Psychological Review
2002, Vol. 109, No. 3, 533–544

0033-295X/02/$5.00 DOI: 10.1037//0033-295X.109.3.533

Acquisition and Extinction in Autoshaping

Sham Kakade and Peter Dayan
University College London

C. R. Gallistel and J. Gibbon (2000) presented quantitative data on the speed with which animals acquire behavioral responses during autoshaping, together with a statistical model of learning intended to account for them. Although this model captures the form of the dependencies among critical variables, its detailed predictions are substantially at variance with the data. In the present article, further key data on the speed of acquisition are used to motivate an alternative model of learning, in which animals can be interpreted as paying different amounts of attention to stimuli according to estimates of their differential reliabilities as predictors.

In autoshaping experiments on pigeons, birds acquire a classically conditioned peck response to a lighted key associated, irrespective of their actions, with the delivery of food (Brown & Jenkins, 1968). As stressed persuasively by Gallistel and Gibbon (2000), there is substantial experimental evidence in favor of a simple quantitative relationship between the speed of acquisition in autoshaping and the three critical variables shown in Figure 1A. The first is *I*, the length of intertrial interval; the second is *T*, the time during the trial for which the conditioned stimulus (CS; a light in this case) is presented; and the third is the training schedule, 1/*S*, which is the fractional number of deliveries per light (for those birds that were only partially reinforced). Here, acquisition speeds are typically measured in terms of the number of trials it takes until a certain behavioral criterion is met, such as pecking during the time the light is illuminated on three out of four successive trials (Gallistel & Gibbon, 2000; Gibbon, Baldock, Locurto, Gold, & Terrace, 1977).

The data in Figure 1B show that the speed of acquisition is approximately inversely proportional to *I*/*T*. More precisely, the median number, *n*, of rewards that must be presented until the behavioral acquisition criterion is met is as follows:

$$n \approx 300\left(\frac{I}{T}\right)^{-1}. \qquad (1)$$

This implies that the relatively shorter the presentation of light, the faster the learning. Gallistel and Gibbon (2000) made two key points about this relationship. First, the number of trials until acquisition depends on the ratio of *I*/*T* and not on *I* and *T* separately—experiments reported for the same *I*/*T* are actually performed with *I* and *T* differing by more than an order of magnitude (Gibbon et al., 1977). Second, Figure 1C shows that partial reinforcement has almost no effect when measured as a function of the number of reinforcements (rather than the number of trials), because although it takes *S* times as many trials to acquire, there are reinforcements on only 1/*S* trials. This effect holds over at least an order of magnitude in *S*. Changing *S* does not change the effective *I*/*T* when measured as a function of reinforcements, so this result might actually be expected on the basis of Figure 1B.

Conversely, when rewards are no longer provided with the light, responding slowly stops, a process called *extinction.* The data show that about 50 rewards must be omitted before the satisfaction of an extinction criterion that the preextinction response rate be halved. The speed of extinction does not depend on the *I*/*T* ratio, by sharp contrast with the speed of acquisition.

These quantitative data provide a most seductive target for models of learning. Indeed, one of Gallistel and Gibbon's (2000) most important contributions is to place the behavior of the animals firmly in the domain of statistical normativity. Normative models suggest that the decisions of the animals to start and stop responding are actually correct according to a well-specified statistical model, given the information they have received about the relationship between stimuli and rewards, and prior expectations. Gallistel, Mark, King, and Latham (2001) showed directly that animals can be statistically optimal detectors of rate changes such as those that underlie autoshaping.

Gallistel and Gibbon (2000) suggested a model for acquisition and extinction data that they called *rate estimation theory* (RET). In RET, animals are estimating rates of reward delivery and start responding to the light when it reliably signals an increase in the reward rate over the context. Equally, in extinction, they stop responding when they have observed enough omitted rewards that they can be adequately certain the reward rate associated with the light has changed.

In this article, we argue that it is not possible to fit the normative statistical model underlying RET to the relevant acquisition data. Further, we construct a normative model that does match these quantitative data and also the relevant results of other experiments. In the new model, stimuli compete to predict the delivery of reward on the basis of estimates of the reliabilities with which they make predictions. Our model incorporates quantitative versions of

Sham Kakade and Peter Dayan, Gatsby Computational Neuroscience Unit, University College London, London, England.

Funding was provided by a National Science Foundation Graduate Research Fellowship awarded to Sham Kakade and by the Gatsby Charitable Foundation. We are most grateful to Randy Gallistel and John Gibbon for freely sharing, prior to publication, their many ideas about timing and conditioning and to Nathaniel Daw for extensive discussion. We also thank them and David Shanks for comments.

Correspondence concerning this article should be addressed to Sham Kakade, Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR, England. E-mail: sham@gatsby.ucl.ac.uk

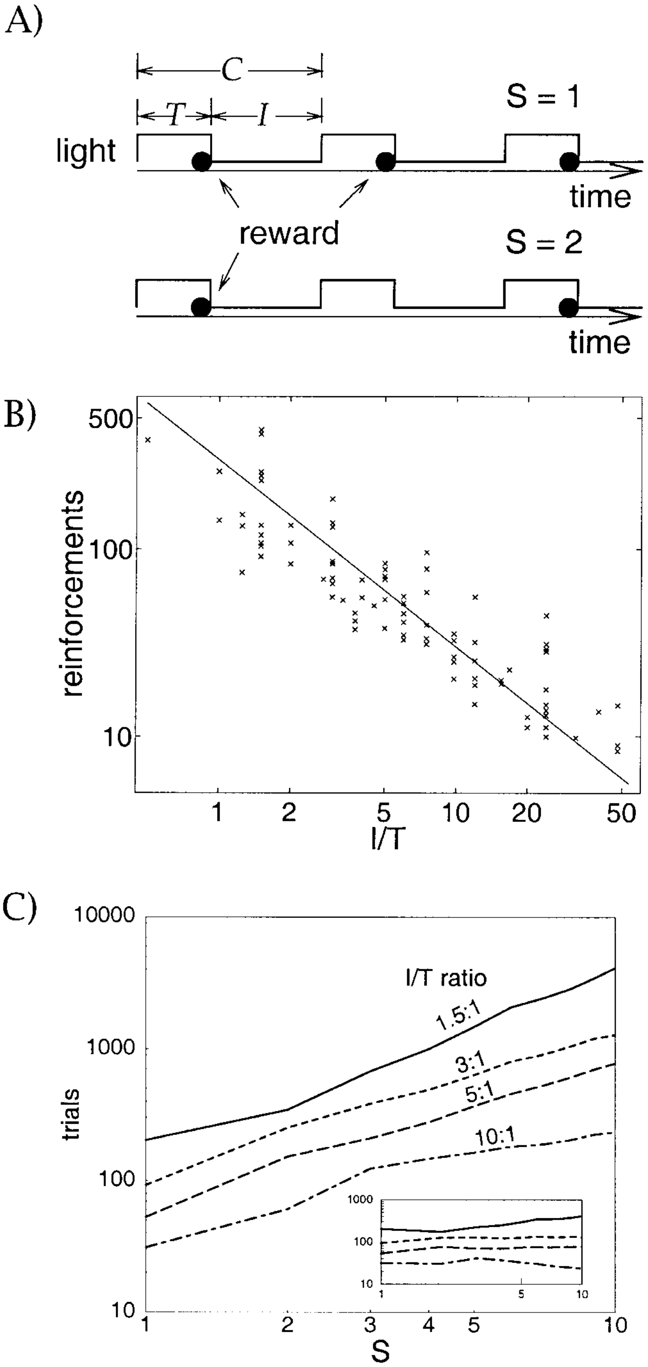

Figure 1. Autoshaping. A: Experimental paradigm. Top: The light is presented for *T* s every *C* s and is always followed by the delivery of food (filled circle). Bottom: The food is delivered with probability $1/S = 1/2$ per trial. In some cases, the length of the intertrial interval *I* is stochastic, with the appropriate mean. B: Log-log plot of the number of reinforcements to a given acquisition criterion versus the *I*/*T* ratio for $S = 1$. The data are median acquisition times from 12 different laboratories. A linear fit to log *I*/*T* is shown. C: Log-log acquisition curves for various *C*/*T* ratios and *S* values. The main graph shows trials versus *S;* the inset shows reinforcements versus *S.* From "Time, Rate, and Conditioning," by C. R. Gallistel and J. Gibbon, 2000, *Psychological Review, 107,* p. 298. Copyright 2000 by the American Psychological Association. Adapted with permission of the authors.

existing theoretical ideas about stimulus competition in classical conditioning (e.g., Grossberg, 1982; Mackintosh, 1975; Pearce & Hall, 1980).

Many of the key aspects of the data that motivated our model have been noted previously and, indeed, have motivated the critical experiments such as those on fast contextual conditioning by Balsam and Schwartz (1981).

Rate Estimation Theory

Gallistel and Gibbon (2000) are among the strongest proponents of the quantitative relationships shown in Figure 1. To account for them, Gallistel and Gibbon suggested that animals are estimating the underlying rates of reward in the world, that is, the number of rewards provided per unit time in the presence of the stimuli.

We call the "true" rate associated with the light λ_l, and that associated with the *background context* λ_b. The background context is the ever-present experimental chamber. RET uses an additive model for the delivery of reward, in that when the light is on, the true rate of reward is the sum $\lambda_l + \lambda_b$. With just the background alone, the rate is λ_b. The underlying rates are not directly observable by the animal, and so it faces the inference problem of estimating them based on the rewards and stimuli that are presented. Under RET, the rewards that are presented are allocated or apportioned between the light and the background, and the rates are estimated by dividing the total number of apportioned rewards by the total time of stimulus presentation.

To put this more formally in the context of the experiment of Figure 1, we can say that the rate of reward during the background is estimated by the mean rate $\bar{\lambda}_b = n_b/t_b$, where t_b is the cumulative exposure to the background alone and n_b is the number of rewards apportioned to the background during this time. In the experiment, no rewards are actually apportioned to the background, because their occurrence is tied to the light. However, it cannot safely be concluded that the background's mean rate is 0. Rather, this only justifies that the estimated mean rate is no higher than the reciprocal of the total exposure to the background, $\bar{\lambda}_b = 1/t_b$, and RET uses this conservative estimate for $\bar{\lambda}_b$. In *n* trials, $t_b = nI$. A standard Bayesian treatment of inference in these circumstances would lead to a similar conclusion.

Because RET uses an additive model for the delivery of reward, the rate of reward when the light is on (and the background is also present) is $\bar{\lambda}_l + \bar{\lambda}_b = n_l/t_l$, where n_l is the number of rewards delivered when the light is on and t_l is the cumulative exposure to the light. Here, if there is no partial reinforcement ($S = 1$), then $n_l = n$ and $t_l = nT$ (see Figure 1A). Thus,

$$\bar{\lambda}_l + \bar{\lambda}_b = \frac{n}{nT} = \frac{1}{T}, \quad \bar{\lambda}_b = \frac{1}{t_b} = \frac{1}{nI}. \tag{2}$$

Notice that as the number of trials grows, and so more rewards are observed when the light is on, the estimated rate with the light is constant, whereas the background rate continually drops as $1/n$.

Gallistel and Gibbon (2000) took the further important step of relating the rates λ_l and λ_b to the decision of the animals to start responding. Gallistel and Gibbon suggested that acquisition should occur when the animals have strong evidence that the fractional increase in the reward rate while the light is on is greater than some threshold. Thus, acquisition should occur when there is sufficient evidence that

$$\frac{\lambda_l + \lambda_b}{\lambda_b} > \beta, \tag{3}$$

where β is the threshold (which should be greater than 1). As the actual rates are unobservable, inferred estimates of them must be used to decide whether or not the criterion has been met. RET uses mean rates for this purpose, asking when

$$\frac{\bar{\lambda}_l + \bar{\lambda}_b}{\bar{\lambda}_b} > \beta. \tag{4}$$

Substituting the estimates given by Equation 2 gives $(1/T)/(1/nI) > \beta$, or equivalently,

$$n > \beta\left(\frac{I}{T}\right)^{-1}, \tag{5}$$

which is evidently linear. This conforms to the empirical data in Figure 1B and Equation 1, for $\beta \approx 300$.

Under a partial reinforcement schedule, the number of trials it takes to observe *n* rewards becomes *nS.* However, the total time that the light is observed during these trials and the total time the background is observed by itself during these trials both go up by the same factor of *S.* Thus, partial reinforcement should have no effect on the number of rewards it takes for the acquisition criterion to be satisfied. As Gallistel and Gibbon (2000) noted, the irrelevance of *S* should be expected on the basis of Equation 1, because changing *S* does not change the effective *I*/*T* when measured as a function of reinforcements. For the remainder of the article, we therefore no longer explicitly consider partial reinforcement.

RET suffers from two main problems, which motivated our search for an alternative. First, the value of $\beta = 300$ is inconsistent with data on the effects of presenting rewards when the background is present by itself (Jenkins, Barnes, & Barrera, 1981). Second, the model is silent on why acquisition should be dramatically faster if the context is extinguished prior to autoshaping (Balsam & Gibbon, 1988; Balsam & Schwartz, 1981). The value $\beta = 300$ can also be ecologically questioned, as it suggests an inordinate statistical conservatism.

First, under RET, rewards that are presented during the intertrial interval (when the light is off) are apportioned to the background, giving it a nonzero rate. Of course, this rate may nevertheless be lower than that for the light. In the case of Figure 2A, for instance, $\lambda_l = 9\lambda_b$, and acquisition is found empirically to occur in 30 rewards, which is comparable to the speeds shown in Figure 1 (Jenkins et al., 1981). Under RET, acquisition can only occur at all if $(\lambda_l + \lambda_b)/\lambda_b > \beta$, and so this experiment provides an upper bound of $\beta \leq 9$. This bound is greatly at variance from the estimate $\beta = 300$ that comes from Figure 1. To spell it out: Either, if $\beta = 300$, then we would not expect acquisition at all in the Jenkins, Barnes, and Barrera study, or, if $\beta = 9$, then the acquisition speeds shown in Figure 1 should be 30 times faster (as shown in Figure 2B). Even $\beta = 9$ is likely to be well above the actual threshold for response, because acquisition occurs relatively quickly for this *I*/*T* ratio compared with its speed for other ratios.

Second, more evidence for this same conclusion comes from cases in which the context is extinguished prior to autoshaping. In most experiments on autoshaping, animals are given prior experience of reinforcements in the context alone (usually in the form of hopper training) to help familiarize them with the environment. The seemingly simple manipulation of extinguishing the context prior to autoshaping produces extremely rapid learning to the light (Balsam & Gibbon, 1988; Balsam & Schwartz, 1981). More generally, the speed with which responding to the light is acquired is strongly affected by the provision of rewards with the background alone, prior to autoshaping (see Figure 3A; Balsam & Schwartz, 1981). Figure 3B shows the results of a study intended to examine this quantitatively. Balsam and Schwartz measured the acquisition speed during standard autoshaping as a function of the number of rewards provided in the context in the time before the light was introduced and autoshaping began (see Figure 3A). As a control, the context was first extinguished to erase associations formed in hopper training before the rewards were provided in the context and without the light. Reconciling the data in Figure 3 and Figure 1 (i.e., Equation 1), it would seem that about 30 prior rewards must have been given in the context prior to conditioning to the light (for the studies in Figure 1), which is roughly consistent with the experimental procedures used (Gibbon et al., 1977; Gibbon & Balsam, 1981).

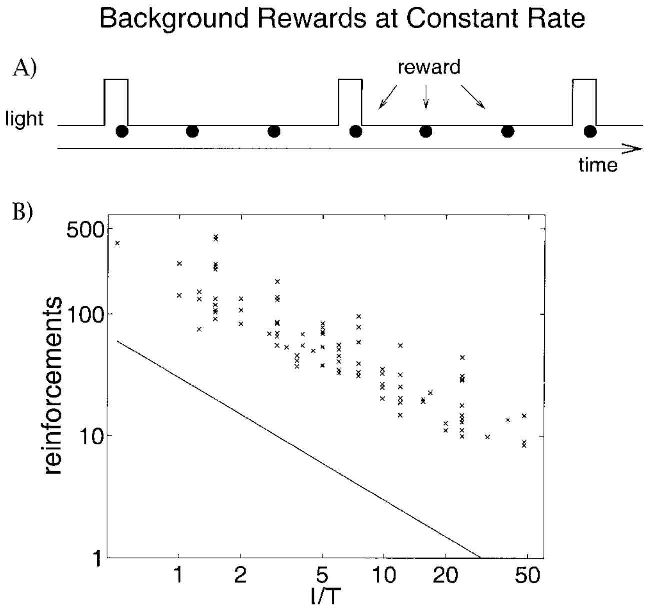

Figure 2. A: The figure is not drawn to scale. Rewards given in the intertrial interval (background rewards) at a constant rate of 1/75 s, while the rate with light is 1/8 s. Acquisition occurs in about 30 rewards for this case (Jenkins, Barnes, & Barrera, 1981). B: The threshold β must satisfy $\beta \leq 10$ for acquisition to occur in A. The solid line shows the predicted acquisition speed for $\beta = 10$ for various *I*/*T* values, and, for comparison, the crosses show the actual acquisition speeds as in Figure 1B. Acquisition is predicted to be 30 times too fast.

Equally important, and pointed out by Gibbon and Balsam (1981), is that the rate at which these prior context rewards are presented seems to have little effect on the subsequent speed of acquisition to the light. Gibbon and Balsam considered acquisition times from a different set of experiments in which about 60 rewards were given in hopper training, prior to autoshaping, in each case at different rates. The differences in the rates of prior rewards had no significant effect on the speed of subsequent acquisition.

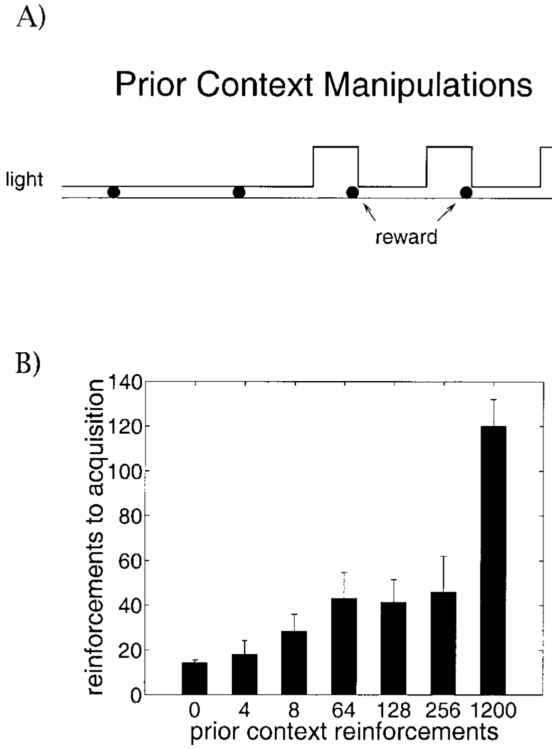

Figure 3. Context manipulations. A: Following extinction of the context (not shown), a fixed number of rewards is given in its presence, and without the conditioned stimulus. Subsequent autoshaping pairs the rewards with the light. B: The effects of such prior context reinforcements on subsequent acquisition speed. The data are taken from two experiments, both with $I/T = 6$, where I is the length of the intertrial interval and T is the time during the trial for which the conditioned stimulus is presented. (The data are from Balsam & Schwartz, 1981, p. 385, except for 0 and 1,200 prior context rewards, which are from Balsam & Gibbon, 1988, p. 407.) Vertical lines depict standard errors of the means.

These data pose a tricky conundrum for RET for two reasons. First, Gibbon and Balsam (1981) interpreted their data as implying that the rate associated with the context adapts quickly, so that the prior context rate does not affect the time it takes to acquire responding to the light. However, RET does not have this explanation open to it. Consider the case that the rates of delivery of reward before and after the light is introduced are the same. A natural mechanism in RET that allocates rewards between the light and the context would allocate all the rewards to the context, because no change in rate is consequent on the introduction of the light. Therefore, the animal should never acquire responding to the light. However, given other rates of delivery of reward to the light, RET would be forced to allocate rewards to the light, and so acquisition could happen. RET therefore predicts an essential dependence of acquisition on the rate at which the prior context rewards are delivered, contrary to Gibbon and Balsam's data. A similar argument can be made that, under RET, there should be a dependence of acquisition times on the absolute difference between the reward rates before and after the light is introduced. No such dependence is observed.

The second problem posed by the prior context rewards has to do with the strong dependence shown in Figure 3B on the number of rewards that are provided. Gibbon and Balsam's (1981) explanation that the speed of learning about the context is fast does not accommodate this aspect of the data. RET is similarly silent. Worse, however, under a statistical interpretation of RET, the dependence of acquisition rate on prior context rewards should be exactly the reverse of what is observed. To see this, consider the statistical effect of seeing more prior context rewards on the estimate of the rate of provision of rewards associated with the context. The more rewards that are observed, the less uncertain the subject should be about this rate (because the estimate of the rate will be based on more examples). Given a more certain prior rate of reward, a statistical test distinguishing this prior rate from a current rate of reward will be more powerful—thus, the more readily RET should be able to decide that the rate has changed, if it does so. Therefore, if the rate of reward changes when the light is introduced, the subject should be able to decide this after fewer trials. Consequently, it should more quickly decide to allocate the rewards to the light rather than the context and so acquire responding more quickly. Thus, under this interpretation of RET, the more prior context rewards, the faster acquisition should be to the light. The data in Figure 3B show exactly the opposite.

A further concern for RET comes from considering the certainty that the reward rate with the light is greater than that of the background at the time of acquisition. Consider a simple example with $T = 4$ s and $I = 20$ s, for which the animals take about 60 rewards to commence responding. Finding the reward during the 4-s light rather than the 20-s context is just like rolling a six-sided die and getting the answer 1 rather than 2–6, as $T/(I + T) = 1/6$. Observing the consistent relation between the light and reward and deciding that the light is associated with an increased reward rate over the background is like throwing the die, observing a 1 each time, and deciding that the die is unfairly biased. It would likely take us a maximum of 5 or 6 rolls to draw the conclusion that the die is loaded (with a chance of error of 10^{-4}). Taking 60 die rolls leaves a minuscule chance of around 10^{-45} that the die is fair. It could be, of course, that animals are extremely conservative in autoshaping compared with their much more reasonable inferential behavior in other paradigms. However, the data from the prior context manipulations (Figure 3B) show that the animals can make detections at more reasonable speeds. Collectively, these data suggest that RET's inferential model is incorrect and inspire a consideration of alternatives.

The Competitive Model

The problems just outlined with Gallistel and Gibbon (2000) form the constraints governing our new model. Furthermore—and this turns out to provide an important hint as to the construction of the model—we used the extra constraint that pecking rates should follow the form of data such as in Figure 4A. This plot shows the development of responding to the light over the course of conditioning. Gallistel and Gibbon's acquisition criterion is satisfied right at the beginning of learning (where the dotted line crosses the solid line); RET is mute on the approach of the rate of responding to its asymptote. Unfortunately, the latter phase of learning is not well explored experimentally and is often obscured by the use of a liberal measure of behavioral response (see the figure caption). For comparison, Figure 4B shows similar curves from the experiment of Balsam and Schwartz (1981) in which the number of prior

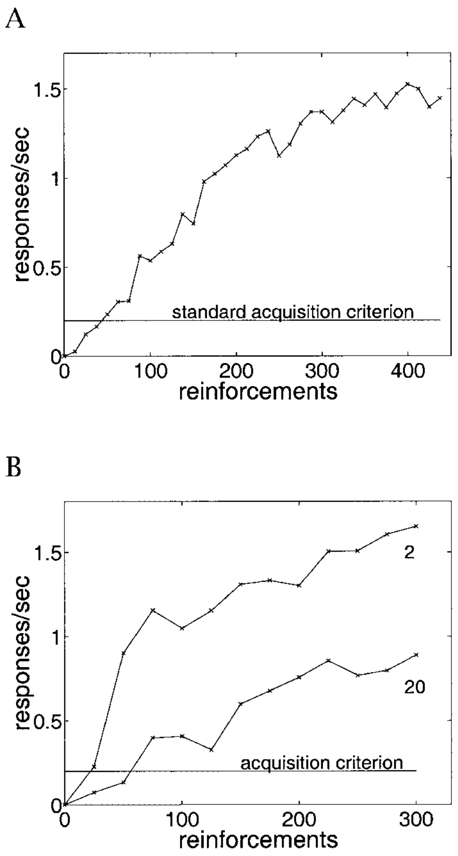

Figure 4. Acquisition of key pecking showing response rate versus reinforcements. A: The standard acquisition criterion is satisfied when the animal responds on three out of four consecutive trials when the response curve crosses the acquisition criterion line shown. Data are averaged from 11 animals, so the sharper transition from no conditioned response to the initial responding is smoothed. Most often these data are obscured by using a permissive measure of behavioral response such as the percentage of trials within a session in which at least one response occurs. The percentage of trials with at least one response often saturates, close to 100%, relatively quickly after the standard acquisition criterion is met. As shown in the figure, the animals usually continue to increase their response rates over a longer time scale. B: Longer term behavioral response for two different pretraining cases. The top curve is for 2 prior context rewards; the bottom curve is for 20 prior context rewards. Pretraining affects not only the short time to acquisition but also the slower phase of learning. Data are from Balsam and Schwartz (1981, p. 389). From "Associative Factors Underlying the Pigeon's Keypecking in Autoshaping Procedures," by E. R. Gamzu and D. R. Williams, 1973, *Journal of the Experimental Analysis of Behavior, 19,* p. 227. Copyright 1973 by the Society for the Experimental Analysis of Behavior. Adapted with permission.

rewards presented to a preextinguished context was controlled (to be 2 or 20).

In order to satisfy all these constraints, we take a different tack from RET. Our theory has two parts. The first part, described in the *Expert Predictions* section, is that each stimulus (the light and the context) is treated as an "expert," learning independently about the world and making an independent prediction of the rate of reward delivery. These predictions are based on a model that is a statistical generalization of standard conditioning theories such as the Rescorla–Wagner (1972) rule, is designed to be consistent with scalar expectancy theory (SET; Gibbon, 1977), and is closely related to well-understood statistical and engineering methods for prediction. The second part of the theory, described in the *Combining the Experts' Predictions* section, suggests that the predictions made by different stimuli should compete (Grossberg, 1982; Pearce & Hall, 1980) according to how reliable each stimulus is.

In our full model, learning is slow in standard autoshaping paradigms because the context acts as a more reliable expert and, under the model, blocks the expression of the prediction made by the less established light. Extinguishing the context puts it and the light on a more equal footing, allowing for much faster expression of the learning associated with the light.

Expert Predictions

In this section, we specify how learning works for each stimulus or expert by considering a simplified case that assumes that each predicts the total rate at which rewards are delivered while that stimulus is present. We start from the case that there is only one stimulus, namely, just the context, and consider how it predicts the reward rate, why this prediction is uncertain, and how the uncertainty can satisfy the constraints suggested by SET.

The model (Dayan, Kakade, & Montague, 2000; Kakade & Dayan, 2000) views the process of making predictions as one of reverse engineering the way that the experimenter has programmed the experimental apparatus, using the observations of rewards provided. That is, at time t, the subjects consider there to be a parameter, $\lambda_c(t)$, under the control of the experimenter, that governs the rate at which rewards will actually be delivered to the animal in the context. For convenience, we suppress the time dependence where the dependence is clear. Here, the rate is just the probability per unit time of delivery of reward. This rate is a function of time, to reflect the possibility that the experimenter might change the reward contingencies over the course of the experiment. Rewards are presumed to be delivered statistically independently of each other.

The subject has to use the actual deliveries of rewards in order to figure out λ_c as best it can and thereby predict the current relationship between the context and the rewards. Obviously, the subject's prediction, which we call $\hat{\lambda}_c$, can be uncertain. For instance, when the subject first encounters the context, it does not know whether it is to be given a low or high rate of reward. This uncertainty should be reduced as it makes observations about the actual reward rate. Appendix A provides a full treatment of the model. We consider just a didactic example, called a *window model,* which has similar properties but is easier to understand. Unlike the full model, the window model uses the notion of a trial, and the estimate, $\hat{\lambda}_c$, is the observed rate of delivery of the last η rewards (the "window"). That is, $\hat{\lambda}_c = \eta/t_\eta$, where t_η is the length of time before t that it took for the previous η rewards to be delivered. Note that this estimate will adapt to contingency changes, because no information from before t_η is being used to construct the estimate. The estimate $\hat{\lambda}_c$ is based on η individual reward times. If the reward rate is actually fixed at $1/C$ during this time (where $C = I + T$ is the total length of a trial), then it is a standard result that the estimate $\hat{\lambda}_c$ is within a small factor of $1/C$ and the standard deviation of the estimate is within a small factor of $1/(C\sqrt{\eta})$. This means that the coefficient of variation of the estimate, which is the standard deviation of the estimate divided by its mean, is roughly $1/\sqrt{\eta}$, independent of the underlying rate λ_c.

The window model (and the full model of Appendix A) shows how the subject might construct an estimate $\hat{\lambda}_c$ of the reward rate and that this estimate is uncertain to the tune of $1/\sqrt{\eta}$. This form of the uncertainty was inspired by SET, which notes that the uncertainty of estimates of time intervals has an asymptotically constant coefficient of variation. That is, no matter how many trials are measured, subjects do not become arbitrarily certain about time intervals but rather have an uncertainty that satisfies an interval-independent relationship. The window model achieves the same result for the rates by measuring the time for a fixed number of rewards. The asymptotic uncertainty comes from the possibility that the experimenter might change the rate, invalidating the information from very old trials. Here, η determines how many trials are included in the total and so should be set in line with the expected rate of change of λ_c. A small value of η will lead to quickly adapting estimates. The price to be paid for a small η is that each estimate is based on only a few rewards, and so the estimates are highly variable.

Because the light and the context are treated as making independent predictions, we similarly have a prediction $\hat{\lambda}_l = 1/T$ with asymptotic standard deviation $\sigma_l = 1/(T\sqrt{\eta})$. What remains to be specified is how the conflicting estimates $\hat{\lambda}_c$ and $\hat{\lambda}_l$ are combined, as both provide predictions when the light is on.

Combining the Experts' Predictions

RET makes an assumption of additivity, that is, that the estimates of reward rates of all stimuli present are added. By contrast, our model assumes instead that the stimuli compete. This idea is conventional in classical conditioning; however, it is most commonly applied to how much learning should be accorded to stimuli (Mackintosh, 1975; Pearce & Hall, 1980) rather than how predictions from different stimuli should be combined (but see Grossberg, 1982).

When only the context is present, that is, in the absence of competition, the prediction of the mean overall reward rate is $\hat{\lambda}_c^* = \hat{\lambda}_c$. However, when both the light and the context are present, we consider a model in which the joint prediction is a weighted average of the individual predictions rather than the sum. As an analogy, consider the task of estimating some fictitious value χ. Expert 1 tells us the value of χ is χ_1, whereas Expert 2 tells us the value is χ_2. In constructing an estimate of χ, we must average the two estimates based on how much we trust each expert. If we have equal trust in both experts, then our estimate of χ should just be the equally weighted average, $(1/2)\chi_1 + (1/2)\chi_2$. If we know Expert 2 is highly unreliable, then our estimate should give little weight to Expert 2 and should be roughly χ_1.

When both the light and the context are present, the combined prediction of reward, which we call $\hat{\lambda}_l^*$, is the weighted average

$$\hat{\lambda}_l^* = \pi_l\hat{\lambda}_l + (1 - \pi_l)\hat{\lambda}_c, \tag{6}$$

where $0 \leq \pi_l \leq 1$ controls the degree of competition between light and context. We use an asterisk to denote the joint prediction made by all present stimuli. Again, we are suppressing the time dependence of the weighting π_l and the rates.

There are various ways that π_l might be determined (Dayan & Long, 1997; Kruschke, 1997). We (Dayan et al., 2000; Dayan & Long, 1997) have used the factorial experts model suggested by Jacobs, Jordan, and Barto (1991). In this, π_l is derived from underlying quantities called reliabilities, ρ_c and ρ_l for the context and light, respectively, as

$$\pi_l = \frac{\rho_l}{\rho_l + \rho_c}. \tag{7}$$

The term *reliability* is used because of the statistical roots of the rule (Dayan et al., 2000; Dayan & Long, 1997; Jacobs, 1995). One can interpret $1/\rho_l$ as the expected distance of the true rate associated with the light from the true overall rate—the more reliable the light as a predictor, the larger ρ_l and the smaller the distance. Note how the prediction made by the context can block the prediction made by the light, provided that it is much more reliable ($\rho_c \gg \rho_l$). This is a *representational* form of blocking (Grossberg, 1982) rather than a learning form of blocking.

Unreliabilities and uncertainties are different, though related. Even if the subject was completely certain about $\hat{\lambda}_l$, it could accord the light little reliability as a predictor, on the basis of past experience, and so have a small value of ρ_l. To be strictly accurate, the uncertainties in the estimates $\hat{\lambda}_c$ and $\hat{\lambda}_l$ should decrease the terms ρ_c and ρ_l in Equation 7, but this is typically only a small correction. Ideally, we would be able to specify a statistically normative model governing the setting of the reliabilities. Unfortunately, as becomes apparent in the next section, there is presently little evidence about the constraints on the model (unlike, for instance, the evidence from SET about the uncertainties). Therefore, we content ourselves with a phenomenological model for them.

Modeling Acquisition and Extinction

We can now combine the two parts of the model to account for the data on acquisition and extinction. Crudely, the predictions associated with the context and the light both adapt quickly, within $\eta = 25$ rewards. There is thus no blocking in the learning of the prediction of the light because of the prediction made by the context. However, the expression of the prediction associated with the light happens slowly, because the context, particularly in the face of hopper training, is treated as being more reliable. We do not present a normative account of how the reliabilities change, because of a lack of data on the slower phase of learning. Rather we show that the model is capable of fitting the data with an assumption of how the reliabilities do change. During extinction, the reliability of the light is almost constant, so the speed is determined just by η.

Acquisition

A fully normative model would come from a statistically correct account of how the reliabilities should change over time. This, in turn, would come from a statistical model of the expectations the animal has of how the predictabilities of stimuli and rewards change in the world. The best data for this come from the slow phases of learning in Figure 4, because it is during this period that the light is coming to be treated as more reliable. Unfortunately, the slow phase is widely ignored in experiments. Further, as Gallistel and Gibbon (2000) pointed out, the existing data presented are averaged over subjects, thus obscuring the behavior of individual subjects. We are therefore forced to make an assumption to fit the acquisition data. We state this assumption in terms of the combination weights $\pi_1(n)$ rather than the reliabilities themselves. Here, *n* indexes the time the *n*th reward is presented, and, for simplicity, $\pi_1(n)$ is assumed constant between rewards.

Rather than using something purely arbitrary, we assume that the animal's response rate during the slow, postacquisition phase of learning follows the estimated reward rate. As can be seen in Figure 4, the response rate is sigmoidal and so is nicely modeled by a tanh function. If we work backward, for this to be the form of the estimated reward rate, the relative responsibility $\pi_1(n)$ should follow

$$\pi_1(n) = \tanh \pi_0 n, \tag{8}$$

where π_0 is a constant that is independent of *I*/*T* (which is essential for the model to fit the data). Figure 5 shows the comparison of the rate of key pecking in the model with the data from Figure 4A. Slow acquisition to the light comes from slow changes in the importance accorded to the long established predictions made by the light.

A decision criterion for a subject turns its estimates of the rates of reward into a time at which it starts responding reliably. Following Gallistel and Gibbon (2000), we assume that responding commences when subjects expect a sufficiently higher response rate with than without the light, that is, when they have good reason to believe that

$$\lambda_l^* > \beta \lambda_c^*. \tag{9}$$

From the *Expert Predictions* section, the estimated means of these rates are

$$\hat{\lambda}_c^* = \frac{1}{C}, \quad \hat{\lambda}_l^* = \frac{1}{C} + \pi_1(n)\left(\frac{1}{T} - \frac{1}{C}\right). \tag{10}$$

If we use these in the decision criterion (which is a good approximation), the threshold value becomes

$$n > \left(\frac{\beta - 1}{\pi_0}\right)\left(\frac{I}{T}\right)^{-1}, \tag{11}$$

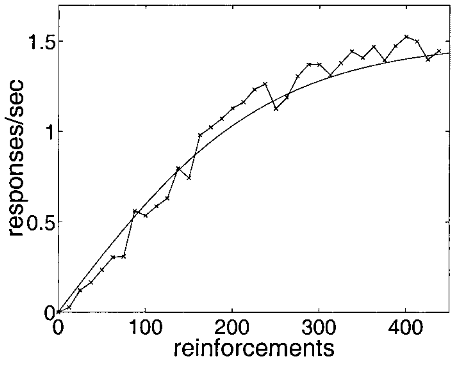

Figure 5. Fit to the behavioral response curve (Figure 4A), using Equation 8 with the constant $\pi_0 = 0.0004$. Here, the response to the light is modeled as being proportional to the increased reward rate with the light, that is, the difference between the estimated reward rate with the light, $\hat{\lambda}_l^*(n)$ and the estimated reward rate with the context, $\hat{\lambda}_c^*(n)$. sec = seconds.

because $\pi_1(n) \approx \pi_0 n$ for early *n*, which gives the correct linear dependency on *I*/*T*. Note that the constant of proportionality is $(\beta - 1)/\pi_0$ rather than the factor of β predicted by RET (Equation 5). Comparing Equations 1 and 11, we see that $(\beta - 1)/\pi_0 = 300$. For π_0 obtained by fitting the postacquisition behavioral response curve (see Figure 4A), this gives a value of $\beta \approx 2$. This is a more reasonable value of β that is not inconsistent with the data on the provision of rewards during the context (Jenkins et al., 1981) and implies that detection will occur for almost all reward rates associated with the light that are larger than those for the context.

The data on preexposure of the context show that the number of rewards provided before the light is shown exerts a strong effect on the speed of acquisition. In the model, these prior rewards serve to increase the reliability of the context in the early stages of conditioning. From Equation 7, because $\rho_l(n)$ is small, the effect of this on the weight $\pi_1(n)$ of the light is to change the slope π_0. Figure 6A shows the effect on the predicted behavioral response of varying π_0. The center dashed line is the same as the solid line in Figure 5. The numbers labeling the curves are the values of π_0 as multiples of the value of π_0 for the center curve. This figure shows that by varying the initial weight by a multiplicative factor of between 0.3 and 3, the acquisition speeds (judged at the criterion line shown) due to between 1,200 and 0 prior context rewards can be obtained, making the model consistent with the data in Figure 3.

Prior context manipulations also strongly affect the postacquisition response curves as shown in Figure 4B. Figure 6A shows the strong effects on the long-term response curves of varying π_0. Figure 6B fits the overall response curves of Figure 4B (using $\pi_0 = 0.0023$ and $\pi_0 = 0.0085$). Note that the lower curve looks substantially far from its asymptote. These different values of π_0 are assumed to come from manipulations to the reliability of the context and are affected by extinguishing the context and providing precisely controlled numbers of rewards in the context. The rate at which the prior context rewards are provided has little effect (Gibbon & Balsam, 1981), because the speed of adaptation to the new rate of the prediction associated with the context is fast (within around $\eta = 25$ rewards) compared with the speed of change of the reliabilities.

Extinction

Figure 7 shows experimental data indicating that the number of reinforcements that must be omitted, $\bar{n}$, to reach an extinction criterion that the rate in responding to the light should halve, satisfies

$$\bar{n} \approx 50. \tag{12}$$

In striking contrast to acquisition, the *I*/*T* ratio has no effect on extinction.

Because our model of acquisition explicitly allows contingencies to change over time, it lends itself naturally to extinction. In the model, the prediction of the current reward rate with the light constantly decreases when the light is no longer reinforced. Consider associating a decline to 50% of the preextinction rate of responding as occurring when the current rate estimate of the animal has decreased to being of the order of 50% of its preextinction estimate. We make the crucial assumption that the reliability of the light does not change significantly over the early phase of extinction.

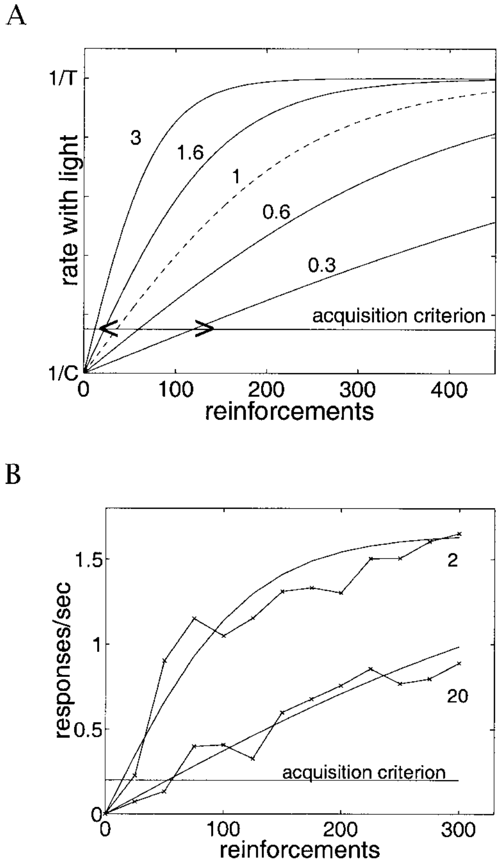

Figure 6. Satisfaction of the constraints. A: Possible acquisition curves showing the estimated reward rate with the light, $\hat{\lambda}_l^*(n)$, versus *n. I* is the length of the intertrial interval, *T* is the time during the trial for which the conditioned stimulus is presented, and $C = I + T$. The dashed curve is the same as in Figure 5. The parameters displayed are values for the constant π_0 in Equation 8 in multiples of π_0 for the center curve. As only π_0 is varied and not *I* or *T*, the speed of acquisition can be measured when the curve crosses the acquisition criterion line shown. The ↔ on the criterion line denotes the range of acquisition speeds (between 15 to 120 reinforcements) shown in Figure 3, as due to between 0 and 1200 prior context rewards. B: The postacquisition response curves of Figure 2B are fit just as in Figure 5. Here, $\pi_0 = 0.0023$ for the top curve and $\pi_0 = 0.0085$ for the bottom curve. sec = seconds.

Consider the simple window model in which rates are assessed by calculating the total time that elapsed for the last η rewards (extinction in the full model is treated in Appendix B). According to this measure, the preextinction reward rate is $\hat{\lambda}_l^{\text{pre}} = 1/T$. After $\bar{n}$ omitted rewards during extinction (i.e., at time $\bar{n}T$ after the beginning of extinction), the postextinction reward rate is $\hat{\lambda}_l(\bar{n}) = \eta/(\eta + \bar{n})T$, because the last η rewards occurred in time $\eta T + \bar{n}T$. The ratio between $\hat{\lambda}_l^{\text{pre}}$ and $\hat{\lambda}_l(\bar{n})$ is 1/2 when

$$\bar{n} = \eta, \tag{13}$$

which shows that the number of omitted rewards until extinction is approximately the number of remembered rewards η. This is clearly independent of *I*/*T*. Note that measuring the window size by rewards, itself chosen in the light of SET, is crucial to obtain this correct dependency. Equations 12 and 13 can be combined to provide an estimate based on extinction of $\eta \approx 50$ that can be compared with the estimate based on acquisition. This provides an independent check on the model. This value is a little higher than that implied by the data on acquisition. One potential source of error is the assumption that there is a direct relationship between the behavioral response and the estimated association between light and reward. If the 50% decline in response occurs when the current rate is about 30% of its preextinction rate, then we obtain $\eta \approx 25$, which is just the same as that for acquisition. The general dependence of this fractional drop of the estimated rate is discussed in Appendix B.

With $\eta \approx 25$, the same set of parameters are used to model both acquisition and extinction. However, for acquisition, the crucial parameter is π_0, which reflects the speed at which the reliabilities change, whereas for extinction, the crucial parameter is η, which reflects the speed at which the estimated contingency between the light and reward could change.

Discussion

The speeds with which animals acquire and extinguish conditioned responses in autoshaping conform to a set of simple quantitative relationships. Normative models suggest that such relationships arise when animals make *optimal* inferences based on their underlying statistical assumptions. Different statistical assumptions (which amount to different ecological expectations) lead to different normative models and different predictions about the speed of acquisition and extinction.

Gallistel and Gibbon (2000) were the first to suggest a normative account for the autoshaping data. Unfortunately, although their model correctly captures the nature of the dependency of the speed of learning on various parameters, its quantitative predictions on the speed of acquisition and extinction are inconsistent with the data. Further, the model has little to say about longer term behavior during acquisition and extinction or about the effects of extinguishing the context prior to the initiation of autoshaping.

We have suggested an alternative normative model to account for both the data underlying Gallistel and Gibbon (2000) and the results of other experiments. Although our model's focus is on acquisition and extinction, it is governed by a rich set of quantitative constraints. First, because animals can be ideal detectors of rates in some circumstances (e.g., Gallistel, Mark, King, & Latham, 2001), we required an account under which their acquisition of responding could be given a rational statistical basis. In this sense, the rate of responding during learning should be based on an optimal evaluation of the contingencies given the observations so far provided. Second, as in Gallistel and Gibbon (2000), the number of reinforcements to acquisition should be governed by the relationship $n \approx 300(1/T)^{-1}$ as shown in Figure 1B. This also implies that *S*, the partial reinforcement schedule, should be irrelevant. Third, pecking rates after the acquisition criterion is satisfied should rationally follow the form of Figure 4. However, as insufficient data exist on this phase of learning, our model makes specific assumptions to capture this constraint—and thus falls short of providing a completely normative account. Fourth, the overall learning speed should be strongly affected by the number of prior context rewards (see Figure 2B), but not by the rate at which they are presented. That is, regardless of the rate it predicts, the context, as an established predictor, should be able substan-

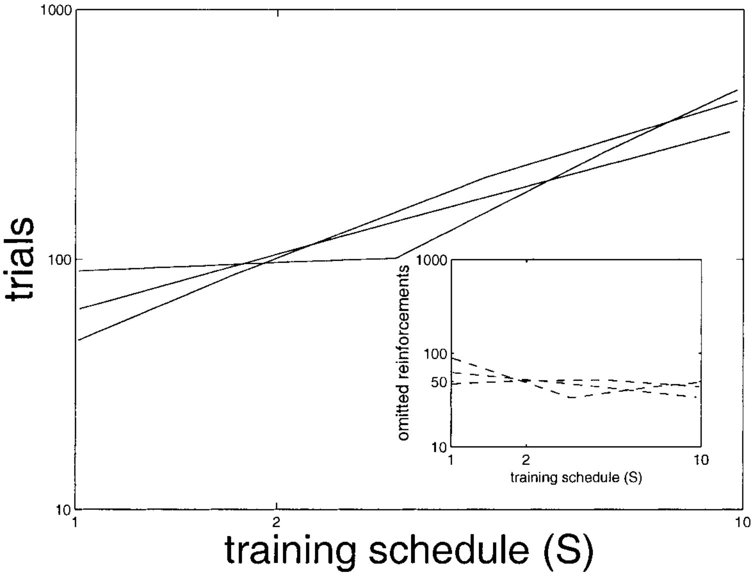

Figure 7. Log-log extinction curves for various *I*/*T* ratios, where *I* is the length of the intertrial interval and *T* is the time during the trial for which the conditioned stimulus is presented, and values of *S*. The main graph shows trials to extinction versus *S* (different curves are for different *I*/*T* ratios); the inset shows omitted reinforcements versus *S*. This shows that the number of omitted reinforcements until the extinction criterion is approximately 50. *I*/*T* ratios go from 1.5:1 to 25:1. From "Time, Rate, and Conditioning," by C. R. Gallistel and J. Gibbon, 2000, *Psychological Review, 107,* p. 305. Copyright 2000 by the American Psychological Association. Adapted with permission of the authors.

tially to block a less established predictor. Fifth, the rate estimates must not become arbitrarily accurate, and their asymptotic uncertainty should be consistent with SET.

Our model is built on a set of statistical assumptions that take explicit account of two central concerns for estimation, uncertainty and unreliability. Uncertainty, also a facet of scalar expectancy theory, comes because information from observations of reward is traded off against the possibility that the reward rates might change. Uncertainty governs the observed rate of extinction. Unreliability, a higher order property of a stimulus, comes as part of an adaptive filtering strategy and changes only slowly. The most unsatisfying aspect of the model is the lack of a normative account of how unreliabilities should change—this is largely due to a lack of experimental data on the slow phase of responding.

The model of unreliability derives from Dayan and Long (1997), who, following Grossberg (1982), used it to account for data on a set of paradigms, including downward unblocking (e.g., Holland, 1988). Dayan and Long's model does not conform to the quantitative timing relationships studied here. Kruschke (1997) suggested an alternative competitive account derived from a different statistical model called the *mixture of Gaussians model* (Jacobs, Jordan, Nowlan, & Hinton, 1991). Models such as ours and Kruschke's, in which stimuli are perfect competitors for each other and their predictions are combined using a weighted average, lie at one end of a spectrum and capture these blocking effects. At the other end of the spectrum lies the Rescorla–Wagner rule (and also the rate additivity assumption of RET), in which stimuli are perfect cooperators and their predictions are simply summed. Additive models cannot simply account for paradigms such as downward unblocking (Holland, 1988). However, there is also experimental evidence in favor of additive combination, particularly in paradigms involving conditioned inhibition (see Rescorla, 1988) or those involving signaled background reinforcers (Durlach, 1983), which our competitive models cannot simply account for. Understanding the rules governing competitive and additive combination is a major task for future theoretical work.

The model is also incomplete. For instance, data from Balsam and Schwartz (1981) suggest the possibility of a sustained difference in maintenance response rates as a function of the amount of pretraining. Our model leaves such a possibility open depending on the asymptotic value of the weight, π_l, of the light's prediction. Equation 8 assumes that the asymptotic value of the combination weight is one, which implicitly assumes that the context is treated as being asymptotically much less reliable than the light (and so the expression of its prediction is completely blocked). If this is not true, then sustained differences in responding might occur, as the combination weight's asymptote will be lower. There are three further complications in interpreting the long-run, maintenance key-peck rates. First, Figure 5 and Figure 6B show that the number of trials before asymptotic response behavior is apparent may be extremely long, outside the realm of many experiments. Second, data from Gibbon et al. (1977) argue that maintenance key-peck rates depend on *T* more strongly than on *I*. This may also play a role in the results of some autoshaping experiments (e.g., Holland, 2000) in which the linear dependence of acquisition times to *I*/*T* is not valid across as wide a range as is evident in Figure 1, B and C. Third, Pearce and Collins (1987) and Swan and Pearce (1987) have suggested that these rates also reflect an orienting response to the light whose strength depends on the accuracy with which its consequences can be predicted. This last effect is of significance primarily in the cases of partial reinforcement (for which $S > 1$). Also, effects of changing the magnitude of reward (which can take various forms; see, e.g., Allan & Zeigler, 1994; Balsam & Payne, 1979; Ploog & Zeigler, 1996) are not modeled.

Finally, we have so far not considered the strong evidence that the animals can perform interval timing, that is, predicting the time after the illumination of the light that the reward will be delivered. This capacity is evident in the average pattern of responding, as well as from other experiments such as the peak procedure (see Gallistel & Gibbon, 2000, for a discussion). This implies that the animals may be making predictions not about a single reward rate during the light but rather about multiple reward rates during separate portions of the light.

Other experiments could test the assumptions of the model. For instance, one expectation is that the asymptotic prediction variance, which is closely tied to the learning rate, should depend on the expected speed of change in the environment. It would be interesting to present animals with a series of autoshaping tasks with different values of I/T, changing between them either quickly or slowly. We would predict that the animals should show fast and slow adaptation to the rates accordingly. There is some evidence for this sort of "metalearning" in other cases (e.g., Krebs, Kacelnik, & Taylor, 1978).

References

Allan, R. W., & Zeigler, H. P. (1994). Autoshaping the pigeon's gape response: Acquisition and topography as a function of reinforcer type and magnitude. *Journal of the Experimental Analysis of Behavior, 62,* 201–223.

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering.* Englewood Cliffs, NJ: Prentice Hall.

Balsam, P. D., & Gibbon, J. (1988). Formation of tone–US associations does not interfere with the formation of context–US associations in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 401–412.

Balsam, P. D., & Payne, D. (1979). Intertrial interval and unconditioned stimulus durations in autoshaping. *Animal Learning & Behavior, 7,* 477–482.

Balsam, P., & Schwartz, A. L. (1981). Rapid contextual conditioning in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 382–393.

Brown, P. L., & Jenkins, H. M. (1968). Autoshaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior, 11,* 1–8.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience, 3,* 1218–1223.

Dayan, P., & Long, T. (1997). Statistical models of conditioning. *Neural Information Processing Systems, 10,* 117–124.

Durlach, P. J. (1983). Effect of signaling intertrial unconditioned stimuli in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 374–389.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107,* 289–344.

Gallistel, C. R., Mark, T. S., King, A., & Latham, P. E. (2001). The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior, 27,* 354–372.

Gamzu, E. R., & Williams, D. R. (1973). Associative factors underlying the pigeon's keypecking in autoshaping procedures. *Journal of the Experimental Analysis of Behavior, 19,* 225–232.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84,* 279–325.

Gibbon, J., Baldock, M. D., Locurto, C., Gold, L., & Terrace, H. S. (1977). Trial and intertrial durations in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 264–284.

Gibbon, J., & Balsam, P. (1981). Spreading associations in time. In C. M. Locurto, H. S. Terrace, & J. Gibbon (Eds.), *Autoshaping and conditioning theory* (pp. 219–254). New York: Academic Press.

Grossberg, S. (1982). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review, 89,* 529–572.

Holland, P. C. (1988). Excitation and inhibition in unblocking. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 261–279.

Holland, P. C. (2000). Trial and intertrial durations in appetitive conditioning in rats. *Animal Learning & Behavior, 28,* 121–135.

Jacobs, R. A. (1995). Methods for combining experts' probability assessments. *Neural Computation, 7,* 867–888.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science, 15,* 219–250.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3,* 79–87.

Jenkins, H. M., Barnes, R. A., & Barrera, F. J. (1981). Why autoshaping depends on trial spacing. In C. M. Locurto, H. S. Terrace, & J. Gibbon (Eds.), *Autoshaping and conditioning theory* (pp. 255–284). New York: Academic Press.

Kakade, S., & Dayan, P. (2000). Acquisition in autoshaping. In S. A. Solla, T. K. Leen, & K.-R. Muller (Eds.), *Advances in neural information processing systems* (Vol. 12, pp. 24–30). Cambridge, MA: MIT Press.

Krebs, J. R., Kacelnik, A., & Taylor, P. (1978, September 7). Test of optimal sampling by foraging great tits. *Nature, 275,* 27–31.

Kruschke, J. K. (1997, November). *Relating Mackintosh's (1975) theory to connectionist models and human categorization.* Paper presented at the Eighth Australasian Mathematical Psychology Conference, Perth, Australia.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

Pearce, J. M., & Collins, L. (1987). An evaluation of the associative strength of a partially reinforced serial CS. *Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology, 39,* 273–293.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variation in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review, 87,* 532–552.

Ploog, B. O., & Zeigler, H. P. (1996). Effects of food-pellet size on rate, latency, and topography of autoshaped key pecks and gapes in pigeons. *Journal of the Experimental Analysis of Behavior, 65,* 21–35.

Rescorla, R. A. (1988). Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience, 11,* 329–352.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: Vol. 2. Current research and theory* (pp. 64–69). New York: Appleton-Century-Crofts.

Swan, J. A., & Pearce, J. M. (1987). The influence of predictive accuracy on serial autoshaping: Evidence of orienting responses. *Journal of Experimental Psychology: Animal Behavior Processes, 13,* 407–417.

Appendix A

The Kalman Filter Model

In our model, the subject attempts to estimate the parameters governing the experimental situation based on observations. Although the most natural parameters are the rates $\lambda_c(t)$ and $\lambda_l(t)$ themselves, we consider a model couched in terms of the inverse rates $s_l(t) = 1/\lambda_l(t)$ and $s_c(t) = 1/\lambda_c(t)$, which are the times between rewards. We have confirmed with simulations that this does not affect our conclusions. Because the light and context act independently, we consider just the context and drop the subscript. Because of the scalar property enshrined in SET, we assume that we measure estimation time in terms of numbers of rewards rather than clock time and so write $s(n)$ rather than $s(t)$.

To specify the model, we have to indicate how the current value of the inverse rate $s(n)$ leads to an observed interval $o(n + 1)$ between rewards (this is called the *output model*) and then how the inverse rate can change (the *dynamic model*). We make simple Gaussian assumptions:

$$o(n+1) \sim G\big(s(n),\, s(n)^2\big) \quad \text{(the output model)}, \tag{A1}$$

$$s(n+1) = s(n) + \epsilon(n) \quad \text{(the dynamic model)}, \tag{A2}$$

$$\epsilon(n) \sim G\big(0,\, (\sigma s(n))^2\big) \quad \text{(the fluctuation model)}. \tag{A3}$$

The scalar property governs the form of the output model (Equation A1) and also the variance of the fluctuations $\epsilon(n)$ in the dynamic model. Here, $\sigma < 1$ is a unitless parameter that sets the scale for how fast the rates can change. Strictly speaking, $s(n + 1)$ is truncated at 0, so that the inverse rate is prevented from being negative, but this is in any case unlikely.

If the variances in Equations A1 and A3 were constant, then a completely standard Kalman filter (Anderson & Moore, 1979) would exactly specify the probability distribution of the current estimate of $s(n)$, in light of the above equations and the observed rewards. We make the extra approximation of using the standard Kalman filter but estimating the variances using just the means, $\hat{s}(n)$, and not worrying about the recursive effects of uncertainty in $\hat{s}(n)$ on the variance. Because the distribution for $s(n)$ is Gaussian, the Kalman filter only specifies update rules for the mean $\hat{s}(n)$ and variance $v(n)$ of the estimate of $s(n)$:

$$\hat{s}(n+1) = \hat{s}(n) + \alpha(n)\big(o(n) - \hat{s}(n)\big), \tag{A4}$$

$$v(n+1) = \big(v(n) + (\sigma\hat{s}(n))^2\big)\big(1 - \alpha(n)\big). \tag{A5}$$

Notice that the update rule for $\hat{s}(n)$ is a delta rule, just like a Rescorla–Wagner (1972) rule with $\alpha(n)$ acting as the learning rate parameter. This learning rate is given by the following:

$$\alpha(n) = \frac{\sigma^2 + v(n)/\hat{s}(n)^2}{\sigma^2 + 1 + v(n)/\hat{s}(n)^2}. \tag{A6}$$

The dependence of the learning rate on the output and fluctuation variances is necessary to match the speed with which the world is expected to change and the degree of certainty in the current estimate.

In this approximate model, the asymptotic variance is a constant multiple of the mean. This satisfies the SET constraint of an asymptotically constant coefficient of variation (as can be verified from Equation A5). In particular, the choice of

$$\sigma^2 = \frac{1}{\eta(\eta - 1)} \tag{A7}$$

gives a terminal coefficient of variation of $1/\sqrt{\eta}$. This gives an asymptotic learning rate of $\alpha = 1/\eta$, as can be easily verified from Equation A6.

After observing n rewards, at time $t = nC$, the posterior distributions for the inverse rates become

$$p\big(s_c(n)|\text{data}\big) \sim G\big(C,\, v_c(n)\big), \quad p\big(s_l(n)|\text{data}\big) \sim G\big(T,\, v_l(n)\big) \tag{A8}$$

and, in about η rewards, $\sqrt{v_c(n)} \rightarrow C/\sqrt{\eta}$ and $\sqrt{v_l(n)} \rightarrow T/\sqrt{\eta}$. This satisfies the constraint suggested by SET and also allows rates to adapt quickly under appropriate circumstances.

Appendix B

Extinction

The approximate model in Appendix A closely matches the full model in the text, but it only provides estimates of the time between rewards at the times that rewards are themselves delivered. In extinction, we are interested in the full model's estimate of the reward rate at times when rewards are not presented. We now specify an extension of this approximation that estimates the rate in the period between rewards, and which also closely matches the full model (again, as verified by simulations).

As discussed in the text, we assume that the light's reliability changes insignificantly over the early phase of extinction, and so the overall predicted rate of reward during extinction while the light is present continues to be given by the light's prediction alone.

Like Gallistel and Gibbon (2000), we measure time in terms of the number of omitted rewards. We treat an observation of $\bar{n}$ of these in time $\bar{n}T$ as an observation of a current reward rate of $1/\bar{n}T$ and use this observation in conjunction with the previous estimate from the last actual delivery (of $\hat{t}_l^{\text{pre}}$, and the previous asymptotic learning rate of $1/\eta$ according to Equation A5):

$$\hat{t}_l(\bar{n}) = \hat{t}_l^{\,\text{pre}} + \frac{1}{\eta}\left(\bar{n}T - \hat{t}_l^{\,\text{pre}}\right), \tag{B1}$$

where $\hat{t}_l^{\,\text{pre}}$ is $1/\hat{\lambda}_l^{\text{pre}}$ and $1/\eta$ is the asymptotic learning rate. From Equation B1, an approximate estimate of the mean rate $\big(1/\hat{t}_l(\bar{n})\big)$ is as follows:

$$\hat{\lambda}_l(\bar{n}) = \left(\frac{\eta}{\eta + \bar{n} - 1}\right)\frac{1}{T}. \tag{B2}$$

Notice this also agrees with the estimate from the didactic window model described in the text.

As argued in the text, we model Gallistel and Gibbon's (2000) extinction criterion as occurring when

(*Appendix continues*)

$$\hat{\lambda}_1(\bar{n}) < \beta_e \hat{\lambda}_1^{\text{pre}}, \tag{B3}$$

where β_e is a threshold less than one (which represents the fraction of the preextinction rate the current estimate must fall below to justify a 50% decline in the response rate). Of course, this is not a statistical decision criterion for the animal, but instead it comes from measuring the extinction speed using a comparison of the current response rate to the preextinction response rate.

Solving this for $\bar{n}$ using Equation (B2) gives

$$\bar{n} > \eta\left(\frac{1}{\beta_e} - 1\right) + 1. \tag{B4}$$

For $\beta_e = 1/3$ and $\eta = 25$, this makes the estimate of $\bar{n} \approx 50$ as shown in Figure 7.

Received May 26, 2000
Revision received July 19, 2001
Accepted July 20, 2001 ■