

# On the role of responses in Pavlovian acquisition

Stefano Ghirlanda

Brooklyn College, CUNY Graduate Center, and Stockholm University

Magnus Enquist

Stockholm University

## Abstract

A defining feature of Pavlovian conditioning is that the unconditioned stimulus (US) is delivered whether or not the animal performs a conditioned response (CR). This has led to the question: Does CR performance play any role in learning? Between the 1930's and 1970's, a consensus emerged that CR acquisition is driven by CS-US experiences, and that CRs play a minimal role, if any. Here we revisit the question and present two new quantitative methods to evaluate whether CRs influence the course of learning. Our results suggest that CRs play a decisive role in Pavlovian acquisition, in such paradigms as rabbit eyeblink conditioning, pigeon autoshaped key pecking, and rat autoshaped lever pressing and magazine entry.

*Keywords:* Pavlovian conditioning, instrumental conditioning, animal learning, computational modeling

## Introduction

In Pavlovian conditioning, an initially neutral conditioned stimulus (CS) precedes an unconditioned stimulus (US), to which the experimental animals have an established reaction, the unconditioned response (UR). Learning is measured by the appearance of a response to the CS, the conditioned response (CR). In Pavlov's (1927) paradigmatic experiments with dogs, CSs were visual or auditory stimuli, the US was food, and the UR and CR were salivation. A defining feature of Pavlovian experiments is that the US is delivered *whether or not* the animal performs a CR. This has led quite naturally to the question: Does the animal have to perform the CR in order for CR acquisition to proceed, or are experiences of CS-US pairings sufficient?

The role of CRs in Pavlovian acquisition has been debated extensively between the 1930's and 1970's, during which time a consensus emerged that acquisition is driven by CS-US experiences, with CRs playing a minimal role, if any. Mackintosh (1974, 1983) effectively summarized the three major arguments in favor of this view as pertaining sensory

preconditioning, response prevention studies, and omission training. These arguments, however, were originally conceived within distinct (albeit related) debates, namely, regarding the role of reinforcement in learning and the distinction between stimulus-stimulus vs. stimulus-response associations. In the following paragraphs, we contend that traditional arguments do not exclude that CRs may have an important role in Pavlovian acquisition.

In **sensory preconditioning** experiments, a first phase of CS<sub>1</sub>-CS<sub>2</sub> pairings precedes a second phase of CS<sub>2</sub>-US pairings. Subsequent tests with CS<sub>1</sub> often reveal a CR, despite no CR having been previously performed to this stimulus (Brogden 1939; Kimmel 1977, but see Reid 1952). These experiments demonstrate that some learning can occur in the absence of both a US and a CR, which is generally conceptualized as a CS<sub>1</sub>-CS<sub>2</sub> association developing during CS<sub>1</sub>-CS<sub>2</sub> pairings. However, CS<sub>1</sub>-CS<sub>2</sub> pairings alone do not establish a CR, which requires CS<sub>2</sub>-US pairings in the second phase. It remains possible that performing CRs during CS<sub>2</sub>-US pairings is important to acquire the CR to CS<sub>2</sub> and, therefore, to CS<sub>1</sub>. Thus sensory preconditioning is compatible with the view that acquiring a Pavlovian CR requires performance of the CR itself.

In **response prevention studies**, subjects experience CS-US pairings while putative CRs are temporarily blocked. After a number of experiences normally sufficient for acquisition, the block is removed and the CS is presented. The occurrence of CRs is then taken as an indication that CS-US experiences are sufficient for learning. Table 1 summarizes classic response-prevention studies, most of them reviewed by Solomon and Turner (1962, few later studies seem to exist). An early, relatively large study by Harlow and Stagner (1933) failed to demonstrate CR acquisition when CS-US experiences were given to animals paralyzed by curare, seemingly supporting the notion that CR performance is necessary for acquisition. However, in a remarkable series of studies, Girden established the following: 1) under curare, muscular contractions *can* be produced, although they are minimal and often do not cause overt movements; 2) these contractions can be conditioned similarly to normal contractions; 3) contractions conditioned under curare disappear once the animal recovers, but reappear if it is again drugged (Girden, 1940, 1942, 1947; Girden & Culler, 1937). Given these findings, Harlow and Stagner's (1933) failure came to be attributed to drugged animals performing *different* CRs than non-drugged animals, rather than no CRs, and the original question remained unsettled.

Solomon and Turner (1962) and Leaf (1964) attempted to make progress by using d-tubocurarine, a curare variety unavailable to Girden that, based on reports from brave human volunteers, was not expected to impair transfer between the drugged and non-drugged states (Prescott, Organe, & Rowbotham, 1946; Smith, Brown, Toman, & Goodman, 1947). While successful at face value, these experiments also trained the CR instrumentally either before or after the animal was drugged. Specifically, Solomon and Turner (1962) first trained dogs to press a panel to postpone a shock US. They then drugged the animals and exposed them to CS<sub>1</sub>-US and CS<sub>2</sub>-No US contingencies. Upon recovery from the drug, four of five dogs pressed the panel more to CS<sub>1</sub> than to CS<sub>2</sub>. Leaf (1964) employed the same experimental phases in the reverse order. In his tests, all eight dogs responded to CS<sub>1</sub> more than to CS<sub>2</sub>. These experiments do demonstrate that some learning can occur without CRs (for example, a

CS<sub>1</sub>-US association). However, because the CR had been trained under conditions in which it could be performed freely, they do not establish that CR acquisition is independent of CR performance. In this sense, these studies are similar to sensory preconditioning experiments.

Table 1 also includes two experiments in which salivation was blocked with atropine (Crisler, 1930; Finch, 1938) and three experiments in which leg flexion was prevented by reversible damage of peripheral nerves (Beck & Doty, 1957; Kellogg, Scott, Davis, & Wolf, 1940; Light & Gantt, 1936). Crisler (1930) successfully conditioned salivation under atropine in five dogs, but according to Finch (1938) his result was confounded by the use of a morphine US, which can itself condition anticipatory nausea and consequent salivation. For this reason, Finch (1938) used an acid solution as a US, obtaining salivary conditioning of very small magnitude in two dogs (< 1 ml/minute over baseline). Of the nerve damage studies, Light and Gantt (1936) and Beck and Doty (1957) demonstrated CRs after nerve regeneration, while Kellogg et al. (1940) failed to observe CRs in four of six subjects, and noted that the two subjects who performed CRs used muscles different from those that had been paralyzed during CS-US experiences. Finally, two experiments with the paralytic agent bulbocapnine resulted in four of five cats acquiring a CR (Beck & Doty, 1957, one of these experiment also included nerve damage).

In summary, the evidence from response prevention studies appears mixed. Girden's work shown that curare is not a good method of response prevention, as it promotes learning of drug-specific CRs that differ from normal CRs. Nevertheless, two small studies suggest that conditioning may be possible under paralysis (Beck & Doty, 1957; Light & Gantt, 1936). Other studies found no CR (Kellogg et al., 1940), or found a CR only when the response was also trained in the non-drugged state (Leaf, 1964; Solomon & Turner, 1962). Even if we interpret these results as demonstrating that CRs can be acquired without overt performance, a general problem with response prevention studies is that blocking the peripheral expression of the CR does not preclude the central nervous system from attempting to initiate the CR (Mackintosh, 1974). It remains possible that this "decision to respond" be necessary, during training, for successful CR acquisition even if the overt CR is blocked.

In **omission training**, a Pavlovian relationship is arranged between a CS and US, but the US is withheld if the animal performs a CR (Sheffield, 1965; D. R. Williams & Williams, 1969). Such an omission contingency often decreases CR occurrence, compared to standard Pavlovian training, but the extent of this decrease depends on many factors (Holland, 1979; Mackintosh, 1974, 1983; Sanabria, Sitomer, & Killeen, 2006). What matters here, however, is that a CR emerges at all from omission contingencies. While this indicates that the consequences of the CR are not all-important for its acquisition (Mackintosh, 1974), it does not rule out that CR performance plays an important role during acquisition, as subjects in omission studies are free to perform CRs.

In conclusion, the findings reviewed above leave at least some room for continued investigation into the role of CRs in Pavlovian acquisition. Indeed, several observations suggest a large role. For example, it is well established that animals can experience many CS-US pairings without producing any CRs, while acquisition proceeds steadily once CRs start to occur (Gallistel, Fairhurst, & Balsam, 2004). Moreover, there is agreement that, if not

Table 1

*Response prevention studies. The Results column indicates how many subjects acquired a CR ('+'), did not acquire a CR ('-'), or provided ambiguous results (labeled '?'). CSs were auditory or visual stimuli, or the experimental apparatus. See the text for discussion of potential confounds. \* Results confounded by additional training of the CR in the non-drugged state. † The CR may have been conditioned by the morphine US. See text for discussion.*

Study	Exp	Species	US	CR	CR block	Results
Harlow and Stagner (1933)	1	Dog, cat	Shock	Jump	Curare	19-
Harlow and Stagner (1933)	8	?	Shock	Leg flexion	Curare	1-
Solomon and Turner (1962)		Dog	Shock	Panel press	Curare	5+, 1? *
Leaf (1964)		Dog	Shock	Jump	Curare	8+ *
Crisler (1930)		Dog	Morphine	Salivation	Atropine	5+ †
Finch (1938)		Dog	HCl in mouth	Salivation	Atropine	2+
Beck and Doty (1957)	1	Cat	Shock	Leg flexion	Bulbocapnine	3+
Beck and Doty (1957)	2	Cat	Shock	Leg flexion	Bulbocapnine, nerve damage	2+, 1-
Light and Gantt (1936)		Dog	Paw shock	Leg flexion	Nerve damage	2+, 2?
Kellogg et al. (1940)		Dog	Paw shock	Leg flexion	Nerve damage	2+, 4-

acquisition, at least CR extinction depends on CR performance (in general: Rescorla 2002; for taste-aversion learning: Balleine and Dickinson 1991; Garcia 1989). For example, when several CRs become conditioned to a CS, and only one of them is extinguished, the others are unaffected (Delamater, 1996; Rescorla, 2002). Lastly, of particular interest is Donahoe and Vegas's (2004) study of pigeons' throat movement responses, which can be elicited as a UR by oral water delivery. The key feature of these responses is their relatively long latency, with a median of around 350 ms. This allowed the authors to arrange trials with a US-CS-UR structure, in which a visual CS followed a water US but preceded the UR. This is akin to backward conditioning, with the difference that responses occur despite the backward CS-US relationship. Surprisingly, Donahoe and Vegas found that throat movement responses thus conditioned were equivalent in frequency and magnitude to those conditioned with the usual trial structure, CS-US-UR (forward conditioning). The authors concluded that the CS-UR temporal relationship is more important than the CS-US relationship. Although aimed at the UR rather than the CR, this study highlights that responses may have an important role in conditioning. (Additionally, the UR and CR had the same form in this study.)

Here we attempt to discern a role for CRs in conditioning by analyzing data from standard acquisition studies, free of experimental interventions other than the presentation of CS-US pairings and the monitoring of CRs. For specificity, we focus on the following models of acquisition:

**Response-independent (R-Indep) model:** CR acquisition is driven by CS-US experiences, regardless of whether CRs occur or not. This is the current consensus. Additionally, it is commonly assumed that overt CRs occur only once a measure of CR potential, such as associative strength, is above a threshold value (Harris, 2011; Spence, 1956).

**Response-dependent (R-Dep) model:** CR probability increases only on conditioning trials in which a CR occurs. Without a CR, CS-US experiences have no effect. In this model, it makes little sense to consider response thresholds. In fact, an initial CR probability below threshold would prevent conditioning altogether, while an initial CR probability above threshold would have no effect.

The R-Indep model predicts that acquisition proceeds steadily as a function of the number of trials, even if no overt CRs occur. The R-Dep model, on the other hand, predicts that CR probability on a given trial depends on the number of CRs performed previously, which results in slower learning if CRs are initially rare. These properties of the R-Indep and R-Dep models result in different statistical features of acquisition records, which we exploit to develop two quantitative methods to discriminate between the models (for related work, see Gallistel et al., 2004; Harris, 2011). The first method considers the probability distributions of the number of trials in between CRs, which can differ considerably under the two models. The second method uses a pattern recognition algorithm to classify acquisition records as typical of one or the other model. We apply these methods to the 5 data sets described in Table 2. We conclude that these data are most compatible with the R-Dep model. We examine the implications of this result in the Discussion.

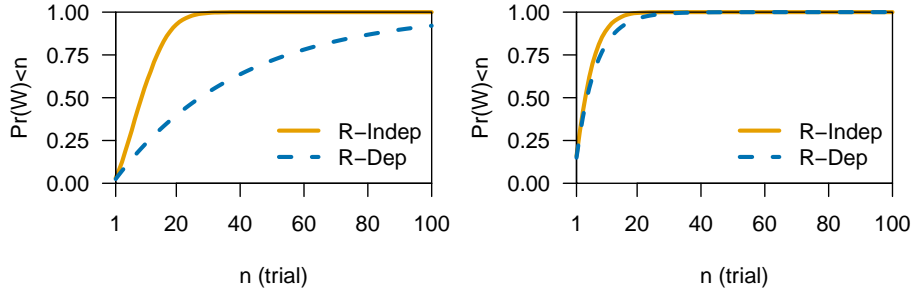
Before proceeding, we note that it is not always feasible, even in principle, to discriminate between the two models with certainty. In fact, each model can generate individual acquisition records that, statistically, are more likely under the other model, or that are equally likely under either model. The situation is similar to the problem of telling a fair coin from a biased coin. By chance, each coin can generate a sequence of heads and tails that mimics the other. Circumventing this statistical difficulty has been our main motivation for developing two very different methods. Encouragingly, the methods lead to the same conclusions.

## Methods

### Waiting time distributions

Our first method focuses on how many trials occur between successive CRs. Let  $W_k$  be the number of trials between the  $(k - 1)$ -th and  $k$ -th CRs, which we refer to as the  $k$ -th waiting time. The R-Indep and R-Dep models predict different probability distributions for waiting times, especially for the first few ones when CR probability is low (Fig. 1). The method leverages these differences to infer which model is a better fit to empirical waiting time distributions. We proceed through the following steps:

1. Use the principle of maximum likelihood to derive the best-fitting R-Indep and R-Dep models for a given data set.
2. Derive the predicted waiting time distributions from the best-fitting models.
3. Compare each model's predictions to empirical waiting time distributions. We make three kinds of comparisons:



*Figure 1.* Example probability distributions of the waiting time until a CR,  $W$ , in two R-Indep R-Dep models of Pavlovian acquisition, with initial probability of  $p_0 = 0.025$  (left) or  $p_0 = 0.15$  (right). The R-Indep model follows Eq. (2) with  $p_\infty = 1$ ,  $a = 0.01$ . In the R-Dep model, CR probability is fixed at  $p_0$  per trial.

- (a) Visual comparisons, plotting the empirical and theoretical distributions to assess their similarity qualitatively.
- (b) Statistical comparison for each waiting time distribution, using the Kolmogorov-Smirnov test to detect differences between empirical waiting time distributions and those predicted by the two models.
- (c) Overall statistical comparison for each data set, using the Wilcoxon paired test to evaluate the null hypothesis that the two models are an equally good fit to the overall distribution of the first 20 waiting times.

The advantage of this method is that it uses the best fitting R-Dep and R-Indep models, and thus it does not require detailed assumptions about learning. For example, it does not assume that learning proceeds according to the Rescorla and Wagner (1972) model or other formal theories of conditioning. The potential disadvantage of the method is that it assumes identical learning parameters across individuals. This assumption is necessary because estimating the empirical distribution of a waiting time requires many observations, while each subject only provides one waiting time. We discuss this assumption in the last part of this section.

**Maximum-likelihood response-dependent model.** In the R-Dep model, CR probability is constant between CRs, which leads to a geometric distribution of waiting times. Let  $p_k$  be the per-trial probability of a CR between  $k - 1$ -th and  $k$ -th CRs. The waiting time distribution of the  $k$ -th CR is then:

$$\Pr(W_k = n) = p_k (1 - p_k)^{n-1} \quad (1)$$

where the first factor is the probability of a CR after  $n$  trials, and the second factor the probability of no CR in the preceding  $n - 1$  trials. Under this model, the expected waiting time for the  $k$ -th CR is  $1/p_k$ . The maximum likelihood estimate of  $p_k$  is simply  $1/\bar{W}_k$ , where  $\bar{W}_k$  is the average waiting time for the  $k$ -th response across all individuals in an experiment.

**Maximum-likelihood response-independent model.** The waiting time distribution for the R-Indep model is not geometric, because CR probability is not constant between CRs. Theios and Brelford (1966) showed that the maximum-likelihood R-Indep model for a given data set has a simple form when stated in terms of CR probability per trial. Namely, the maximum-likelihood R-Indep model is the one in which CR probability at trial  $t$  equals the observed CR frequency on that trial,  $f_t$ . Unfortunately, the corresponding waiting time distributions do not have a simple analytical form such as Eq. (1). We have thus estimated waiting time distributions by running, for each data set, 10,000 simulations of the maximum-likelihood R-Indep model, i.e., using  $f_t$  as CR probability on trial  $t$ , and then calculating the distribution of waiting times observed in the simulated data.

Note that this method applies equally to R-Indep models with and without a response threshold. The maximum-likelihood model, in fact, would simply estimate as 0 the probability of a CR on any trial in which no CRs are observed, which would be interpreted as trials in which response potential is below threshold.

**Homogeneity of learning parameters.** The assumption of identical learning parameters across individuals is necessary to compute empirical waiting time distributions, but it appears to clash with the observation that individuals vary in the onset of responding and in the rate of approach to asymptotic performance. Variability in performance, however, does not *per se* prove heterogeneity of learning parameters, if we assume that stochastic factors can influence learning and responding. For example, under the R-Dep model, a fixed response probability of  $p = 0.05$  per trial leads to a distribution of waiting times with a standard deviation of  $\sqrt{1-p}/p \simeq 19.5$  trials, almost as large as the expected value of  $1/p = 20$ . The question of whether observed variability implies heterogeneity in learning parameters is thus difficult to settle. Here we limit ourselves to testing whether a homogeneous R-Indep or R-Dep model is compatible with empirical waiting time distributions. Namely, we will compare theoretical and empirical waiting time distributions by means of Kolmogorov-Smirnov tests (in its version for discrete distributions, see Arnold & Emerson, 2011). A significant  $p$ -value from such a test would lead to rejecting the hypothesis that a homogeneous model can reproduce the empirical distributions. A non significant  $p$ -value, on the other hand, would be compatible with the hypothesis that different individuals have similar learning parameters.

### Classification of individual acquisition records

Here we present a method that makes stronger assumptions about learning than the previous one, but which allows different learning parameters across individuals. The method relies on training a feed-forward artificial neural network (ANN, see Haykin, 2008) to identify patterns of responding characteristic of the R-Indep and R-Dep models. For example, response records with large gaps between the first few CRs would be more typical of the R-Dep model. We proceeded in three steps:

1. We simulated a large number of acquisition records using the R-Dep and R-Indep models with a variety of parameter values.

2. We trained the ANN to recognize whether simulated acquisition records were generated by the R-Indep or R-Dep models.
3. We used the trained ANN to estimate the probability that actual acquisition records originate from either model.

The first two steps are detailed below. The third step is reported in Results.

We simulated the R-Indep model using the standard error-correction learning equation (Bush & Mosteller, 1955; Rescorla & Wagner, 1972):

$$\Delta p_t = a(p_\infty - p_t) \quad (2)$$

where  $p_t$  is response potential at trial  $t$ ,  $\Delta p_t$  the change in  $p_t$  that occurs at trial  $t$ ,  $p_\infty$  the asymptotic CR probability, and  $a$  the learning rate. At each trial, a CR was generated with probability  $p_t$ , if  $p_t$  was larger than a threshold value  $\theta \geq 0$ . On any trial in which  $p_t \leq \theta$ , no CR was generated. In each simulation, we set the threshold  $\theta$  at a random value between 0 and 0.3. This range ensured that simulated data would include both situations in which the threshold had virtually no effect, and situations in which the presence of the threshold would delay CR onset significantly. (In pilot work, we also simulated the R-Indep model without a threshold, with no change in results.)

We simulated the R-Dep model similarly to the R-Indep model, with the difference that  $p_t$  would change only on trials with a CR and that no response threshold was employed. Specifically, we used the following equation:

$$\Delta p_t = ar_t(p_\infty - p_t) \quad (3)$$

where  $r_t$  indicates whether a CR occurs ( $r_t = 1$ ) or not ( $r_t = 0$ ) on trial  $t$ . At each trial, a CR was generated with probability  $p_t$ . We used Eq. (2) and Eq. (3) to generate 7000 acquisition records from each model, running 200 simulations with each combination of the following parameter values:

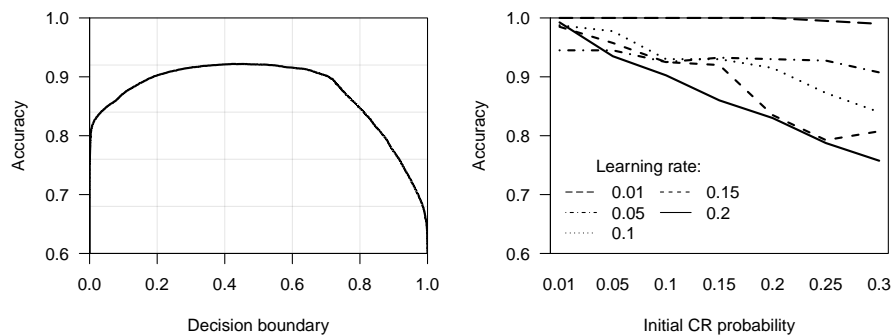
$$\begin{aligned} p_0 &= \{0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3\} \\ a &= \{0.01, 0.05, 0.1, 0.15, 0.2\} \\ p_\infty &= 0.95 \end{aligned} \quad (4)$$

These values were chosen to cover approximately the range of learning speeds, and of initial and asymptotic CR probabilities in empirical data. Each simulation included 150 trials. (In preliminary investigations, we found that classification accuracy would decrease with fewer trials.)

After having generated the simulated acquisition records, we trained the ANN to output zero to records generated by the R-Dep model, and one to records generated by the R-Indep model. The ANN had 150 input units, 10 hidden units, and one output unit. The activation of input unit  $i$  was set to one if a CR had occurred on trial  $i$ , and to zero otherwise. The network was trained by gradient descent as described in Venables and Ripley (2002), including five-fold cross validation (see Software for details). Training resulted in an overall accuracy of



about 92%. The left panel of Fig. 2 shows that accuracy was not overly dependent on the decision boundary, i.e., the ANN output value used to discriminate between the two models. We adopted a boundary of 0.5, meaning that network output above (below) 0.5 was interpreted as identifying the R-Indep (R-Dep) model. The right panel in Fig. 2 shows ANN accuracy as a function of initial CR probability and learning rate  $a$ . Note that model classification degrades as initial CR probability increases, especially if  $a$  is not small. The reason is that high CR probability leads to fast approach to asymptote in both models, reducing the number of trials during which the models behave differently.



*Figure 2.* Artificial neural-network (ANN) classification of simulated acquisition records from the R-Dep and R-Indep models. Data from 7000 simulations for each model. Left: Classification accuracy as a function of decision boundary, i.e., the ANN output value chosen to demarcate the R-Dep and R-Indep models. Right: Classification accuracy as a function of initial CR probability and learning rate, for a decision boundary at 0.5.

## Data sources

Individual acquisition records are seldom published in their entirety. With the exception of one published data set (Theios & Brelford, 1966), we obtained our data by contacting other researchers based on their published work and suggestions from colleagues. Table 2 summarizes the data sets, each of which is identified by a code for brevity. All data are available at <http://dx.doi.org/10.6084/m9.figshare.6494867> with code to reproduce our analyses. From data set Gu, we excluded one subject that was recorded as having undergone only 5 conditioning trials, and one subject that responded only four times in 450 trials. Numbers in Table 2 indicate subjects included in our analyses. We had also received autoshaping data for nine pigeons from Federico Sanabria, which we set aside when we received the larger Gu data set, also from pigeon autoshaping. Results in the two data sets are comparable. Additionally, Gavan McNally had provided us with data from conditioned suppression experiments, which we excluded as their analysis would require different methods.

In data sets T and K, the response measure is whether a CR occurs or not during a trial, which we coded as one and zero, respectively. In data sets H, Go, and Gu, the response measure is the absolute number of CRs during CS presentation within a trial. We

Table 2

*Data sets. The K data is from E. James Kehoe’s laboratory, previously analyzed by Gallistel et al. (2004). The Gu data set was collected with the methods in Yamazaki, Aust, Huber, Hausmann, and Güntürkün (2007).*

Data set	Species	Paradigm	Subjects	Source
T	Rabbit	Eyeblink conditioning	90	Theios and Brelford (1966)
K	Rabbit	Eyeblink conditioning	47	E. J. Kehoe, Gallistel et al. (2004)
H	Rat	Autoshaped lever press	36	Holland et al. (2014)
Go	Rat	Autoshaped magazine entry	55	Gottlieb (2005, 2008)
Gu	Pigeon	Autoshaped keypeck	19	O. Güntürkün

dichotomized this number to zero (no CRs) or one (one or more CRs). It would be ideal to avoid this dichotomization in order to assess the effect of multiple responses in the same trial, but we leave this refinement for future work.

Some data sets include data from similar experiments from the same laboratory, differing in some parameters such as inter-stimulus interval (ISI) and inter-trial interval (ITI). This is not a problem for our classification method, which analyzes individual acquisition records, but it may confound the empirical distribution of waiting times. Variation in experimental parameters, however, had at most slight effects on CR acquisition (see Appendix), and conducting separate analyses for the different parameter values leads to the same conclusions reported below for the pooled data (not shown).

## Software

Analyses were conducted within R version 3.5.0 (2018-04-23) (R Core Team, 2018). The ANN was implemented using the ‘nnet’ package, version 7.3.12 (Venables & Ripley, 2002). Cross validation was implemented by the ‘tune’ function of the ‘e1071’ package, version 1.6.8 (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2017). For the discrete-distribution Kolmogorov-Smirnov test we used the ‘dgof’ package, version 1.2 (Arnold & Emerson, 2011). Analysis code is available at the URL given in the preceding section.

## Results

### Waiting time distributions

Figure 3 shows maximum-likelihood estimates of the first five waiting time distributions for the R-Dep and R-Indep models, and compares them to empirical distributions. The models frequently make similar predictions, but when predictions differ the data typically support the R-Dep model. Examples include  $W_1$  in data sets T, K, Go, and Gu;  $W_5$  in data sets T, K, and Gu data sets;  $W_3$  in data set T. The figure also highlights when either model’s predictions are significantly different from empirical data, according to Kolmogorov-Smirnov tests. Note that failure to reject either model implies that the hypothesis of homogeneous learning parameters cannot be rejected, either.

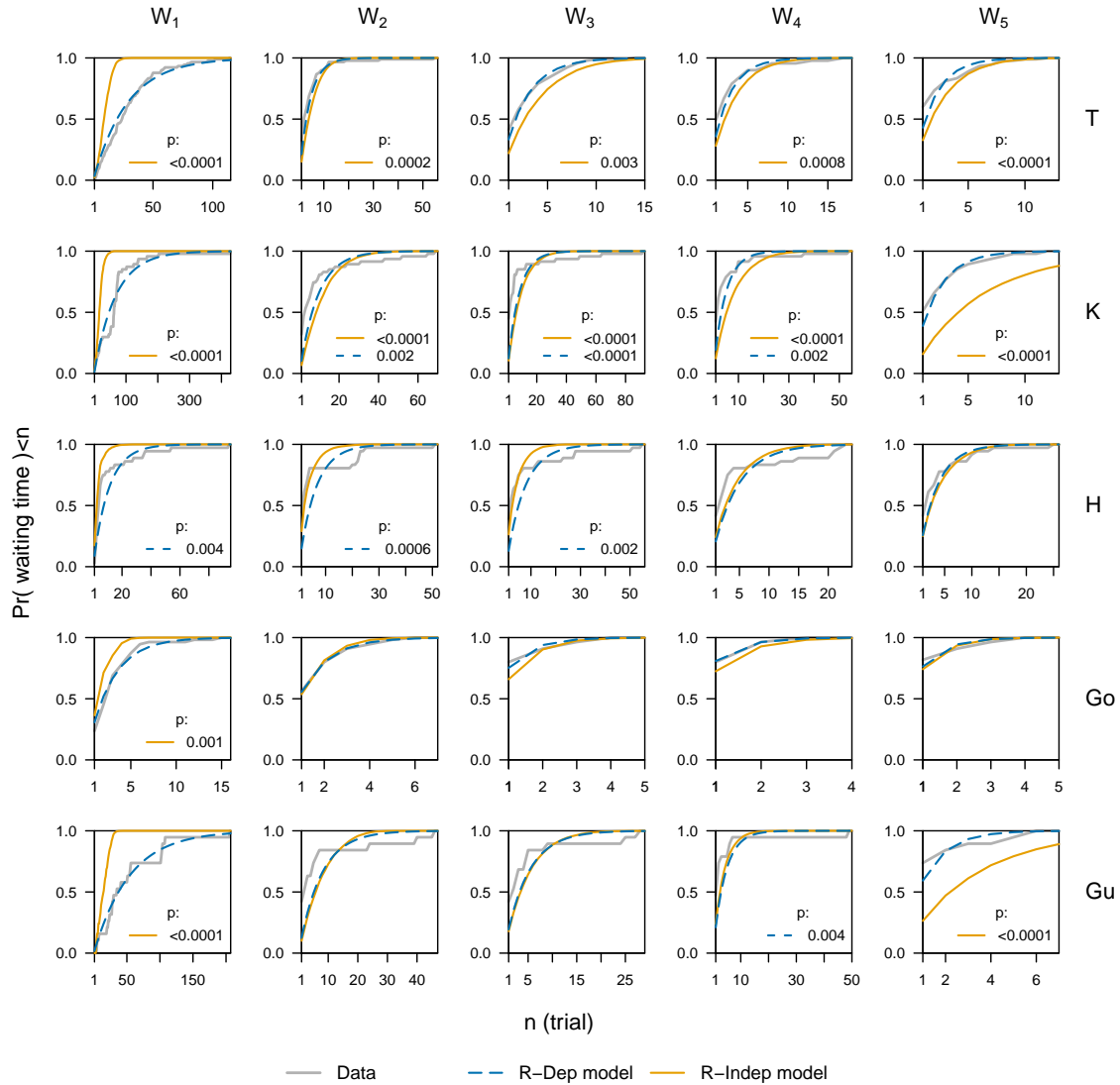
Figure 4 summarizes the result up to the 20<sup>th</sup> waiting time, in terms of each model's mean error. In the T, K, Go, and Gu data sets, the R-Dep model is overall closer to empirical data. The H data set offers similar support for both models. Theios and Brelford (1966) provide further evidence against the homogeneous R-Indep model, by showing that it cannot reproduce the sequential structure of responses in the T data set.

### **Classification of individual acquisition records**

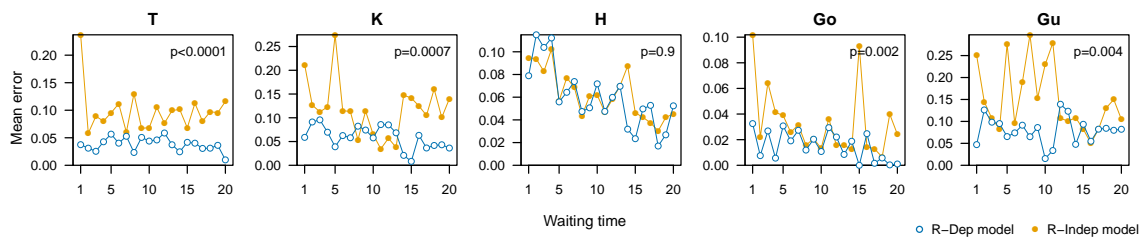
The ANN, trained as described in Methods, classifies the vast majority of individual response records in the T, K, and Gu data sets as typical of the R-Dep model (Fig. 5). In data set H, non-significant majority of records is classified as R-Dep, while records from data set Go are largely classified as R-Indep (we could apply the ANN to only 16 subject of the Go data set, as the others had received fewer than the 150 trials used to train the network).

It is not always easy to understand the classification criteria of ANNs, but in this case visual inspection provides a rather clear picture. Figure 6 shows all individual records analyzed, with those classified as typical of the R-dep model plotted with plus symbols, and those classified as typical of the R-Indep model plotted with circles. Starting from the T data set (top left), we see that the records that are classified as R-Indep often exhibit early onset of responding and, once responding starts, a relatively steady progression toward asymptotic response rates. Records that are ascribed to the R-Dep model, in contrast, have more delayed onset of responding and longer gaps between the initial CRs. These criteria conform well to our intuition of the two models. The K and Gu data set have rather fragmented records, often with delayed onset of responding, and are classified almost entirely as R-Dep. The H data set exhibits a mix of patterns, which are classified as R-Dep or R-Indep in approximately equal number. The Go data set, with very early onset of responding and few breaks, is classified almost entirely as R-Indep.

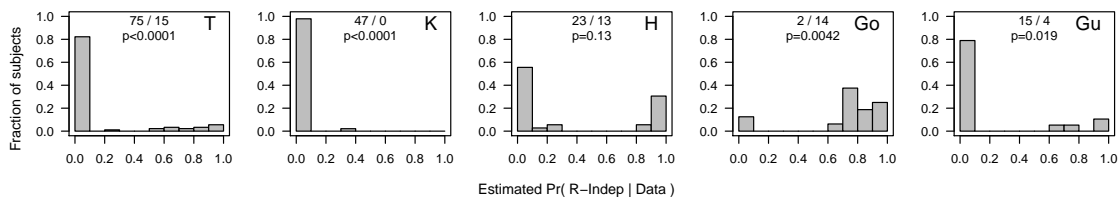
We believe it is possible to explain the the difference in results across data sets by considering procedural differences between studies, rather than assuming different learning processes across data sets or individuals. First, we recall that the R-Indep and R-Dep models are more difficult to tell apart when initial CR probability is high (Fig. 2). Second, we note that animals in the Go data set had a very high CR probability from the very first trials, as evidenced by a mean of only 3 trials until the first CR. These values can be contrasted with 29, 68, and 54 trials in the T, K, and Gu data sets, respectively. The reason for high initial CR probability in the Go data set was likely pre-training that facilitated CR performance. In fact, Go experiments measured autoshaped magazine entries in rats who had been already trained to approach the food magazine. The Go data set also exhibits a high learning rate, as subjects quickly reached asymptotic responding. In conclusion, data set Go has characteristics that are, on average, more typical of the R-Indep model, but which are also compatible with the R-Dep model under conditions of high initial CR probability and high learning rate. A similar argument can be offered for the H data set, which does not clearly support either model, and which also employed a response (presses of a lever close to the food magazine) that was facilitated by previous magazine training, resulting in high initial CR probability (a mean of 11 trials until the first CR).



*Figure 3.* Waiting time distributions predicted by maximum-likelihood R-Dep model (dashed) and R-Indep model (solid), compared to empirical data (gray).  $W_k$  is the  $k$ -th waiting time, i.e., the number of trials between the  $k - 1$ -th and  $k$ -th CRs. Plots in the same row pertain to the same data set (indicated on the right). Plot in the same column pertain to the same waiting time (indicated on top). P-values refer to Kolmogorov-Smirnov tests of the difference between the empirical and the R-Dep or R-Indep distribution. Only tests yielding  $p < 0.01$  are displayed.



*Figure 4.* Prediction error for waiting time distributions by maximum-likelihood R-Dep model (open circles) and R-Indep model (closed circles). The error is computed as the mean distance between predicted and observed distributions. Lower value indicates better fit;  $p$ -values refer to two-tailed Wilcoxon paired tests of the null hypothesis that the two models do not differ.



*Figure 5.* Classification of data from Pavlovian acquisition experiments by an artificial neural network (ANN) trained on simulated data (see Methods). The numbers indicate the number of individual acquisition records classified as typical of the R-Dep and R-Indep R-Dep models, respectively. The  $p$ -value indicates the outcome of a two-tailed binomial test of the null hypothesis that acquisition records are equally likely to be classified as R-Dep or R-Indep.

## Discussion

We developed two methods to gauge the role of CRs in Pavlovian acquisition. In our first method, we used the maximum-likelihood R-Dep and R-Indep models with identical learning parameters across individuals, and we compared the waiting time distributions predicted by these models with empirical distributions. In our second method, we trained an ANN to recognize whether simulated acquisition records had been generated by the R-Dep or R-Indep model, in this case using a wide range of parameter values. A comparison of Fig. 4 and Fig. 5 shows that results from the two methods are in full agreement. Overall, the results favor the R-dep model, indicating that CR performance may be an important trigger of acquisition. This surprising result suggests new directions for research in associative learning, neuroscience, and applied settings. Experimentally, it would be valuable to devise new Pavlovian acquisition paradigms in which the role of CRs can be assessed more directly. Theoretically, it may be possible to devise more effective data analysis strategies. From a neuroscientific point of view, we now have the tools to overcome the limitations of early CR-prevention studies and determine which neural events are necessary for acquisition. For example, our analysis indicate an important role of CRs in rabbit eyeblink conditioning,



Figure 6. Individual response records for the data sets in Table 2. Each row represents an individual, with symbols indicating trials in which a CR occurred. Records are arranged according to trial of the first CR, from bottom to top. Records with CRs marked as + are those classified as typical of the R-Dep model by the ANN (see Methods), records with CRs marked as o are classified as typical of the R-Indep model.

whose neural basis are understood to a considerable degree (Christian & Thompson, 2003). In applied settings, our results may suggest to reconsider the role of CRs in the acquisition of phobias and other maladaptive behavior with a recognized Pavlovian component (Bernstein, 1999; Conklin & Tiffany, 2002; Siegel & Ramos, 2002). We leave these analyses and investigations to future research. Below, we turn to the implications of our results for associative learning theory.

### **What causes the first CRs?**

The R-Dep model holds that CS-US pairings alone do not increase response probability, and thus begs the question of what causes the first response (Mackintosh, 1983). This might not be an insurmountable difficulty. First, as we remarked above when discussing the H and Go data sets, previous training can favor CR performance. For example, rats typically start autoshaping of magazine entry after completing magazine training, hence they are likely to enter the magazine even during the very first trials of CS training. In other experimental paradigms, we speculate that the first CRs occur because of sensitization, which typically facilitates biologically appropriate responses to the US (Overmier & Curnow, 1969; Prokasy & Ebel, 1964; Richardson & Elsayed, 1998). Thus, in eyeblink conditioning the US may increase the frequency of eyeblinks (Grant, 1943a, 1943b), which can bootstrap learning by increasing the probability that an eyeblink will occur in a conditioning trial. Similarly, in autoshaped lever pressing or key pecking, animals may perform the first responses because food delivery (the US) sensitizes consummatory behavior such as approaching, manipulating, or pecking salient objects. This perspective also suggests that only responses that are sensitized by US delivery have a sizable likelihood of becoming CRs, compatible with a perspective in which Pavlovian conditioning is integrated into biologically organized behavior systems (Moore, 1973; Timberlake, 1994; Timberlake & Lucas, 1989). A pioneering recognition of Pavlovian CRs as just one component of a biologically structured behavioral and motivational system can be found in the work of Zener (1937), who observed that salivation is just a fraction of a dog's behavior in a conditioning experiment. When not restrained, the dog also enacts other behavior such as approach to, orientation toward, and investigation of the CS and experimental apparatus, which in nature is appropriate for securing food.

### **Alternative models of Pavlovian acquisition**

The R-Indep model we investigated is representative of accepted principles of Pavlovian acquisition, that is, reliance on CS-US experiences rather than on CR performance, and gradual approach to asymptote similarly to Eq. (2). However, it is possible that other response-independent models may fare better. In this section we briefly discuss some prominent variants of the response-independent hypothesis.

The debate around Pavlovian acquisition is traditionally intertwined with a debate on the shape of acquisition curves. Mean acquisition curves are either uniformly negatively accelerated or have a sigmoid shape with a shorter phase of positive acceleration followed by a longer phase of negative acceleration (Harris, 2011; Mackintosh, 1974; Spence, 1956).

Figure 7 exemplifies these findings in the data sets analyzed above. The classic analysis of mean acquisition curves was conducted by Spence (1956), based on prior work by Hull (1943). In his model, overt CRs occur only when response potential is above a threshold. A fixed threshold, however, would disagree with the observed smooth rise in CR frequency. Spence thus assumed that momentary factors, such as physiological and motivational variables, add normally distributed noise to the CR generating process, which results in nonzero CR probability even when response potential is nominally below threshold. This model accounts well for the shape of acquisition curves, and allows sigmoid acquisition curves to arise from the negatively accelerated curves generated by Eq. (2) and similar models (Bush & Mosteller, 1955; Pearce, 1987; Rescorla & Wagner, 1972). The model, however, produces even shorter waiting times than the R-Indep model we have analyzed, which has a threshold mechanism but no noise, and thus agrees more poorly with empirical data. The reason is as follows. Suppose that response potential is nominally below threshold for the first few trials. It is enough for the noise to bring response potential above threshold once in order to produce a CR and thus record an earlier waiting time than if there were no noise. On the other hand, to postpone a waiting time significantly, the noise would need to bring below threshold an above-threshold response potential for a number of consecutive trials, which is less likely. Thus adding noise reduces waiting times, compared with a model without noise.

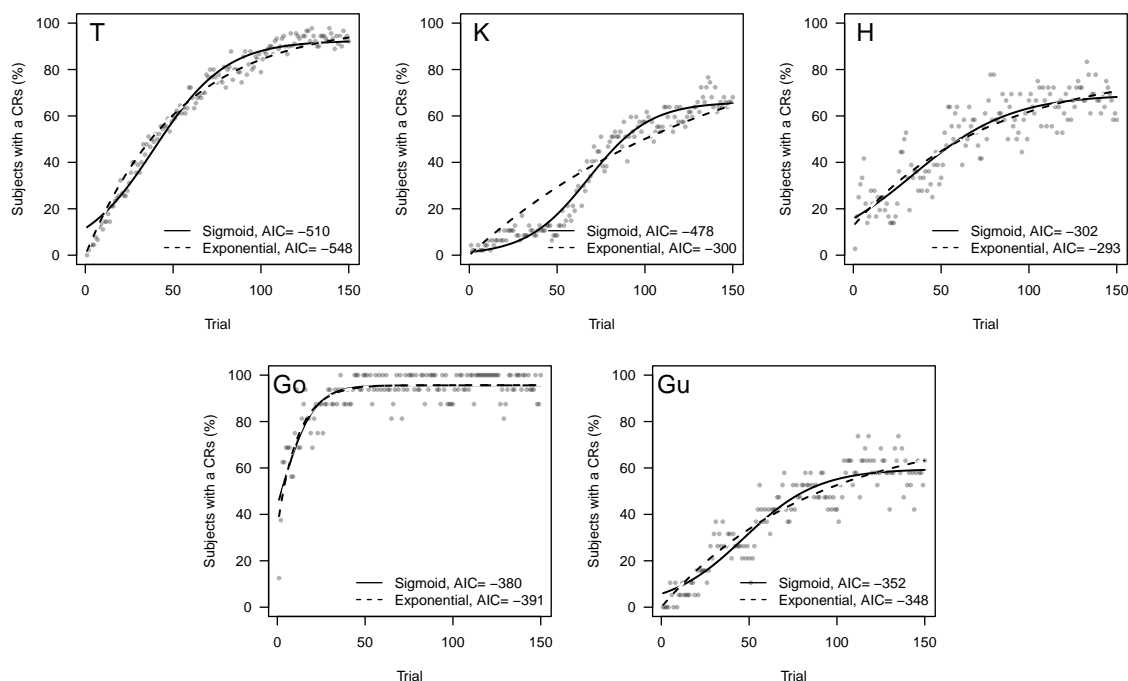
In addition to Spence's model, other models can produce both sigmoid and negatively accelerated mean acquisition curves, including our R-Indep model with a distribution of thresholds across individuals, as used above, or with a non-linear response function. Similarly, the R-Dep model can produce both types of learning curves depending on learning rate and initial response probability. Such flexibility in theoretical models implies that mean acquisition curves are not especially diagnostic of underlying learning processes. For example, according to our analysis, both the T and K data sets strongly favor the R-Dep over the R-Indep model, yet their mean acquisition curves have very different shape (Fig. 7).

Nearly 50 years after Spence's (1956) classic treatment, Gallistel et al. (2004) revitalized the debate on Pavlovian acquisition curves by proposing to conceptualize conditioning as a process of accumulation of evidence, with CRs starting abruptly once enough evidence has accumulated. This model, although radically different in theoretical motivation, makes in practice the same predictions as the R-Indep model with a threshold and a very high learning rate. For example, the response records classified as R-Indep by our ANN (Fig. 6) conform well to Gallistel et al.'s (2004) model by virtue of displaying a relatively abrupt onset of responding, followed by relatively steady responding. Many response records, however, appear too fragmented to conform to this picture. Although we did not include very high learning rates in our analysis, doing so would produce response records that fit the data more poorly than the R-Indep model we used. For example, waiting times beyond the first would be substantially shorter than in the data, because by definition each subject would be above threshold at this point and thus it would respond at asymptotic levels. Harris (2011), in a detailed study of autoshaped magazine approach in rats, reaches a similar conclusion by providing evidence of a gradual, rather than abrupt increase in CR probability. (In Harris' data, CRs were already occurring at the start of conditioning, at the rate of  $\simeq 4$  CR/trial, due



to previous training; see our discussion of the Go data set in Results for implications.)

Finally, we would like to mention that the applicability of the R-Dep model to acquisition records does not prove, *per se*, that CRs are a causal factors in Pavlovian acquisition. CRs might also be manifestations of other causal factors. For example, CR occurrence may correlate with the animal attending to the CS, with attention rather than CR performance as the main causal factor (we owe this observation to a reviewer). Of note, Theios (1963, 1968) and Theios and Brelsford (1966) introduced models in which CS-US experiences cause probabilistic transitions between states of increasing CR probability. In these models, CR onset is just a consequence of a transition having occurred. More work is needed to compare empirical data with the predictions of these models, as well as of attentional models (Le Pelley, 2004; Mackintosh, 1975; Pearce & Hall, 1980). Experimental work is also possible in this domain, for example by attempting to quantify separately the effects of CR performance and orientation toward the CS, which may be used as an operational measure of attention. We hope that our results will provide motivation for both empirical and theoretical work.



*Figure 7.* Average acquisition curves in the five data sets in Table 2. Solid lines are fits of the sigmoid equation  $f(x) = c[1 + \exp(-a(x - b))]^{-1}$ ; dashed lines are fits of the negative exponential equation  $f(x) = c[1 - \exp(-a(x - b))]$ . Akaike information criteria (AIC) for both models are given in figure legends (Akaike, 2011).

### The interaction of Pavlovian and instrumental conditioning

The question of what triggers Pavlovian acquisition is linked to another unresolved issue in associative learning theory, namely the relationship between instrumental and Pavlo-

vian conditioning. Current theory holds that Pavlovian and instrumental contingencies can result in a variety of associations between conditioned/discriminative stimuli, responses, and unconditioned stimuli/reinforcers (Hall, 2002). How these associations interact in determining behavior remains a matter of considerable debate (Bouton, 2016; Dickinson, 1980; Mackintosh, 1983). While there has been a steady trend recognizing many similarities between Pavlovian and instrumental learning (Dickinson, 1980; Miller & Balaz, 1981), most texts in animal learning, cognition, and behavior analysis discuss Pavlovian and instrumental phenomena largely separately (Bouton, 2016; Pearce, 2013; Pierce & Cheney, 2017), and there is no accepted formal model encompassing both Pavlovian and instrumental learning. Recently, we presented an associative learning model drawing from machine learning, experimental psychology, and ethology (Enquist, Lind, & Ghirlanda, 2016). Although our presentation focused on instrumental learning, the model also includes a conceptualization of Pavlovian learning. Here we offer a brief summary in order to clarify the model’s account of Pavlovian acquisition, referring to Enquist et al. (2016) for other phenomena.

The model includes two learning processes. One is the learning of S-R associations. This learning is “instrumental” in that it occurs only when a response is actually performed, regardless of whether experimental contingencies are Pavlovian or instrumental. The second process is the conditioning of stimulus values, that is, the value of stimuli as predictors of reinforcers. This learning is “Pavlovian” in that it occurs independently of responding, based on experienced S-S pairings. The model formalizes the interaction between S-S (“Pavlovian”) and S-R (“instrumental”) learning by positing that learned stimulus values can reinforce responding, that is, that they function as conditioned reinforcers. Both S-R and S-S learning follow simple temporal-difference equations (Sutton & Barto, 1998; Wiering, 2005) that are closely related to Rescorla and Wagner’s (1972) learning equation. Formally, the model is specified as follows. Consider a stimulus  $s$  to which the animal responds with action  $a$ , which in turn is followed by stimulus  $s'$ . In Pavlovian conditioning, the sequence  $s \rightarrow a \rightarrow s'$  corresponds to  $CS \rightarrow CR \rightarrow US$ . Let  $v(a, s)$  be the associative strength between  $a$  and  $s$ , we have:

$$\Delta v(a, s) = \alpha_v [u(s') + w(s') - v(a, s)] \quad (5)$$

where  $\alpha_v$  is a learning rate,  $u(s')$  the unconditioned value of  $s'$ , and  $w(s')$  its conditioned value. Equation (5) is equivalent to the classic Rescorla and Wagner (1972) learning equation if we let  $\lambda = u(s') + w(s')$ , i.e., if the asymptotic associative strength equals the sum of unconditioned and conditioned values. Conditioned values are updated according to:

$$\Delta w(s) = \alpha_w [u(s') + w(s') - w(s)] \quad (6)$$

where  $\alpha_w$  is a second learning rate. Note that this update rule depends only on the succession of stimuli (e.g.,  $CS \rightarrow US$ ), and not on how the animal responds. The formalization of a learning rule for conditioned values enables the model to learn sequences of actions through chaining, as envisaged by Skinner (1936) and later authors (Pierce & Cheney, 2017; B. A. Williams, 1994a, 1994b). Finally, the model posits a probabilistic decision making

rule that selects actions according to their associative strength, in a manner compatible with the matching law (Baum, 1974; Davison & McCarthy, 2016; Herrnstein, 1974). Phenomena of stimulus competition such as blocking and overshadowing are dealt with similarly to the Rescorla and Wagner (1972) model, by assuming that the elements of a stimulus compound can acquire separate associative strength (Enquist et al., 2016; see also Ghirlanda, 2015).

In the case of Pavlovian acquisition, the model predicts that the CS will automatically acquire conditioned value (as demonstrated by second-order conditioning, Pearce 2013), but that CR probability will not increase unless the CR is performed. That response probability changes only when a response is tried is a feature our model shares with similar models from artificial intelligence (Sutton & Barto, 1998; Van Seijen, Van Hasselt, Whiteson, & Wiering, 2009; Watkins & Dayan, 1992), which provided additional motivation for our study. One consequence of this feature is that waiting times between CRs are distributed geometrically, as in the R-Dep model considered above, and compatibly with the empirical data we analyzed. The model also eliminates the asymmetry between acquisition and extinction noted in the Introduction, as both phenomena would require CR performance. Finally, the model takes into account that Pavlovian and instrumental phenomena are part of biologically structured behavior systems (Hogan, 1994; Timberlake, 1994; Zener, 1937) by positing that motivational states, US delivery, and other factors can boost the probability of performing biologically appropriate behavior (Enquist et al., 2016). We do not claim that the model captures perfectly the relationship between Pavlovian and instrumental learning, or indeed that it is a complete model of either. For example, the model's predictions in the case of omission training has still to be worked out in full. In pilot work, we have found omission training to maintain varying degrees of responding, depending on such factors as initial associative strengths, whether alternative behavior is possible or not, and duration of the CS. Whether or not the model can successfully capture Pavlovian-instrumental interactions and other associative learning phenomena, we believe that a clear formalization of both Pavlovian and instrumental processes in a single model can be a meaningful step toward a more harmonic theory of associative learning.

## Acknowledgments

We are grateful to Charles Gallistel, Dan Gottlieb, Onur Güntürkün, Peter Holland, Sara Letzner, Gavan McNally, and Federico Sanabria for sharing data. We thank Andrew Delamater, Guillermo Esber, Federico Sanabria, and two anonymous reviewers for discussion and suggestions. This work was partly supported by grant 2015-0005 from the Knut and Alice Wallenberg Foundation.

## References

- Akaike, H. (2011). Akaike's information criterion. In *International encyclopedia of statistical science* (pp. 25–25). Berlin: Springer.
- Arnold, T. A., & Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3(2), 34–39.

- Balleine, B. W., & Dickinson, A. (1991). Instrumental performance following reinforcer devaluation depends upon incentive learning. *The Quarterly Journal of Experimental Psychology*, *43*(3), 279–296.
- Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching 1. *Journal of the Experimental Analysis of Behavior*, *22*(1), 231–242.
- Beck, E. C., & Doty, R. W. (1957). Conditioned flexion reflexes acquired during combined catalepsy and de-efferentation. *Journal of Comparative and Physiological Psychology*, *50*(3), 211.
- Bernstein, I. L. (1999). Taste aversion learning: A contemporary perspective. *Nutrition*, *15*(3), 229–234.
- Bouton, M. E. (2016). *Learning and behavior: A modern synthesis* (2nd ed.). Sunderland, MA: Sinauer.
- Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, *25*(4), 323–332.
- Bush, R. R., & Mosteller, F. (1955). *Stochastic models of learning*. New York: Wiley.
- Bush, R. R., & Mosteller, F. (1955). *Stochastic models of learning*. New York: Wiley.
- Christian, K. M., & Thompson, R. F. (2003). Neural substrates of eyeblink conditioning: acquisition and retention. *Learning & memory*, *10*(6), 427–455.
- Conklin, C. A., & Tiffany, S. T. (2002). Applying extinction research and theory to cue-exposure addiction treatment. *Addiction*, *95*, 155–167.
- Crisler, G. (1930). Salivation is unnecessary for the establishment of the salivary conditioned reflex induced by morphine. *American Journal of Physiology*, *94*(3), 553–556.
- Davison, M., & McCarthy, D. (2016). *The matching law: A research review*. New York: Routledge.
- Delamater, A. R. (1996). Effects of several extinction treatments upon the integrity of pavlovian stimulus-outcome associations. *Animal Learning & Behavior*, *24*(4), 437–449.
- Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.
- Donahoe, J. W., & Vegas, R. (2004). Pavlovian conditioning: the cs-ur relation. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(1), 17.
- Enquist, M., Lind, J., & Ghirlanda, S. (2016). The power of associative learning and the origin of optimal behaviour. *Royal Society Open Science*, *3*(11). doi: 10.1098/rsos.160734
- Finch, G. (1938). Salivary conditioning in atropinized dogs. *American Journal of Physiology—Legacy Content*, *124*(1), 136–141.
- Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the USA*, *101*(36), 13124–13131.
- Garcia, J. (1989). Food for Tolman: Cognition and cathexis in concert. In T. Archer & L.-G. Nilsson (Eds.), *Aversion, avoidance and anxiety* (p. 45–85). Erlbaum.
- Ghirlanda, S. (2015). On elemental and configural theories of associative learning. *Journal of Mathematical Psychology*, *64–65*, 8–16.
- Girden, E. (1940). Cerebral mechanisms in conditioning under curare. *The American Journal of Psychology*, *53*(3), 397–406.
- Girden, E. (1942). The dissociation of blood pressure conditioned responses under erythroidine. *Journal of Experimental Psychology*, *31*(3), 219.
- Girden, E. (1947). Conditioned responses in curarized monkey. *American Journal of Psychology*, *60*, 571–587.
- Girden, E., & Culler, E. (1937). Conditioned responses in curarized striate muscle in dogs. *Journal of*

- Comparative Psychology*, 23(2), 261.
- Gottlieb, D. A. (2005). Acquisition with partial and continuous reinforcement in rat magazine approach. *Journal of Experimental Psychology: Animal Behavior Processes*, 31(3), 319.
- Gottlieb, D. A. (2008). Is the number of trials a primary determinant of conditioned responding? *Journal of Experimental Psychology: Animal Behavior Processes*, 34(2), 185.
- Grant, D. A. (1943a). The pseudo-conditioned eyelid response. *Journal of Experimental Psychology*, 32, 139-149.
- Grant, D. A. (1943b). Sensitization and association in eyelid conditioning. *Journal of Experimental Psychology*, 32, 201-202.
- Hall, G. (2002). Associative structures in pavlovian and instrumental conditioning. In H. Pashler & G. R. (Eds.), *Stevens' handbook of experimental psychology: Learning, motivation, and emotion* (p. 1-45). Hoboken, NJ: John Wiley & Sons.
- Harlow, H. F., & Stagner, R. (1933). Effect of complete striate muscle paralysis upon the learning process. *Journal of Experimental Psychology*, 16(2), 283.
- Harris, J. A. (2011). The acquisition of conditioned responding. *Journal of Experimental Psychology: Animal Behavior Processes*, 37(2), 151-164. doi: 10.1037/a0021883
- Haykin, S. (2008). *Neural networks and learning machines* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Herrnstein, R. J. (1974). Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, 21(1), 159-164.
- Hogan, J. A. (1994). Structure and development of behavior systems. *Psychonomic Bulletin & Review*, 1(4), 439-450.
- Holland, P. C. (1979). Differential effects of omission contingencies on various components of pavlovian appetitive conditioned responding in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 5(2), 178.
- Holland, P. C., Asem, J. S. A., Galvin, C. P., Keeney, C. H., Hsu, M., Miller, A., & Zhou, V. (2014). Blocking in autoshaped lever-pressing procedures with rats. *Learning & Behavior*, 42(1), 1-21.
- Hull, C. L. (1943). *Principles of Behaviour*. New York: Appleton-Century-Crofts.
- Kellogg, W. N., Scott, V. B., Davis, R. C., & Wolf, I. S. (1940). Is movement necessary for learning? an experimental test of the motor theory of conditioning. *Journal of Comparative Psychology*, 29(1), 43-74.
- Kimmel, H. D. (1977). Notes from "Pavlov's Wednesdays": sensory preconditioning. *American Journal of Psychology*, 319-321.
- Leaf, R. C. (1964). Avoidance response evocation as a function of prior discriminative fear conditioning under curare. *Journal of Comparative and Physiological Psychology*, 58(3), 446.
- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*, B57, 193-243.
- Light, J. S., & Gantt, W. H. (1936). Essential part of reflex arc for establishment of conditioned reflex. formation of conditioned reflex after exclusion of motor peripheral end. *Journal of Comparative Psychology*, 21(1), 19.
- Mackintosh, N. J. (1974). *The psychology of animal learning*. London: Academic Press.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.
- Mackintosh, N. J. (1983). *Conditioning and associative learning*. Oxford: Oxford University Press.

- Mackintosh, N. J. (1983). *Conditioning and associative learning*. Oxford: Oxford University Press.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2017). e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien [Computer software manual]. (R package version 1.6-8, <https://CRAN.R-project.org/package=e1071>)
- Miller, R. R., & Balaz, M. A. (1981). Differences in adaptiveness between classically conditioned responses and instrumentally acquired responses. In N. E. Spear & M. R. R. (Eds.), *Information processing in animals: Memory mechanisms* (p. 49-80). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Moore, B. R. (1973). The role of directed pavlovian reactions in simple instrumental learning in the pigeon. In (eds Hinde, R. A. Stevenson-Hinde, J) *Constraints on learning: Limitations and predispositions*. Academic Press, Oxford, 488 pp.
- Overmier, J. B., & Curnow, P. F. (1969). Classical conditioning, pseudoconditioning, and sensitization in “normal” and forebrainless goldfish. *Journal of Comparative and Physiological Psychology*, 68(2), 193-198.
- Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford: Oxford University Press.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94(1), 61–73.
- Pearce, J. M. (2013). *Animal learning and cognition: an introduction*. Psychology Press.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Pierce, W. D., & Cheney, D., Carl. (2017). *Behavior analysis and learning*. Routledge.
- Prescott, F., Organe, G., & Rowbotham, S. (1946). Tubocurarine chloride as an adjunct to anesthesia. *The Lancet*, 248(6412), 80–84.
- Prokasy, W. F., & Ebel, H. C. (1964). Gsr conditioning and sensitization as a function of intertrial interval. *Journal of Experimental Psychology*, 67(2), 113-119.
- R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Reid, R. L. (1952). A test of sensory pre-conditioning in pigeons. *Quarterly Journal of Experimental Psychology*, 4(2), 49–56.
- Rescorla, R. A. (2002). Extinction. In L. Bäckman & C. von Hofsten (Eds.), *Psychology at the turn of the millenium. vol. 1: Cognitive, biological, and health perspectives* (p. 217-244). Hove, UK: Taylor & Francis.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement [incollection]. In *Classical conditioning: current research and theory*. Appleton-Century-Crofts.
- Richardson, R., & Elsayed, H. (1998). Shock sensitization of startle in rats: The role of contextual conditioning. *Behavioral Neuroscience*, 112(5), 1136-1141.
- Sanabria, F., Sitomer, M. T., & Killeen, P. (2006). Negative automaintenance omission training is effective. *Journal of the experimental analysis of behavior*, 86, 1-10.
- Sheffield, F. D. (1965). Relation between classical conditioning and instrumental learning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium* (pp. 302–322). New York: Appleton-Century-Crofts.
- Siegel, S., & Ramos, B. M. C. (2002). Applying laboratory reseach: Drug anticipation and the treatment of drug addiction. *Experimental and Clinical Psychopharmacology*, 10(3), 162-183.

- Skinner, B. (1936). The reinforcing effect of a differentiating stimulus. *The Journal of General Psychology*, 14(2), 263–278.
- Smith, S. M., Brown, H. O., Toman, J. E. P., & Goodman, L. S. (1947). The lack of cerebral effects of d-tubocurarine. *Anesthesiology*, 8, 1-14.
- Solomon, R. L., & Turner, L. H. (1962). Discriminative classical conditioning in dogs paralyzed by curare can later control discriminative avoidance responses in the normal state. *Psychological Review*, 69(3), 202.
- Spence, K. W. (1956). *Behavior theory and conditioning*. New Have, CT: Yale University Press.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
- Theios, J. (1963). Simple conditioning as two-stage all-or-none learning. *Psychological Review*, 70(5), 403.
- Theios, J. (1968). Finite integer models for learning in individual subjects. *Psychological Review*, 75(4), 292-307.
- Theios, J., & Brelford, J. W. J. (1966). A Markov model for classical conditioning: Application to eye-blink conditioning in rabbits. *Psychological Review*, 73(5), 393-408.
- Timberlake, W. (1994). Behavior systems, associationism, and Pavlovian conditioning. *Psychonomic Bulletin & Review*, 1(4), 405–420.
- Timberlake, W., & Lucas, G. A. (1989). Behavior systems and learning: From misbehavior to general principles.
- Van Seijen, H., Van Hasselt, H., Whiteson, S., & Wiering, M. (2009). A theoretical and empirical analysis of Expected Sarsa. In *Ieee symposium on adaptive dynamic programming and reinforcement learning. adprl'09* (pp. 177–184).
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth ed.). New York: Springer.
- Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, 8, 279-292.
- Wiering, M. (2005). QV( $\lambda$ )-learning: A new on-policy reinforcement learning algorithm. In D. Leone (Ed.), *Proceedings of the 7<sup>th</sup> european workshop on reinforcement learning* (p. 17-18).
- Williams, B. A. (1994a). Conditioned reinforcement: Experimental and theoretical issues. *Behavior Analyst*, 2, 261–285.
- Williams, B. A. (1994b). Conditioned reinforcement: Neglected or outmoded explanatory construct? *Psychonomic Bulletin and Review*, 1(4), 457-475.
- Williams, D. R., & Williams, H. (1969). Auto-maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the experimental analysis of behavior*, 12(4), 511–520.
- Yamazaki, Y., Aust, U., Huber, L., Hausmann, M., & Güntürkün, O. (2007). Lateralized cognition: asymmetrical and complementary strategies of pigeons during discrimination of the “human concept”. *Cognition*, 104(2), 315–344.
- Zener, K. (1937). The significance of behavior accompanying conditioned salivary secretion for theories of the conditioned response. *American Journal of Psychology*, 50(1), 384-403.

## Appendix

### Homogeneity of data sets

We assessed homogeneity of data with respect to experimental parameters by evaluating whether the waiting times until the 1<sup>st</sup>, 10<sup>th</sup>, and 50<sup>th</sup> CR varied significantly across parameter

values.

The K data set includes ISIs between 200 ms and 1200 ms, but this variation did not affect waiting times (Table A1). The Go data set includes data from experiments with different ITIs (20 or 30 s), CS duration (120, 150, 240, 480, or 960 s), and spacing between sessions (0, 1, or 3 days). A full assessment of the effect of these variables is impossible because not all combinations of these values were used. There was, however, little effect of these parameters on acquisition (Table A2, Table A3 Table A4). Acquisition was slower for subjects receiving training every fourth day and with a 480 s CS (Table A2). Only eight subjects were run in this condition, however, and removing their data does not alter the results reported in the main text.

In the Gu data set, variation in US rate (100% or 50%) had a small effect on waiting times (Table A5). Analyzing the two conditions separately, however, yields the same results as pooling all data. Lastly, the T and H data sets were homogeneous with respect to experimental parameters.

Table A1

*Results of ANOVAs relating the number of trials until the 1<sup>st</sup>, 10<sup>th</sup>, and 50<sup>th</sup> CR and the ISI in the K data set.*

Variable	Term	Sum of squares	d.f.	<i>F</i>	<i>p</i>
$W_1$	ISI	11162.51	1.00	2.41	0.13
	Residuals	208597.24	45.00		
$W_{10}$	ISI	5976.43	1.00	0.86	0.36
	Residuals	312342.85	45.00		
$W_{50}$	ISI	907.27	1.00	0.14	0.71
	Residuals	281807.53	44.00		

Table A2

*Results of ANOVAs relating the number of trials until the 1<sup>st</sup>, 10<sup>th</sup>, and 50<sup>th</sup> CR and days between sessions in the Go data set.*

Variable	Term	Sum of squares	d.f.	<i>F</i>	<i>p</i>
$W_1$	Days	120.08	1.00	26.26	0.00
	Residuals	242.35	53.00		
$W_{10}$	Days	92.04	1.00	7.77	0.01
	Residuals	627.89	53.00		
$W_{50}$	Days	42.67	1.00	1.11	0.30
	Residuals	1149.33	30.00		



Table A3

*Results of ANOVAs relating the number of trials until the 1<sup>st</sup>, 10<sup>th</sup>, and 50<sup>th</sup> CR and ITI in the Go data set.*

Variable	Term	Sum of squares	d.f.	<i>F</i>	<i>p</i>
$W_1$	ITI	0.61	1.00	0.09	0.77
	Residuals	361.83	53.00		
$W_{10}$	ITI	0.01	1.00	0.00	0.98
	Residuals	719.91	53.00		
$W_{50}$	ITI	112.67	1.00	3.13	0.09
	Residuals	1079.33	30.00		

Table A4

*Results of ANOVAs relating the number of trials until the 1<sup>st</sup>, 10<sup>th</sup>, and 50<sup>th</sup> CR and CS duration in the Go data set.*

Variable	Term	Sum of squares	d.f.	<i>F</i>	<i>p</i>
$W_1$	CSdur	4.81	1.00	0.71	0.40
	Residuals	357.63	53.00		
$W_{10}$	CSdur	0.00	1.00	0.00	1.00
	Residuals	719.93	53.00		
$W_{50}$	CSdur	16.79	1.00	0.43	0.52
	Residuals	1175.21	30.00		

Table A5

*Results of ANOVAs relating the number of trials until the 1<sup>st</sup>, 10<sup>th</sup>, and 50<sup>th</sup> CR and US rate in the Gu data set.*

Variable	Term	Sum of squares	d.f.	<i>F</i>	<i>p</i>
$W_1$	Reward	9013.73	1.00	4.18	0.06
	Residuals	36679.96	17.00		
$W_{10}$	Reward	6852.01	1.00	2.54	0.13
	Residuals	45936.10	17.00		
$W_{50}$	Reward	36802.63	1.00	6.93	0.02
	Residuals	90308.32	17.00		