Behavioural Processes 169 (2019) 103978

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Increased generalization in a peak procedure after delayed reinforcement

Jonathan Buriticá, Emmanuel Alcalá

Universidad de Guadalajara, Calle Francisco de Quevedo # 180, Col. Arcos Vallarta, 44130, Guadalajara, Jalisco, Mexico

ARTICLE INFO

Keywords:
Timing
Stimulus discrimination
Delay of reinforcement
Generalization
Rats

ABSTRACT

Temporal control of behavior might be impaired by reinforcement devaluation and other motivational operations such as delaying reinforcement of the instrumental response. Here, we report an experiment that assessed the effect of delayed reinforcement on a timing peak procedure. Using a within-subject design with a multiple two-component schedule of reinforcement, we found evidence of flat temporal generalization gradients, along with degraded response-reinforcer contingency, lower response rates and changes in the responding patterns due to delayed reinforcement. This result is consistent with the Learning to Time (LeT) and some versions of Scalar Expectancy Theory (SET).

1. Introduction

Timing is behavior adapted to time regularities of the environment with relevance for fitness. One method to investigate timing is the peak procedure, an experimental arrangement composed of two types of trials: reinforcement and extinction trials. The reinforced trials are the Fixed-Interval (FI) schedule, where the first response after a fixed time produces the outcome. The extinction trials are three times longer than the FI and end without reinforcement. The responses in FI trials begin after a pause and follow a break-run (low-high) pattern until reinforcement. In peak trials, responding decreases after the trained interval (Catania, 1970). Both types of trials are randomly intermixed without signaling which type is presented. The obtained data can be analyzed with two approaches: aggregation of several peak trials or trial-by-trial. Although generally both analyses generate similar results, the trial-by-trial approach is more sensitive to subtle effects, e.g., different levels of motivation affect trial-by-trial measures, but the same effect is not clearly observed in the aggregated analysis (see revision below). The average response-function of several peak trials begins with a steady climb close to the FI duration and then decreases toward zero. In a single peak trial, the performance follows a low-high-low pattern, where both the start (the time at which there is a transition from a low to a high response rate) and the stop (the time of a transition from a high to a low rate) times of responding can be identified, and with such times the middle time (a peak that corresponds to the highest response rate) and the duration (spread) of the high-rate state can be estimated (Church et al., 1994). A suitable explanation for performance in this procedure is the concept of the temporal generalization gradient (de Carvalho et al., 2016).

The temporal generalization gradient explanation regards time as another stimulus controlling behavior. Thus, by such conceptualization, the idea of stimulus control is used to explain how environmental factors produce behavior. The stimulus sets the occasion for, or controls, the instrumental responses. To apply the idea to the peak procedure, we must think of time intervals as a factor controlling instrumental responses, so after training, the elapse of time produces an instrumental response and its associated outcome. Nevertheless, control by the trained stimulus is not perfect, and the response can appear after other similar stimuli, especially if the stimuli are similar in the same dimension (e.g., wavelength, sound, shape, brightness). Therefore, the response should follow a function with the most responses on the trained stimuli and a slight decrease as a function of dissimilarity. Such a bell-shaped pattern of response is observed in the peak trial, where the peak of the response occurs at the trained interval, and the response frequency drops before and after such time; in this case, the stimulus similarity is the closeness to the trained interval.

Weak motivation produces flat generalization gradients when color, brightness intensity and loudness have been trained (for a review see Lotfizadeh et al., 2012). Water satiation and food deprivation, percentage of weight, and prefeeding have been the variables used to decrease motivation. Notwithstanding some procedures to decrease motivation have been used with the peak procedure, it is still unclear how motivation influences time discrimination. The peak procedure is an appropriate method to investigate the relation between motivation and timing because it is a direct form to measure generalization gradients across time (de Carvalho et al., 2016).

In the peak procedure, the evidence suggests a relation between motivation and timing; strong motivation generates shorter and weak motivation lengthens start times. For example, satiation, from the beginning to the end of the session, lengthens start and peak times (Balcı

E-mail address: jjburiticab@unal.edu.co (J. Buriticá).

https://doi.org/10.1016/j.beproc.2019.103978
Received 22 February 2019; Received in revised form 13 September 2019; Accepted 23 September 2019
Available online 30 September 2019
0376-6357/

et al., 2010). The same lengthening of start times has been observed with fewer pellets and reward devaluation with pellet-illness association (Galtress and Kirkpatrick, 2009), pre-feeding with sucrose or food pellets (Galtress et al., 2012), low magnitudes of brain stimulation reward (Ludvig et al., 2007), and duration of access to reinforcement (Ludvig et al., 2011). Additionally, access to reinforcement before the session and delayed reinforcement lengthens post-reinforcement pauses in FI schedules (Buriticá and dos Santos, 2017; Elcoro and Lattal, 2011). Despite earlier reports with aggregated analyses suggesting that timing is not affected by prefeeding (Roberts, 1981), the conclusion is that timing and motivation are related; however, the explanation for such a relation is a topic of debate (Balcı, 2014; Galtress et al., 2012).

Although weak motivation lengthens start times, it is not yet clear how motivation affects time discrimination. Two sets of theories could explain such results: scalar expectancy theories (SET; Balcı, 2014; Galtress et al., 2012) and behavioral theories of timing (BeT and LeT; Killeen and Fetterman, 1988; Machado, 1997). Both theory sets offer explanations for temporal generalization gradients.

Briefly, SET comprises a pacemaker, one memory mechanism and a decision process: if the elapsed time is sufficiently similar to a time sample retrieved from a memory bin of reinforced times, the response occurs, and the process of comparison (usually represented as a ratio) continues until the reinforcement is obtained or the elapsed time has exceeded the remembered time. Motivation could operate in the mechanism at different sites, for example, in the switch between pacemaker and working memory, or changing the threshold to respond (Balcı, 2014; Galtress et al., 2012). In this case, the temporal gradient emerges from the comparison between working memory (of the interval) and reference memory (of previous trained intervals) and the threshold value. If the memory of the elapsed interval (or the trained interval) was affected by weak motivation, we should expect a displacement of the gradient to the right in cases in which the update of the working memory was slow; this would happen if the switch between the pacemaker and working memory occurred at a low rate. However, if the threshold value was reduced due to weak motivation, the generalization gradient would flatten because more values in the comparison between memories would exceed the threshold.

In LeT, timing depends on the association between behavioral states and the instrumental response: if a behavioral state is active and its association with the instrumental response is strong, the response occurs with high probability; otherwise, the response likelihood is low (Machado, 1997). The temporal gradient emerges from the increase in associative strength of the instrumental response and the behavioral states after reinforcement, as well as the reduction of strength between the response and the states during extinction (unreinforced interval durations). One conjecture is that motivation operates in LeT through the speed of succession of states (in BeT this could be the arousal; Killeen and Fetterman, 1988), and weak motivation could lead to a slow wave of activation across states. Thus, if the motivation is weak, then longer times are needed to initiate a response because the states associated with the response require more time to appear. The effect should be transient, and it should disappear after extended training, because the association with new behavioral states that coincide with reinforcement become stronger. Another proposition is that weak motivation (less arousal) generates fewer instrumental responses per interval, decreasing the number of coincidences between instrumental responses and behavioral states. This phenomenon could create a situation in which more behavioral states, far from the time of reinforcement, create links to the instrumental response with sufficient strength to produce a response, in comparison to situations where the motivation is strong. In this case, a flat temporal generalization gradient should be observed.

Delay of reinforcement is an operation that affects the effectiveness of reinforcement, and its effects on the temporal generalization gradient have not yet been fully established. Evidence of a delayed reinforcement effect will strengthen the argument about the relation between timing and motivation, and it will help to elucidate the underlying mechanisms. A first approach to determine how delayed reinforcement affects temporal discrimination is to measure the effect on a time discrimination task. In a temporal bisection task, two durations (short and long) are trained, and intermediate durations are introduced during the test, in which both levers are presented and the subject chooses one lever. The usual result is that the frequency of long lever pressing increases with longer durations. Delayed reinforcement of the training trials reduced stimulus control in this procedure, decreased the range of responses to long when short or long signals were presented, and increased the Weber fraction (Buriticá et al., 2016). If we take a synthetic approach to the problem and suppose two generalization gradients around the short and long durations (de Carvalho et al., 2016), then the observed result could be explained by flat generalization gradients across time. Nevertheless, a direct observation of the generalization gradient could provide direct evidence of such a flattening effect, and the peak procedure is a straightforward measure of such temporal generalization gradients.

According to SET, variations in start, stop, middle time, and spread are indicative of sources of variation in the peak procedure (Gibbon and Church, 1990) that manifest themselves as changes in the generalization gradients across time. For example, a positive correlation between start and stop times indicates one memory of the to-be-timed interval, and a negative correlation between the start time and spread provides evidence of two thresholds, one to start responding and another to stop. A proportional change in start and stop times suggests a change in the memory interval, while changes in one but not the other could be due to changes in thresholds, to start or stop. A detained analysis, trial-by-trial, of start, stop, middle times and spread will provide evidence about the source of observed variations: memory of the intervals, or thresholds to start or stop. The data obtained thus far are in line with the hypothesis of variations in threshold values as the most plausible explanation; however, it is not clear whether the thresholds change in value, variability or both (Balcı, 2014; Buriticá et al., 2016; Galtress et al., 2012).

A closer look at the patterns of behavior, in addition to the analysis of start, stop, middle-times and spread, could elucidate whether other factors not yet considered could explain the longer start times due to weak motivation, and the inconsistent variation in measures such as peak time that are seemingly not always affected (Galtress and Kirkpatrick, 2009; Roberts, 1981). Other measures should provide a quantitative estimate of the consistency of the low-high-low response pattern produced by time discrimination, and the adaptation of other behaviors besides the instrumental response, such as head entries, to time regularities (López and Menez, 2012). Such patterns and their consistency with previous reports could provide a hint about how the whole pattern of behavior is related to trained time intervals.

In addition to its effects on time discrimination, delayed reinforcement also reduces the response rate, and the usual explanation for this effect is the devaluation of reinforcement, or its value reduction (Buriticá and dos Santos, 2017). In this sense, we can think of delayed reinforcement as another way to reduce reinforcement effectiveness or to diminish motivation. Nevertheless, another explanation for the effects of delayed reinforcement is contingency degradation. Increasing the time interval between the response and the reinforcement may diminish the correlation between both events, and such lower contingency between response and reinforcer decreases the response rate in Random Interval schedules (Hammond, 1980). In the Davison and Nevin (1999) theoretical approach to contingency, reinforcement delay affects behavior because it reduces the discriminability of the response-reinforcer relation, and it is likely that such reduction in discriminability weakens contingency, or its control over behavior. To understand how delayed reinforcement affects performance, we measure the response rate, the time between responses and reinforcement, and the correlation between such events.

Finally, the objective of the study was to test the effects of delayed

reinforcement in peak-procedure performance and to establish if delayed reinforcements increases generalization, as suggested by Buriticá et al. (2016) and Lotfizadeh et al. (2012). The evidence of such increased temporal generalization gradients will reveal how motivation affects timing. We used a multiple schedule with two components: at baseline, the same FI was programmed in both components; in the experimental phase, one component maintained the same average programmed delay to reinforcement but with a terminal response-independent delay, while the intervals in the second component were yoked to the first component but with immediate reinforcement. Peak trials were presented in both phases and components, so we could compare temporal generalization gradients between phases and components. At baseline, we expected to observe the same performance in both components, but in the experimental phase we predicted flatter generalization gradients in the delayed-reinforcement component than in the immediate component according to Buriticá et al. (2016) and Buriticá and dos Santos (2017). We also anticipated more variability in behavior during the experimental phase compared with the baseline due to the greater variability in the intervals-to-reinforcement.

2. Methods

2.1. Apparatus

Three rat-conditioning chambers from MED-PC Associates were used. Each chamber had a 0.1-cc water dispenser on the central panel. Each chamber was enclosed in a sound-attenuating box. Two retractable levers were positioned adjacent to the water dispenser 2.5 cm above the grid floor; a 0.14-N press was required to record a response. A light bulb at the rear panel provided general illumination. MED-PC IV and later MED-PC V software were used to schedule and record events.

2.2. Subjects

Nine male Long Evans rats bred in the laboratory, aged eight months at the beginning of training, were used. They had free access to food and only received access to water 30 min after the session. If the animals were below 85% of their *ad-libitum* weight, they had access to water for 25 min; otherwise they had access for 20 min. One subject became ill before the last condition and, after veterinary treatment, was removed from the experiment.

2.3. Procedure

2.3.1. Lever pressing training

Sessions took place five to six days per week. Every training session lasted for 30 min or 150 rewards, whichever occurred first. Lever pressing was trained with a conjoint Fixed-Time (FT) Continuous Reinforcement (CRF) schedule. The cup with water raised after 30 s elapsed since the last reward-offset, and the lever was retracted after reward-onset. Additionally, if one lever press occurred, it was reinforced, and the lever retracted after reward-onset. Both levers were presented pseudo-randomly, one at a time, with no more than three presentations of the same lever in a row, and each lever presentation was ~ 50%. The time between lever presentations was 5 s, the time of access to reinforcement. If the subject failed to obtain 100 reinforcers in a session, FT was increased by 30 s for the next session, up to 120 s. After the acquisition criterion was met, only the CRF was operative in the subsequent sessions until the subjects obtained 100 reinforcers for two consecutive sessions. Then, the subjects were exposed to FI 2 s, 5 s, 10 s, 15 s and 30 s presented on a daily progression. Both levers were presented with the same FI, but only one at a time and randomly chosen. The lever was inserted at the beginning of the interval and retracted with the reward-offset. In this part of the training, each session lasted 60 min or 120 reinforcers, whichever occurred first.

2.3.2. Baseline

After training, a multiple FI 60 s FI 60 s schedule was presented to the subjects. Each FI was signaled by a lever, but only the lever associated with the ongoing component was present. The lever was extended at the beginning of the component and retracted with the reward-offset. Components were presented pseudo-randomly, with no more than three of the same in a row. Each component was presented ~ 50% times. When 60 s had elapsed since the trial onset, the first response to the lever produced access to the cup for 5 s. Every interval was followed by a 10-s intercomponent-interval. The house light was on during the intervals, but it was off during the intercomponent-interval. After 30 sessions, the peak trials were introduced. Peak trials lasted for 180 s and ended without reinforcement. For every 10 presentations of a component, one was sampled randomly as a peak trial. The session finished after 70 min. The baseline, with the peak trials and immediate reinforcement in both components, lasted for 39 sessions.

2.3.3. Experimental phase

In the last condition, a multiple two-component schedule was used (see Buriticá and dos Santos, 2017; Experiment 2), but only the lever associated with the active component was extended at the beginning of the interval and retracted after reinforcement (reward offset). One component was a Tandem FI 54 s FT 6 s (delayed component), and the second component was a Yoked Interval (immediate component). The components were alternated pseudo-randomly, with no more than three in a row, and each component was presented ~ 50% times. For the delayed component, reinforcement was obtained after the first response in FI 54 s, but it was delivered after FT 6 s. The interval between the start of the FI and reinforcement delivery was recorded and yoked to the other component. For the immediate component, the first response after the yoked interval elapsed was reinforced. The peak trial duration and presentation were the same as baseline. This phase lasted for 30 sessions.

2.4. Data analysis

The response-reinforcer contingency was measured as the correlation between responses and reinforcements for every session for all subjects in both phases (see DeRusso et al., 2010). First, we created a common vector from the onset to the offset of the FI trials for every component and session, and we divided the vector in bins of 200 ms each; for every component, we counted the responses and reinforcers for each bin of the common vector, and finally the sample Pearson correlation coefficient was calculated as follows:

$$r_{Rr} = \frac{S_{Rr}}{S_R S_r} \quad (1)$$

where S_{Rr} is the sample covariance of reinforcers (R) and responses (r), and S_R and S_r are the sample standard deviation of R and r, respectively. Fig. 1 shows a schematic of the analysis, in which the correlation measures the coincidence of responses and reinforcers, so if the coincidence of responses and reinforcers is higher in the last bin, then a higher Pearson correlation is obtained. In the example shown in Fig. 1, only one response coincides in the last bin with several reinforcers in the delayed component, while the coincidence of responses and reinforcers is higher in the immediate component. This pattern should lead to a lower Pearson correlation for the delayed component compared with the immediate component. Additionally, contiguity between responses and the next reinforcer was determined as the difference between the time of each response and the time to the next reinforcer in the FI trials of the last six sessions in each phase, as used for the trial-by-trial analysis.

We computed the Kullback-Leibler divergence (D_{KL}, or relative entropy) for the last 30 peak-trials in both phases to measure the similarity between components of the response-rate functions in each phase. D_{KL} measures the similarity between two distributions (Cover and

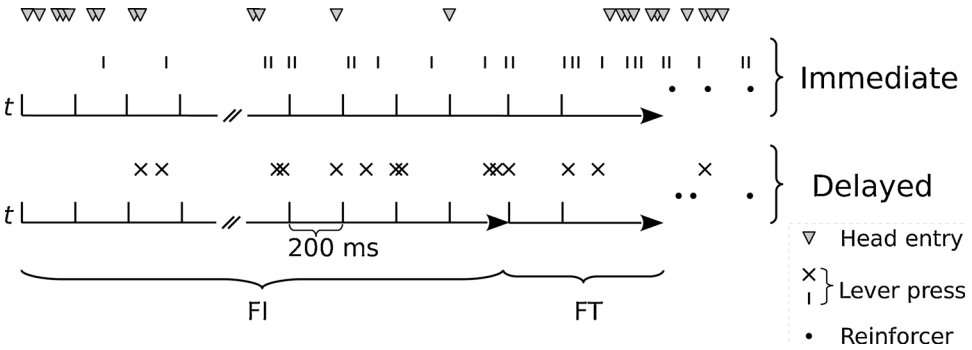

Fig. 1. Diagram showing the relationship between responses (lever presses and head entries) and reinforcers in the timeline of the trials, binned at 200 ms, for both components. The immediate component is in the upper panel; the delayed component is in the lower panel. The lever presses were specific to the component, while the head entries were not because there was only one water dispenser.

Thomas, 2006) and is calculated using the following equation:

$$D_{KL} = \sum_{i=1}^{n} P(i)\log_b\left(\frac{P(i)}{Q(i)}\right) \quad (2)$$

Here, P and Q are the probability distribution functions of the immediate and delayed components, respectively; i is the i-th bin in the probability comparison; $\log_b$ takes the logarithm of the ratio $P(i)/Q(i)$; and b is the base of the logarithm (we used base 2, so D_{KL} is measured in bits). When $P(i) = Q(i)$, the Kullback-Leibler divergence is 0 (because $\log_2(1) = 0$). Thus, as P diverges from Q, D_{KL} strays from zero, and if $P(i) > Q(i)$, $D_{KL} > 0$, i.e., D_{KL} increases. A value of divergence equal to zero suggests similar response functions, and a departure from zero suggests a more dissimilar shape between compared functions. If delayed reinforcement generates flat temporal generalization gradients, we should observe higher divergences in the experimental phase than at baseline.

We used the last 30 peak-trials per component to measure stable-state performance, with the trial-by-trial analysis; the included trials had at least nine responses. The algorithm segmented performance along the trial in three sections or states: the first with a low response-rate, the second with a high rate and the third with a low rate (see Fig. 1 in the Supplementary material). The points of partition between the states were the start times (between the first and second state) and stop times (between the second and third state). In individual trials, we determined the start and stop times by computing an index that maximized the area A (Eq. (3)): the sum of the absolute deviations between response rates in the states and in the whole trial multiplied by its durations (Church et al., 1994):

$$A = d_1\ |r-r_1| + d_2\ |r-r_2| + d_3\ |r-r_3| \quad (3)$$

d_1, d_2 and d_3 were the durations of states with first-low, high and second-low response rates; r_1, r_2 and r_3 were the respective response rates and r was the rate during the whole trial; d_1 was the interval between the beginning of the component and start of the high-rate state r_2 ; d_2 was the interval between start and stop times; and d_3 was the interval between stopping and the end of the trial. Fig. 1 in the Supplementary material shows a graphical representation of Eq. (3). The algorithm was used to compute the start time for every trial, using all responses except the last; the stop time was then computed for the remaining responses, with the following constraint: the stop must be longer than the start. We allowed the start times to be greater than 60 s and the stop times to be shorter than 60 s (c.f., Church et al., 1994, which refers to start times < 60 s as "good starts" and stops > 60 s as "good stops"). Such criteria were employed to allow as much variability in the measure of both components as it could be generated by delayed reinforcement in the experimental phase.

We also determined the response rate as a function of time in the trial with the same 30 peak-trials, for both lever pressing and head entries, grouped by component and phase. The response rates were normalized to constrain response rates between 0 and 1, using the unity-based normalization as follows:

$$r_N = \frac{r - r_{min}}{r_{max} - r_{min}} \quad (4)$$

where r_N is the normalized response rate, and r_{min} and r_{max} are the minimum and maximum response rates, respectively.

We used the same 30 peak-trials of the trial-by-trial analysis to determine the averaged analysis. We pooled the trials per session to acquire more data (approximately two trials per session), and we obtained probability density functions (pdf) with the *density()* function in the R package *stats* (R Core Team, 2018). With the density functions, we obtained estimations of the middle time and spread. With the probability density functions, we calculated the Full Width at Half Maximum (FWHM; Hinton and Rao, 2004), the distance between two points t_1 and t_2 at half of the maximum (or peak) value of the density functions. This measure avoids the ramp part of the response rate functions, increasing the accuracy of the spread estimation. The middle time was the middle point of the FWHM (see Fig. 5 in the Supplementary material for an example of this measure for one session of a representative subject). The middle time measures the timing accuracy (how far the behavior was from the trained interval), and the spread measures the precision (the closeness of different attempts to gain reinforcement). We did not expect large differences in the middle times between components because temporal generalization gradients should have similar centers; however, a flat gradient should lead to less precision and, therefore, to a longer (and variable) spread in the delayed compared with the immediate component.

We implemented a cluster analysis with the k -medoids method using the partitioning around medoids (PAM) algorithm, which is robust to outliers (Kaufman and Rousseeuw, 1990), to estimate the frequency of the patterns of the response rate states (low or high) in the peak trials. Briefly, the medoid is a member of a dataset with minimal dissimilarity from the remaining data points in this dataset, that is, it is the most central data point. The PAM algorithm starts with an initial set of k data points, which serves as the initial medoid of the clusters, which in turn are constructed by assigning the data points to the closest medoids. It then iteratively selects other k data points as the new medoids and compares the sum of distances (usually the Euclidean distance between the medoids and the data points) of the previous medoids with the new ones; if the total distance of the new medoids is less, it selects the new data points as the medoids; otherwise, it retains the previous values. Finally, the algorithm stops when there is no change in total distance. This cluster analysis, applied to the times of every response, produced an estimation of the number of response-states transitions that occurred during the trial. We computed the gap statistic for every cluster number to determine the optimal number of clusters between one and four. Gap statistics were used to compare the within-cluster sum of squares for every cluster with a reference null distribution of data with a single cluster, and then the number of clusters with the maximum value for this statistics was selected (see Tibshirani et al., 2001).

We used linear mixed modeling to test hypotheses of differences between phases and components. Because the comparisons included

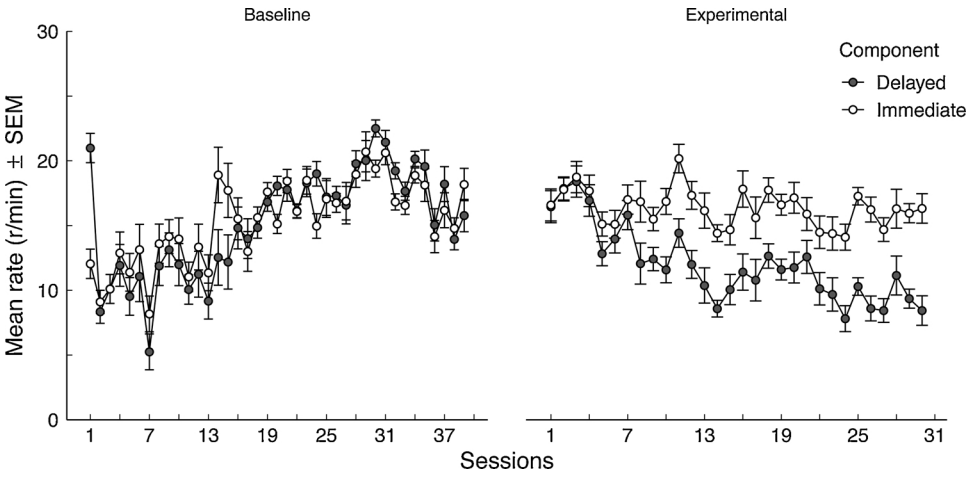

Fig. 2. Average response rate by session (± S.E.M). The baseline for both components is shown on the left. The right panel shows the response rate under the experimental condition. Gray and white circles correspond to delayed and immediate components in the experimental condition and the corresponding lever at baseline. At baseline, both components had the same response rate, but the rate was higher in the immediate component throughout the experimental phase.

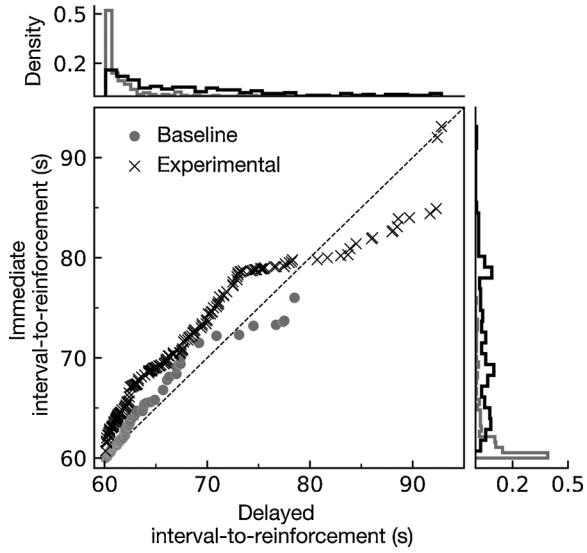

Fig. 3. QQ-plot of the interval-to-reinforcements in the immediate component (*y* axis) vs. delayed (*x* axis) components in the baseline phase (gray points) and experimental phase (×). The dashed line shows a 1:1 relation between the components. The marginal histograms show the densities of the intervals by phase. The upper histograms denote the delayed component, and the immediate component is shown on the right, with their colors matching the colors of the QQ-plot.

repeated measures (the same subjects between phases, and within components of the multiple schedule), we incorporated random effects in the model to quantify individual subject variability and to obtain estimations of the effects of the change of phase, the component (delay of reinforcement) and its interactions (see DeHart and Kaplan, 2019). Data analysis was conducted with R 3.5.0 (R Core Team, 2018) and the following libraries: *tidyverse* (Wickham, 2017), *data.table* (Dowle and Srinivasan, 2018), *cluster* (Maechler et al., 2018), *gridExtra* (Auguie, 2017), and *nlme* (Pinheiro et al., 2018). A repository of the data and code used in this work can be found at https://github.com/jealcalat/Generalization_decrement_data-analysis.

3. Results

At baseline, the response rate was equal in both components (Fig. 2, left panel). In the experimental phase, the response rate in the delayed component was lower in the final sessions compared with the immediate component (Fig. 2, right panel). The statistical models for comparison included the subject, session (in subject) and component (in session) as random factors, and the phase and component as fixed effects; unless otherwise reported, the models for other dependent measures were similar. The phase influenced the response rate $\chi^2(6) = 7.35$, $p = .006$, as did the component $\chi^2(7) = 28.22$, $p < .001$, and the interaction phase by component $\chi^2(8) = 32.67$, $p < .001$. The factor phase showed a higher response rate due to the experimental phase, $b = 1.11$, $t(478) = 2.10$, $p = .03$, so an increase of ~ 1 r/min could be attributed to the change of phase. Curiously, the delayed component had no statistical effect, $b = -0.11$, $t(299) = -0.23$, $p = .81$; if we consider all the data, the response rate was similar, most likely due to the same rate in both components observed at baseline. However, the interaction phase by component was significant, $b = -4.29$, $t(478) = -5.77$, $p < .001$, which indicated that the change of phase, the introduction of the interval between the reinforced response and the delivery of reinforcement in the delayed component could explain the reduction in response rate of ~ 4 r/min. Thus, the component by itself had no effect, but the phase and interaction phase by component seemed to explain most of the observed changes in response rate. The response rate increased from baseline to the experimental phase, but the response rate decreased in the delayed component. Thus, reinforcement delay decreased the response rate, which is consistent with previous reports in FI and VI schedules (Buriticá and dos Santos, 2017; Elcoro and Lattal, 2011; Sizemore and Lattal, 1978, 1977).

The reduced response rate in the delayed component could not be explained as result of longer intervals-to-reinforcement (in a free operant procedure, these intervals would be Inter Reinforcement Intervals) because the obtained distributions should produce the opposite results, as shown in Fig. 3. Such intervals were the time between the onset of light and the lever until the delivery of reinforcement. The QQ-plot (Wilk and Gnanadesikan, 1968) in Fig. 3 shows the intervals-to-reinforcement of the baseline (gray points) and experimental (black × symbol) phases, and the marginal histograms show the densities of the intervals by phase. These plots show that the intervals-to-reinforcement became elongated beyond the maximum value of the baseline, providing values of 78.5 s and 76 s for the delayed and immediate components in the experimental phase, 92.8 s and 93.1 s for the delayed and immediate components, and an increase (nonlinear, as shown in Fig. 3) of 14.3 and 17 s, respectively. In the experimental phase, the intervals-to-reinforcement in the immediate component was longer than in the delayed component. After 80 s, this relation was shifted below the dashed line; however, most of the percentiles clustered before this point. Thus, in the experimental phase, subjects experienced longer intervals-to-reinforcement in the immediate than in the delayed component. If the response rate was determined solely by intervals-to-reinforcement, the rate should be higher at the delayed component because the intervals-to-reinforcement there were shorter,

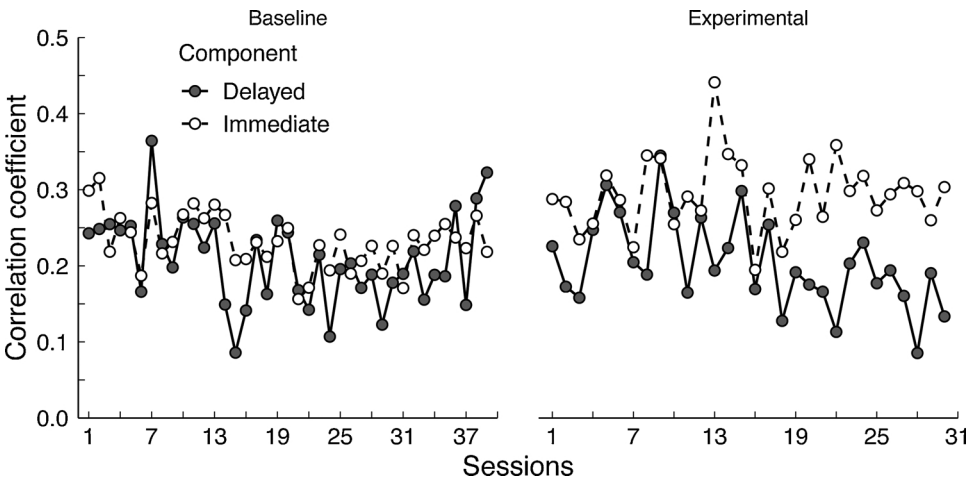

Fig. 4. Pearson correlation between responses and reinforcers. Trends of the Pearson correlation coefficients are similar at baseline but different between components in the experimental phase. During the experimental phase, the correlation of the response-reinforcer was higher for the immediate component compared with the delayed component.

but the opposite results were obtained. The response rate was lower in the delayed component; thus, it was sensitive to the delay between the response and the reinforcer, notwithstanding the longer intervals-to-reinforcement in the immediate component.

The Pearson coefficients were similar for both components at baseline, but the correlation was lower in the delayed component in the experimental phase, as shown in Fig. 4. The coefficient was different between components from session 17 onwards in the experimental phase, with lower correlations for the delayed component; this result suggested that delayed reinforcement decreased the response-reinforcer contingency. Such a reduction of contingency could explain the lower responses rates, but in our case, we only measured such contingency and did not manipulate its value. Consequently, this conclusion should be considered with caution until an empirical demonstration confirms the statement.

Fig. 5 shows the distributions of the intervals between every response and the next reinforcer in FI trials for both components. The time window between the responses and scheduled reinforcement was wider in the delayed than in the immediate component, probably because responding after 54 s was not necessary for reinforcement, thus allowing greater temporal distances between responses and reinforcement; with bins of 200 ms, the frequency of coincidences between responses and reinforcers decreased in the delayed compared with the immediate component (see Fig. 1). Based on this observation, we can expect that the contiguity between responses and reinforcers was greater for the immediate than for the delayed component in the experimental phase. At baseline (Fig. 5), there was a high density of intervals between response and reinforcement at 1 s, and another peak immediately after 10 s in both components, although the bar indicating a high density at 1 s disappeared for the delayed component in the experimental phase. Additionally, the density of temporal windows showed two peaks for the delayed component: one at 6 s and another at 20 s, suggesting a more variable expectation of the time of reinforcement due to the delayed reinforcement (see Figs. 3 and 4 in the Supplementary material. The effect was more pronounced at the individual level for most of the subjects except M332 showing no effect).

Fig. 6 shows the normalized response rate as a function of time in the last 30 peak trials averaged across subjects (panels A and B). We compared the peaks of the functions between phases and components (arrows in the graph). At baseline, the response functions for both components overlapped, but in the experimental phase, the function of the delayed component was flat and had a shorter peak than the immediate component; the function of the delayed component peaked at the same time as both functions of the baseline. The arrows show a peak shift to the right in the immediate component compared with the peaks at baseline. This result is at odds with the trial-by-trial analysis of the middle times (see below). The lack of consistency trial-by-trial and aggregated analysis suggested that the peak of the aggregated function could be an artifact of averaging. In the peak trials, the behavior followed the low-high-low pattern and the transitions in response rate were usually abrupt, as well as variable in time between trials and related to the trained interval. However, the averaged results showed a smooth change of rate, so the results in the aggregated analysis may represent a general trend in behavior (when responses are common) but do not accurately describe the behavior pattern.

The similarity measure of the response distributions functions

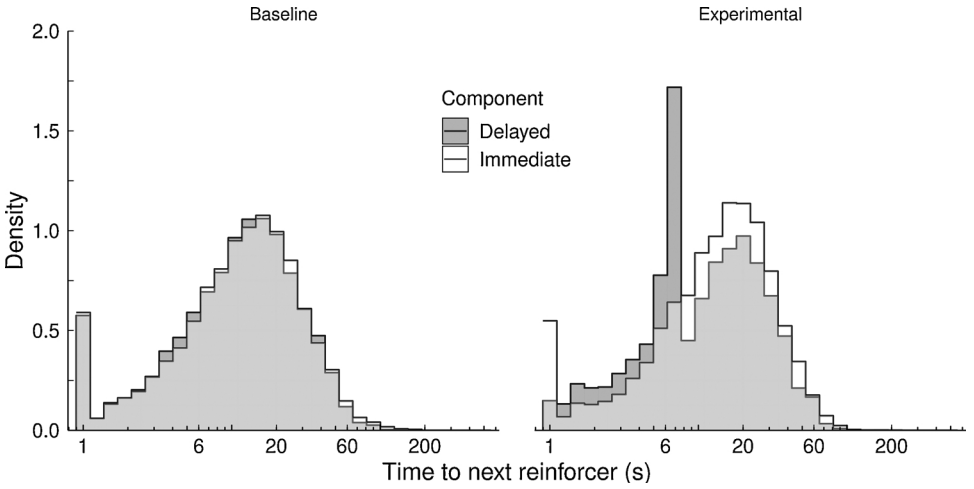

Fig. 5. Distributions of time intervals between the response and the next reinforcer, i.e., the difference between the time of the next reinforcer and the time at which every response occurred. The delayed component is shown in gray, and the immediate component is shown in white. The distributions were similar at baseline but differed in the experimental phase. The frequency of time intervals less than 1 s diminished dramatically in the delayed component compared with the immediate component.

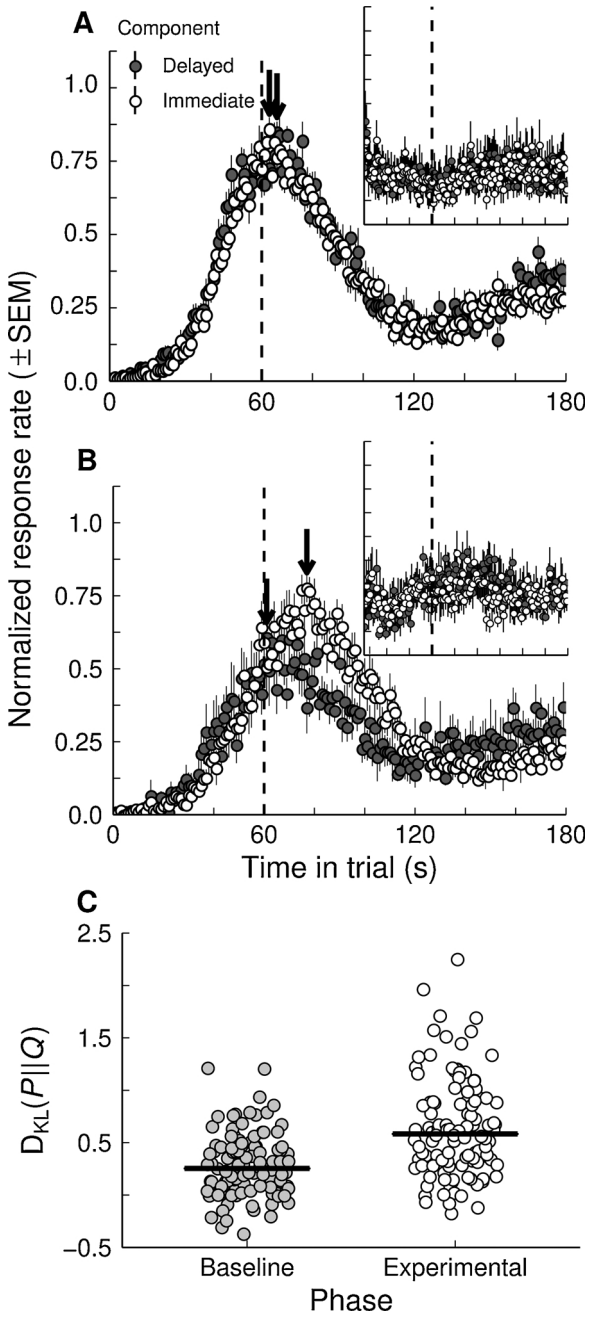

Fig. 6. Panels A and B: Normalized response rate as a function of time in peak trials. An average of 30 trials per component of eight subjects in the baseline (A) and experimental (B) phases. Gray and white circles correspond to delayed and immediate components, respectively. Arrows shows the time at which the response rate reached its maximum value, i.e., the peak times. Dashed lines show the time of scheduled reinforcement in both components. The inset plots are the head entry rates. C: Kullback-Leibler divergence (D_{KL}) between immediate and delayed components; the black line shows the median.

(Fig. 6C) allowed us to compare the probability mass functions (Eq. (2)) of the components in both phases. Each individual point in Fig. 6C resumes the D_{KL} comparison of both components of the last 30 trials of the experimental phase. The closer D_{KL} was to zero, the higher was the similarity between distributions. At baseline, the average D_{KL} was 0.30 (SD = 0.30); in the experimental phase, it was 0.66 (0.47). The divergence in functions was then closer to zero at baseline compared with the experimental phase. A linear model with subject and session modeled as random factors and phase as a fixed effect was used to compare the average means, and the difference between phases was significant: t (46) = 7.16, $p < .001$. Thus, the analysis confirmed that the dissimilarity between response functions of the immediate and delayed components increased from baseline to the experimental phase due to the delayed reinforcement. This analysis supports the conclusion obtained from Fig. 6B where both functions seemed to be different, so the delayed reinforcement seemed to create a flat generalization gradient around the trained interval.

Another behavior that is susceptible to temporal control is the pattern of head entries to the water dispenser (López and Menez, 2012). The head-entry rate (inset plots in Fig. 6, panels A and B) at baseline had a higher value and decayed to a constant rate over the trial, but in the experimental phase, head entries started at a high rate and then decayed and again increased approximately 30 s and peaked at 60 s; there were no differences in head-entry patterns between components. The pattern of head entries suggests that delayed reinforcement may affect behaviors other than the instrumental response. The patterns of head entries changed as a function of time from baseline to the experimental phase, which suggested that delayed reinforcement expanded the share of behaviors controlled by the to-be-timed interval.

Fig. 7A shows similar distributions of start times for both components at baseline. The distribution width was greater in the experimental phase compared with baseline, accompanied by right tail elongation in the delayed component. In the experimental phase in the delayed component, the start times increased beyond 60 s ("bad starts"). The statistical analysis showed that phase influenced the start times, $\chi^2(6) = 57.48$, $p < .001$, and the interaction phase by component, $\chi^2(8) = 6.95$, $p < .008$, but the component was not significant, $\chi^2(7) = 0.60$, $p = .43$. The model factors in the interaction phase by component showed lengthened start times due to the experimental phase, $b = 22.64$, $t(449) = 4.92$, $p < .001$, and the component showed no statistical effect, $b = 2.30$, $t(162) = 1.19$, $p = .23$; however, the interaction phase by component was significant, $b = -7.51$, $t(449) = -2.63$, $p = .008$. Briefly, the lengthened start times were due to delayed reinforcement. The long right tail of the start time distribution suggested an increased threshold to initiate responses; the rats waited longer times to initiate a high rate of lever pressing when the delayed component was active. An increase in variability of the threshold to start should be accompanied by an increase in frequency of shorter and longer values, not just an increase in longer starts; thus, delayed reinforcement appears to change only the value of the threshold but not its variability.

Fig. 7B shows similar distributions for stop times at baseline, but the average of the immediate component was greater than the average of the delayed component in the experimental phase (see Table 1). Additionally, the distribution of stop times in the delayed component was bimodal, showing another peak close to 120 s. For stop times, only the phase had a statistical effect, $\chi^2(6) = 10.40$, $p = .001$, but not the component, $\chi^2(7) = 0.72$, $p = .39$, or the interaction phase by component, $\chi^2(8) = 0.17$, $p = .67$. The phase lengthened the stop times, $b = 5.81$, $t(450) = 3.23$, $p = .001$. The stop times increased between phases, most likely due to the change in intervals-to-reinforcement, which lengthened in both components of the experimental phase and became even longer in the immediate component (see Fig. 3).

The spread is shown in Fig. 7C. The median of FWHM (black horizontal line) was similar in both components at baseline, but FWHM was larger for the delayed compared with the immediate component in the experimental phase. A statistical effect was not observed for FWHM in phase, $\chi^2(6) = 0.07$, $p = .78$, but it was observed for the component, $\chi^2(7) = 11.32$, $p < .001$, and the interaction phase by component, $\chi^2(8) = 4.80$, $p = .02$. The factors in the interaction model phase by component showed an increase in FWHM due to the experimental phase, $b = 19.50$, $t(105) = 2.17$, $p = .03$. The component showed no statistical effect, $b = -3.23$, $t(174) = -0.79$, $p = .43$; however, the interaction phase by component was significant, $b = -12.51$, $t(105) = -2.19$, $p = .03$. Delayed reinforcement increased the spread of the function; this may imply less precision and less control by the temporal

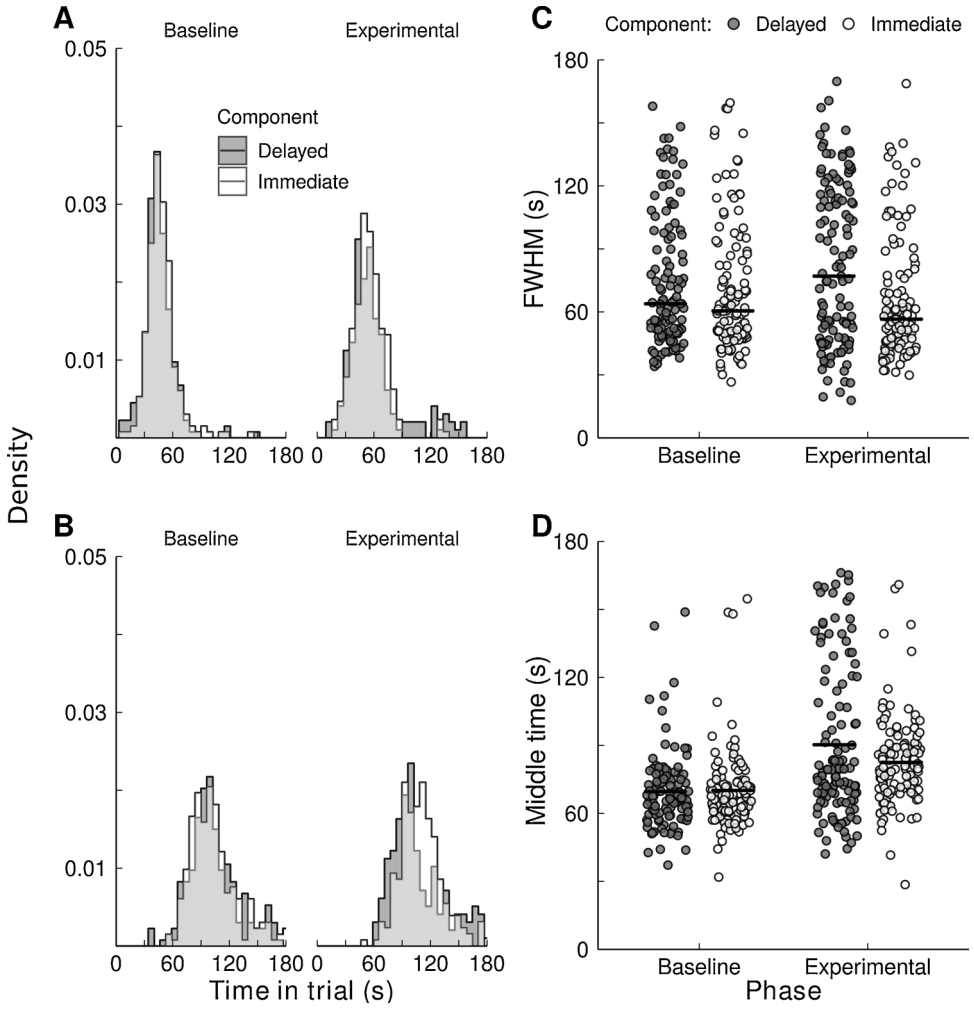

Fig. 7. Panels A and B: Probability density of start and stop times, respectively, of individual peak trials obtained with Eq. (1). Baseline data are shown on the left, and experimental phase data are shown on the right sides of the panels. C: Full Width at Half Maximum (FWHM). D: Peak time obtained as the middle point of FWHM.

Table 1
Average for start, stop, middle times, and spread (FWHM) in a trial-by-trial analysis.

Measure	Component	Baseline	Experimental
Start times	Delayed	45.58 (17.23)	60.45 (28.84) *
	Immediate	48.02 (16.51)	55.66 (18.02)
Stop times	Delayed	103.13 (26.44)	106.85 (28.50)
	Immediate	103.72 (25.40)	109.94 (21.89)
Middle times	Delayed	69.65 (17.07)	89.86 (33.71)
	Immediate	70.11 (17.48)	82.39 (19.82)
Spread (FWHM)	Delayed	74.23 (31.31)	81.19 (36.73) *
	Immediate	71.22 (31.29)	65.86 (29.30)

Note. Numbers in parentheses are standard deviations. The asterisk (*) marks the effects of delayed reinforcement; the effects of the interaction phase by component.

stimulus, suggesting increased generalization across time.

Fig. 7D shows the middle-time median for both components. As expected, the measure was similar between components at baseline, but in the experimental phase, the middle time was longer in both components compared with baseline, and it was even longer in the delayed component (see Table 1). For the peak time, the phase showed a statistical effect, $\chi^2(6) = 56.19$, $p < .001$, but no statistical effect was observed for the component, $\chi^2(7) = 3.12$, $p = .08$, or the interaction phase by component, $\chi^2(8) = 3.58$, $p = .06$. The factors in the interaction model phase by component (although without statistical significance, this model had the lowest AIC) showed an increased middle time due to the experimental phase, $b = 28.28$, $t(105) = 4.29$, $p < .001$. No statistical effect was observed for the component, $b = 0.37$, $t(174) = 0.12$, $p = .90$, or the interaction phase by component, $b = -7.88$, $t(105) = -1.87$, $p = .06$. The middle times increased with the introduction of the experimental phase, but the delayed reinforcement did not differentially affect the measure, so the observed differences might be explained as a result of the longer intervals-to-reinforcement values in such phase. The standard deviations of the middle times in the delayed component were greater compared with the immediate (see Table 1), which is consistent with a wider generalization gradient that could result from delayed reinforcement. Such a reduction of stimulus control could generate extreme values in some trials, affecting the reliability of the average estimation. A summary of the results obtained in the trial-by-trial analysis and the obtained interaction effects (phase by component) are presented in Table 1.

We computed the optimal number of clusters in which the responses could be grouped to quantify the response patterns during the peak trials. Fig. 8 shows the frequency of optimal clusters per subject and component; that is, how performance for every subject in different trials can be classified (maximizing the gap statistic) in one of four categories (x axis) based on the number of clusters in each trial (one cluster = constant rate; two clusters = low-high rate; three

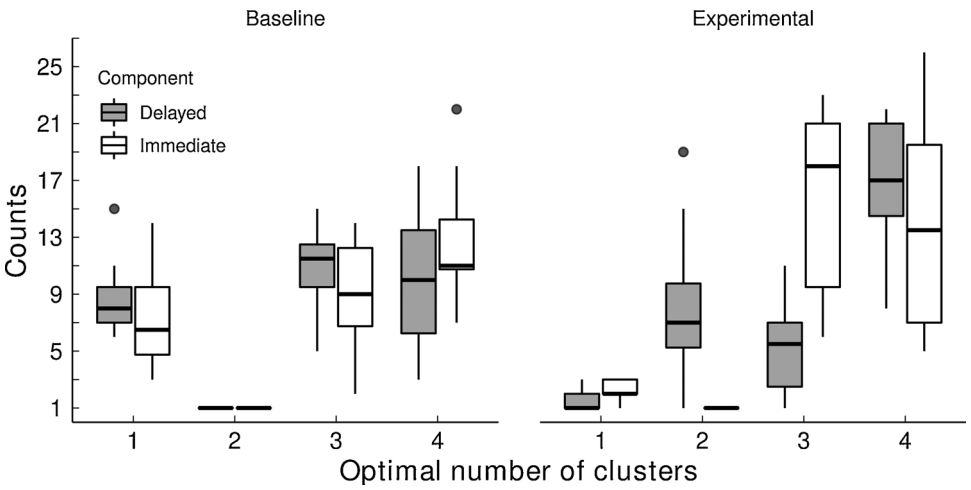

Fig. 8. The optimal number of clusters with the *k*-medoids algorithm shows different patterns of response-states in peak trials. The boxplots are similar between components at baseline; however, the immediate component has a greater frequency among the three clusters, while the delayed component has more clusters with two states in the experimental phase.

clusters = low-high-low rate; four clusters = low-high-low-high rate). For example, if performance never had, say, a low-high or high-low pattern (two clusters) for a subject across trials and components, there would be zero counts in the two optimal clusters for that subject. If the most common pattern was the low-high-low (or break-run-break) pattern, we expected three clusters to be most frequent. At baseline, we did not observe large differences in frequency between components. The pattern with four clusters showed the same frequency as the pattern with three, and the pattern with one cluster between two and four trials is shown below; there were no patterns of two clusters. In the experimental phase, the introduction of delayed reinforcement disrupted the patterns. The frequency of the patterns of one cluster diminished, and the pattern of four clusters increased in both components. However, differences between components appeared in two and three clusters. The delayed component showed more two-cluster patterns but fewer patterns with three clusters. These results suggest departure from the patterns observed at baseline and the immediate component during the experimental phase. The three-cluster pattern exhibited the expected low-high-low typical performance; the four clusters might be the pattern originating from the ramp component in the aggregated functions. The frequency of the usual patterns in the peak procedure changed after delayed reinforcement, and a similar phenomenon was observed in the FI schedules after delayed reinforcement, in which the percentage of intervals with only one response increased (Buriticá and dos Santos, 2017). The emergence of unusual response patterns may evidence more stimulus durations triggering a burst of responses, which is expected with increases in generalization across time.

4. Discussion

Changing an FI schedule to a tandem FI-FT schedule with delayed reinforcement decreased both the response rate and contingency discriminability. Similar decreases in these measures were not observed under a control component that moved from an FI schedule to a yoked schedule. Delayed reinforcement diminished the response rate, response-outcome contingency and increased the intervals between responses and reinforcement. Additionally, the change of phase increased the intervals-to-reinforcement and changed the temporal distributions of head entries in both components. The longer intervals-to-reinforcement in the experimental phase may explain the longer stop and middle-times when the phase changed. Notwithstanding the observed lower response rate, longer start times, longer spread, differences in the function of the response (D_{KL}) and changes in the frequency of response patterns (cluster analysis) may be a consequence of delayed reinforcement independent of longer intervals-to-reinforcement. The statistical effects of the interaction phase by component in start times and spread (FWHM) suggest an effect of delayed reinforcement (component) beyond the longer intervals-to-reinforcement (phase). These data agree with the idea that delayed reinforcement decreases temporal control of behavior, creating flat generalization gradients across time.

In the SET perspective, longer start times in the component with delayed reinforcement were consistent with the evidence that weak motivation increases the thresholds to initiate responding. Stop times were similar between components in the experimental phase, so it could be concluded that the interval timing was similar between components (Gibbon and Church, 1990), with no memory differences due to delayed reinforcement. Nevertheless, the increase in spread suggested less precision in the time estimation, in which the spread was not just longer but highly variable, suggesting less certainty regarding when to expect the reinforcement. The cluster analysis led to the same conclusion: the low-high-low pattern was less frequent in the component with delayed reinforcement and unusual patterns when two or four clusters appeared frequently compared with the baseline and immediate component.

The explanation that motivation affect timing through arousal effects in LeT does not apply to our results because the effects are sustained in time (Buriticá et al., 2016). Nevertheless, LeT may explain the increased generalization if we assume that more states become associated with the instrumental response when delayed reinforcement is used compared with immediate reinforcement, which may be because the time between the instrumental response and reinforcement increased (see Fig. 5), leading to an increase in the associative strength between more states and the instrumental response (lever press). Additionally, the general reduction in associative strength between states (due to competition among them) and the instrumental response may explain the reduction in response rate. Thus, LeT may explain both results, the reduction in response rate and the increased generalization as a result of a reduction of control by the trained interval. Thus, LeT could operate through contingency degradation; if the contingency is low, the associative strength could be divided among different states to create a flat generalization gradient across time. Additionally, because the model is blind regarding the instrumental response, is likely that head entry could also gain some associative strength, because its coincidence with reinforcement in the experimental phase should be higher than the lever press (see Fig. 1). Thus, this characteristic of the model could explain the appearance of such a temporal pattern in head entries in the experimental phase (see Appendix A for a simulation based on Machado et al., 2009).

The reduction of contingency between response and reinforcement may be relevant to the effect of delayed reinforcement, when the contingency is low response rate decreases in VI schedules (Hammond, 1980). Additionally, increased generalization of the LeT explanation arises from the notion that the instrumental response is associated with sufficient strength with more states than usual, a side effect of the decrease of contingency between the instrumental response and

reinforcement. In LeT, the instrumental response has links with more than one state, but delayed reinforcement may diminish the strength with the states close to the FI and spread the associative strength throughout several states. Nevertheless, a strong conclusion regarding the effect of response-reinforcer contingency in timing schedules will require an experiment that manipulates such a contingency. Additionally, a similar analysis of contingency in studies with other forms or reinforcement devaluation will provide evidence about how general this reduction of response-reinforcement contingency could work. In theoretical models of contingency such as those reported by Davison and Nevin (1999), the delay of reinforcement reduced the response-reinforcer discriminability (parameter d_{brij} in the model). If we use the components of the multiple schedule (the tandem FI with delay, and the yoked component) as terminal links in concurrent chained schedules with a similar VI as the first link, we can estimate the effect of reinforcement delay on the discriminability of the response-reinforcer relation.

Delayed reinforcement increased the generalization gradients in the peak procedure. Nevertheless, the mechanism underlying this effect remains unclear. According to SET, the most suitable explanation is a change in the threshold to initiate the response. Our threshold data suggest that changes in value may explain the pattern of results. In accordance with our proposition for LeT, the increased generalization would arise as a decreased association between the instrumental response and the normal set of states, and as a concomitant increase in associative strength between such a response and more previous and subsequent states. Nevertheless, a definitive conclusion would require the modeling of such affirmations and testing of the predictions of such models against parametric variations of the procedure; the parametric variations may detect whether the timing follows a consistent pattern of change due to increased generalization across time or depends on something else, and if the effects are dependent on the parameters used: trained intervals, delay durations, etc.

5. Conclusion

Behavior in the peak procedure can be explained as result of time discrimination or temporal generalization gradients. Low motivation weakens stimulus control and flattens the generalization gradients. Our results are consistent with such an interpretation, which indicates that the delay of reinforcement reduces control over time, creating flat generalization gradients around the trained interval. Such temporal generalization gradients can be generated by two sets of theories: SET models or behavioral models such as LeT. We offer interpretations concerning how both sets of theories could generate flat temporal generalization gradients, but without further evidence, simulations and more data, a conclusive recommendation regarding the utility of any theory concerning this point could be unjustified. The data obtained thus far in the literature agree with our hypothesis of flat generalization gradients across time. The strongest prediction of our hypothesis is an increased variability in performance, as evidenced by the increase in divergence between functions (D_{KL}) and the spread (FWHM).

Finally, we explain our results as an effect of increased generalization due to delayed reinforcement. We determine that despite similar experience times in both components, the generalization gradient across time in FI with delayed reinforcement should be flatter than the gradient in the yoked component. In case training with delayed reinforcement has the effect we suggest, the schedule with the flatter temporal gradient should show less control by the passage of time. In this case, such a schedule is FI with added FT. Our data also suggest that a delay of reinforcement reduces the contingency, or discriminability (of the relation), in response-reinforcement. To establish whether the same reduction in response-reinforcement contingency operates in other manipulations that affect reinforcement effectiveness or motivation, a similar analysis of contingency would be required with such manipulations, and the contingency should be directly manipulated.

Acknowledgments

Thanks to Diana Moran for help in data collection. The research was funded by CONACyT INFR-281265.

Appendix A

We performed two simulations of multiple two-component schedule with a tandem FI 54 s FT 6 s and a Yoked Interval using LeT. The parameters, symbols and values used in the simulation are in Table A1. In Simulation 1, the parameter λ was sampled from the same distribution in both components; in Simulation 2, we sampled λ from two distributions with different σ for each component.

Description of the model

Here we present an algorithm of LeT to simulate the multiple two-component Tandem FI 54 s FT 6 s (delayed component) with a Yoked Interval (immediate component). The LeT implementation follows Machado et al. (2009). Readers interested in the mathematical details can see that

Table A1
Parameters, symbols and values used in both simulations.

Parameter		Values
λ	Rate of activation, sampled every trial from a normal distribution with mean μ and standard deviation σ	Delayed: $\mu = 1$, $\sigma = 0.3$ Immediate: $\mu = 1$, $\sigma = 0.2$
σ	Extinction parameter	1
β	Learning parameter	0.2
n	Behavioral states	Variable
n^+, n^-	States active at the end of FI and peak trials, respectively	Variable
$W_0(n)$	Initial strength of association from every state n to R	0.12
$W_i(n)$	The strength of the association of state n in the i-th trial.	Variable
θ	Threshold decision rule	0.1
π	Proportion of FI trials	0.75
$1 - \pi$	Proportion of peak trials	0.25
T_{del}	Programmed duration of delayed component. It is the FI + FT durations. Also used for the peak trial which are three times longer T_{del}, i.e., $3T_{del} = 3 \times 60$	FI 54 s + FT 6 s
T_{imm}	Duration of the immediate component. Depends on the duration of reinforced trials of the delayed component.	Variable. See step 5 in the algorithm description

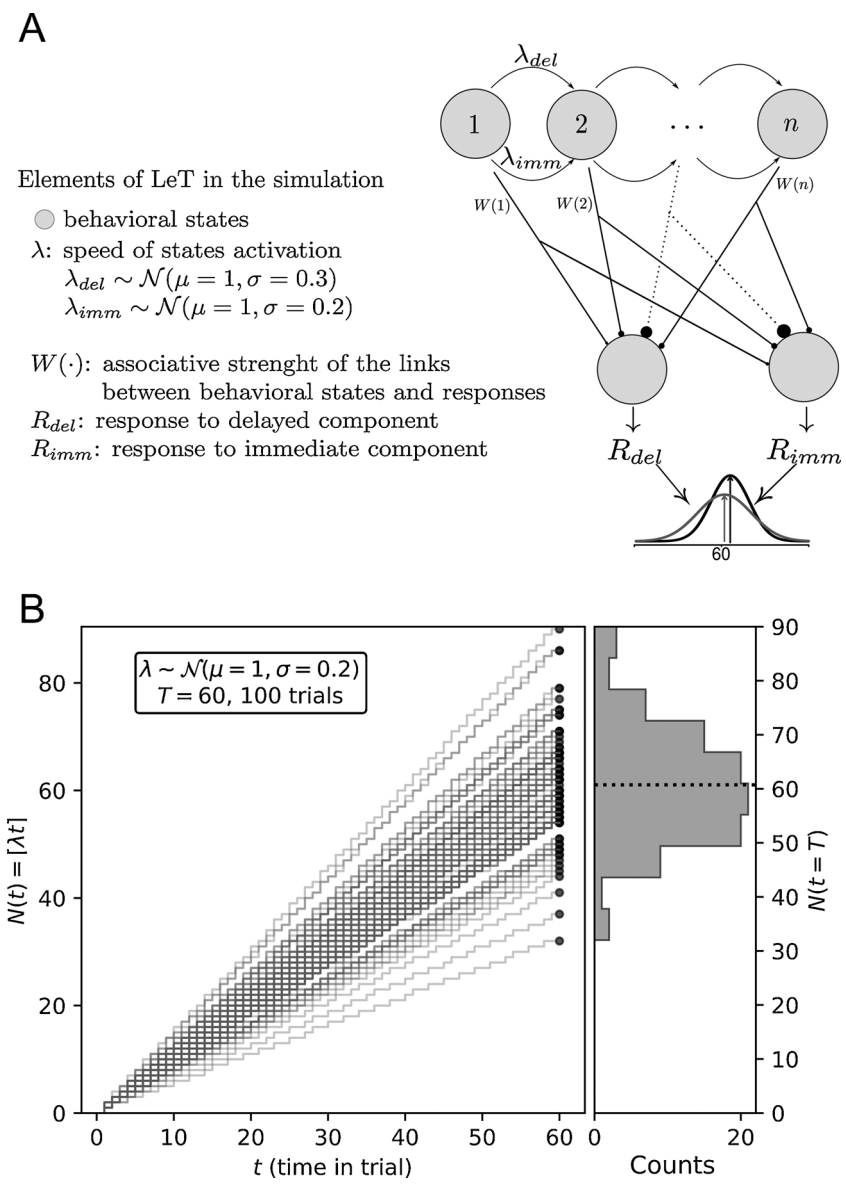

Fig. A1. Panel A: Diagram of the LeT implementation used in the simulations. The text in the left describes briefly the elements of the diagram. Panel B: Simulation of 100 trials with λ sampled from a normal distribution with $\mu = 1$, $\sigma = 0.2$, and using a T of 60 s. The step lines shows the spreading of states as a function of time, each one with a different λ sampled. The dot at the end of every step line is the reinforced state, when $N(t = T)$. The histogram shows the final distribution of reinforced states, with a mean close to 60 (dotted line).

reference.

LeT comprises three components (see panel A of Fig. A1 for a diagram of those components and their relations):

1) A set of behavioral states {1, 2, ...,n}.
2) An operant response R.
3) A vector of associative links between the behavioral states and R.

According to Machado et al. (2009), at the trial onset, states are activated serially from 1 to n at rate λ states/s. Thus, at time t, the state active is the smallest integer greater than the product λt. This is symbolized as $\lceil \lambda t \rceil$, the symbol $\lceil \cdot \rceil$ is the ceiling function. A value of λ is sampled every trial from a normal distribution with mean μ and standard deviation σ. Each state controls the operant response R with strength W. States gain strength if they are reinforced at the end of a trial (T), or lose strength if they were active during the trial but unreinforced (states active at $t < T$). This strength is translated into performance by a simple rule: if the strength of the link of state n is above a threshold, a response occurs, if it is below, there's no response. As trials proceed, different λ are sampled, and the distribution of states active during reinforcement resembles a normal distribution with mean and standard deviation, approximately, of $\mu' = \mu \times T$ and $\sigma' = \sigma \times T$ respectively (see panel B of Fig. A1 for a simulation with a single value of σ).

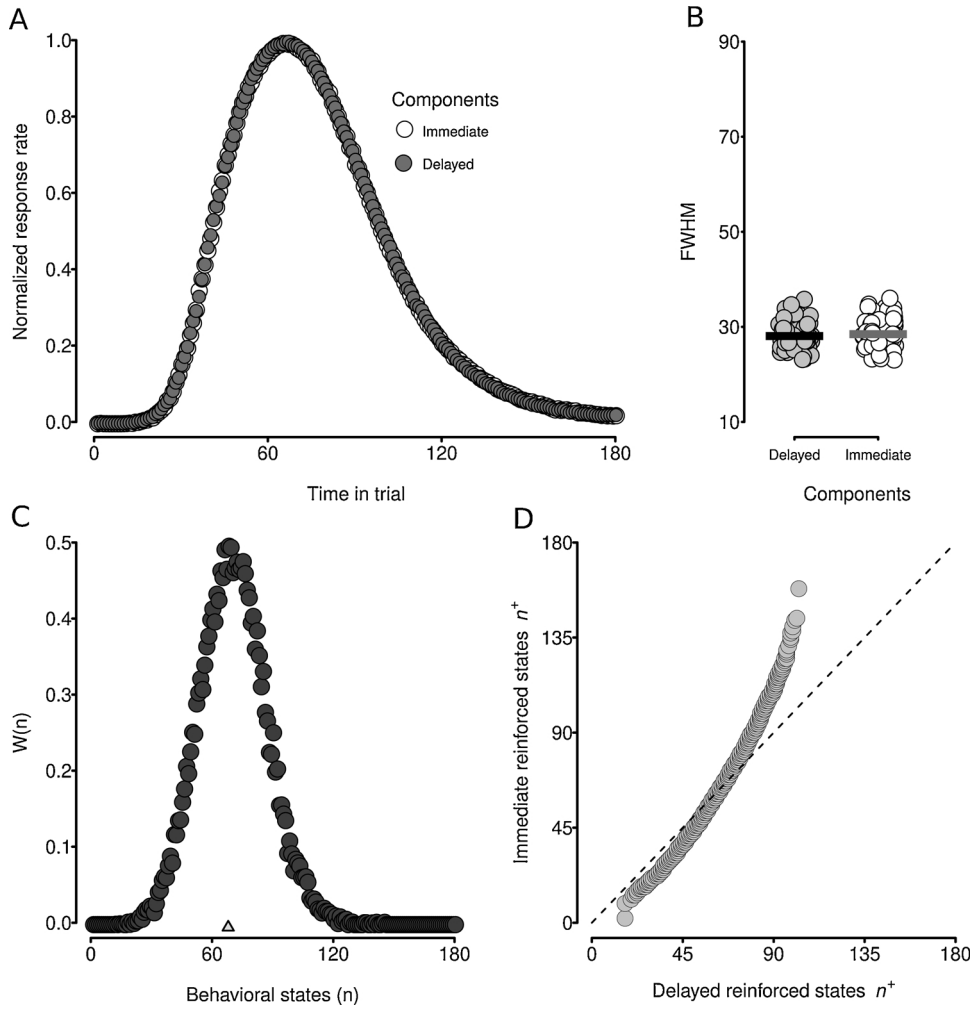

Fig. A2. Simulation 1 of LeT sampling the values of λ from the same distribution. Panel *A:* Response rate functions of both components. Panel B: FWHM for every session (50 trials) and subject, as described in Data analysis section. Panel C: Strength of associative links from every state to the operant response. The small triangle shows the state of the maximum value of $W(n)$. Panel D: QQ-plot of reinforced states of the immediate component as a function of the delayed component. The dashed line shows the one-to-one correspondence of the quantiles.

Simulation 1

Both simulations had 50 trials, 8 subjects and 50 sessions each. For modeling the master component as a Tandem FI 54 s FT 6 s, we assumed that a response will occur when the state active is at the end of the FI, which in turn will depend on the sampled λ value. Because responses are not required during the FT, the reinforcer will be set up for the state active at the end of the FI plus the FT (the number of states after the FI were exactly the value of the FT). The state active at the end of the FI, when a response will occur, plus the FT was taken as the interval-to-reinforcer for the yoked component, so at every master (delayed) trial, we saved the value of the reinforced state and used it as the interval-to-reinforcer on the next yoked (immediate) trial. We did not assume anything else in this first simulation. All parameters were the same for both components, including the spreading rate of states activation. The results of this simulation (Fig. A2) did not show the pattern of responding observed in Figs. 6 and 7 (peak shift to the right in the immediate yoked component, less responding and increased generalization in the delayed component). Because of this, we did another simulation with a different assumption.

Simulation 2

For Simulation 2, we assumed that delayed reinforcement allows more states to control the operant response. One way to do this is by sampling λ from different distributions; for example, taking λ from a distribution with a σ greater for the delayed component than for the immediate. Similar to Simulation 1 the intervals to reinforcement for the yoked (immediate) component were the state active at the FI 54 s FT 6 s. The only difference was in λ, as we describe below.

Algorithm

1 Set parameters values (see Table A1) and create a vector of trials types, randomly sampled from a Bernoulli distribution, with π probability of success (FI trials), and $1 - \pi$ of no reinforcement (peak trials). Hence, if Tr is the number of trials, the number of F (for FI trials) and P (for peak trials), are $F = \pi Tr$ and $P = (1 - \pi)Tr$. Every trial counts for two, one for every component, and computations are performed serially: First for

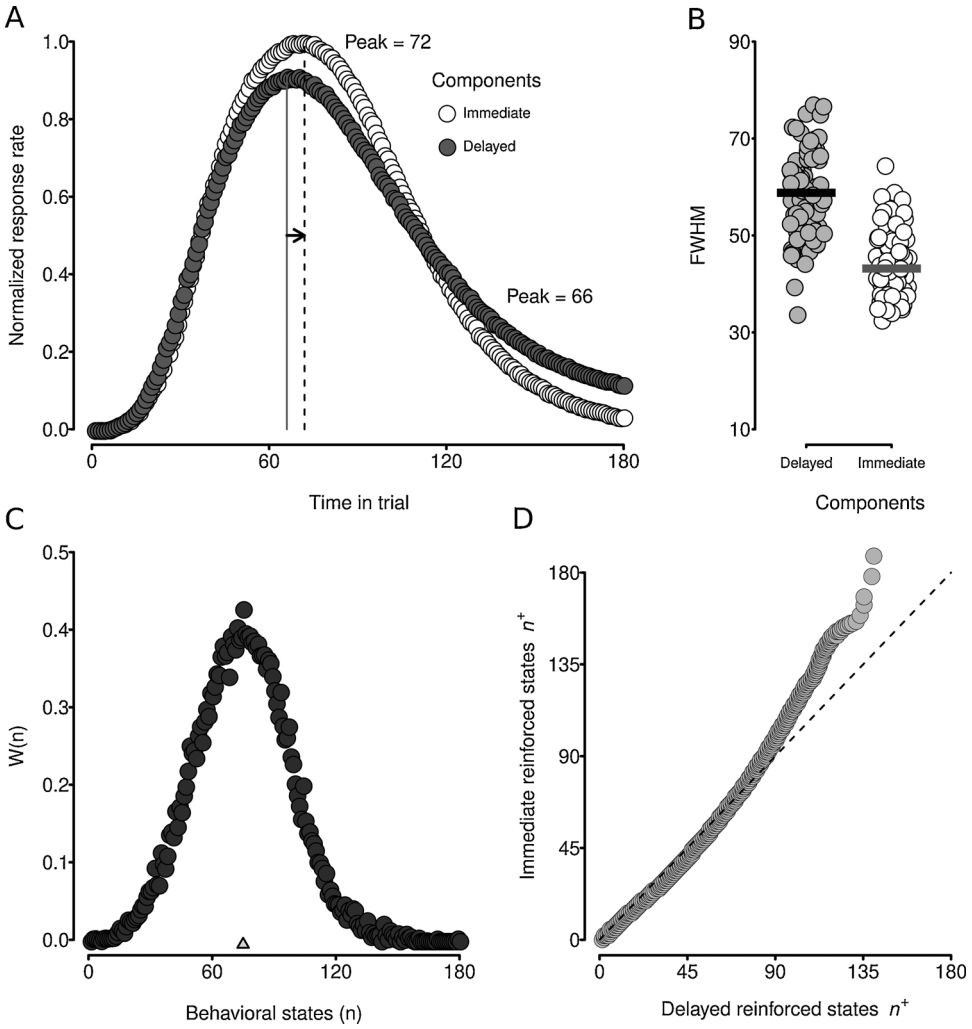

Fig. A3. Simulation 2 of LeT sampling the values of λ for both components from different distributions. Panel *A:* Response rate functions of both components. The continuous and dashed lines show the time of the mean response rates for the delayed and immediate components, respectively. The arrow illustrates the shift of the peak time. Panel B: FWHM for every session and subject, greater for the delayed component than for the immediate. Panel C: Strength of associative links from every state to the operant response. The small triangle shows the state of the maximum value of $W(n)$. Panel D: QQ-plot of reinforced states of the immediate component as a function of the delayed component. The dashed line shows the one-to-one correspondence of the quantiles.

the master component and then for the yoked.

2 For every trial
  a Determine the trial type (FI or peak trial).
  b Sample a value of λ for each component.
3 If the trial type is P, go to step 10. If is F, then for every t second in 1,2, ...,54.
  a Establish the state active at time t, $N(t)$, as the smallest integer greater than λt, or $N(t) = \lceil \lambda t \rceil$.
  b Decide if there is a response at time t, $R(t)$. If the strength of the associative link at $N(t)$ is above θ, that is, $W\{N(t)\} > \theta$, $R(t) = 1$, otherwise $R(t) = 0$.
4 Determine the state n^+ to be reinforced of the delayed component. Its value is $n^+ = \lceil \lambda 54 \rceil + 6$.
5 Set n^+ of the delayed component as the duration T_{imm} for the immediate (yoked) component.
6 For every t in 1,2, ..., T_{imm} repeat step 3 with the λ value for the immediate component.
7 Determine the state n^+ to be reinforced of the immediate component. Its value is $n^+ = (T_{imm}) = \lceil \lambda T_{imm} \rceil$.
8 Determine the strength of the links as follows
  a For n^+, the increase is $W_i(n^+) = W_{i-1}(n^+) + \beta\{1 - W_{i-1}(n^+)\}$.
  b For all $n < n^+$, that is, for every state active and not reinforced during the trial, decrease the link strength according to $W_i(n) = W_{i-1}(n) - (\alpha/n^+)W_{i-1}(n)$.
9 End of F trial. Save trial statistics and go to step 2.
10 For both components, if the trial type is P, then for every t second in 1,2, ..., $3T_{del}$ do step 3.a and 3.b.
11 Determine the state active n^- at the end of the nonfood trial as $n^- = N(3T_{del}) = \lceil \lambda 3T_{del} \rceil$, with the corresponding λ sampled in step 2.
12 Decrease link strength for all $n \leq n^-$ according to $W_i(n) = W_{i-1}(n) - (\alpha/n^-)W_{i-1}(n)$.
13 End of a P trial. Save trial statistics and go to step 2.

The R-code for the algorithm can be found in the same repository mentioned in the Data analysis section. https://github.com/jealcalat/

Generalization_decrement_data-analysis.

Results

Fig. A2 shows results for Simulation 1 where we sampled λ from the same distribution for both components. The response rate functions (panel A) for both components are superimposed and the spread (FWHM - panel B) is the same. Panel C of Fig. A2 shows the strength of associative links with a maximum value close to 60. Increases in reinforced states n^+ for the immediate component were bigger as a function of n^+ of the delayed component (panel D of Fig. A2). This simulation did not show the results found in the experiment.

Fig. A3 shows the results of Simulation 2 when we sampled λ from different distributions for each component. Response rate functions are different between components (panel A). There is a (rather) small peak shift to the right in the immediate component, and a decrease in response rate for the delayed component. FWHM is wider for the delayed component (panel B), and wider than in Simulation 1, which is not surprising considering that we assumed just one set of states for both components, so that the changes at every trial in the associative links affected the performance of both components. This can be seen in panel C of Fig. A3 compared to Fig. A2. The distribution of $W(n)$ is flatter in the second simulation, reflecting the effects of sampling λ from a distribution with a bigger value of σ.

Taken together, the results of the simulations show that reinforcing a large number of behavioral states mimics the results in our experiment. A wider distribution of behavioral states associated with the operant response can be thought, we suggest, as greater uncertainty of when to respond, thus degrading the response-reinforcer contingency.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.beproc.2019.103978.

References

Auguie, B., 2017. gridExtra: Miscellaneous Functions for “Grid” Graphics.

Balcı, F., 2014. Interval timing, dopamine, and motivation. Timing Time Percept. 2, 379–410. https://doi.org/10.1163/22134468-00002035.

Balcı, F., Ludvig, E.A., Brunner, D., 2010. Within-session modulation of timed anticipatory responding: when to start responding. Behav. Processes 85, 204–206. https://doi.org/10.1016/J.BEPROC.2010.06.012.

Buriticá, J., dos Santos, C.V., 2017. Reinforcement value and fixed-interval performance. J. Exp. Anal. Behav. 108, 151–170. https://doi.org/10.1002/jeab.279.

Buriticá, J., Vilchez, Z., dos Santos, C.V., 2016. Temporal discrimination and delayed reinforcement. Behav. Processes 130, 71–74. https://doi.org/10.1016/j.beproc.2016.07.009.

Catania, A.C., 1970. Reinforcement schedules and psychophysical judgments. In: Schoenfeld, W.N. (Ed.), The Theory of Reinforcement Schedules. Appleton Century Crofts, New York, pp. 1–42.

Church, R.M., Meck, W.H., Gibbon, J., 1994. Application of scalar timing theory to individual trials. J. Exp. Psychol. Anim. Behav. Process. https://doi.org/10.1037/0097-7403.20.2.135.

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory, 2nd ed. John Wiley & Sons, Hoboken, NJ.

Davison, M., Nevin, J.A., 1999. Stimuli, reinforcer, and behavior: an integration. J. Exp. Anal. Behav. 71, 439–482. https://doi.org/10.1901/jeab.1999.71-439.

de Carvalho, M.P., Machado, A., Vasconcelos, M., 2016. Animal timing: a synthetic approach. Anim. Cogn. 19, 707–732. https://doi.org/10.1007/s10071-016-0977-2.

DeHart, W.B., Kaplan, B.A., 2019. Applying mixed-effects modeling to single-subject designs: an introduction. J. Exp. Anal. Behav. 111, 192–206. https://doi.org/10.1002/jeab.507.

DeRusso, A., Fan, D., Gupta, J., Shelest, O., Costa, R., Yin, H., 2010. Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. Front. Integr. Neurosci. 4, 17. https://doi.org/10.3389/fnint.2010.00017.

Dowle, M., Srinivasan, A., 2018. data.table: Extension of `data.frame`.

Elcoro, M., Lattal, K.A., 2011. Effects of unsignaled delays of reinforcement on fixed-interval schedule performance. Behav. Processes 88, 47–52. https://doi.org/10.1016/J.BEPROC.2011.07.001.

Galtress, T., Kirkpatrick, K., 2009. Reward value effects on timing in the peak procedure. Learn. Motiv. 40, 109–131. https://doi.org/10.1016/j.lmot.2008.05.004.

Galtress, T., Marshall, A.T., Kirkpatrick, K., 2012. Motivation and timing: clues for modeling the reward system. Behav. Processes 90, 142–153. https://doi.org/10.1016/j.beproc.2012.02.014.

Gibbon, J., Church, R.M., 1990. Representation of time. Cognition 37, 23–54. https://doi.org/10.1016/0010-0277(90)90017-E.

Hammond, L.J., 1980. The effect of contingency upon the appetitive conditioning of free-operant behavior. J. Exp. Anal. Behav. 34, 297–304. https://doi.org/10.1901/jeab.1980.34-297.

Hinton, S.C., Rao, S.M., 2004. “One-thousandone …one-thousandtwo …”: Chronometric counting violates the scalar property in interval timing. Psychon. Bull. Rev. 11, 24–30. https://doi.org/10.3758/BF03206456.

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, Hoboken, NJ.

Killeen, P.R., Fetterman, J.G., 1988. A behavioral theory of timing. Psychol. Rev. 95, 274–295. https://doi.org/10.1037/0033-295X.95.2.274.

López, F., Menez, M., 2012. Transference effects of prior non-contingent reinforcement on the acquisition of temporal control on fixed-interval schedules. Behav. Processes 90, 402–407. https://doi.org/10.1016/J.BEPROC.2012.04.007.

Lotfizadeh, A.D., Edwards, T.L., Redner, R., Poling, A., 2012. Motivating operations affect stimulus control: a largely overlooked phenomenon in discrimination learning. Behav. Anal. 35, 89–100.

Ludvig, E.A., Balcı, F., Spetch, M.L., 2011. Reward magnitude and timing in pigeons. Behav. Processes 86, 359–363. https://doi.org/10.1016/J.BEPROC.2011.01.003.

Ludvig, E.A., Conover, K., Shizgal, P., 2007. The effects of reinforcer magnitude on timing in rats. J. Exp. Anal. Behav. 87, 201–218. https://doi.org/10.1901/jeab.2007.38-06.

Machado, A., 1997. Learning the temporal dynamics of behavior. Psychol. Rev. 104, 241–265. https://doi.org/10.1037/0033-295X.104.2.241.

Machado, A., Malheiro, M.T., Erlhagen, W., 2009. Learning to time: a perspective. J. Exp. Anal. Behav. 92, 423–458. https://doi.org/10.1901/jeab.2009.92-423.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2018. cluster: Cluster Analysis Basics and Extensions.

Pinheiro, J., Bates, D., Debroy, S., Sarkar, D., 2018. Nlme: Linear and Nonlinear Mixed Effects Models.

R Core Team, 2018. R: A Language and Environment for Statistical Computing.

Roberts, S., 1981. Isolation of an internal clock. J. Exp. Psychol. Anim. Behav. Process. https://doi.org/10.1037/0097-7403.7.3.242.

Sizemore, O.J., Lattal, K.A., 1978. Unsignaled delay of reinforcement in variable-interval schedules. J. Exp. Anal. Behav. 30, 169–175. https://doi.org/10.1901/jeab.1978.30-169.

Sizemore, O.J., Lattal, K.A., 1977. Dependency, temporal contiguity, and response-independent reinforcement. J. Exp. Anal. Behav. 27, 119–125. https://doi.org/10.1901/jeab.1977.27-119.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B (Statistical Methodol.) 63, 411–423. https://doi.org/10.1111/1467-9868.00293.

Wickham, H., 2017. tidyverse: Easily Install and Load the “Tidyverse”.

Wilk, M.B., Gnanadesikan, R., 1968. Probability plotting methods for the analysis of data. Biometrika 55, 1–17. https://doi.org/10.2307/2334448.