste than to a light was not simply the result of greater salience of the taste as a conditioned stimulus?

5 How might classical conditioning contribute to an animal's survival? Why might it be better *not* to associate a US with all the stimuli that precede it?

6 How did Kamin account for blocking?
7 What is the difference between signal and substitution accounts of conditioning? What evidence supports each?
8 What is the two-system hypothesis? How does it account for the conflicting evidence on whether a CS functions as a signal or a substitute for the US?
9 The two-system hypothesis suggests that conditioning can occur even when we are not consciously aware of the relationship between the CS and the US. What kind of evidence has been used to test this claim, and what has it shown?
10 How did Rescorla and Wagner build on Kamin's work? What changes did they make to his explanation of blocking?
11 What equation did they use to predict learning? What does each symbol represent?
12 Why didn't Rescorla and Wagner try to determine the real values of the parameters c and V_{max}? What approach did they take instead?
13 What are the three main criteria usually used to evaluate theories? How does the Rescorla–Wagner model fare on each?
14 Why did the evidence for latent inhibition pose a problem for the Rescorla–Wagner model?