to realize that both groups in this experiment received exactly the same number of pairings of T and shock. Because of the equal number of pairings, the frequency principle predicts that conditioning to T should be equally strong in the two groups. However, Kamin obtained a strikingly different result: Whereas he observed a strong CR to T in the control group, he recorded essentially no conditioned responding at all to T in the blocking group. Because the only difference between the two groups was that the blocking group received conditioning trials with L in Phase 1 but the control group did not, Kamin concluded that this prior conditioning with stimulus L somehow "blocked" the later conditioning of stimulus T. Since Kamin's pioneering work, the blocking effect has been demonstrated in numerous experiments using a variety of conditioning situations, with both animal and human subjects (for example, Goddard & Jenkins, 1988; Martin & Levey, 1991).

An intuitive explanation of the blocking effect is not difficult to construct: To put it simply, stimulus T was redundant in the blocking group; it supplied no new information. By the end of Phase 1, subjects in the blocking group had learned that stimulus L was a reliable predictor of the US—the US always occurred after L, and never at any other time. The addition of T to the situation in Phase 2 added nothing to the subject's ability to predict the US. This experiment suggests that conditioning will not occur if a CS adds no new information about the US.

This experiment demonstrates that conditioning is not an automatic result when a CS and a US are paired. Conditioning will occur only if the CS is informative, only if it is predictive of something important, such as an upcoming shock. This view seems to imply that the subject has a more active role in the conditioning process than was previously thought—the subject is a selective learner, learning about informative stimuli and ignoring uninformative ones. For two psychologists, Robert Rescorla and Allan Wagner (1972), the blocking effect and related findings underscored the need for a new theory of classical conditioning, one that could deal with these loose notions of informativeness and predictiveness in a more rigorous, objective way. The results of their collaborative efforts was the Rescorla-Wagner model, now one of the most famous theories of classical conditioning.

WWW

You can try an interactive simulation of salivary conditioning that demonstrates acquisition, extinction, and blocking at *http://www.uwm.edu/People/johnchay/cc.htm.*

The Rescorla-Wagner Model: Basic Concepts

The Rescorla-Wagner model is a mathematical model about classical conditioning, and because of its technical nature, it can be quite challenging to understand. The next section will present the quantitative details of the model. However, the basic ideas behind the theory are quite simple and reasonable, so let us begin by examining these ideas in an informal way. This section is designed to give you a good understanding of the concepts behind the model without using any equations.

Classical conditioning can be viewed as a means of learning about signals (CSs) for important events (USs). The Rescorla-Wagner model is designed to predict the outcome of classical conditioning procedures on a trial-by-trial basis. For any trial in which one or more CSs is presented, the model assumes that there can either be excitatory conditioning, inhibitory conditioning, or no conditioning at all. According to the model, two factors determine which of these three possibilities actually occurs: (1) the strength of the subject's expectation of what will occur, and (2) the strength of the US that is actually presented. The model is a mathematical expression of the concept of *surprise:* It states that learning will occur only when the subject is surprised—that is, when what actually happens is different from what the subject expected to happen.

You should be able to grasp the general idea of the model if you learn and understand the following six rules:

1. If the strength of the actual US is greater than the strength of the subject's expectation, all CSs that were paired with the US will receive excitatory conditioning.
2. If the strength of the actual US is less than the strength of the subject's expectation, all the CSs that were paired with the US will receive some inhibitory conditioning.
3. If the strength of the actual US is equal to the strength of the subject's expectation, there will be no conditioning.
4. The larger the discrepancy between the strength of the expectation and the strength of the US, the greater will be the conditioning (either excitatory or inhibitory) that occurs.
5. More salient (more noticeable) CSs will condition faster than less salient (less noticeable) CSs.
6. If two or more CSs are presented together, the subject's expectation will be equal to their total strength (with excitatory and inhibitory stimuli tending to cancel each other out).

We will now examine several different examples to illustrate how each of these six rules applies in specific cases. For all of the examples below, we will imagine that a rat receives a conditioning procedure in which a CS (light, tone, or similar stimulus) is paired with the presentation of food as a US. In this conditioning situation, the CR is activity, as measured by the rat's movement around the conditioning chamber (which can be automatically recorded by movement detectors). In actual experiments using this procedure, the typical result is that as conditioning proceeds, the rat becomes more and more active when the CS is presented, so its movement can be used as a measure of the amount of excitatory conditioning.

Acquisition. Consider a case in which a light (L) is paired with one food pellet. On the very first conditioning trial, the rat has no expectation of what will follow L, so the strength of the US (the food pellet) is much greater than the strength of the rat's expectation (which is zero). Therefore, this trial produces some excitatory conditioning (Rule 1). But conditioning is rarely complete after just one trial. The second time L is presented, it will elicit a weak expectation, but it is still not as strong as the actual US, so Rule 1 applies again, and more excitatory conditioning occurs. For the same reason, further excitatory conditioning should take place on trials 3, 4, and so on. However, with each conditioning trial, the rat's expectation of the food pellet should get stronger, and so the difference between the strength of the expectation and the strength of the US gets smaller. Therefore, the fastest growth in excitatory conditioning occurs on the first trial, and there is less and less additional conditioning as the trials proceed (see Rule 4). Eventually, when L elicits an expectation of food that is as strong as the actual food pellet itself, the asymptote of learning is reached, and no further excitatory conditioning will occur with any additional L-food pairings.

Blocking. Continuing with this same example, now suppose that after the asymptote of conditioning is reached, a compound CS of L and tone (T) are presented together and are followed by one food pellet. According to Rule 6, when two CSs are presented, the subject's expectation is based on the total expectations from the two. T is a new stimulus, so it has no expectations associated with it, but L produces an expectation of one food pellet. One food pellet is in fact what the animal receives, so the expectation matches the US, and no additional conditioning occurs (Rule 3). That is, L retains its excitatory strength, and T retains zero strength.

This, in short, is the model's explanation of the blocking effect: No conditioning occurs to the added CS because there is no surprise—the strength of the subject's expectation matches the strength of the US.

Extinction and Conditioned Inhibition. Let us now think about a slightly different example. Suppose that after conditioning with L has reached its asymptote, the rat receives trials in which L and T are presented together, but no food pellet is delivered on these trials.

This is an example in which Rule 2 applies: The strength of the rat's expectation will exceed the strength of the actual US. This is because the previous training with L will give the rat a strong expectation of food, yet the strength of the actual US is zero (since no US is presented on these extinction trials). According to Rule 2, both CSs, L and T, will acquire some inhibitory conditioning on these extinction trials.

Let us be clear about how this inhibitory conditioning will affect L and T. Because L starts with a strong excitatory strength, the trials without food (and the inhibitory conditioning they produce) will begin to counteract the excitatory strength. This is merely an example of extinction: Presenting an excitatory CS without the US will cause the strength of the CS to weaken. In contrast, T begins this phase with zero strength, because it has not been presented before. Therefore, the trials without food (and the inhibitory conditioning they produce) will cause T's strength to decrease below zero—it will become a conditioned inhibitor.

Overshadowing. In a conditioning experiment with a compound CS consisting of one intense stimulus and one weak one, Pavlov discovered a phenomenon he called **overshadowing.** After a number of conditioning trials, the intense CS would produce a strong CR if presented by itself, but the weak CS by itself would elicit little if any conditioned responding. It was not the case that the weak CS was simply too small to become an effective CS, because if it were paired with the US by itself, it would soon elicit CRs on its own. However, when presented in conjunction with a more intense CS, the latter seemed to mask, or overshadow, the former. Overshadowing has been observed in experiments with both animal and human subjects (Spetch, 1995).

The Rescorla-Wagner model's explanation of overshadowing is straightforward. According to Rule 5, more salient stimuli will condition faster than less salient stimuli. If, for example, a dim light and a loud noise are presented together and followed by a food pellet, the noise will acquire excitatory strength faster than the light. When the total expectation based on both the noise and the light equal the strength of the food pellet, excitatory conditioning will stop. Because the noise is more salient, it will have developed much more excitatory strength than the light. If the dim light is presented by itself, it should elicit only a weak CR.

The Overexpectation Effect. Besides being able to account for existing data, another characteristic of a good theory (called fruitfulness in Chapter 1) is the ability to stimulate new research by making novel predictions that have not been previously tested. The Rescorla-Wagner model deserves good grades on this count, because hundreds of experiments have been conducted to test the model's predictions. Its prediction of a phenomenon known as the **overexpectation effect** is a good case in point.

Table 5-2 presents the design of an experiment that tests the overexpectation effect. Two CSs, L and T, are involved. For Phase 1, the notation "L^+, T^+" is used to indicate that on some trials L is presented by itself and followed by a food pellet, whereas on other trials T is presented by itself and followed by a food pellet. The two types of trials, L^+ and T^+, are randomly intermixed in Phase 1. Consider what should happen on each type of trial. On L^+ trials, the strength of the expectation based

TABLE 5-2 Design of an Experiment on the Overexpectation Effect

Group	*Phase 1*	*Phase 2*	*Test Phase*	*Result*
Overexpectation	L^+, T^+	LT^+	L, T	Moderate CRs
Control	L^+, T^+	No stimuli	L, T	Strong CRs

on L will continue to increase and eventually approach the strength of one food pellet. Similarly, on T^+ trials, the strength of the expectation based on T will grow and also approach the strength of one food pellet. Note that because L and T are never presented together, the conditioned strengths of both stimuli can individually approach the strength of one food pellet.

In Phase 2, rats in the control group receive no stimuli, so no expectations are changed. Therefore, in the test phase, these rats should exhibit a strong CR to both L and T on the first several test trials (which are extinction trials).

The results should be quite different for rats in the overexpectation group. In Phase 2, these rats receive a series of trials with the compound stimulus, LT, followed by one food pellet. On the first trial of Phase 2, a rat's total expectation, based on the sums of the strengths of L and T, should be roughly equal to the strength of two food pellets (because each stimulus has a strength of about one food pellet). Loosely speaking, we might say that the rat expects a larger US (two food pellets) on the compound trial because two strong CSs are presented, but all it gets is a single food pellet. Thus, compared to what it actually receives, the animal has an overexpectation about the size of the US, and Rule 2 states that under these conditions both CSs will experience some inhibitory conditioning (they will lose some of their associative strength).

With further trials in Phase 2 for the overexpectation group, the strengths of L and T should continue to decrease, as long as the total expectation from the two CSs is greater than the strength of one food pellet. When tested in the next phase, the individual stimuli L and T should exhibit weaker CRs in the overexpectation group because their strengths were weakened in Phase 2. Experiments have confirmed this prediction that CRs will be weaker in the overexpectation group than in the control group (Khallad & Moore, 1996; Kremer, 1978).

The model's accurate prediction of the overexpectation effect is especially impressive because the prediction is counterintuitive. If you knew nothing about the Rescorla-Wagner model when you examined Table 5-2, what result would you predict for this experiment? Notice that subjects in the overexpectation group actually receive more pairings of L and T with the US, so the frequency principle would predict stronger CRs in the overexpectation group. Based on the frequency principle, the last thing we would expect from more CS-US pairings is a weakening of the CS-US associations. Yet this result is predicted by the Rescorla-Wagner model, and the prediction turns out to be correct. The overexpectation effect is only one of several counterintuitive predictions of the Rescorla-Wagner model that have been supported by subsequent research.

Brief tutorials on blocking, overshadowing, and other topics of classical conditioning can be found at *http://brembs.net/classical.*

The Rescorla-Wagner Model: Equations and Mathematical Examples

Having examined the Rescorla-Wagner model in a nonmathematical way, we are now in a better position to tackle the more difficult task of learning the mathematical details.

Notation. In the model, the strength of a US is signified by A, and a subscript can be used to identify exactly what US is presented. For example, A_1 could represent the strength of one food pellet, A_2 could represent the strength of two food pellets, and A_0 could represent the strength of a trial with no food pellets. The letter V is used to represent the conditioned strength of a CS, and again subscripts are used to indicate which CS is being discussed. For example, V_L could be the conditioned strength of a light, and V_T the conditioned strength of a tone. V is positive if the CS is excitatory and negative if the CS is inhibitory. Because the subject's expectation is said to be based on the total strength of all CSs that are presented on a given trial, a special term, V_{sum}, is used to represent this total. The salience of each CS is designated by S. For example, S_L could represent the salience of a light, and S_T the salience of a tone. The salience of a CS must be a number between 0 and 1. Finally, the notation ΔV (pronounced "delta V") refers to the change in strength of a CS that occurs on a single conditioning trial. (ΔV_L is the change in strength of the light, and ΔV_T is the change in strength of a tone.)

According to the Rescorla-Wagner model, on any conditioning trial, the following equation can be used to describe the change in strength of a CS on a single trial:

$$\Delta V_i = S_i \times (A_j - V_{sum})$$

The subscript i refers to any single CS, and the subscript j refers to any single US. Notice that the quantity in parentheses, $(A_j - V_{sum})$, represents the difference between the strength of the US and the total strength of the subject's expectation (based on the sum of all the CSs presented on a given trial). ΔV will be positive whenever this quantity is positive, it will be negative when the quantity is negative, and it will be zero when the quantity is zero. This quantity in parentheses is simply multiplied by the salience parameter for the CS to determine how much excitatory or inhibitory conditioning is predicted for a single trial.

Although almost anyone could memorize this equation, for most people it will take more work to understand how the equation is actually applied to specific cases. The best way to gain such an understanding is to use the equation to make predictions for a variety of conditioning situations. We will now work through several such examples, using numerical and graphic aids to make the predictions concrete.

Because both A, the strength of a US, and V, the conditioned strength of a CS, are hypothetical quantities that cannot be directly observed, we can use any convenient scale of numbers to represent these quantities. To keep the calculations as simple as possible, we will arbitrarily assign a strength of 100 to one food pellet.

Acquisition. Consider the first conditioning trial on which L is paired with one food pellet. If L is the only CS present, then $V_{sum} = V_L$, and $V_L = 0$ because there has been no prior conditioning with L. For the purposes of this example, we will set the salience value of the light, S_L, equal to .2. To calculate the amount of conditioning on this trial, we need to solve the following equation:

$$\Delta V_L = S_L \times (A_1 - V_{sum})$$

Inserting the values we have chosen, we get

$$\Delta V_L = .2 \times (100 - 0) = 20$$

Therefore, on this first trial, V_L should grow by 20 units. This process of growth is depicted graphically in Figure 5-1.

On trial 2, V_L (and therefore V_{sum}) begins at 20, so the equation becomes

$$\Delta V_L = .2 \times (100 - 20) = 16$$

This equation states that on trial 2, V_L will increase by another 16 units, so after two trials, $V_L = 20 + 16 = 36$ (see Figure 5-1). Notice because of the smaller discrepancy between A_1 and V_{sum} on trial 2, the amount of learning is smaller than on trial 1 (16 units instead of 20). Figure 5-1 shows that the increase in V_L should be 12.8 on trial 3, but only 2.7 by trial 10. By the end of trial 10, V_L has risen to 89.3, and with additional trials it would get closer and closer to the asymptote of 100. To summarize, the Rescorla-Wagner model predicts that in simple acquisition, the initial increases in V_L will be the largest, and the increments will become smaller and smaller as the asymptote is approached.

Overshadowing. It is easy to show how the Rescorla-Wagner model accounts for the phenomenon of overshadowing. Let us assume that we begin a new conditioning experiment with two CSs, the same light used in the previous example ($S_L = .2$) and a very loud noise (salience of the noise = $S_N = .5$). Figure 5-2 shows the results of several conditioning trials with this compound CS. On trial 1, $V_{sum} = V_L + V_N = 0$, so the discrepancy between A_1 and V_{sum} is 100, as in the previous example. Unlike the previous example, however, there

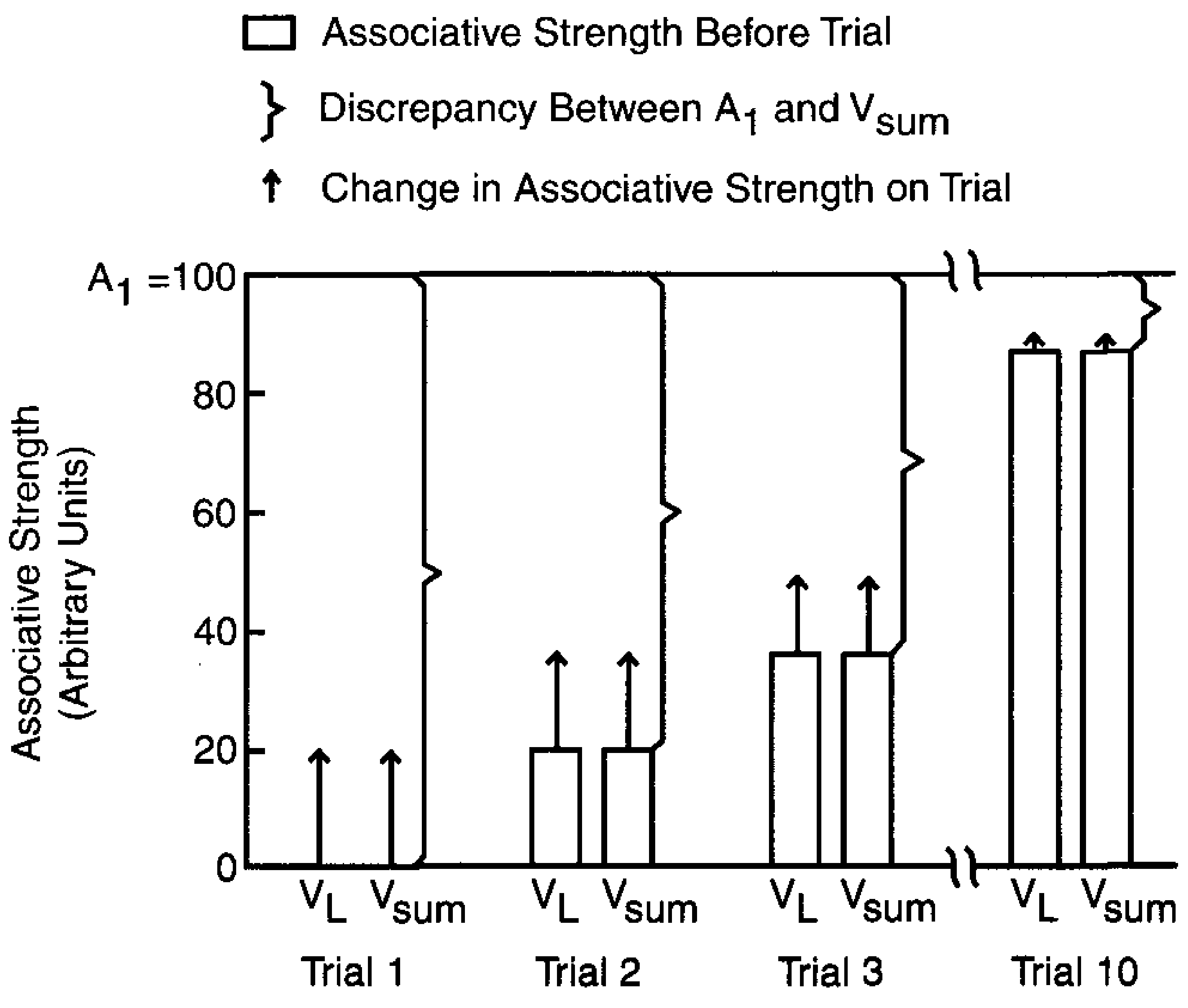

FIGURE 5-1 Predictions of the Rescorla-Wagner model for simple acquisition with a single CS. Parameter values used were $A_1 = 100$, $S_L = .2$.

are two CSs whose associative strengths must be incremented, so we need to solve two equations:

$$\Delta V_L = .2 \times (100 - 0) = 20$$

$$\Delta V_N = .5 \times (100 - 0) = 50$$

Notice that ΔV_L is the same as in trial 1 of the first example. For the more salient noise, however, the increment in conditioned strength is 50 units. Thus, after trial 1, $V_{sum} = 20 + 50 = 70$. At the start of trial 2, therefore, the difference between A_1 and V_{sum} has already been reduced to 30. The equations for trial 2 are

$$\Delta V_L = .2 \times (100 - 70) = 6$$

$$\Delta V_N = .5 \times (100 - 70) = 15$$

Therefore, total increment in V_{sum} on trial 2 is 21, so that after two trials, $V_{sum} = 70 + 21 = 91$.

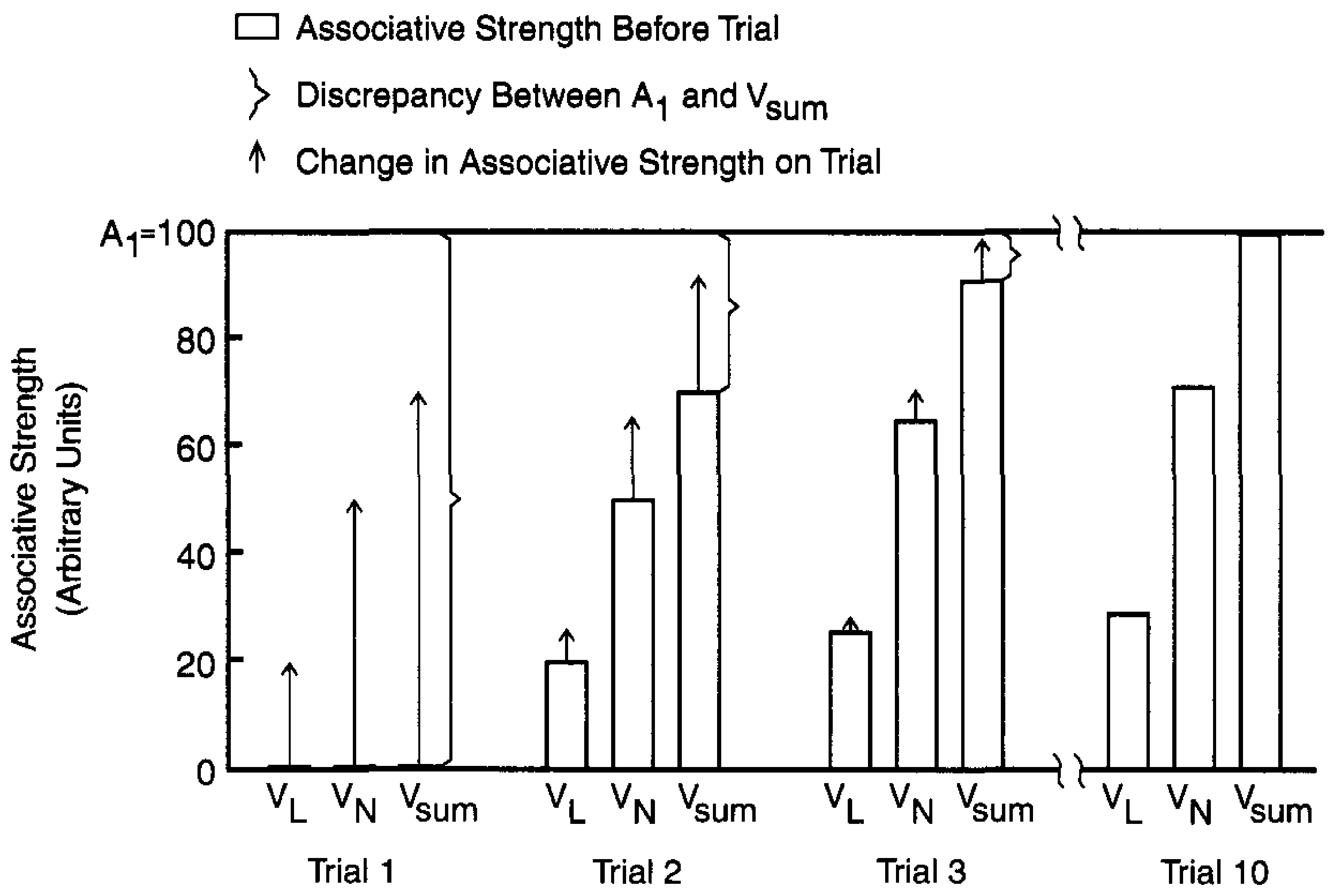

FIGURE 5-2 Predictions of the Rescorla-Wagner model for a case where an intense noise overshadows a light. Parameter values used were $A_1 = 100$, $S_L = .2$, $S_N = .5$.

Figure 5-2 also shows the predictions for trials 3 and 10. Notice that with the two CSs, V_{sum} approaches A_1 much more rapidly than in Figure 5-1, and by trial 10 the increments in strength are too small to show in the graph.

The model's prediction of overshadowing can be seen clearly by comparing the course of V_L in Figures 5-1 and 5-2. The only difference between these two conditioning situations is the addition of the noise in the second example. In Figure 5-1, V_L has reached a strength of 89.3 after 10 trials, and with further trials it will approach 100. In Figure 5-2, V_{sum} has nearly reached 100 by trial 10, but because the more salient noise has usurped over 70 units of strength, V_L will never rise above 30. In short, because the total strength of both CSs in the compound can never rise above 100 in this example, the model predicts that the light will be overshadowed—the level of conditioning will never be what it would be in the absence of the noise.

Blocking. The model's explanation of the blocking effect is similar to that of overshadowing. Suppose that in a blocking group, stimulus L receives many pairings with one food pellet in Phase 1, so that by the end of this phase, V_L is approximately 100 (assuming once again that the strength of one food pellet is 100). At the start of Phase 2, the quantity ($A_1 - V_{sum}$) will be close to zero, and so there will be no further changes in conditioned strength for either stimulus: V_L will remain near 100, and V_T will remain at 0.

Conditioned Inhibition. Let us return to the simple case in which a single CS, L, is paired with one food pellet ($A_1 = 100$). Suppose there have been enough acquisition trials to bring V_L to a value of 90. Now, in the second phase of this experiment, a second CS, T, with a salience the same as L ($S_T = .2$) is presented in addition to L, but no food pellets are delivered. Figure 5-3 shows that according to the model, T should become a conditioned inhibitor during these extinction trials. The reason is that, despite the presence of T on these extinction trials, the US will be overpredicted because of the conditioned strength of L. The model states that if the US is overpredicted (that is, if A_j is less than V_{sum}), then the

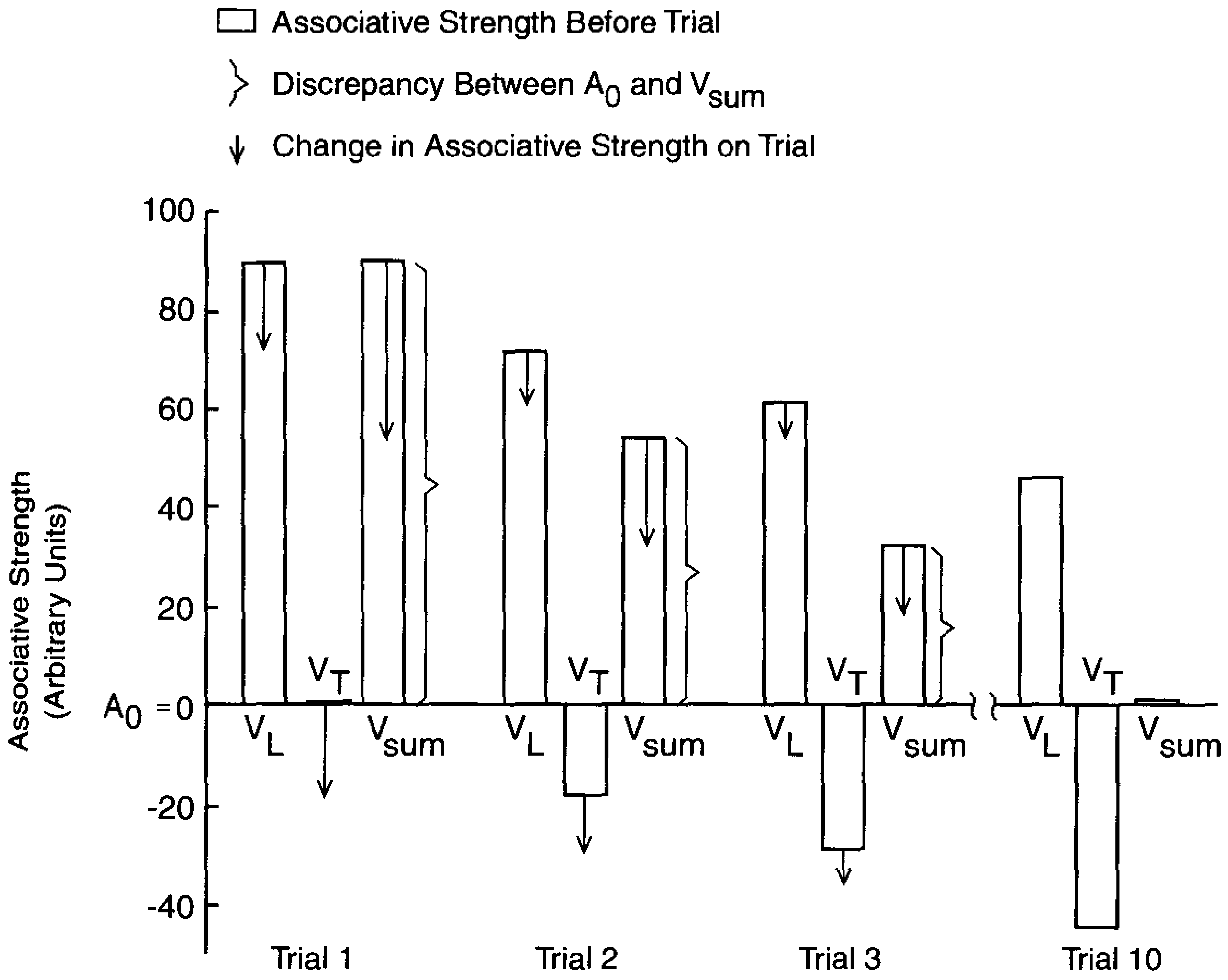

FIGURE 5-3 Predictions of the Rescorla-Wagner model for a case where T should become a conditioned inhibitor. Parameter values used were $A_0 = 0$, $S_L = .2$, $S_T = .2$, starting value for $V_L = 90$.

strength of all the CSs present on the trial will be decremented. Since V_T is initially 0, any decrements in strength will push V_T into the negative range, making it a conditioned inhibitor. To be more precise, on trial 1, the following equations apply:

$$\Delta V_L = .2 \times (0 - 90) = -18$$

$$\Delta V_T = .2 \times (0 - 90) = -18$$

That is, on this first extinction trial, V_L will lose 18 units of strength (from 90 to 72), and V_T will also lose 18 units (from 0 to −18).

The second trial shown in Figure 5-3 is the first example we have encountered in which both an excitatory CS and an inhibitory CS are present. With $V_L = 72$ and $V_T = -18$, Figure 5-3 shows that $V_{sum} = 54$. The following equations apply to the second extinction trial:

$$\Delta V_L = .2 \times (0 - 54) = -10.8$$

$$\Delta V_N = .5 \times (0 - 54) = -10.8$$

Therefore, on the second extinction trial, V_L will lose 10.8 units of strength (from 72 to 61.2), and V_T will lose 10.8 units (from −18 to −28.8). With additional extinction trials, V_L will become less positive, V_T will become more negative, and V_{sum} will approach an asymptote of zero. But notice that at this asymptote (which Figure 5-3 shows is nearly reached by trial 10), V_L retains an excitatory strength of approximately 50. The inhibitory strength of V_T is approximately −50, and because there is almost no discrepancy between A_0 and V_{sum}, there will be almost no further changes in the associative strength of either stimulus with additional extinction trials.

Summary. The Rescorla-Wagner model might be called a theory about US effectiveness: It states that an unpredicted US is effective in promoting learning, whereas a well-predicted US is ineffective. As the first formal theory that attempted to predict when a US will promote associative learning and when it will not, it is guaranteed a prominent place in the history of psychology. The model has been successfully applied to many conditioning phenomena, but it is not perfect. Some well-established phenomena are difficult for the model to explain. For this reason, other psychologists have proposed alternative theories of classical conditioning that are based on fairly different assumptions about the learning process. We will examine two types of alternative theories in the following sections.

Theories of CS Effectiveness

The main assumption of this class of theories is that the conditionability of a CS, not the effectiveness of the US, changes from one situation to another. A phenomenon called the **CS preexposure effect** provides one compelling piece of evidence for this assumption.

The CS Preexposure Effect. Consider a simple conditioning experiment with two groups of subjects. The control group receives simple pairings of one CS with a US. The only difference in the CS preexposure group is that before the conditioning trials, the CS is presented by itself a number of times. The comparison of interest concerns how quickly conditioned responding develops in these two groups. The common finding, which has been obtained with both animal and human subjects, is that conditioning proceeds more rapidly in the control group than in the CS preexposure group (Lubow & Moore, 1959; Lipp, Siddle, & Vaitl, 1992; Zalstein-Orda & Lubow, 1995). A common sense explanation of this result is that a sort of habituation occurs in the CS preexposure group—because the CS is presented repeatedly but initially predicts nothing, the subject gradually pays less and less attention to this stimulus. We might say that the subject learns to ignore the CS because it is not informative, and for this reason the subject takes longer to associate the CS with the US when conditioning trials begin and the CS suddenly becomes informative.

Although it is a well-established phenomenon, the Rescorla-Wagner model does not predict the CS preexposure effect. Let us examine what the model has to say about the first preexposure trial, on which the CS is presented by itself. Since there have been no prior conditioning trials, the CS elicits no expectation at all, and since no US is presented, the strength of the US is zero. Because the strength of the subject's expectation equals that of the US