

Journal of Experimental Psychology: Animal Learning and Cognition

Optimal Response Rates in Humans and Rats

David M. Freestone, Fuat Balci, Patrick Simen, and Russell M. Church

Online First Publication, December 1, 2014. <http://dx.doi.org/10.1037/xan0000049>

CITATION

Freestone, D. M., Balci, F., Simen, P., & Church, R. M. (2014, December 1). Optimal Response Rates in Humans and Rats. *Journal of Experimental Psychology: Animal Learning and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xan0000049>

Optimal Response Rates in Humans and Rats

David M. Freestone
Brown University

Fuat Balci
Koç University

Patrick Simen
Oberlin College

Russell M. Church
Brown University

The analysis of response rates has been highly influential in psychology, giving rise to many prominent theories of learning. There is, however, growing interest in explaining response rates, not as a global response to associations or value, but as a decision about how to space responses in time. Recently, researchers have shown that humans and mice can time a single response optimally; that is, in a way that maximizes reward. Here, we use the well-established differential reinforcement of low rates (DRL) timing task to show that humans and rats come close to optimizing reinforcement rate, but respond systematically faster than they should.

Keywords: decision making, DRL, interval timing, response rate, reward rate

Response rates are studied heavily in experimental psychology. Often, response rates are viewed as a measure of an underlying internal variable like associative strength (see Pearce & Bouton, 2001, for an overview), subjective value (Baum, 1974; Herrnstein, 1961), or confidence (Blough, 1967; Yi, 2009), to name a few. In most theoretical frameworks, there is a directional flow in which reinforcement updates the internal variable, and the internal variable is transformed into a response rate. Often, this is where the flow stops because it represents one of the simplest models that allows us to make useful inferences about those internal variables by measuring the response rate. But the picture is somewhat more complicated because the response rates often modulate reinforcement in instrumental situations. There is a response-reinforcer feedback loop in which response rates dictate how much reinforcement is received, and reinforcement may influence the response rate (Baum, 1973; Rachlin, 1978).

Rachlin (1978) approached this as a maximization problem: The response rate should be chosen such that some value function (e.g., reinforcement rate) is maximized. Nowhere is this idea more evident than in the Matching Law literature from the late 1970s and early 1980s (e.g., Baum, 1981; Herrnstein & Heyman, 1979; Heyman & Luce, 1979; Heyman & Tanz, 1995; Houston & McNamara, 1981; Staddon & Motheral, 1978). This view has not driven nearly as much enthusiasm or research in single-operant paradigms.

One recent example, though, comes from the reinforcement learning literature (Niv, Daw, & Dayan, 2005; Niv, Daw, Joel, & Dayan, 2007; Niv, Joel, & Dayan, 2006). Normally, reinforcement learning models either assume a monotonic transformation from value (or association) to response rate (see Miller, Barnet, & Grahame, 1995), or they are silent on the transformation altogether (Sutton & Barto, 1998). But Niv and colleagues view the response rate as the outcome of a sequence of decisions about which action to perform, and for how long (which they term *vigor*). How long to perform an action is amenable to optimality analysis: Given the schedule of reinforcement, what is the optimal time between actions needed to maximize the reinforcement rate? They show that this time is proportional to the cost of making an action, and more importantly for the present discussion, but perhaps not surprisingly, inversely proportional to the average reinforcement rate. That is, it is optimal to perform all actions faster (a higher response rate) when the average reinforcement rate is higher. They derived exact results for vigor for ratio schedules, but not for interval schedules, nor for more specific reinforcement schedules.

The differential reinforcement of low rates (DRL) task was originally developed to test the hypothesis that rats could withhold responding (Wilson & Keller, 1953). In this task, rats are rewarded for spacing their responses in time. A lever press is reinforced if, and only if, the time between it and the last response (the interresponse time; IRT) is greater than some experimenter defined target duration, often called the *DRL schedule*. This task is interesting

David M. Freestone, Cognitive, Linguistics, and Psychological Sciences, Brown University; Fuat Balci, Department of Psychology, Koç University, Istanbul, Turkey; Patrick Simen, Department of Neuroscience, Oberlin College; Russell M. Church, Cognitive, Linguistics, and Psychological Sciences, Brown University.

David M. Freestone is now at the Center for Neural Science, New York University.

We thank Federico Sanabria, Vladimir Orduna, and Peter Balsam for sharing data with us. The results from the analysis of these data are presented in the discussion. We also thank Laura deSouza for aiding us in data collection, and Andra Geana for invaluable discussions on earlier versions of the manuscript. Last, we thank two anonymous reviewers for their helpful suggestions. Parts of this work was supported by an FP7 Marie Curie PIRG08-GA-2010-277015, TÜBİTAK 1001 111K402, and a BAGEP Grant from Bilim Akademisi—The Science Academy, Turkey to Fuat Balci.

Correspondence concerning this article should be addressed to David M. Freestone, Center for Neural Science, New York University, 4 Washington Place, New York, NY 10003. E-mail: freestone@nyu.edu

because it imposes a well-defined feedback loop between the response rate and the reinforcement rate.

Wearden (1990) set the foundation for the work presented here by analyzing what the IRT should be in order to maximize reinforcement on the DRL task. He assumed (as we do) that the correct quantity to maximize is the reinforcement rate (see Balci, Freestone, and Gallistel (2009) for a task in which reinforcement per trial is maximized). Wearden's initial simulations suggested that the variability of the IRT distribution was a key component in the feedback loop and should be accounted for by any model that seeks to maximize reinforcement rate on this task (in contrast to earlier "minimize the average delay" models that did not account for variability; see Zeiler, Scott, & Hoyert, 1987). Çavdaroglu, Zeki, and Balci (2014) laid the critical mathematical foundation by using statistical decision theory (Girshick & Blackwell, 1954; Wald, 1950) to rigorously specify how the IRT should depend on the variability. They showed that humans, without explicit task instructions, learned the target intervals and nearly maximized reinforcement rate. The present article seeks to replicate these findings in humans and extend the findings to rats.

Optimal Response Rates in the DRL Task

The mathematical treatment of the IRTs in the DRL task requires a few definitions and assumptions. We define the optimal response rate as the response rate that maximizes the reinforcement rate. The reinforcement rate R is

$$R(t, \gamma) \equiv p(r; t, \gamma)t^{-1} \quad (1)$$

where r represents reinforcement or not (1 or 0), t is the mean IRT, γ is the variability of the IRT distribution (we will formalize this in a moment), so that $p(r; t, \gamma)$ is the probability of a reinforcement given the IRT distribution. It is an empirical fact that animals are imperfect timers; that is, $\gamma > 0$ (Gibbon, 1977; Gibbon, Malapani, Dale, & Gallistel, 1997).

Because of this fact, there is a speed-accuracy trade-off in this task. The probability of reinforcement (accuracy) grows with t because the longer the rat waits before responding, the more likely it is that a lever press will be reinforced. The highest possible accuracy is when the entire IRT distribution falls longer than the target interval T . By definition, high accuracy results in a high reinforcement rate per response. The problem is that high accuracy leads to a low rate of return per unit of time. As t grows, the time between reinforcers grows, leading to a smaller overall reinforcement rate per unit of time.

The first assumption that our model makes is that animals *do not* control their own variability. This is an oversimplification (see Schoenfeld, Harris, & Farmer, 1966), but a useful one. The second assumption is that animals *do* control their own mean IRT. In other words, the response-reinforcer feedback loop depends on both the mean and the variability of the response distribution, but the animal can only adjust the mean to increase reinforcement rate. How the animal does this, either through hill-climbing the reinforcement rate gradient or through computations on representations (etc.), is outside the scope of this article, but see Kheifets and Gallistel (2012) and Trommershäuser, Landy, and Maloney (2006) for some thoughts and data on the topic.

The response rate that maximizes the reinforcement rate is obtained by differentiating Equation 1 with respect to time, setting it to zero, and solving for t (the response rate is $1/t$).

We can test for optimal response rates in two ways. The first is empirical: Shift the actual IRT distribution until the reinforcement rate would be maximized. The amount the data needed to be shifted gives an index of optimality (the smaller the shift, the closer to optimal the response times). The second is theoretical: Fit the IRT distribution with a suitable generative model and analyze the parameters of the fit. We believe the second method is more informative, but we will leave that judgment to our readers. However, we will explain that method in detail now so we can make concrete statements about Equation 1.

When Wearden (1990) performed his initial simulations, he assumed the IRT distribution was Gaussian. Sanabria and Killeen (2008) also assumed normally distributed IRTs. Our empirical data, and others like it, suggest that the IRT distribution is positively skewed and bimodal (Richards, Sabol, & Seiden, 1993). The full IRT distribution has both short (seemingly untimed) and long (seemingly timed) responses. We will formalize the full bimodal model in a moment, but for simplicity and clarity, we focus on the long (timed) responses here. Taking our cues from Simen, Rivest, Ludvig, Balci, and Killeen (2013), we assume the IRTs come from an inverse Gaussian distribution (in our Results section, we show this is superior to the Gaussian assumption). The probability density function for the inverse Gaussian distribution is given by

$$IG(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x - \mu)^2}{2x\mu^2}} \quad (2)$$

This equation gives the probability of obtaining x , given scale and shape parameters μ and λ , respectively. This allows us to write the probability of a reinforcement given the IRT distribution as

$$p(r; t, \gamma) = \int_{\tau=T}^{\infty} IG(\tau; t, \lambda) d\tau \quad (3)$$

with λ given by

$$\lambda = \frac{t}{\gamma^2} \quad (4)$$

and γ is the coefficient of variation. In words, the probability of reinforcement is simply the fraction of IRTs greater than the target interval, T , and this fraction depends on γ . Our choice of parameterizing the model with the coefficient of variation γ instead of the inverse Gaussian scale parameter λ is important. Timescale invariance (that response time distributions overlap when plotted on a normalized time-axis) is a well-established result in the timing literature, including the DRL task (e.g., Wearden, 1990), and underlies all timing models. This empirical fact allows us to normalize the observed data from different DRL schedules and then compare them against a single optimum. Equation 4 must hold for the inverse Gaussian to be timescale invariant across different values of t .

Solving for the optimal response rate analytically in this case is difficult. To give some intuition, Çavdaroglu et al. (2014) showed that the optimal IRT is the solution to the equation

$$0 = 1 - p(r; t, \gamma) - T \times IG(T; t, \lambda) \quad (5)$$

when solved for t . We will denote the solution t_o ; it is the optimal IRT needed to maximize reinforcement rate. The result is a set of optimal solutions—an *optimal performance curve*—that depends on the animal's coefficient of variation γ .

Figure 1 shows the step-by-step process toward creating the optimal performance curve. The target interval is set to 1 (dotted vertical line), reflecting the fact that the IRT distributions can be normalized by the target time because of timescale invariance. The

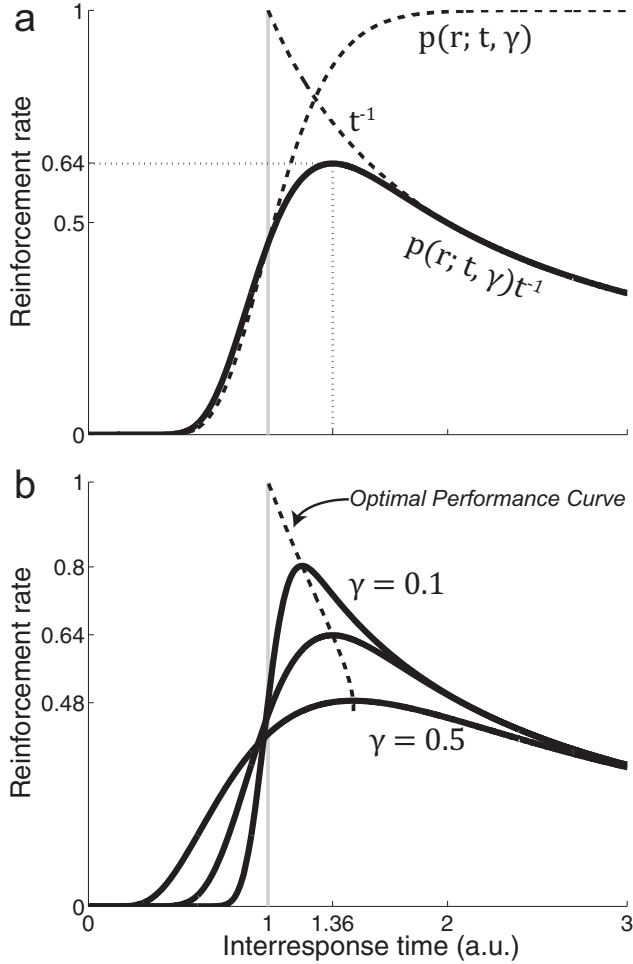


Figure 1. The model. Panel a (top) shows how the optimal model is constructed. The gray vertical line at 1 shows the DRL schedule in arbitrary units (a.u.). The rising dashed line shows the probability of a reward, given the statistics of the interresponse time distribution (the mean, t , and coefficient of variation γ). The falling dashed line shows how much the reinforcement rate declines simply because of the passage of time ($1/t$). The solid black curve is the element-wise multiplication of the two dashed lines and gives the reinforcement rate for a single coefficient of variation for any interresponse time ranging from 0 to 3 (i.e., 0 to 3 times the target interval). The light dashed horizontal and vertical lines intersect at the maximum of the reinforcement rate curve, showing the optimal interresponse time. Panel b shows 3 more reinforcement rate curves ($\gamma = 0.1, 0.25$, and 0.50). The dashed line goes through the maximum of all possible reinforcement rate curves when γ is varied. It describes the optimal relationship between the mean and variability of the interresponse time distribution.

rising dashed line displays the probability of reinforcement (Equation 3). The falling dashed line displays $1/t$. This line shows, holding the probability of reinforcement constant, how decreases in the response rate decrease the reinforcement rate. The product of these two curves (element-wise multiplication) gives the reinforcement rate as a function of t for a given γ , shown as the solid black curve. The maximum of this curve is the maximum reinforcement rate possible (given γ), and its location on the time-axis gives the optimal mean IRT t_o . Normally distributed IRTs do not change the main result, only the details.

Panel b shows how the reinforcement rate curve depends on the animal's coefficient of variation. Each solid black curve shows the reinforcement rate curve for a different coefficient of variation. Keep in mind that the coefficient of variation is a constraint in our model; the animal cannot optimize it. The first thing of note is that as γ increases, the entire reinforcement rate function decreases. All things being equal, animals with higher variability will necessarily get less reward than animals with lower variability. This is because animals with higher variability necessarily make more errors (i.e., early responses that go unreinforced). The dashed line in the figure displays the optimal performance curve that runs through the maxima of all possible reinforcement rate curves when γ is varied. One way to think about the optimal performance curve is that it specifies the ideal relationship between an individual animal's variability and its mean IRT. This is one of the critical predictions we test in this article: As the variability increases, so should the mean IRT.¹

This prediction is tested in both rats and humans. The take-home results are that although both species show some critical features of optimality, their response rates are systematically faster than optimal.

Method

Sixty rats and 15 humans were tested on this task. Rats were required to press a lever to obtain a food pellet, and humans pressed the spacebar on a computer keyboard to obtain a point that was later converted into money (between \$10 and \$20 for a session, depending on performance). Both rats and humans were tested in a between-subjects design in which each participant saw only one DRL target interval. Participants were only rewarded if the time between two consecutive responses was greater than a particular target interval. For rats, the target intervals were 7, 10, 14, 28, and 56 s (12 rats per schedule). For humans, the schedules were 5, 8, 10, 12, and 15 s. Rats were tested for 41 daily sessions, and humans were tested in a single session.

Portions of the human and rat data from this study were used to construct Figure 4 in Balci et al. (2011) and briefly discussed there.

Subjects

Humans. Fifteen adults (seven males and eight females between 18 and 30 years of age) participated. They were recruited

¹ This is true for plausible values for the coefficient of variation γ . For larger values of γ (roughly >0.7), the optimal performance curve turns back toward the schedule. This is because the reinforcement rate curve flattens to the point where waiting longer no longer compensates for having high variability.

from online postings and fliers posted around the Princeton University campus. All participants provided written consent for their participation, and the Institutional Review Panel for Human Subjects of Princeton University approved the experiment. One participant in the DRL 12-s group reported that s/he did not follow the instructions in the task, so the data are not included in the analysis.

Rats. Sixty male Sprague–Dawley rats (Taconic Laboratories, Germantown, NY) were used in this experiment. The colony room was on a 12:12 light–dark cycle (lights off at 8:30 a.m.). During the dark cycle, dim red lights provided light in the colony room and testing rooms. Upon arrival, the rats were roughly 4 weeks old and weighed between 75 and 100 g. During the first week, the rats were on a free-feeding schedule. After a week, their daily food (FormuLab 5008) was rationed to 16 g per day. During each experimental session, the rats were fed 45 mg Noyes pellets (Improved Formula A) as a reinforcer. Water was available *ad libitum* in both the home cage and the testing chamber. Forty-eight of the rats were previously used in a fixed-interval experiment with gaps that used lights and sounds as stimuli (these rats were approximately 6 months old at the start of this experiment). They were assigned to each DRL schedule randomly so that their previous training did not impact the results. Over the course of the experiment, 1 rat rarely pressed the lever, and 2 rats died. This left 57 rats for data analysis. All procedures were approved by the Brown University Institutional Animal Care and Use Committee.

Apparatus

Humans. The visual stimulus was a white square on black background. The stimuli were coded in MATLAB and displayed on a Macintosh computer using the Psychophysics Toolbox extension (Brainard, 1997). Participants indicated their responses with a standard computer keyboard.

Rats. Twenty-four experiment chambers (Med Associates; dimensions 25 × 30 × 30 cm) were situated in two separate experiment rooms (12 in each room). Each chamber was contained in a sound-attenuating box (Med Associates; dimensions 74 × 38 × 60 cm) with a fan for ventilation. Each experimental chamber was equipped with a pellet dispenser (Med Associates, ENV-203) on the front wall that delivered the reward into a food cup. A head entry into this cup interrupted a photobeam (Med Associates, ENV-254). On both sides of the food cup, there were two retractable levers. On the opposite wall, a water bottle protruded into the chamber allowing *ad libitum* access to water during the session. A lick on the spout of the water bottle completed an electric circuit. Four Gateway Pentium III/500 computers running Med-PC for Windows (version 1.15) controlled the experiments and recorded the data. The interruption of the photobeam and the completion of the lick and lever circuits were recorded in time-event format with 2-ms accuracy.

Procedure

Humans. Participants attended single-session experiments. One of the following DRL schedules was assigned to each session: 5 s, 8 s, 10 s, 12 s, or 15 s. Participants were told that they would earn money for their responses if, and only if, they waited for a minimum duration since their previous response (but they were not explicitly told the duration). They were also told that each response

would reset the trial clock. Participants were told how much they would earn for each correct response, that the session time was fixed, and they were encouraged to make as much money as possible. Participants could earn at most around \$20 per session, and the reward magnitude per correct response was adjusted for different DRL schedules (higher magnitude for longer DRL schedules). This aimed to equate the reinforcement rate for different schedules. Participants were also asked not to count, tap, or adopt any rhythmic activity to time the intervals.

Participants were first presented with the minimum wait time (the DRL schedule) 3 times. This duration was signaled by a white square presented on a black background. Participants then reproduced this interval roughly 50 times over two blocks. During this phase, a white square appeared in the middle of the screen to signal the onset of the interval and participants were instructed to press the space key when they thought the target interval elapsed.

The square disappeared upon pressing the space key, and participants were provided feedback on a fixed-length horizontal line. The feedback graphically indicated how far their reproduction time was from the target interval in that trial. A white vertical line represented the target interval, and the red vertical line represented the reproduction interval in that particular trial. The length of the horizontal line was normalized by the DRL schedule to ensure that the same number of pixels referred to the same proportion of discrepancy for every schedule. The point of the feedback was to give the participants time to learn their own variability, if they did not already have an estimate (the rats had 41 sessions in which to do so).

After the reproduction blocks, participants were tested in eight 5-min long blocks of DRL testing. Test blocks were initiated with the appearance of a white square in the middle of the screen. Participants could respond at any time and rate they wished during test blocks.

If the IRT was greater than or equal to the target interval, the square turned green, a brief beep sounded, and the amount of money earned that trial was displayed on the screen. If the IRT was earlier than the target interval, the square turned red and was accompanied by a brief buzzer sound. The square turned white again after the feedback, and a new interval began. Cumulative earning was always present on top of the screen during the test blocks.

To prevent explicit counting, a secondary task was used during both familiarization and DRL testing. Participants were presented with a 4-digit number at the beginning of each block, and a single-digit number was presented at the end of that block. Participants were asked if that single digit was 1 of the 4 digits observed at the beginning of that block. Participants were told at the beginning of the experiment that they would walk away with the total earnings from the timing tests multiplied by the proportion of correct responses in the working memory task. For example, if they earned \$20 from the timing portion of the experiment, but only got 50% of the memory questions correct, they would earn a total of \$10.

Rats. Forty-eight of the 60 rats were tested daily in 1-hr sessions. Twelve of the 60 rats were run nightly in three 1-hr sessions per night (with a 3-hr break in between sessions). The rats were placed in the operant chamber, and a lever was inserted. Rats were rewarded for spacing their lever presses (IRT) by at least the target interval. If a rat pressed the lever before the target interval

had elapsed, food was not delivered, and the interval started over. There were 5 groups (12 rats per group), each with a different target interval: 7, 10, 14, 28, and 56 s (the 10-s group was run several months later, and overnight). Unlike the human task, the reinforcement rate was not held constant across groups. The amount of reinforcement per session for an optimal animal ranged from about 60 (target interval = 56) to about 515 (target interval = 7) in a 1-hr session. The rats that were tested daily were fed 16 g of food in their home cages after the session, whereas the rats that were tested nightly earned all their food in the testing chambers.

Results

As with many data sets on the DRL task, our data showed a bimodal IRT distribution (one mode near 0 and the other near the target interval). Given this, there are a few approaches toward analysis (e.g., Richards et al., 1993; Sanabria & Killeen, 2008). We employed two methods. The first method follows Sanabria and Killeen (2008) and is in spirit of building a probabilistic model. We model the data as coming from two processes: one fast and untimed (e.g., impulsive), and a slower and timed. We model the fast process as an exponential distribution; animals emit untimed responses as a Poisson process with some rate κ . These make up a fraction p of all responses. The slower, timed process is modeled as an inverse Gaussian distribution as described earlier. The full model for describing the IRT distribution is

$$p(IRT = x) = p\kappa e^{-\kappa x} + (1 - p)IG(x; t, \lambda) \quad (6)$$

where $IG(x; t, \lambda)$ was given previously. The four parameters κ , p , t , and λ were fit with maximum likelihood estimation using *fminsearch* in Matlab, which implements the Nelder-Mead simplex method. We bounded the parameters within plausible limits, $p \in (0, 1)$; $\kappa \in (0, 100)$; $t \in (0, 100)$; $\lambda \in (0, 4000)$, and ran the algorithm with 50 seeds uniformly distributed over these bounds. The average fits are displayed in Table 1 (mean \pm SEM; note that γ is shown instead of λ), and two examples of individual data (1 rat, 1 human, same DRL target interval) is shown in Figure 2 and will be described in more detail below.

The second method for dealing with bimodal IRTs is to use the empirical data exactly to compute the shift necessary to obtain optimal performance. We do not dwell heavily on this method, but we include it for completeness. Here, the actual IRTs (normalized by T) are shifted by Δt until the reinforcement rate is maximized. The value of the shift gives an index of optimality. Smaller shifts indicate closer to optimal (reinforcement maximizing) performance. The shift Δt for each animal was fit using *fminsearch* (50 seeds). In general, the fitting procedure used (e.g., Nelder-Mead simplex fitting vs. BFGS quasi-newton fitting) does change the exact fitted values somewhat, but it does not change any of the conclusions drawn from the fits. We chose the Nelder-Mead algorithm here because it does not require approximating the gradient of the likelihood function.

The untimed responses are not optimal in the sense described by our model. They may be adaptive in some other sense, but we do not explore the issue here. The probabilistic model analysis (Equation 6) factors out the untimed responses so that the statistics of the timed responses can be isolated. The shift analysis uses both the timed and untimed responses. This allows us to use the entire data set to assess optimality (all the IRTs, not just the timed responses). Using more than one analysis technique helps to show that the results we obtained are robust (and not specific to our particular probabilistic model assumptions). And because our model considers the untimed responses suboptimal, the shift analysis helps determine how heavily these untimed responses impact performance.

For most of the analyses shown here, the last 10 sessions from each rat were pooled. This resulted in $1,897.04 \pm 107.87$ (mean \pm SEM) responses per rat. Each human was only run for a single session. This resulted in 266.07 ± 27.05 responses. For the analyses during learning (rats only), each session was fit individually. Occasionally, the rats made no lever presses for long periods of time (e.g., they were drinking or sleeping). These outliers do not represent task performance and can heavily skew the fits. They were removed using a cutoff for responses greater than $q_3 + w(q_3 - q_1)$, where q_1 and q_3 are the first and third quartiles, respectively. For normally distributed data, $w = 1.5$ is

Table 1
Parameter Estimates of Equation 6

Rats Schedule	$n = 12$ 7	$n = 12$ 10	$n = 11$ 14	$n = 10$ 28	$n = 12$ 56	$F(4, 56)$	p
p	0.29 (0.04)	0.21 (0.02)	0.17 (0.04)	0.17 (0.02)	0.22 (0.03)	2.15	.09
κ	10.75 (0.88)	9.31 (1.05)	7.56 (1.90)	2.57 (1.20)	9.18 (3.44)	2.37	.15
t	8.54 (0.19)	11.2 (0.16)	17.0 (0.53)	30.9 (1.22)	51.8 (3.31)	121.7	9.7×10^{-26}
γ	0.23 (0.02)	0.19 (0.01)	0.32 (0.04)	0.40 (0.03)	0.42 (0.02)	15.5	2.0×10^{-8}
Humans Schedule	$n = 4$ 5	$n = 4$ 8	$n = 4$ 10	$n = 1$ 12	$n = 1$ 15	$F(4, 13)$	p
p	0.02 (0.01)	0.003 (0.003)	0.008 (0.004)	3.4×10^{-11} (n/a)	0.007 (n/a)	1.27	.35
κ	26.5 (24.5)	20.9 (13.6)	16.84 (14.75)	87.0 (n/a)	0.001 (n/a)	0.90	.50
t	5.78 (0.18)	9.92 (0.96)	11.92 (0.21)	14.37 (n/a)	19.07 (n/a)	34.8	1.8×10^{-5}
γ	0.13 (0.02)	0.16 (0.06)	0.19 (0.01)	0.22 (n/a)	0.20 (n/a)	0.52	.72

Note. The rows are the parameters (the proportion of untimed responses, p ; the rate parameter of the untimed responses, κ ; the mean of the timed responses, t ; and the coefficient of variation γ). The first five columns are the groups (the DRL schedule) and the elements in those columns are the mean (\pm SEM) parameter estimates. The sixth column is the F statistic of the analysis of variance comparing the estimate over groups, and the last column is the p value of the statistics (effect sizes are given in the text). The table is split by rats (top) and humans (bottom).

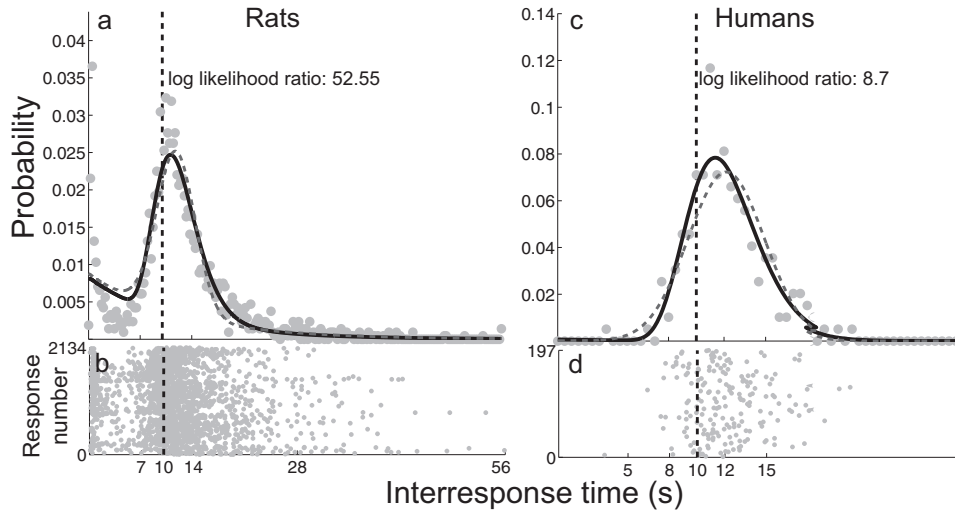


Figure 2. Example subjects. Panels a and b (first column) show the interresponse times of a single randomly chosen rat from the DRL 10 second group (indicated by the vertical dashed line at 10). Panel b (bottom left) shows every raw interresponse time (x-axis) as a function of response number (y-axis) for the last 10 sessions of this rat. Panel a (top left) shows the probability density computed from the raw interresponse times shown in panel b. The solid curve through the data is the best fitting exponential-inverse Gaussian mixture distribution, and the dashed curve through the data is the best fitting exponential-Gaussian mixture distribution. Panels c and d (right column) show a randomly chosen human from the DRL 10-s group. Positive likelihood ratios indicate that for these two subjects, the inverse Gaussian model fit the data better than the Gaussian model.

often used, which results in a criterion about 2.7σ higher than the mean. For our data, we tripled this to $w = 4.5$ to exclude on the most extreme IRTs.

For completeness, we also fit the mixture model assuming Gaussian distributed timed responses. Examples of the fits are shown in Figure 2. The inverse Gaussian fit the data better in 47/57 (82.5%) rats; the log-likelihood ratios significantly favored the inverse Gaussian assumption, $t(56) = 6.04$, $p \leq .001$, $M = 44.76 \pm 7.4$, $d = 0.81$. In humans, the inverse Gaussian fit the data better in 14/14 (100%), and the log-likelihood ratios significantly favored the inverse Gaussian assumption, $t(13) = 3.86$, $p \leq .001$, $M = 9.38 \pm 2.4$, $d = 1.07$.

Descriptive Results

Raw IRT data, and the resulting estimated probability distributions, are shown in Figure 2. The first column (Figure 2, a and b) shows a single rat for which the target interval was 10 s. Panel b (bottom) shows every IRT from the last 10 sessions as a gray dot (schedule shown in black). These were the data used in the maximum likelihood fits described earlier (each subject was fit individually). Panel a (top) shows the same data plotted as a probability distribution, along with the IG fit (black) and Gaussian fit (dark gray dashed). A single human participant is shown in panels c and d (second column). Both the human and rat examples were drawn randomly from the same target interval (10 s). These random examples show a strong similarity between the IG and Gaussian fits, although this is not true for all subjects (the IG fits significantly better, as described earlier). In some of the analysis that follows, the normalized IRT is used. This refers to t/T , where t is the estimate of the mean of the IG component of the distribu-

tion, and T is the target interval. It is reasonable to think of this as simply scaling the entire raw IRTs by T . Note that it is inappropriate to use a z-scores to normalize the data because it changes the variability of the distribution—the exact thing that needs to vary across subjects to test the hypothesis that the mean of the distribution precisely correlates with the variability.

Learning

We assessed learning over sessions by evaluating power law regressions of the form $\log y = \beta_0 + \beta_1 \log x$, where y is the parameter of interest (p , γ , and t), and x is session number. The value of β_0 is the constant (in logs) of the regression, and a β_1 significantly different from 0 indicates a change in $\log y$ over sessions. We used a mixed-effects model letting β_0 include both fixed and random effects (β_1 was fixed for a group), although this assumption is not necessary for reproducing our main results.

The coefficient of variation γ significantly decreased over sessions for all groups of rats ($\min \beta_1 = 0.07 \pm 0.02$), $t(408) = -3.26$, $p \leq .005$. That is, over sessions, the width of the timed responses became narrower. Colloquially, the rats became better at timing their responses. The mean \pm SEM over sessions for each group are shown in Figure 3a.

The proportion of untimed responses p decreased over sessions in three of the five groups of rats, 7 s: $\beta_1 = -0.25 \pm 0.04$, $t(490) = -6.40$, $p < 10^{-10}$; 14 s: $\beta_1 = -0.41 \pm 0.07$, $t(449) = -5.67$, $p < 10^{-7}$; 28 s: $\beta_1 = -0.23 \pm 0.08$, $t(408) = -3.4$, $p < 10^{-3}$. But p did not significantly change for DRL 10- and 56-s groups, 14 s: $\beta_1 = 0.006 \pm 0.13$, $t(538) = 0.04$, $p = 0.96$; 56 s: $\beta_1 = -0.09 \pm 0.07$, $t(490) = -1.20$, $p = .23$. The mean \pm SEM over sessions for each group is shown in Figure 3b.

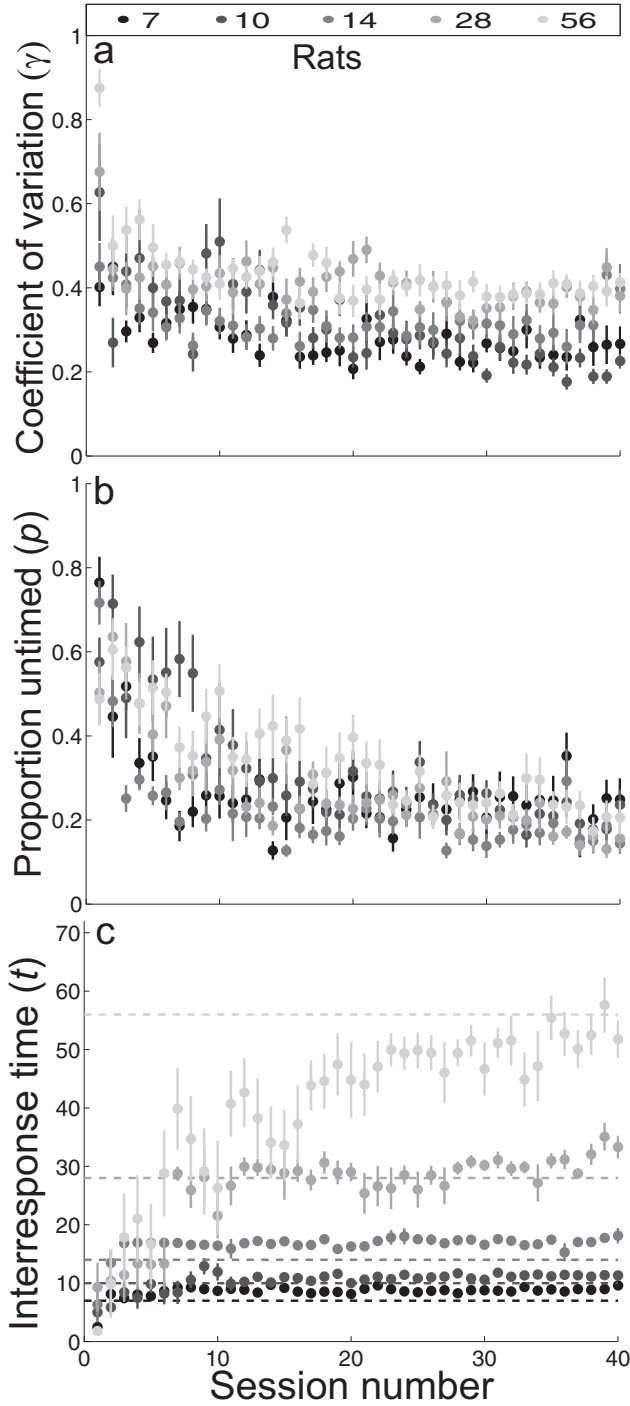


Figure 3. Learning in rats. Panel a (top) shows the mean coefficient of variation for each of the five groups of rats. Each data point was computed by fitting the exponential-inverse Gaussian model to every session for every rat, and then averaging over rats in a group for a given session (with standard error bars). Panel b (middle) shows the proportion of untimed responses declined over sessions. And panel c (bottom) shows the mean interresponse time from the inverse Gaussian fit over sessions. For shorter schedules (DRL 7 s, 10 s, and 14 s) the rats learn to withhold responding within a few sessions. For the longer schedules (DRL 28 s and 56 s), the rats take considerably longer to learn to withhold responding and it is unclear that they ever reached steady state (see text for details).

The mean of the timed response times t significantly increased over sessions. For the shorter DRL intervals (7 s, 10 s, and 14 s) the sessions coefficient hovers around 0.27 (7 s: $\beta_1 = 0.25 \pm 0.03$; 10 s: $\beta_1 = 0.30 \pm 0.03$; 14 s: $\beta_1 = 0.27 \pm 0.03$, all $p < 10^{-17}$). For the two longer DRL intervals, the sessions coefficient is much higher with the DRL 28-s group at $\beta_1 = 0.81 \pm 0.06$ ($p \leq 10^{-32}$) and the DRL 56-s group at $\beta_1 = 1.12 \pm 0.06$ ($p \leq 10^{-52}$). The mean \pm SEM over sessions for each group is shown in Figure 3c. The figure and the analysis suggests that it takes significant time for rats to learn longer intervals. It is possible that the rats in the DRL 28- and 56-s groups did not fully reach steady state.

Steady State

The proportion of untimed responses p did not significantly differ by DRL schedule in either rats or humans, rats: $F(4, 56) = 2.15$, $p = .09$, $\eta^2 = 0.14$; humans: $F(4, 13) = 1.27$, $p = 0.35$, $p = 0.35$, $\eta^2 = 0.36$.

There were no significant differences in the rate parameter of untimed responses κ in either rats or humans, rats: $F(4, 56) = 2.37$, $p = .06$, $\eta^2 = 0.15$; humans: $F(4, 13) = 0.90$, $p = .50$, $\eta^2 = 0.29$. Note, however, that the κ estimates in humans have large SEs, probably because it is difficult to properly identify the κ parameter when the proportion of untimed responses is very small.

For humans, there were no significant differences in the coefficient of variation γ , $F(4, 13) = 0.52$, $p = .73$, $\eta^2 = 0.18$. But there were differences in rats, $F(4, 56) = 15.5$, $p < 2 \times 10^{-8}$, $\eta^2 = 0.54$. Post hoc tests revealed that this was because the rats with longer DRL intervals (DRL 28 s and 56 s) differed significantly from the DRL 7-s and DRL 10-s groups. The DRL 10-s group also differed from the DRL 14-s group. The means \pm SEM are shown in Table 1.

Of course, the fact that the t parameter differs across groups is trivial because both rats and humans were grouped by the target interval. An important prerequisite for optimal performance is that the IRTs be greater than the target interval. This is true for the rats in the DRL 7-s, 10-s, and 14-s groups, smallest effect in the DRL 14-s group, $t(10) = 5.63$, $p \leq .001$, $d = 1.78$, and every rat in these three groups had a mean IRT t greater than the schedule. The DRL 28-s group showed a marginally significant result, $t(9) = 2.38$, $p = .041$, $d = 0.79$, and 20% (2/10) of the rats had a mean IRT less than 28 s. The IRTs in the DRL 56-s group did not significantly differ from 56 s, $t(11) = -1.27$, $p = .23$, $d = -0.38$, and 75% (9/12) of the rats had a mean IRT less than 56 s. Figure 4b shows that the normalized IRT (t/T) tends to decrease as the interval increases (black dots; the figure will be described in more detail). That is, the longer the target interval, the less likely it is that rats withhold responding for the entire interval. This may be because of the incomplete learning in those groups, as described earlier, or a natural feature of longer DRL intervals (e.g., Richardson & Loughhead, 1974).

Every human in our study had mean IRTs greater than their target interval. The sample size in two of the groups (12 s and 15 s) is too small for group level statistics ($n = 1$). The participants in the 5- and 10-s groups had mean IRTs significantly greater than the schedule, minimum in 5-s group: $t(3) = 4.32$, $p = .02$, $d = 2.49$. In the 8-s group, one of the subjects had an IRT much higher than the others, which resulted in an insignificant t test, $t(3) = 2.0$,

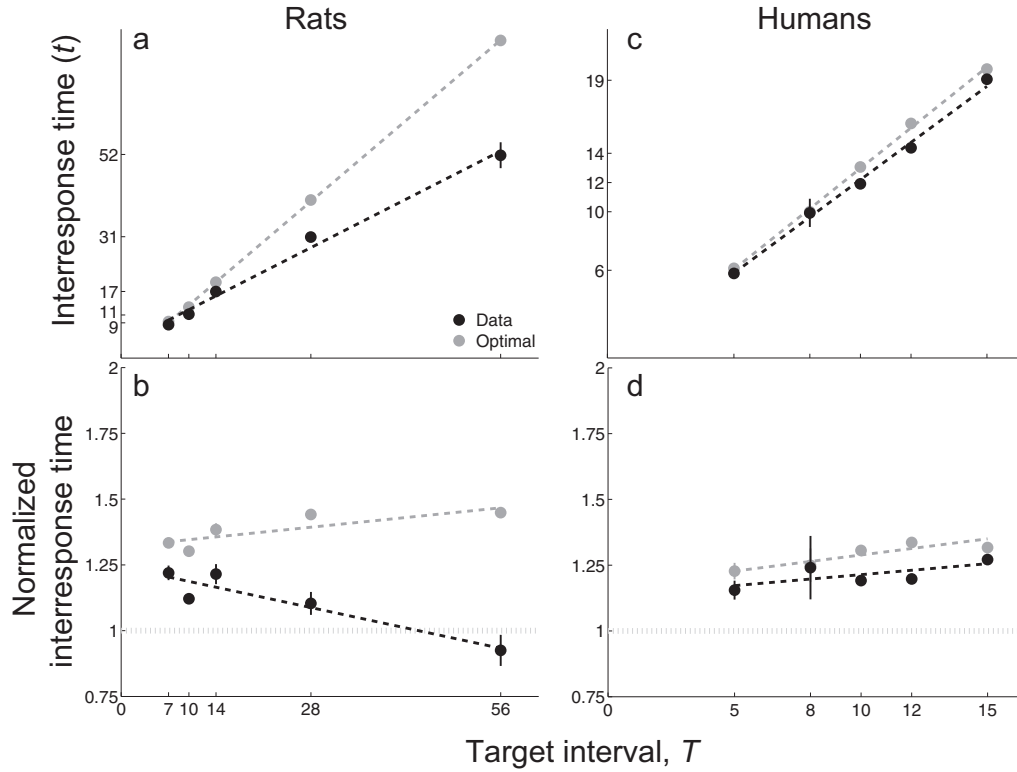


Figure 4. Interresponse times by schedule. The black dots in top row (panels a and c) show the mean interresponse times, from the inverse Gaussian fit, averaged over rats in a group (mean \pm SEM). The black dashed line is the best fitting line. The gray dots and dashed line show the optimal interresponse time for that group (the average of the optimal for each rat in a group). The bottom row (panels b and d) shows the same thing, except normalized by the schedule. The left column shows rats, the right column shows humans. Note that both the x- and y-axes are different on the four plots.

$p = .14$, $d = 1.15$. The mean IRTs for the four participants in this DRL 8-s group are 8.12, 9.39, 9.43, and 12.68.

Because the rats in the DRL 28- and 56-s groups (a) tended to have a higher coefficient of variation γ , (b) tended to respond earlier than their schedule, and (c) likely failed to asymptote (as indicated by the large session coefficient in the power regression), we do not expect them to be optimal in the sense described previously. Indeed, Figure 4 shows them deviating substantially from the optimal IRTs (black dots compared with the gray dots). For completeness, we will continue to analyze their data alongside the others, but because of these caveats, we will show the single subject data from these rats in open circles rather than closed. We will discuss this further in the Discussion section.

Optimality Results

Figure 4 shows the mean IRTs for each group (black dots) and the associated optimal IRT (gray dots). The top panels show the raw (nonnormalized) means, and the bottom panels show the normalized means (t/T). The dashed lines show the least squares linear fit. For humans (Figure 4, c and d) the IRTs closely track the optimal IRTs. For rats (Figure 4, a and b), this is true for shorter intervals, but the result falls apart for longer intervals, where the normalized IRT deviates heavily from optimality (in Figure 4a, the

slopes are different, and in Figure 4b the slope of the black line is negative). Humans were not tested at these longer intervals, so it is an open question whether their performance would also break down in the same way. In all groups, the mean IRT is shorter than optimal; that is, the response rate is faster than optimal. We will unpack these results.

The critical feature of optimality discussed above is that, to maximize reinforcement rate, animals should account for the variability of their responses. Figure 5 shows the normalized IRT plotted against the coefficient of variation for every rat (Figure 5a) and human (Figure 5b) in our experiment. The optimal performance curve is shown as a black line. The dashed line shows how much the optimal performance curve has to shift in order to obtain the least squares fit to the data (rats: -0.15 ± 0.01 , $R^2 = 0.59$; humans: -0.07 ± 0.02 , $R^2 = 0.66$). The fit has a single parameter τ . Note that this is similar but not identical to Δt . Here, we shift the optimal curve until it fits the data, while the analysis that produced Δt shifts the raw data (both timed and untimed responses) until the reinforcement rate is maximized. Across human participants, the mean IRT is highly correlated with the coefficient of variation, $F(1, 12) = 54.55$, $p \leq 8.5 \times 10^{-5}$, $R^2 = 0.82$. This is also true for shorter DRL intervals in rats, $F(1, 33) = 57.31$, $p \leq 1.1 \times 10^{-8}$, $R^2 = 0.63$; pooled over the 7-, 10-, and 14-s groups). When longer DRL intervals are

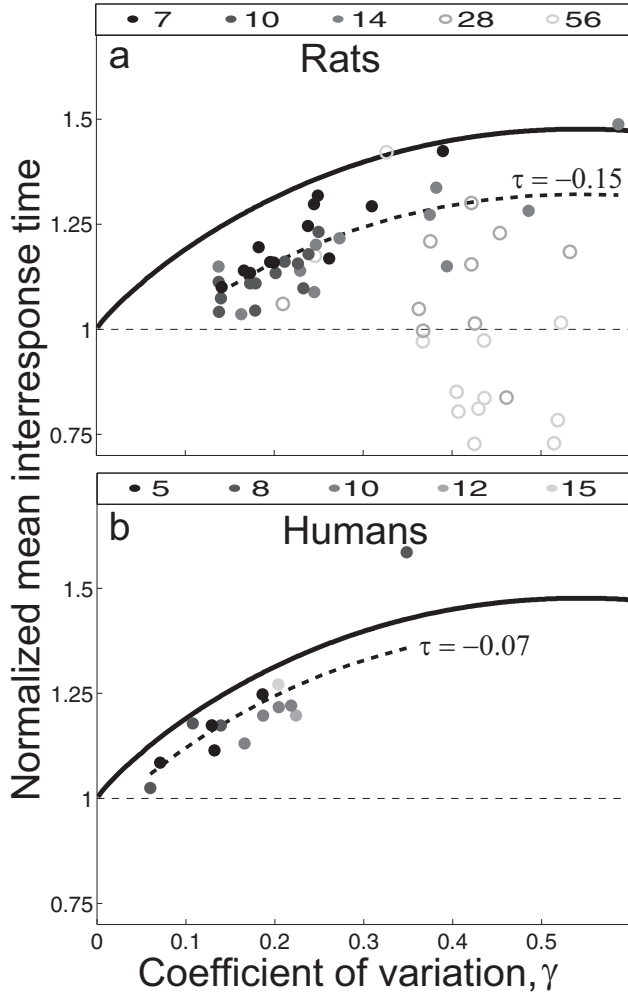


Figure 5. Interresponse time as a function of the coefficient of variation. Panel a (top) shows the normalized interresponse time from the fit against the coefficient of variation for every rat in the experiment. The dashed horizontal line shows the normalized target interval. The solid curve is the optimal performance curve (i.e., it is the same curve as the dashed optimal performance curve in Figure 1b, just displayed on different axes). The dashed curve through the data shows how far the optimal performance curve has to shift in order to fit the data (it is an index of optimality such that values closer to zero are closer to optimal performance). The open circles show the rats from the DRL 28-s and 56-s groups, because their performance differs from the performance of rats in shorter intervals in many ways (see text for details). Panel b (bottom) shows the results for humans.

included, there is no significant main effect of schedule, $F(1, 55) = 2.11$, $p = .15$, $R^2 = 0.04$. This result is clearly seen in the figure. The rats in the longer DRL schedules cannot be optimal in the sense we described previously, even in principle, because there is no reliable relationship between the mean and variability of their IRT distribution. But for the humans and the rats with shorter DRL schedules, the mean and variability are correlated as prescribed by the model.

From a practical point of view, it is probably more important whether the mean IRTs tracked the optimal IRT across animals (Figure 6). Across human participants, the mean IRTs were highly

correlated with the optimal IRTs, $F(1, 12) = 28.78$, $p \leq 2.0 \times 10^{-4}$, $R^2 = 0.71$. In rats, this was also true in the shorter DRL schedules, $F(1, 33) = 50.02$, $p \leq 4.3 \times 10^{-8}$, $R^2 = 0.60$, but not the longer ones ($p = .23$). Of course, because the optimal IRT depends on the variability, finding a significant regression coefficient for both γ and t_o does not give completely independent pieces of evidence that animals track optimality, but they are not completely redundant measures either because their relationship is nonlinear.

The previous analyses tell us that both humans and rats (at shorter intervals) have mean IRTs that depend on (are highly correlated with) the variability of those IRTs. But that does not mean they are maximizing reinforcement rate; it only suggests

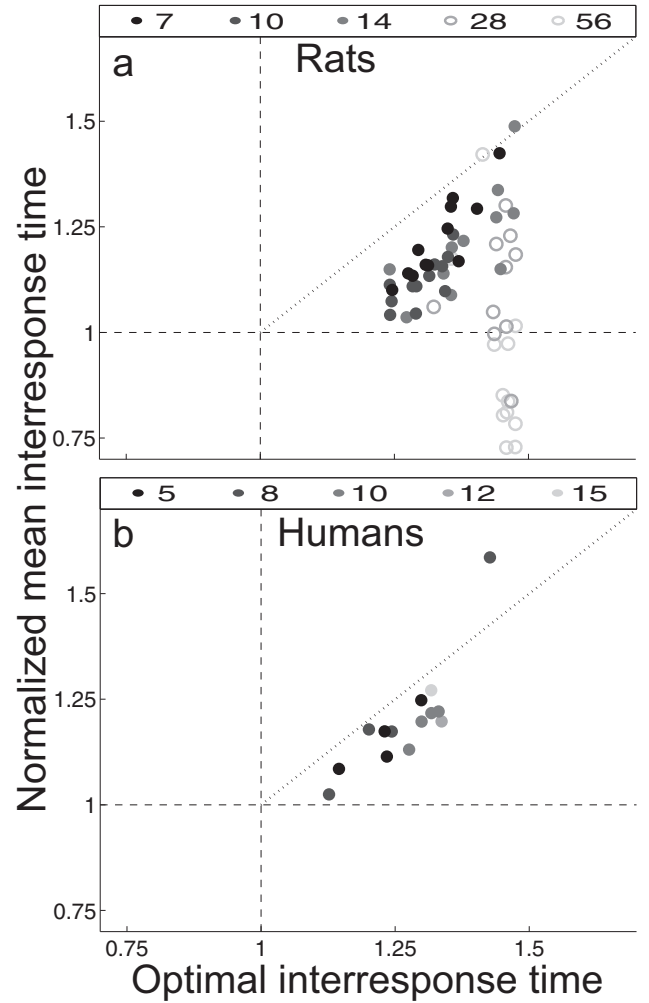


Figure 6. Interresponse time as a function of the optimal interresponse time. Panel a (top) shows the normalized interresponse time as a function of the optimal interresponse time for every rat in the experiment. The vertical and horizontal dashed lines show the target interval at 1 arbitrary units. The diagonal line is the identity line. Points below this line show faster than optimal responses, and points above this line show slower than optimal responses (points on the line show optimal responses). The rats with longer target intervals (DRL 28 and 56 s) are still shown as open circles. Panel b shows the results for humans. The normalized interresponse times increases as the optimal increases. But nearly all data points are below the diagonal line.

they could be. To check for optimality, we have to assess how close the mean IRTs are to the optimal, and how much the animals earn (their reward rate) against the optimal reward rate they could earn given their variability (recall that different levels of variability lead to different reinforcement rates even for rate maximizing animals). We define the temporal distance ratio as the mean IRT divided by the optimal IRT: t/t_o . Note this can take on any positive value. Numbers less than one indicate the animal's IRT is shorter than it should be, and their response rate is too high. Numbers higher than one indicate the opposite.

Nearly all rats and humans had faster-than-optimal IRTs. This was significant in every group of rats, minimum effect size in DRL 14 s, $t(3) = -6.48$, $p \leq 7.1 \times 10^{-5}$, $d = -0.65$. This was true for all groups of humans, minimum effect size in DRL 5 s, $t(10) = -4.47$, $p \leq .02$, $d = -1.49$, except for the DRL 8 s, $t(3) = -0.31$, $p = .78$, $d = -0.10$. Recall that the DRL 8-s group had a participant with much longer IRTs than the others. In humans, the temporal distance ratio did not significantly depend on the schedule, $F(1, 12) = 0.19$, $p = .67$, $R^2 = 0.02$. For shorter intervals in rats, the ratio didn't significantly depend on the schedule, either, $F(1, 33) = 2.87$, $p = .1$, $R^2 = 0.08$. But when the longer intervals were included for rats, the result was highly significant, $F(1, 55) = 71.77$, $p \leq 1.5 \times 10^{-11}$, $R^2 = 0.57$. This again points to a key difference between the rats in the shorter and longer intervals. In fact, across all animals (human and rats) only 3 of 71 animals (4.2%) had a mean IRT greater than or equal to the optimal. For rats, the temporal distance ratios (mean \pm SEM, median \pm IQR (interquartile range) in parenthesis) were 7: 0.91 ± 0.01 (0.91 ± 0.05), 10: 0.86 ± 0.008 (0.86 ± 0.03), 14: 0.88 ± 0.02 (0.88 ± 0.09), and 28: 0.77 ± 0.03 (0.80 ± 0.14), 56: 0.64 ± 0.04 (0.58 ± 0.14). For humans, the temporal distance ratios were 5: 0.94 ± 0.01 , (0.95 ± 0.03), 8: 0.99 ± 0.04 (0.96 ± 0.12), 10: 0.91 ± 0.008 (0.92 ± 0.02), 12: 0.90 (0.90), 15: 0.97 (0.97).

Animals that have close to optimal IRTs will have a higher reinforcement rate and gain more food. But how much they gain depends on the curvature of the gain surface (how drastically changes in IRTs change the reinforcement rate). Animals with higher variability have shallower reinforcement rate functions (Figure 1); relatively large changes in their IRT may result in relatively small changes in their overall reinforcement rate. We defined the gain ratio as the reinforcement rate earned (based on the fits of the timed IRT distribution) divided by the maximum reinforcement rate for that particular subject, g/g_o . Unlike the temporal distance ratio, the gain ratio is bound by 0 and 1 (subjects cannot do worse than nothing or better than the best).

In general, the gain ratio does not significantly depend on the schedule in humans or rats, humans: $F(1, 12) = 0.15$, $p = .70$, $R^2 = 0.01$; rats: $F(1, 33) = 0.53$, $p = 0.47$, $R^2 = 0.01$, except, like before, when the longer interval schedules are included for rats, $F(1, 55) = 43.14$, $p \leq 1.92 \times 10^{-8}$, $R^2 = 0.44$. For rats, the gain ratios were 7: 0.95 ± 0.01 (0.95 ± 0.06), 10: 0.88 ± 0.02 (0.89 ± 0.08), 14: 0.96 ± 0.02 (0.96 ± 0.07), and 28: 0.86 ± 0.03 (0.87 ± 0.15), 56: 0.69 ± 0.05 (0.62 ± 0.26). For humans, the gain ratios were 5: 0.96 ± 0.02 (0.96 ± 0.06), 8: 0.92 ± 0.046 (0.98 ± 0.14), 10: 0.96 ± 0.01 (0.97 ± 0.03), 12: 0.96 (0.96), 15: 0.995 (0.995). It is perhaps interesting to note that, even though the performance in the DRL 28-s group tended to be quite different from the performance with shorter intervals (as described at length earlier),

the rats still obtained $86 \pm 3\%$ of the best they could have gained. There is no rigorous way to say what a good gain ratio is, partly because there are no suitable null-hypotheses, so we leave it to our reader to decide whether $\sim 90\%$ in rats and $\sim 95\%$ in humans is suitably close to the optimal gain of 100%.

But to give some intuition, we can compare the gain ratio the subjects received with the gain ratio they would have received if they had responded at the schedule. The DRL task requires that subjects respond at or after the target interval to be reinforced. But it does not require that the subjects actually do so (as shown by the DRL 56-s group) or even that a substantial proportion of responses are longer than the target interval (both the DRL 28-s and 56-s group). The gain ratio at the schedule is the baseline we would expect from a subject who learned the interval and chose an optimal mean IRT without regard to the response time variability. We define the marginal gain ratio as how much better the subjects do by having a mean IRT that depends on (is correlated with) their variability as $(g - g_T)/(g_o - g_T)$, where g_T is the gain the subject would have obtained given their actual coefficient of variation, but with a mean at the schedule. This value must be smaller than the gain ratio because g_T is positive, but how much smaller depends on both how close the subject is to the optimal IRT and how shallow the curvature of the reinforcement rate function is (which depends on how large the coefficient of variation is). For rats, the marginal gain ratios were 7: 0.86 ± 0.03 (0.86 ± 0.19), 10: 0.64 ± 0.05 (0.68 ± 0.24), 14: 0.75 ± 0.07 (0.85 ± 0.27), and 28: 0.34 ± 0.17 (0.48 ± 0.69), 56: -0.53 ± 0.24 (-0.85 ± 1.08). For humans, the marginal gain ratios were 5: 0.89 ± 0.05 (0.92 ± 0.15), 8: 0.81 ± 0.14 (0.93 ± 0.32), 10: 0.87 ± 0.04 (0.91 ± 0.08), 12: 0.86 (0.86), 15: 0.98 (0.98). There are two conclusions. The trivial conclusion is that overall both the rats and humans still maintain a relatively high ratio indicating closeness to optimality. The more interesting conclusions are that both the DRL 10-s and 28-s group's marginal gain is quite a bit lower than their gain ratio. This indicates those groups either had relatively earlier mean IRTs (Figure 4 suggests they did) or the curvature of the reinforcement rate function is shallow, or both (the likely culprit for the DRL 28-s group). On the whole, these results suggest that the rats were about 75% better and humans about 85% better than if they had optimized their mean response time without regard to their variability.

Last, instead of using the parameters of the fit to assess optimality, we shifted the raw (normalized) IRTs until the resulting reinforcement rate was maximized. This analysis includes all the IRTs, including the untimed responses. For rats, the shifts were 7: 0.13 ± 0.01 (0.12 ± 0.06), 10: 0.16 ± 0.01 (0.16 ± 0.05), 14: 0.22 ± 0.03 (0.22 ± 0.12), and 28: 0.36 ± 0.08 (0.31 ± 0.21), 56: 0.66 ± 0.10 (0.67 ± 0.59). For humans, the shifts were 5: 0.05 ± 0.01 (0.05 ± 0.03), 8: 0.05 ± 0.01 (0.05 ± 0.04), 10: 0.09 ± 0.01 (0.08 ± 0.02), 12: 0.14 (0.14), 15: 0.015 (0.015). To give some intuition about what a shift of, for example, 0.15 means, note that the optimal shift is zero and the maximal positive shift is one (because at that point, every single data point is greater than the normalized schedule of one). For a shift of 0.15, this amounts to needing to shift the IRTs by 15% to maximize the reinforcement rate. In humans, the shift did not significantly depend on the schedule, $F(1, 12) = 0.62$, $p = .45$, $R^2 = 0.05$. But it did in rats, $F(1, 55) = 64.86$, $p \leq 8 \times 10^{-11}$, $R^2 = 0.54$, even when the longer schedules were not included, although with a smaller effect size, $F(1, 33) = 10.6$, $p \leq .01$, $R^2 = 0.24$. The

shift is significantly positive in all groups in both rats and humans, smallest effect size in rats (DRL 56 s): $t(11) = -3.41$, $p \leq .01$, $d = -0.31$; humans (DRL 8 s): $t(3) = -84.09$, $p \leq 4 \times 10^{-6}$, $d = -28.03$). These results show what we already showed previously in various analyses: that the longer the schedule, the more suboptimal the IRTs. In fact, this is just another way to measure the bias optimal performance curve shown in Figure 5. But in Figure 5, we used the parameter fits to compute it, and here we used the raw data.

Discussion

Our model was designed to be as simple as possible. The model is a set of three key assumptions. The first is that animals try to maximize reinforcement rate. This imposed the restriction that our model use the equation for the reinforcement rate (Equation 1). The second assumption is that animals do not control their variability. This imposed the restriction that γ could not be optimized. The last assumption is that animals do control their mean IRT. Together, these three assumptions imposed the final form of the model: maximize the reinforcement rate equation with respect to t constrained by γ .

We also assumed, for simplicity, that the subjective value of a reinforcement does not change over time. This restriction can be relaxed—we do not see it as a key assumption of our model—but we do not do so here because it requires adding assumptions about how subjective value changes over time, and would likely add free parameters to our model.

Nevertheless, our model captures some key aspects of performance on the task. The model gives the mean IRT that maximizes the reinforcement rate given the coefficient of variation. This specifies an ideal (reinforcement-maximizing) relationship between the mean and the variability. This relationship was robust in our data (except for the longer intervals in rats). And the optimal IRT (t_o) predicted the actual IRTs quite well in a simple regression ($R^2 > 0.60$). The model generates an optimal benchmark for both the mean IRT and the reinforcement earned, and the animals in our experiment, both humans and rats, were roughly $\sim 90\%$ or more toward that optimal benchmark.

To see whether the same results were obtained in different labs, we performed simple secondary data analysis on DRL data sets in rats from three different labs (Orduña, Valencia-Torres, & Bouzas, 2009; Sanabria & Killeen, 2008; Taylor & Balsam, 2009). In the combined data set, the target intervals were 5 s, 10 s, and 16 s. And in each case, the performance was close to optimal; the rats median gain ratios were 93%, 94%, and 92%, respectively.

What Does a Generative Model Give Us That an Empirical Description Cannot?

We noted throughout this article that there were at least two ways to go from the raw IRTs to measures that can be compared with the model's predictions. The first way is to use the raw IRT distribution and shift it until reinforcement rate is maximized. This method is computationally efficient and has the nice property of working directly with the actual IRTs at all times. The second, and our focus, was to build a generative model of the IRT distribution, fit it to the raw IRTs, and analyze the parameters of the fit. The drawback of this approach is that parameter estimates are analyzed, rather than the raw data itself.

But what we gain are summary measures of the data that give rise to theoretical questions. Our data, and others like it, suggested two processes generated the data. A fast untimed one, which we modeled as an exponential distribution, and a slow timed one, which we modeled as an inverse Gaussian distribution. The inverse Gaussian distribution fit our data better than the Gaussian. This is in line with the recent push to bring the drift-diffusion model of perceptual decision making into the timing domain (Simen, Balci, de Souza, Cohen, & Holmes, 2011; see also Brunton, Botvinick, & Brody, 2013 for a different model in a related context). Here, the decision to respond is based on accumulating evidence toward a single response threshold (similar to the way Scalar Expectancy Theory is conceptualized; Gibbon, Church, & Meck, 1984), and where the accumulation rate is related to the target interval (similar to the way the behavioral theory of timing (BET) is conceptualized; Killeen & Fetterman, 1988; Simen et al., 2013). The drift-diffusion model predicts inverse Gaussian response times (Luce, 1986). The inverse Gaussian assumption is also used as the IRT distribution in Modular Theory (Guilhardi, Yi, & Church, 2007).

Fitting a generative model to our IRT data also allowed us to consider the fast, seemingly untimed responses in a principled way (as others had; e.g., Richards et al., 1993; Sanabria & Killeen, 2008). We modeled these as an exponential distribution that embodies the notion that animals emit operant responses at a random rate. Informally, we did not find anything that predicted the random rate parameter at a subject specific level, and it did not significantly change based on the target interval. These responses are not optimal in the sense described by our model because all they do is reduce the probability of reinforcement and skew the total mean IRTs toward even faster-than-optimal response rates. But that does not mean they are completely without purpose. The simplest explanation may be that they are exploratory responses. A few, quick responses may be cheap compared with the potential loss incurred by failing to detect a change in the environment (Cohen, McClure, & Yu, 2007). Of course, our data cannot rule out the hypothesis that they are impulsive and maladaptive, nor can we rule out the hypothesis that animals just have a tendency to respond in stereotypical bouts in many task settings (Kirkpatrick, 2002), which is simply too engrained to easily overcome.

Deviations From Our Optimal Model

There are two glaring deviations from optimality in our data set. The first is that the rats with longer target intervals show no signs of optimality. Their mean IRTs were not significantly correlated with their variability, and they did not seem to track the optimal IRT. For many of the measures we discussed, pooling the rats in the longer DRL schedules in with the shorter target schedules resulted in a skewed picture of the performance in shorter schedules. Our results suggest the rats should be split into two groups: those with shorter target intervals that showed the necessary characteristics of optimality, and those with longer target schedules that showed no signs of optimality. Why this seemingly arbitrary split? We speculate that the culprit is buried in the learning results: It took the rats in the DRL 28-s and 56-s group far longer to learn to withhold responding. Indeed, it is unclear whether they ever reached asymptote in the 41 days of training in our experiment. Given that the rats in those groups were rewarded several hundred times during the experiment (28: 607 ± 23 , 56: 222 ± 19), and

that learning temporal intervals often requires only a few trials (Drew, Zupan, Cooke, Couvillon, & Balsam, 2005), we find it unlikely that those rats did not adequately learn the fixed DRL requirement, although we cannot rule it out. Instead, we believe the issue is that it takes considerable time to learn to withhold a response that results in reinforcement, and there may be schedules for which this is extremely difficult (e.g., Doughty & Richards, 2002; Richardson & Loughhead, 1974). Machado and Rodrigues (2007) present a process model, designed for a similar task, in which two of the parameters together determine how much a reinforcement will change behavior. According to this view, the longer schedules in our study may result in reinforcement that is less effective at driving behavioral change.

The second glaring deviation of our data set is that 68 of the 71 subjects responded at a faster than optimal rate. Figures 4–6 suggest that simply subtracting a constant from the optimal time predicts performance relatively well. The constant we fit depended on species (rats: -0.15 ± 0.01 ; humans: -0.07 ± 0.02) but was a strikingly reliable result in our data. This is in contrast to our recent article that showed no systematic deviation from optimality in humans (Çavdaroglu et al., 2014). There are a few methodological differences between the studies. For instance, the participants in that study were not given any information about the task structure, nor were they pretrained with any intervals. Furthermore, the participants in that study were trained for 8 sessions with the same target interval, while the participants in the present study were only trained for a single session. We cannot rule out the possibility that the bias would have decreased or disappeared in these data as well. Further studies are needed to determine the sources of these systematic deviations from optimality.

Conclusion

The take-home message is that animals choose a response rate that largely maximizes reinforcement rate on this task, except that they are consistently faster than optimal. Our simplistic model is a good starting place for more in-depth models that can account for these deviations. For example, it may be that animals put a premium on speed over accuracy in this task that simply did not go away with training, unlike in other tasks (Balci et al., 2011; Simen et al., 2009).

We chose to remain silent about the underlying mechanism(s) that might give rise to the behavior. In Marr's (1982) levels of analysis, our model is firmly in the computational camp. Our model specifies a set of goals and constraints for an organism (maximizing the reinforcement rate given timing variability). These assumptions give rise to precise predictions about what we should see in the averaged data. There are likely to be many algorithmic (e.g., process) models that could give rise to both the optimal behavior and the actual behavior. For example, a reinforcement learning approach would posit that the animal's IRTs are adjusted by hill-climbing a reinforcement rate gradient (Sutton & Barto, 1998). A representational approach would posit that the animal keeps track of the task-relevant parameters (i.e., the target interval, and the statistics of the IRT distribution) and computes the optimal aim point (Trommershäuser et al., 2006). In each of these approaches, there are several ways to formulate the problem. And, of course, there are other approaches altogether. In

turn, these algorithmic models constrain and are constrained by the underlying neural circuitry.

In contrast, our goal was to further explore the notion that both rats and humans choose a response rate that maximizes the reinforcement rate (regardless of the mechanism). The DRL task is particularly well designed to tackle this problem because the task is relatively simple and the IRT statistics are well defined, making it straightforward to write out the equations that relate the response rate to the reinforcement rate. Here, we showed that both rats and humans show some of the key characteristics of optimality. Expanding this approach to other tasks, as well as focusing on the deviations from optimality in this task, may give insight into the underlying mechanisms that give rise to response rates.

References

- Balci, F., Freestone, D., & Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences of United States of America*, 106, 2459–2463. <http://dx.doi.org/10.1073/pnas.0812709106>
- Balci, F., Freestone, D., Simen, P., Desouza, L., Cohen, J. D., & Holmes, P. (2011). Optimal temporal risk assessment. *Frontiers in Integrative Neuroscience*, 5, 56.
- Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J., Holmes, P., & Cohen, J. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception & Psychophysics*, 73, 640–657.
- Baum, W. M. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior*, 20, 137–153. <http://dx.doi.org/10.1901/jeab.1973.20-137>
- Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, 22, 231–242. <http://dx.doi.org/10.1901/jeab.1974.22-231>
- Baum, W. M. (1981). Optimization and the matching law as accounts of instrumental behavior. *Journal of the Experimental Analysis of Behavior*, 36, 387–403. <http://dx.doi.org/10.1901/jeab.1981.36-387>
- Blough, D. S. (1967). Stimulus generalization as signal detection in pigeons. *Science*, 158, 940–941. <http://dx.doi.org/10.1126/science.158.3803.940>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. <http://dx.doi.org/10.1163/156856897X00357>
- Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340, 95–98. <http://dx.doi.org/10.1126/science.1233912>
- Çavdaroglu, B., Zeki, M., & Balci, F. (2014). Time-based reward maximization. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369, 20120461. <http://dx.doi.org/10.1098/rstb.2012.0461>
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 362, 933–942. <http://dx.doi.org/10.1098/rstb.2007.2098>
- Doughty, A. H., & Richards, J. B. (2002). Effects of reinforcer magnitude on responding under differential-reinforcement-of-low-rate schedules of rats and pigeons. *Journal of the Experimental Analysis of Behavior*, 78, 17–30. <http://dx.doi.org/10.1901/jeab.2002.78-17>
- Drew, M. R., Zupan, B., Cooke, A., Couvillon, P. A., & Balsam, P. D. (2005). Temporal control of conditioned responding in goldfish. *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 31–39. <http://dx.doi.org/10.1037/0097-7403.31.1.31>

- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325. <http://dx.doi.org/10.1037/0033-295X.84.3.279>
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52–77. <http://dx.doi.org/10.1111/j.1749-6632.1984.tb23417.x>
- Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, 7, 170–184. [http://dx.doi.org/10.1016/S0959-4388\(97\)80005-0](http://dx.doi.org/10.1016/S0959-4388(97)80005-0)
- Girshick, M. A., & Blackwell, D. (1954). *Theory of games and statistical decisions*. London, England: Wiley.
- Guilhardi, P., Yi, L., & Church, R. M. (2007). A modular theory of learning and performance. *Psychonomic Bulletin & Review*, 14, 543–559. <http://dx.doi.org/10.3758/BF03196805>
- Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4, 267–272. <http://dx.doi.org/10.1901/jeab.1961.4-267>
- Herrnstein, R. J., & Heyman, G. M. (1979). Is matching compatible with reinforcement maximization on concurrent variable interval variable ratio? *Journal of the Experimental Analysis of Behavior*, 31, 209–223. <http://dx.doi.org/10.1901/jeab.1979.31-209>
- Heyman, G. M., & Luce, R. D. (1979). Operant matching is not a logical consequence of maximizing reinforcement rate. *Animal Learning & Behavior*, 7, 133–140. <http://dx.doi.org/10.3758/BF03209261>
- Heyman, G. M., & Tanz, L. (1995). How to teach a pigeon to maximize overall reinforcement rate. *Journal of the Experimental Analysis of Behavior*, 64, 277–297. <http://dx.doi.org/10.1901/jeab.1995.64-277>
- Houston, A. I., & McNamara, J. (1981). How to maximize reward rate on two variable-interval paradigms. *Journal of the Experimental Analysis of Behavior*, 35, 367–396. <http://dx.doi.org/10.1901/jeab.1981.35-367>
- Kheifets, A., & Gallistel, C. R. (2012). Mice take calculated risks. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 8776–8779. <http://dx.doi.org/10.1073/pnas.1205131109>
- Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, 95, 274–295. <http://dx.doi.org/10.1037/0033-295X.95.2.274>
- Kirkpatrick, K. (2002). Packet theory of conditioning and timing. *Behavioural Processes*, 57, 89–106. [http://dx.doi.org/10.1016/S0376-6357\(02\)00007-4](http://dx.doi.org/10.1016/S0376-6357(02)00007-4)
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (No. 8). New York, NY: Oxford University Press.
- Machado, A., & Rodrigues, P. (2007). The differentiation of response numerosities in the pigeon. *Journal of the Experimental Analysis of Behavior*, 88, 153–178. <http://dx.doi.org/10.1901/jeab.2007.41-06>
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Freeman.
- Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117, 363–386. <http://dx.doi.org/10.1037/0033-2909.117.3.363>
- Niv, Y., Daw, N. D., & Dayan, P. (2005). How fast to work: Response vigor, motivation and tonic dopamine. *NIPS Proceedings*, 18, 1019–1026.
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191, 507–520. <http://dx.doi.org/10.1007/s00213-006-0502-4>
- Niv, Y., Joel, D., & Dayan, P. (2006). A normative perspective on motivation. *Trends in Cognitive Sciences*, 10, 375–381. <http://dx.doi.org/10.1016/j.tics.2006.06.010>
- Orduña, V., Valencia-Torres, L., & Bouzas, A. (2009). DRL performance of spontaneously hypertensive rats: Dissociation of timing and inhibition of responses. *Behavioural Brain Research*, 201, 158–165. <http://dx.doi.org/10.1016/j.bbr.2009.02.016>
- Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, 52, 111–139. <http://dx.doi.org/10.1146/annurev.psych.52.1.111>
- Rachlin, H. (1978). A molar theory of reinforcement schedules. *Journal of the Experimental Analysis of Behavior*, 30, 345–360. <http://dx.doi.org/10.1901/jeab.1978.30-345>
- Richards, J. B., Sabol, K. E., & Seiden, L. S. (1993). DRL interresponse-time distributions: Quantification by peak deviation analysis. *Journal of the Experimental Analysis of Behavior*, 60, 361–385. <http://dx.doi.org/10.1901/jeab.1993.60-361>
- Richardson, W. K., & Loughhead, T. E. (1974). Behavior under large values of the differential-reinforcement-of-low-rate schedule. *Journal of the Experimental Analysis of Behavior*, 22, 121–129. <http://dx.doi.org/10.1901/jeab.1974.22-121>
- Sanabria, F., & Killeen, P. (2008). Evidence for impulsivity in the spontaneously hypertensive rat drawn from complementary response-withholding tasks. *Behavioral and Brain Functions*, 4, 7.
- Schoenfeld, W. N., Harris, A. H., & Farmer, J. (1966). Conditioning response variability. *Psychological Reports*, 19, 551–557. <http://dx.doi.org/10.2466/pr0.1966.19.2.551>
- Simen, P., Balci, F., de Souza, L., Cohen, J. D., & Holmes, P. (2011). A model of interval timing by neural integration. *The Journal of Neuroscience*, 31, 9238–9253. <http://dx.doi.org/10.1523/JNEUROSCI.3121-10.2011>
- Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: Empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1865–1897. <http://dx.doi.org/10.1037/a0016926>
- Simen, P., Rivest, F., Ludvig, E. A., Balci, F., & Killeen, P. (2013). Timescale invariance in the pacemaker-accumulator family of timing models. *Timing and Time Perception*, 1, 159–188.
- Staddon, J. E., & Motheral, S. (1978). On matching and maximizing in operant choice experiments. *Psychological Review*, 85, 436–444. <http://dx.doi.org/10.1037/0033-295X.85.5.436>
- Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. Cambridge, MA: MIT Press.
- Taylor, K., & Balsam, P. (2009). [Unpublished raw data]. Raw interresponse times from each rat in their study.
- Trommershäuser, J., Landy, M. S., & Maloney, L. T. (2006). Humans rapidly estimate expected gain in movement planning. *Psychological Science*, 17, 981–988. <http://dx.doi.org/10.1111/j.1467-9280.2006.01816.x>
- Wald, A. (1950). *Statistical decision functions*. Oxford: Wiley.
- Wearden, J. H. (1990). Maximizing reinforcement rate on spaced-reinforcing schedules under conditions of temporal uncertainty. *Behavioural Processes*, 22(1–2), 47–59. [http://dx.doi.org/10.1016/0376-6357\(90\)90007-3](http://dx.doi.org/10.1016/0376-6357(90)90007-3)
- Wilson, M. P., & Keller, F. S. (1953). On the selective reinforcement of spaced responses. *Journal of Comparative and Physiological Psychology*, 46, 190–193. <http://dx.doi.org/10.1037/h0057705>
- Yi, L. (2009). Do rats represent time logarithmically or linearly? *Behavioural Processes*, 81, 274–279. <http://dx.doi.org/10.1016/j.beproc.2008.10.004>
- Zeiler, M. D., Scott, G. K., & Hoyert, M. S. (1987). Optimal temporal differentiation. *Journal of the Experimental Analysis of Behavior*, 47, 191–200. <http://dx.doi.org/10.1901/jeab.1987.47-191>

Received July 17, 2013

Revision received September 4, 2014

Accepted September 13, 2014 ■