

Integrating Timing and Conditioning Approaches to Study Behavior

Kathryn L. Kalafut, David M. Freestone, Mika L. M. MacInnis, and Russell M. Church
Brown University

Skinner and Pavlov had innovative ways to measure both the times of their subject's responses, as well as the rate of their responses. Since then, different subfields within the study of animal behavior have prioritized either the rate or timing of responses, creating a divide in data and theory. Both timing and conditioning fields have proven fruitful, producing large bodies of empirical data and developing sophisticated models. Despite their individual successes, a unified view of simple behavior is still lacking. This may be caused, at least in part, by the differential emphasis on data collection and analysis techniques. The result is that these subfields produce models that fit their data well, but fail to translate to the other domain. This is startling given the fact that both subfields use nearly identical experimental procedures. To highlight similarities within the subfields, and provide empirical data in support of this integration, 18 Sprague-Dawley rats were trained on trace, delay, and backward conditioning procedures. Using these empirical data we discuss how traditional summary measures used by these subfields can be limiting, and suggest methods that may aid in the integration of these subfields toward common goals.

Keywords: timing, conditioning, response rate, transition times, discrimination ratio

The principles of classical conditioning were first explored by Pavlov, as described in his book *Conditioned Reflexes* (1927). In it, he discusses some of the conditions necessary to establish the stimulus-response connection that has become the basis for many modern theories of conditioning. Pavlov recorded the secretion of saliva in drops as a function of time since the presentation of a stimulus. Each 0.01-cc drop of saliva was marked by a tick on a piece of paper. The faster the rate of salivation, the more tightly packed the tick marks. Using this method, Pavlov and his collaborators were able to report how the distribution of drops depended on the time between food presentations (p. 37), as well as how this distribution changed as a function of experimental manipulations such as delay, trace, and backward conditioning procedures (p. 40).

Only a decade later, Skinner (1938) published *The Behavior of Organisms*. He too recorded both the time and rate of behavioral responses, but did so on a cumulative recorder. The rate at which the cumulative recorder scrolled allowed Skinner to determine the time of each behavioral response. The slope between any two points on the cumulative record were used to estimate the response rate. Ferster and Skinner (1957) qualitatively described the behavior on many reinforcement schedules. The reader could easily see

the patterns they described in the cumulative records (e.g., the postreinforcement pause, fixed-interval scallop, steady response rate, etc.).

Although both Pavlov and Skinner recorded the times of critical events and responses, over the years, researchers began to focus on either the time of responses or the rate of responses, but rarely both. This led to the development of separate subfields within animal learning. The conditioning subfield prioritized the frequency of responses, while the timing subfield prioritized the time at which responses occurred.

Different measures and analysis tools led to different theories. Within the subfield of conditioning, the Rescorla-Wagner model is often used to describe how associative strength (measured via response rates) changes as a function of conditional stimulus (CS)-unconditional stimulus (US) pairings (Rescorla & Wagner, 1972). Within the subfield of timing, Scalar Expectancy Theory (SET) is often used to describe the time at which responses occur as a function of the CS-US interval (Gibbon, 1977). The Rescorla-Wager model does not predict response times, and SET does not predict response rates. These are just two examples of timing and associative models and their predictions of behavior. Noticing this, researchers have begun to bridge the gap by discussing commonalities between timing and conditioning processes (Kirkpatrick, in press; also see the special issue in *Behavioral Processes: Associative & Temporal Learning*, 2013) and by building hybrid models like Modular Theory (Guilhardi, Yi, & Church, 2007), the Temporal Coding Hypothesis (Matzel, Held, & Miller, 1988; Savastano & Miller, 1998), and others (Machado, Malheiros, & Erlhagen, 2009).

Although both conditioning and timing subfields have flourished independently for several decades, the work they have produced has not led to a general theory of behavior. A more integrative approach would allow for both conditioning and timing subfields to benefit from each other's data and analysis tools. Published data might be more accessible for testing new and

This article was published Online First July 28, 2014.

Kathryn L. Kalafut, David M. Freestone, Mika L. M. MacInnis, and Russell M. Church, Department of Cognitive, Linguistics, and Psychological Sciences, Brown University.

David M. Freestone is now at the Department of Neural Science, New York University.

This research was supported by the National Institute of Mental Health Grant MH44234 to Brown University.

Correspondence concerning this article should be addressed to Kathryn L. Kalafut, Department of Cognitive, Linguistics, and Psychological Sciences, Brown University, 190 Thayer Street, Providence, RI 02906. E-mail: Kathryn_Kalafut@brown.edu

existing models of behavior. Common problems might be easier to identify, and, hopefully, easier to study. In an attempt to advance the integration process, we would like to propose a list of methods of analysis methods that may provide useful information to both conditioning and timing researchers. We encourage others to add to this list. The methods should be familiar to researchers in both subfields and should aid in the mutual understanding of behavior on both basic and theoretical levels. Some analysis methods are likely to be simple descriptive or graphical representations of the data, and others are likely to be quantitative summary measures that can be used for statistical analyses. There are likely to be many such analysis methods, and although a comprehensive list of them should be compiled, that is beyond the scope of this article, so we suggest only a few here.

To describe this approach, we collected data from 18 rats on three standard conditioning procedures (trace, delay, and backward conditioning). These procedures have been used by conditioning and timing researchers, but are liked for different reasons. For the conditioning researchers, these procedures have been paramount in establishing results that help define CS-US associations. For the timing researchers, these procedures have been paramount in establishing results that identify how intervals between events are perceived. First, a description of a classic conditioning measure (the discrimination ratio) and a classic timing measure (the start times) are presented, followed by a discussion of the information these measures provide, and how this information has been used to drive the theories within the subfields. This direct comparison of these measures highlights how each is limiting outside of the subfield in which they are commonly used. Several other methods of visual presentation and analysis are discussed in the hope of maintaining the benefits for each subfield, while furthering the possibility of integration of these two subfields in the future.

Method

Animals

Eighteen experimentally naive male Sprague-Dawley rats were used in the experiment (Taconic Laboratories, Germantown, NY). They were 30 to 37 days old upon arrival, and were handled daily from arrival to the onset of the experiment (Taconic Laboratories, Germantown, NY). Testing began one month after arrival. The rats were housed individually in a colony room on a 12:12 hour light/dark cycle (lights off at 9:30 a.m.). Dim red light illuminated the colony room and the testing room. In addition to the reinforcements obtained in the experimental chambers, the rats were fed 15 g of FormuLab 5008 food daily. Water was available ad libitum in both the home cages and the experimental chambers. The rats were tested daily in two squads, starting at approximately 7:00 and 9:30 a.m.

Apparatus

Twelve chambers ($25 \times 30 \times 30$ cm), each located inside a ventilated, noise-attenuating box ($74 \times 38 \times 60$ cm), were used in testing. Each chamber was equipped with a food cup ($5 \times 5 \times 2$ cm) and a water bottle. A magazine pellet dispenser (Model ENV-203, Med Associates, St. Albans, VT) delivered 45-mg Dustless Precision Pellets (Bio-Serv, Rodent Grain-Base Formula, Frenchtown, NJ) into the food cup. Each head entry into the food

cup was transduced by a LED-photocell and was recorded. The water bottle was mounted on the outside of the experimental chamber. Water was available through a tube that passed through the middle of one of the walls. The food cup was located in the center of the wall opposite the water bottle. The CS, a 70-dB white noise, with an onset rise time and termination fall time of 10 ms, was generated by an audio amplifier (Model ANL-926). Two Gateway Pentium computers, running the Med-PC Medstate Notation Version 2.0 (Tatham & Zurn, 1989), controlled experimental events and recorded the time at which events occurred with 2-ms resolution.

Procedure

In all conditions the CS (a 70-dB white noise) began at 0 s and ended at 20 s (shaded gray). The six conditions differed only in the time that food was available. For three conditions, food was available at 0, 10, or 20 s after CS onset (*delay conditioning*). For two conditions, food was available at either 30 or 40 s after CS onset (*trace conditioning*). And for one condition (*backward conditioning*) food was made available 10 s before CS onset (or, equivalently, 110 s after). Food was delivered contingent on the first head-entry after food availability. All of these conditions are shown in Figure 1. Half of the cycles (intermixed) ended without food, regardless of the rat's behavior (nonfood cycles). There were approximately 60 cycles in each 2-hr session.

The 18 rats were randomly assigned to one of six groups, each group consisting of three rats. A random Latin Square design was used in which each group received each condition for 20 sessions (a phase), but in different orders. Thus, in each of the six phases, each condition was presented. Although an instrumental procedure is used (i.e., a head entry was required to deliver the food), the language used throughout the article will cross instrumental and classical conditioning literatures. This is to aid the reader. Direct comparisons of classical and instrumental procedures have demonstrated that, after controlling for the number of stimuli, both the

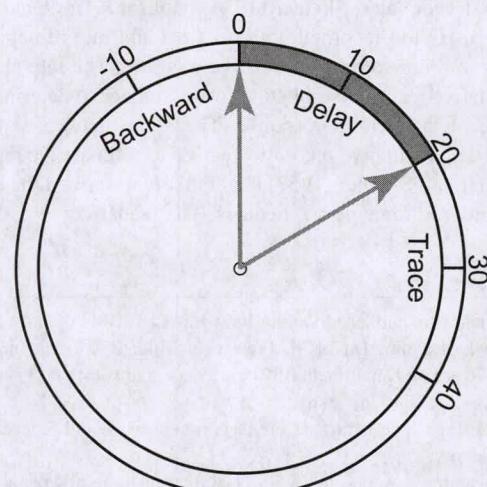


Figure 1. A schematic of each procedure used in this experiment. Time zero indicates the onset of the conditional stimulus, shown by the gray shadowing. The time food is made available in each procedure is indicated by each tick mark, and labeled with its corresponding name.

response rate and temporal structure of the responses are nearly identical (Freestone, MacInnis & Church, 2013). A fixed interval (120 s) was chosen in this preparation to highlight the cyclical nature of the methodology, to highlight how both subfields can use the same data sets, and has been used previously in the literature (Farmer & Schoenfeld, 1966; Han et al., 2003; Plotkin & Oakley, 1975). Choosing a random interval would change our results and conclusions somewhat, but would not at all alter the measures we suggest as tools for integration.

The last 10 sessions of each condition were used in the analyses. Only cycles preceded by food cycles were used. This ensured that the number of cues presented in each cycle was the same for all cycles used in the following analyses (previous food delivery, CS onset, and CS termination).

Results

A Popular Conditioning Measure

When conducting a delay, trace, or backward conditioning procedure, a conditioning researcher may focus on the strength of association between the CS and the US, measured by the response rate (or relative response rate) during the CS. Although response rates are likely the result of many factors, they have been most consistently used as a measure of the strength of an association (Hull, 1943; Skinner, 1938; Rescorla & Wagner, 1972; Hall & Pearce, 1979; Domjan, 2014). Other popular measures, such as resistance to change and outcome devaluation, use a transformed response rate as measures of strength of association (e.g., Nevin, Mandell, & Yarensky, 1981; Rescorla, 1992). The conditioning researcher would note that the absolute number of responses does not matter; individual differences among subjects and procedures can easily obscure true effects and hurt interpretation. Instead, one might report a discrimination ratio: the ratio of responses during the CS to the total number of responses during the CS and during an equivalent baseline interval (in our case, 20 s immediately before the onset of the CS on no food cycles). A ratio of 0.5 indicates no difference in the number of presses during the CS (relative to before). A value higher than 0.5 generally indicates excitatory conditioning, and a value less than 0.5 generally indicates inhibitory conditioning (Moscovitch & Lolordo, 1968). This measure would be applied across procedures and the results would be used to interpret the level of associative strength between the CS and the US. The textbook result is that, in general, delay procedures produce excitatory conditioning, trace procedures produce less excitatory conditioning, and backward procedures produce either no conditioning, or inhibitory conditioning (e.g., Domjan, 2004).

Of course, a discrimination ratio is not the only measure of interest to a conditioning researcher. This measure was chosen as a popular example. The top panel in Figure 2 shows the discrimination ratio across the six procedures, ordered by the time of food availability since CS onset, using nonfood cycles. Our results support the textbooks conclusions that the discrimination ratio is strongly influenced by the procedure, $F_{(5, 85)} = 111, p < .001$, partial eta squared = .902; mixed within-subject ANOVA. Inhibitory conditioning occurs when the US is presented 10 s before the CS (*backward conditioning*). Strong excitatory conditioning occurs when the US occurs during the CS (*delay conditioning*), and

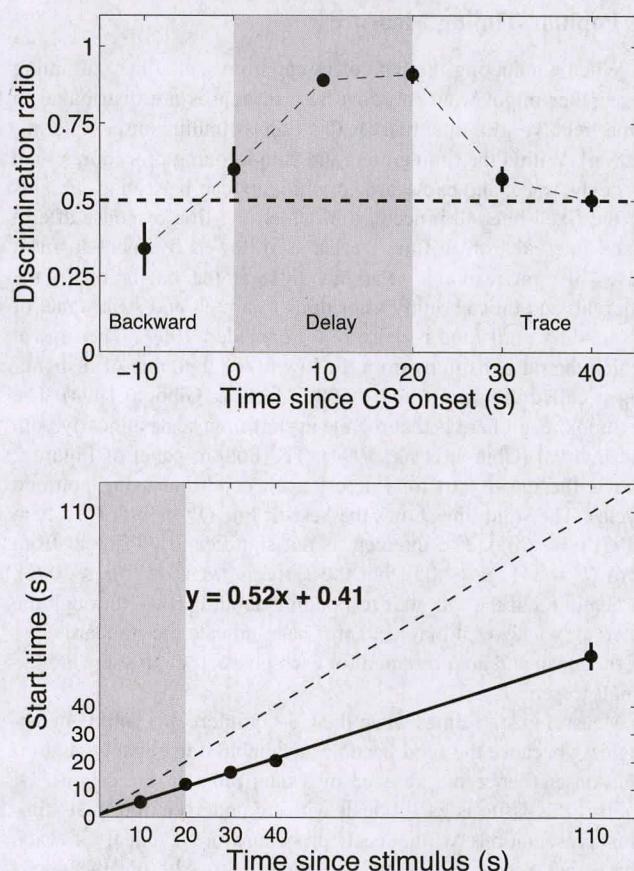


Figure 2. Top panel: The discrimination ratio was calculated by dividing the number of responses during the conditional stimulus (CS) to this number plus the number of responses made in the 20 s before the CS onset for each subject before being combined. Only nonfood cycles were used. Bottom panel: The mean start time calculated across each procedure. In both panels the gray shading indicates the time the CS was present, and the error bars indicate 95% confidence intervals (though some may be difficult to see). Only nonfood cycles were used.

little or no excitatory conditioning occurs when the US occurs long after the CS (*trace conditioning*). When the data are organized in this way, a clear quadratic trend is shown, $F_{(1, 18)} = 476, p < .001$, partial eta squared, .975. The order in which the rats were tested in each condition did not appear to influence the discrimination ratio (no main sequence effect, $F_{(5, 12)} = 1.52, p = .257$, partial eta squared, .387) and there were no post hoc order differences ($p > .293$). Error bars indicate the standard error of the mean for each phase. (Note that one of them was smaller than the filled circles.)

The discrimination ratio demonstrates the differences of associative strength of the CS (relative to the absence of the CS) across these procedures. Although this measure is quite popular in the conditioning subfield as a measure of what the animal has learned, it does not provide much information to those outside the subfield. For example, the discrimination ratio provides no information regarding how time may also serve as a conditioned stimulus, or how subjects respond during other parts of the cycle. This limits the view of behavior to a summary measure of a ratio of two 20-s intervals.

A Popular Timing Measure

When conducting the type of research presented here, a timing researcher might want to know how responses are distributed in time, relative to a time marker (e.g., any stimulus onset or termination). Within the timing literature, conditioning procedures such as delay, trace, and backward conditioning can be seen as variants of the fixed-interval procedure in which the first response after a fixed interval from a time marker is followed by food. In most cases, the rat responds relatively little at the beginning of the interval, and then abruptly transitions to a high and steady rate of responding until food is delivered (Schneider, 1969). The time at which the rat transitions from the low to the high rate of responding is called the *start time* (Church, Meck, & Gibbon, 1994). The textbook result here is that the mean start time scales linearly with the interval (Church et al., 1994). The bottom panel of Figure 2 shows the mean start time across each condition using nonfood cycles. The solid line shows the best-fit line ($R^2 = .94$, $F_{(1, 81)} = 1197$, $p < .001$). The intercept is not significantly different from zero ($b = .41$, $p > .05$), but the slope is ($m = .52$, $p < .001$), indicating that the rats start responding about halfway through the interval, whatever it may be. Error bars indicate the standard error of the mean and are presented for each phase, though some are too small to see.

Measuring start times when the CS-US interval is zero is meaningless¹ because the food becomes available immediately at stimulus onset (hence the absence of a data point in the origin). Of course, this issue is easily dealt with by understanding that stimulus presentations within each procedure are cyclical (in each phase they repeat in exactly the same way). Altering the x axis from the time since the CS onset, for example, to an axis spanning the time from food-to-food allows for transition times to be calculated for each procedure. But there is a bigger issue; the start time provides no information about the response rate (among other things). Viewing the time of responses within a cycle provides valuable information for those within the timing subfield, but very little to those outside of it, who may be more interested in the rate of behavior across each cycle.

Methods Toward Integration

Summary measures that are unique, or are important to only a subset of those studying behavior isolate behavioral findings to that particular group. Allowing theory to dictate data analysis and graphical representation is incredibly useful, but only if all readers want to know how the data change the interpretation of that particular theory. Those readers interested in the research for a different reason (e.g., they are interested in the procedure) might be left frustrated with how much the article leaves out. Most conditioning and timing experiments utilize the same general procedures in terms of presenting stimuli and recording responses. How a scientist uses these data may be dictated by the specific question being asked. The results from almost any experiment, if prepared and presented in a more integrative manner, could aid other scientists (from multiple subfields of behavior) in tackling similar questions. Of course, there is a careful tradeoff between being too specific and being too general. Identifying methods that are both specific and of use to both subfields is a difficult but important problem.

The goal of this article is to describe three measures and discuss how they may mutually benefit both subfields. The measures are the response gradient (graphical and quantitative), the raster plot (graphical), and the change point (quantitative). Some readers may already be familiar with some (or all) of these. We are not attempting to break new ground; we are attempting to integrate existing foundations.

Method 1: The Response Gradient

Reporting the entire function form rather than just reporting descriptive statistics is emphasized to all new scientists in their behavioral statistics course. One reason for this is that descriptive statistics are misleading if the distributional assumption is wrong (e.g., assuming normality when the data are in fact, exponential). Another reason is that, the distribution form (plus its descriptive statistics) often suggests underlying psychological mechanisms. Different transformations of the data (like averaging) may help identify these underlying mechanisms, as long as the transformation does not obscure the true effect (as averaging often does).

The response gradient is a useful integrative measure for these same reasons. It provides a basis for the function form of the responses as it unfolds over time, and it shows the number of responses in each interval (usually 1-s bins) since some event (such as stimulus onset). The measure has been around for decades, but often goes unreported. It is easily calculated as a scaled histogram of the response times, and indicates not only the number of responses, but also how these responses are distributed over time.

Figure 3 shows the response gradients during delay, trace, and backward conditioning procedures separately for both food and non-food cycles. For this figure, the mean response rate in each 1-s bin is plotted as a function of time since the CS onset. The filled circles represent the response rate during cycles in which food was delivered; the open circles represent the response rate during cycles in which no food was delivered. The vertical black line indicates the time when food was available (on food cycles) or would have been available (on nonfood cycles).

Like a statistical distribution function, a response gradient allows one to estimate many descriptive measures. For example, a discrimination ratio can be estimated across conditions relatively easily (compare the response rates in the gray region to response rates at the end of the interval). It can be formalized by simply taking ratios of the total number of responses in different parts of the interval. This can be done in most software packages (like Matlab) in a single line of code. In symbols, the discrimination ratio is still $A/(A + B)$, but now A and B can be computed directly from the response gradient:

$$A = \sum_{t=0}^{20} r(t)$$

and

¹ Calculating the start times for this group is possible, and returns a bimodal distribution with two modes at time zero and at 76 seconds. These times capture the burst of responding when the CS is presented (at time zero) and anticipatory responding to the next impending CS presentation. The means of these times (as the start times for the other conditions are calculated) does not accurately represent the underlying processes seen in the distribution.

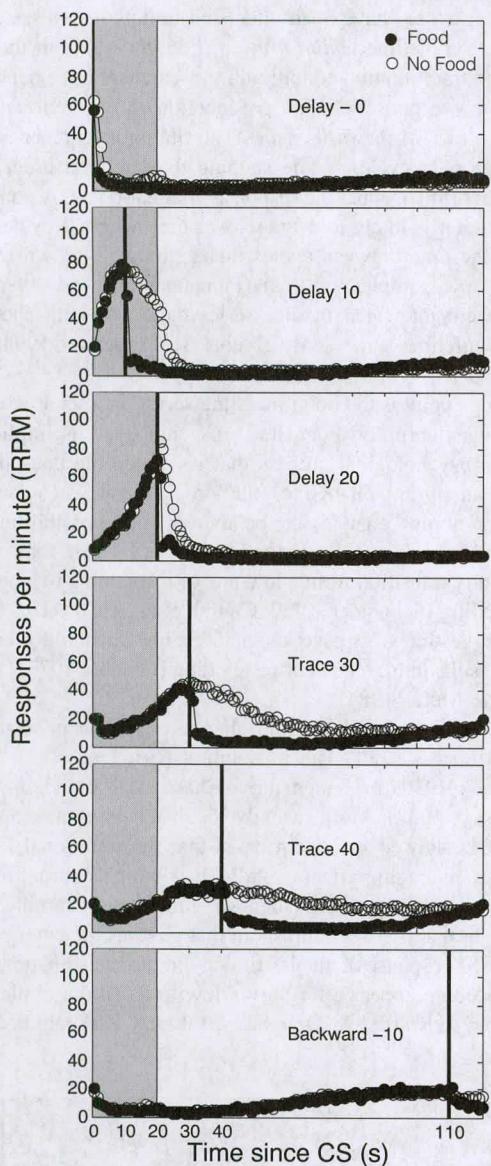


Figure 3. Response gradients observed during food (black circles) and nonfood (white circles) cycles when food was made available (indicated by vertical black line) at 0, 10, and 20 s after conditional stimulus (CS) onset (delay conditioning), 30 and 40 s after CS onset (trace conditioning), and at 110 s after CS onset (backward conditioning).

$$B = \sum_{t=100}^{120} r(t)$$

where $r(t)$ is the response rate at time t . That is, A is the sum of the responses that make up the gradient from 0 to 20 s (the CS interval). And B is the sum of the responses that make up the gradient in the 20 s before the CS onset (which, in this case, happens to also be the last 20 s of the response gradient). Any multiplicative scale factor introduced to transform the gradient (e.g., into responses per minute), does not affect the discrimination ratio.

Some summary statistics representing the start times can be estimated from the response gradient, too. This can either be

approached empirically (compute the ratio of the slope of the function to the location of the midpoint), or it can be done with curve fitting. Guilhardi et al. (2007) make the case that if the start times are normally distributed, then the response gradient will approximate the cumulative normal distribution (the logic is that the response gradient is largely made from averaging individual trial step functions whose step-times are normally distributed). This suggests that fitting a cumulative normal distribution to the response gradient returns the summary statistics of the underlying start time distribution. In general, the best curve for the data depends on the procedure, and whether the procedure produces start times at all.

In other words, from the response gradient, we can obtain both panels of Figure 2. Moreover, we could, either empirically or through curve fitting, obtain estimates of the peak response rate, the time of the peak response, low operant rate, the high posttransition rate (Schneider, 1969), and many others. We suggest, where possible, that discrimination ratios be replaced with their corresponding response gradients. The discrimination ratios can be given in the text (or in a table). This way, those who want the specific discrimination ratios will get them by looking at the table or text, and those who may want to calculate other values from the response gradient, or simply see the gradient itself, will be given the opportunity to do so.

Presenting these data as response gradients emphasizes the similarities among the conditions. Each conditioning procedure produces response gradients that increase as a function of time following CS onset; the magnitude of the response gradient (i.e., the response rate) is related to the interval between the CS onset and food availability for the given condition (Farmer & Schoenfeld, 1966). The shorter the CS-US interval, the faster the response rate rises. In all cases, the response rate increases until the time of food availability, when the response rate either drops off suddenly (on food cycles) or more gradually (on nonfood cycles). Both cycle types clearly indicate that regardless of the condition (the duration of the CS-US interval), rats time the interval accurately. Each have the same number of time markers, and subjects use these to estimate the time of food availability.

The procedural figure (see Figure 1) is represented as a clock to showcase the cyclical nature of the rat's experimental environment. Presumably, to the animal, the world is not divided into discrete, well-separated cycles as neatly as experimenters have designed their behavioral tasks. To an animal, the experimental world moves in real time, and is cyclical. Thus, choosing to start the x axis of Figure 3 at CS onset was (somewhat) arbitrary; it was predetermined by underlying behavioral theory that suggests that CS onset was paramount. Different theories may lead to different start points. In Figure 4, the data from food cycles are redrawn relative to the time of the last food. That is, the exact same data was used to generate Figures 3 and 4, and the difference in appearance is due solely to the starting point of the x axis. All that differs is the arbitrary (theory defined) cycle start time. To save space and help with visual comparisons, we reduced the number of panels in Figure 4 relative to Figure 3.

Method 2: The Raster Plot

The raster plot shows the data in one of its most raw forms. The x axis is time since some experimenter-defined event (here the CS

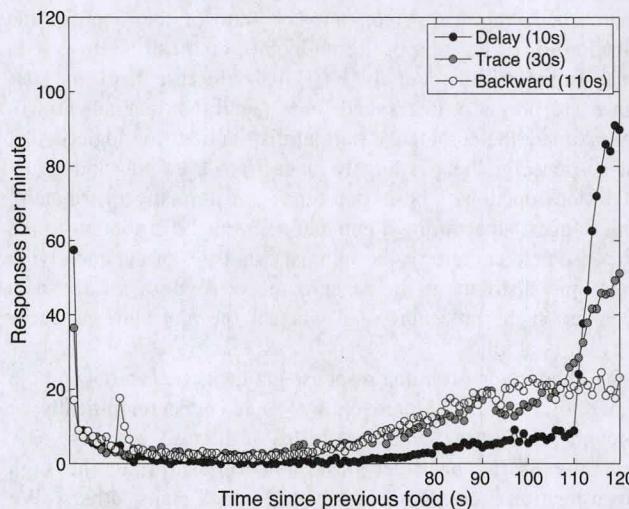


Figure 4. Response gradients when food was available 110 s after the stimulus onset in white (backward conditioning), 10 s after the stimulus onset in black (delay conditioning), and when food was available 30 s after stimulus onset in gray (trace conditioning) plotted as a function of time since food delivery.

onset), and the *y* axis is cycle number. Every individual response is represented as a single black dot. Scanning a single cycle by moving your eyes left to right gives information about the time of each response in that cycle, and visually averaging over a small temporal window gives a sense of the response rate in that window. Visually averaging the data over cycles by moving your eyes up and down the figure gives a sense of the response rate on a second-by-second basis, providing a quick, albeit crude, estimate of the response gradient.

Figure 5 shows a raster plot for a representative individual rat on nonfood cycles across different phases of the experiment (the black vertical bars show when food would be available on food cycles, the horizontal bars show the phase boundaries). A quick visual inspection of the times of responses shows the highest density of responses occur at or near the time of food availability.

The nice thing about the raster plot is that different scientists can read it to find answers to many qualitatively different questions. The conditioning researcher interested in looking at the CS-US association can simply scan up and down the gray portion of the interval (the time when the CS is present). The timing researcher can see how the responses are distributed over time. Both can mentally compute what the response gradient would look like. Further, it serves to convince others of the quality of the data at the level of a single subject. Of course, the raster plot is purely graphical, similar (but not identical) to Ferster and Skinner's use of cumulative records in *Schedules of Reinforcement* (1957). It serves as a powerful springboard from which more rigorous methods can be used.

Method 3: View Both Timing and Conditioning as a Time-Series Problem

Perhaps the most important feature of the raster plot is that it illustrates that both timing and conditioning researchers are fundamentally confronted with a time-series problem. Traditionally,

timing researchers have been interested in how responses change as a function of time *within a cycle* (i.e., the *x* axis in the raster plot). And traditionally, conditioning researchers have been interested in how responses change as a function of time *between cycles* (i.e., the *y* axis of the raster plot). The timing researcher is likely to average over cycles, while keeping the temporal order of the cycle intact (this is what the response gradient does). A conditioning researcher is likely to average over time within a cycle, while keeping the time between cycles intact (the well-known learning curve is an example of this). Fundamentally, however, both subfields are interested in time-series data, and both should be familiar with time-series analysis tools (see Anderson, 1994; Chatfield, 2003).

A change point is the point in a time-series data set at which the data changes abruptly. For data in psychology, this implies that either the psychological process that generated the data also underwent an abrupt change or the process itself is capable of producing abrupt changes in behavior. Both possibilities help constrain plausible models. In the Fixed-Interval task, for example, rats abruptly transition from a low rate of responding to a high rate of responding (Schneider, 1969; Church et al., 1994). This has led to the suggestion of a psychological timing threshold that, once passed, results in high rates of responding (Gibbon, 1977; Gibbon, Church, & Meck, 1984).

In the conditioning literature, change points occur during the learning process itself. The presolution period has been known since the 1930s (Heinemann, Chase & Mandell, 1968; Krechevsky, 1932). More recently, Gallistel, Fairhurst, and Balsam (2004) showed in certain cases that the traditional learning curve is an averaging effect of individual subjects abrupt changes in behavior early in training (suggesting that either learning occurs abruptly, or that there is a threshold that governs the emergence of conditioned responses). In the timing literature, stop times in a peak procedure appear after only a few trials (Balci et al., 2009; see Abner, Edwards, & Douglas, 2001 and Kirkpatrick-Steger,

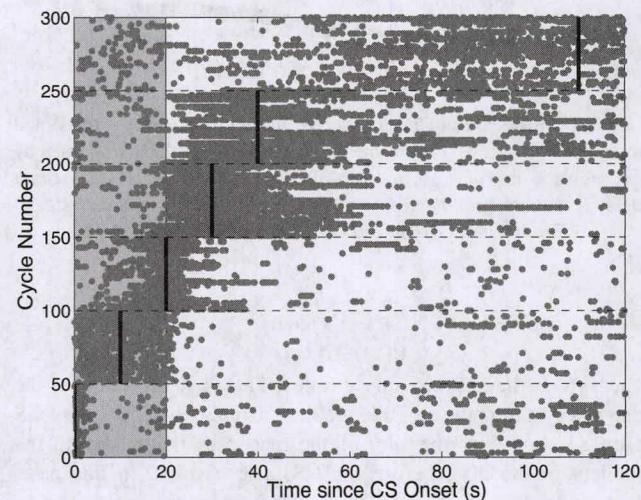


Figure 5. A raster plot indicating the times individual responses were made in seconds on individual nonfood cycles for a single subject. The black bars indicate the time of food availability, moving from the bottom to the top: 0, 10, 20, 30, 40, and 110 seconds from conditional stimulus (CS) onset (shown by gray shading), respectively.

Miller, Betti, & Wasserman, 1996, for instances when more gradual changes in behavior have been observed). Rats seem to quickly and abruptly change their choice behavior when reinforcement rates suddenly change (Gallistel, 2001). Reversal learning (Schoenbaum, Saddoris, & Stalnaker, 2007) and the learning-to-learn phenomena (Harlow, 1949) lend themselves to questions about how abruptly the animal changes its behavior after the conditions change.

There are many algorithms of estimating change points (see Steyvers & Brown, 2006), and the one discussed here was chosen for its simplicity. The method is attributed to Ploberger and Kramer (1992), but the general idea goes back further (Brown, Durbin, & Evans, 1975). The most popular method for finding transitions in Fixed-Interval data is a special case (Church et al., 1994), and so is the method used by Gallistel et al. (2004) to detect abrupt changes in conditioned responding during learning. The OLS-CUSUM (ordinary least squares cumulative sum) test statistic is computed by cumulating the residuals

$$W(t) = \frac{1}{\sigma\sqrt{T}} \sum_{i=0}^t u_i$$

for $t = 1, 2, \dots, T$ (total number of data points), and u_i is the i^{th} residual. The residuals can either be obtained using linear regression over the time series data set (the most general case), or it can simply be the residuals relative to the mean of the data set (the special cases mentioned above use this). $W(t)$ is a function of time, and will have maxima and minima. The point of these maxima and minima are the best estimates of the change points in the data. The scale factor $1/\sigma\sqrt{T}$ plays a similar role as a z score and allows us to use it as a test statistic and compare it to a suitable criterion (1.22 for $\alpha = 0.10$, 1.36 for $\alpha = 0.05$, and 1.63 for $\alpha = 0.01$; see Ploberger & Kramer, 1992).

The top panel of Figure 6 shows two CS difference score distributions from one representative rat using all cycles. The difference score is similar to the discrimination ratio, except it is defined as $A - B$, instead of $A/(A + B)$. Anything less than zero indicates the rat responded more before the CS, and anything greater than zero indicates the rat responded more during the CS. Like the discrimination ratio, the difference score can be calculated at a cycle-by-cycle level, which results in a distribution of difference scores for a single rat per condition. Shown in light gray is the backward conditioning difference score distribution. This distribution is tightly centered on zero ($M = 0.33$, $SD = 5.29$). The dark gray distribution shows delay conditioning at 20 s. The distribution is centered to the right of zero, indicating excitatory conditioning ($M = 10.1$, $SD = 9.9$). Further, the distribution is bimodal, suggesting that the rat's behavior itself may be drawn from two underlying distributions.

The middle panel of Figure 6 shows the CS difference score for every cycle of the experiment for this rat. Vertical dashed lines show the condition boundaries (the food availability time is displayed at the top of each condition), and the light gray and dark gray bars match the conditions from which the distributions in the top panel were drawn. Showing the entire experiment allows us to visualize the difference score distributions for each condition, and more importantly, how they change over time.

The bottom panel of Figure 6 shows the scaled cumulative residuals (using the mean to compute the residuals). Each maxima

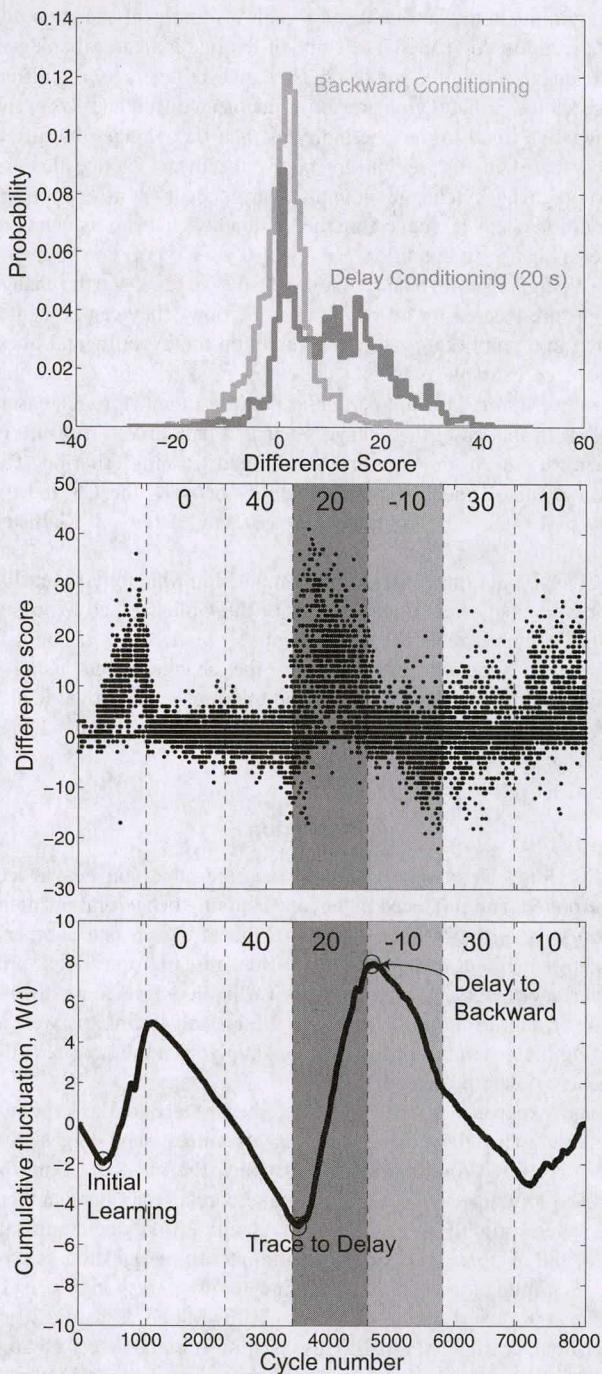


Figure 6. Top panel: The distribution of difference scores (the number of responses prior to conditional stimulus (CS) subtracted from the number of responses made during the CS) for the backward and delay (20 s) conditioning procedures. Middle panel: the difference scores as a function of cycles across the experiment for an individual subject. Dashed lines indicate a change in condition. Bottom panel: $W(t)$ (analogous to a cumulative z score) for the same subject across all conditions of the experiment. Dashed lines indicate a change in condition.

and minima in the bottom panel signals a significant change in the data from the top panel. The slope of the line indicates the degree of change. A positive trend indicates that, on average, the difference scores in that region are above the mean difference score, and a negative trend in the line indicates that, on average, the difference scores in that region are below the mean. Notice that the abrupt changes (change points) always occur shortly after the condition changes, suggesting the rat quickly changed its behavior after a change in condition.

This figure shows data from one illustrative rat. Once the change points are located for all rats in all conditions, they can be aggregated in a number of ways, depending on the experimental question. One example is to plot the cumulative fraction of rats that have transitioned by any particular point in time (More & Jensen, 2014). In this case, the change point is a data-driven measure of when rats reach the experimenter defined learning criterion. The inverse can be measure of associability between the CS and the US, given the new contingencies (Balsam, Drew, & Gallistel, 2010).

Change point methods can be computed in Matlab fairly easily. One version is freely available from the Gallistel Lab Web site (<http://cognitivegenetic.rutgers.edu/ptn/>). Matlab has a built-in *segment.m* function that computes the change points using a slightly different algorithm than the one presented above. R also has a handful of change point algorithms in the *Structchange* package (Zeileis et al., 2002).

Discussion

Providing an integrative set of measures that can be viewed, interpreted, and produced by anyone studying behavioral phenomenon will allow for each subfield to benefit from one another's findings, without changing their current research questions, procedures, or theory. At the very least it will allow for a more diverse pool of data to be used for model fitting and testing, as well as testing the generality of findings; or simply, as a tool to assess the quality of data presented.

The experimental data presented here attest to these points. When viewing these data, a timing researcher may suggest that there are three time markers from which the rat could time the food-to-food interval. A conditioning researcher may say that there are several stimuli from which the rat could form associations that drive the response rate. For both, the interesting question of how these stimuli compete and combine to drive responding is of interest (Kehoe, Ludvig, & Sutton, 2013; Matell & Kurti, 2013; McMillan & Roberts, 2010). This question is immediately obvious when looking at the food-to-food response gradient in Figure 4; although subjects are able to estimate the time to food in each procedure, the rate of responding up to the time of food availability is very different. Although our particular data set does not allow for it, other topics such as extinction or spontaneous recovery would also benefit from additional data and analyses from experiments conducted by both conditioning and timing researchers.

Of course, providing data in more general (i.e., less summarized) forms is not always appropriate for publication. One simple solution is to display the more raw form of the data in a figure, and provide the summarized measures in a table or in the text. Cycle-by-cycle measures, as well as distributions, are good candidates for figures, and the summary statistics derived from them are good

candidates for tables or text. For example, instead of plotting bar graphs, show the response distributions and report descriptive statistics in the text or a table. Another simple solution is to make full data sets available online via lab, university, or personal Web sites. With massive storage space becoming more readily available, many labs have already committed to making their data sets freely available. Our repository, for example, can be found at <http://www.brown.edu/Research/Timelab/>.

By the mid 1970s, it became common to control experimental protocols with computers. In that era, programming languages were unintuitive, hard drives were small (or nonexistent), and researchers had to be mindful of the amount of data they were storing. With a premium on space, and limited computing capabilities, scientists collected the data necessary to answer their questions, but little else. But now things have changed. Storage space is no longer an issue, and statistical software packages are common (Matlab, R, and Python are heavily used in most branches of science). Med Associates software (MedPC) allows for the storing of complex time-series data sets in time-event format. Most events in the experimental chamber can now be recorded with millisecond precision and cameras can record the rest (Jhuang et al., 2010). Modern software packages handle these large data sets with ease and, in some cases, can even be used to control the experimental chambers.

We started our efforts by appealing to the methods of Pavlov and Skinner. They found innovative ways to record the times of the critical events and extracted summary measures from them. Both conditioning and timing subfields have benefitted tremendously from their work. The pressure to focus on and record only a few measures (however important) has alienated the subfields from each other. The outcome has been the development of separate subfields that use similar procedures to answer similar questions but with very different theoretical frameworks and analysis tools. The existence of hybrid models suggests this may be changing; our goal is to aid in this change.

References

- Abner, R. T., Edwards, T., & Douglas, A. (2001). Pharmacology of temporal cognition in two mouse strains. *International Journal of Comparative Psychology*, 14, 189–210.
- Anderson, T. W. (1994). *The statistical analysis of time series*. New York, NY: Wiley-Interscience.
- Balci, F., Gallistel, C. R., Allen, B. D., Frank, K. M., Gibson, J. M., & Brunner, D. (2009). Acquisition of peak responding: What is learned? *Behavioural Processes*, 80, 67–75.
- Balsam, P. D., Drew, M. R., & Gallistel, C. R. (2010). Time and Associative Learning. *Comparative Cognition and Behavior Reviews*, 5, 1–22.
- Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the consistency of regression relationships over time. *Journal of the Royal Statistical Society, Series B (Methodological)*, 37, 149–192.
- Chatfield, C. (2003). *The analysis of time series: An introduction*. New York, NY: Chapman & Hall.
- Church, R. M., Meck, W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *Journal of Experimental Psychology: Animal Behavior Processes*, 20, 35–155.
- Domjan, M. (2004). *The essentials of learning and conditioning* (3rd ed.). Stamford, CT: Cengage Learning.
- Domjan, M. (2014). *The principles of learning and behavior* (7th ed.). Stamford, CT: Cengage Learning.

- Farmer, J., & Schoenfeld, W. N. (1966). Varying temporal placement of an added stimulus in a fixed-interval schedule. *Journal of Experimental Analysis of Behavior*, 9, 369–375.
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Englewood Cliffs, NJ: Prentice Hall.
- Freestone, D., MacInnis, M. L., & Church, R. M. (2013). Timing and contingency. *Timing and Time Perception*, 1, 3–20.
- Gallistel, C. R. (2001). Mental representations, psychology of. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopaedia of the social and behavioural sciences* (9691–9695). New York, NY: Elsevier.
- Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Science*, 7, 13124–13131.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325.
- Gibbon, J., Church, R. M., & Meck, W. R. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52–77.
- Guilhardi, P., Yi, L., & Church, R., (2007). A modular theory of learning and performance. *Psychonomic Bulletin and Review*, 14, 543–559.
- Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during a CS-US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 31–42.
- Han, C. J., O'Tuathagh, C. M., van Trigt, L., Quinn, J. J., Fanselow, M. S., Mongeau, R., . . . Anderson, D. J. (2003). Trace but not delay conditioning requires attention in the anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 13087–13092.
- Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, 56, 51–65.
- Heinemann, E. G., Chase, S., & Mandell, C. (1968). Discriminative control of 'attention'. *Science*, 160, 553–554.
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Oxford, England: Appleton-Century.
- Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., . . . Serre, T. (2010). Automated home-cage behavioral phenotyping of mice. *Nature Communications*, 1, 68.
- Kehoe, E. J., Ludvig, E. A., & Sutton, R. S. (2013). Timing and cue competition in conditioning of the nictitating membrane response of the rabbit (*Oryctolagus cuniculus*). *Learning and Memory*, 20, 97–102.
- Kirkpatrick, K. (in press). Interactions of timing and prediction error. *Behavioral Processes*.
- Kirkpatrick-Steger, K., Miller, S. S., Betti, C. A., & Wasserman, E. A. (1996). Cyclic responding by pigeons in the peak timing procedure. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 447–460.
- Krechevsky, I. (1932). Hypotheses in rats. *Psychological Review*, 39, 516–532.
- Machado, A., Malheiro, M. T., & Erlhagen, W. (2009). *Learning to time: A perspective Journal of the Experimental Analysis of Behavior*, 92, 423–458.
- Matell, M. S., & Kurti, A. N. (2013). Reinforcement probability modulates temporal memory selection and integration processes. *Acta Psychologica*, 147, 80–91.
- Matzel, L. D., Held, F. P., & Miller, R. R. (1988). Information and expression of simultaneous and backward associations: Implications for contiguity theory. *Learning and Motivation*, 19, 317–344.
- McMillan, N., & Roberts, W. A. (2010). The effects of cue competition on timing in pigeons. *Behavioural Processes*, 84, 581–590.
- More, L., & Jensen, G. (2014). Acquisition of conditioned responding in a multiple schedule depends on the reinforcement's temporal contingency with each stimulus. *Learning and Memory*, 21, 258–262.
- Moscovitch, A., & Lolordo, V. M. (1968). The role of safety in the Pavlovian backward conditioning procedure. *Journal of Comparative and Physiological Psychology*, 66, 673–678.
- Nevin, J. A., Mandell, C., & Yarensky, P. (1981). Response rate and resistance to change in chained schedules. *Journal of Experimental Psychology: Animal Behavior Processes*, 7, 278–294.
- Pavlov, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). London, UK: Oxford University Press.
- Ploberger, W., & Kramer, W. (1992). The CUSUM test with OLS residuals. *Econometrica*, 60, 271–285.
- Plotkin, H. C., & Oakley, D. A. (1975). Backward conditioning in the rabbit (*Oryctolagus cuniculus*). *Journal of Comparative and Psychological Psychology*, 88, 586–590.
- Rescorla, R. A. (1992). Depression of an instrumental response by a single devaluation of its outcome. *The Quarterly Journal of Experimental Psychology*, 44, 123–136.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York, NY: Appleton-Century-Crofts.
- Savastano, H. I., & Miller, R. R. (1998). Time as content in Pavlovian conditioning. *Behavioural Processes*, 44, 147–162.
- Schneider, B. A. (1969). A two-state analysis of fixed-interval responding in the pigeon. *Journal of the Experimental Analysis of Behavior*, 12, 677–688.
- Schoenbaum, G., Saddoris, M. P., & Stalnaker, T. A. (2007). Reconciling the roles of the orbitofrontal cortex in reversal learning and the encoding of outcome expectancies. *Annals of the New York Academy of Sciences*, 1121, 320–335.
- Skinner, B. F. (1938). *The behavior of organisms*. Cambridge, MA: Copley Publishing Group.
- Steyvers, M., & Brown, S. (2006). Prediction and change detection. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 1281–1288). Cambridge, MA: MIT Press.
- Tatham, T. A., & Zurn, K. R. (1989). The Med-PC experimental apparatus programming system. *Behavioral Research Methods, Instruments, and Computers*, 21, 294–302.
- Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). Strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7, 38.

Received November 24, 2013

Revision received June 6, 2014

Accepted June 6, 2014 ■

Copyright of Journal of Experimental Psychology: Animal Learning & Cognition is the property of American Psychological Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.