Title
Control of variation by reward probability

Journal
Journal of Experimental Psychology-Animal Behavior Processes, 30(4)

ISSN
0097-7403

Authors
Gharib, A
Gade, C
Roberts, Seth

Publication Date
2004-10-01

Peer reviewed

Journal of Experimental Psychology: Animal Behavior Processes, in press

Control of Variation by Reward Probability

Afshin Gharib, Christopher Gade, and Seth Roberts

Dominican University of California, San Rafael

and

University of California, Berkeley

Running head: Effect of reward probability on variation

Control of Variation by Reward Probability
29 March 2004 version

Abstract

Two bar-press experiments with rats tested the rule that reducing expectation of reward increases the variation from which reward selects. Experiment 1 used a discrete-trial random-interval schedule, with trials signalled by light or sound. One signal always ended with reward; the other ended with reward less often. The two signals were randomly mixed. Bar-press duration (how long the bar is held down) varied more during the signal with the lower probability of reward. Experiment 2 closely resembled Experiment 1 but used a random-ratio schedule rather than a random-interval schedule. Again, bar-press duration varied more during the signal with the lower probability of reward. The results support the rule–the first well-controlled comparisons to do so.

Control of Variation by Reward Probability

Instrumental learning requires both variation and selection -- variation of action, selection by reward of what works (Staddon & Simmelhag, 1971; Hull, Langman, & Glenn, 2001) -- but few experiments have studied variation (Chance, 1999; Domjan, 1998; Lieberman, 1990; Pearce, 1997; Staddon & Cerutti, 2003). Yet it is plausible that variation depends on recent events, just as response rate does. If an animal's actions vary too little, it will not find better ways of doing things; if they vary too much, rewarded actions will not be repeated. So at any time there is an optimal amount of variation, which changes as the costs and benefits of variation change. Animals that do instrumental learning would profit from a mechanism that regulates variation so that the actual amount is close to the optimal amount. If such a mechanism exists, we know little about it.

Experiments about variation have been rare partly because variation has been hard to measure. Antonitis (1951), for example, took "6,600 photographs of nose-thrusting responses" (p. 275) to study variation in nose position. Herrnstein (1961) used a special apparatus with ten switches to measure the location of key pecks. Although most studies of variation have measured spatial variation (e.g., Balsam, Deich, Ohyama, & Stokes, 1998; Eckerman & Lanson, 1969; Ferraro & Branch, 1968; Neuringer, 2002), temporal variation may be much easier to measure. A few studies (Lacter & Corey, 1982; Margulies, 1961; Millenson, Hurwitz, & Nixon, 1961) have measured the variation of bar -press duration (how long the bar is held down). Rats can be trained to make bar presses of specified durations (e.g., Notterman & Mintz, 1965;

Platt, Kuch, & Bitgood, 1973), so the variation measured by bar-press duration includes the variation from which reward selects.

Our use of bar-press duration to study variation began with a puzzle involving rats trained with the peak procedure (Gharib, Derby, & Roberts, 2001). On most trials, food was given for the first bar press more than 40 sec after the start of the trial, at which point the trial ended. On some trials, however, no food was given for the entire trial. The puzzle was that on trials without food bar-press duration increased sharply in the middle of the trial, starting with the second bar press after Second 40. The upper panel of Figure 1 shows an example. We originally thought the increase was due to frustration (Amsel, 1992) but a test of this idea indicated otherwise (Gharib, Derby, & Roberts, 2001).

Data analysis suggested a different explanation. Gharib et al. recorded several million bar-press durations, which made it easy to see how their distribution changed from early to late in the trial (lower panel of Figure 1). The mean increase seen in the upper panel of Figure 1 was not due to an upward shift of the whole distribution, as we had expected. Instead, it reflected a widening of the distribution–that is, an increase in variation. Why would the variation increase? One clue was that the increase began exactly when the rats could realize that further bar presses would not produce reward. Another was that variation remained high for the rest of the trial, during which no food was given. This led Gharib et al. to propose the following rule: reducing expectation of reward increases variation of form.

This was plausible. When the probability of reward goes down, so does the cost

of variation, because less is lost when an action falls outside the boundary of what is rewarded. When the cost of variation goes down, the optimal amount of variation goes up. So a mechanism that regulated variation *should* increase variation when expectation of reward decreased.

The rule correctly predicted other results. It correctly predicted two effects in another experiment by Gharib et al.: (a) the direction of the signal/intertrial-interval difference in duration (reward density was greater during the signal than during the intertrial interval, thus reward expectation was greater; and duration varied less during the signal), and (b) the direction of the effect of a reward during the intertrial interval on the duration of responses 30-40 sec later (a reward increased expectation of reward; and a reward decreased variation). The rule correctly predicted the well-established increase of variation during extinction (Balsam, Deich, Ohyama, & Stokes, 1998; Neuringer, 2002). In several studies measuring peck location, reductions in reward probability short of extinction increased variability (e.g., Eckerman & Lanson, 1969; Ferraro & Branch, 1968; Millenson, Hurwitz, & Nixon, 1961; see Ferraro & Branch, 1968, for other examples). Using fixed-ratio schedules, Cherot, Jones, and Neuringer (1996) found that the variation of bar-press sequences decreased as reward came closer and reward expectation presumably increased. Several other fixed-ratio experiments have observed similar results (e.g., Schwartz, 1982; see Cherot, Jones, & Neuringer, 1966, for a review). The only inconsistent result, as far as we know, is Herrnstein's (1961) finding that variability of peck location was greater with continuous reinforcement (CRF) than with a variable-interval 3-minute schedule (VI-3 min).

However, Ferraro and Branch (1968) and Eckerman and Lanson (1969) found the opposite result. Moreover, the two conditions that Herrnstein compared differed in several ways besides reward expectation.

None of the Gharib et al. results nor any other previous result is conclusive evidence for (or against) the rule because none were designed to test it. All involve confoundings–comparisons between conditions that differed in more than reward expectation–so results that support the rule can be explained in other ways. For example, the increase in variation during extinction supports the rule because reward expectation is less during extinction than training. But there are other differences between training and extinction.

The two experiments reported here were designed to test the rule. They tried to vary reward expectation while keeping other things constant. Because we assumed that reward expectation is generated by an associative process similar to Pavlovian conditioning, we especially tried to minimize non-associative differences between conditions being compared. The present experiments are the first experiments about variation to do so.

Experiments 1 and 2 were very similar. Both used a discrete-trials procedure in which two signals were associated with different probabilities of reward. Trials with the two signals were randomly mixed. To reduce the influence of non-associative effects of reward and nonreward, all rewards happened at the end of a trial (after the bar was released), and there was 1 min between trials. Thus expectation differed between signals while other properties, such as overall density of reward, time since last reward,

and time since last nonreward, were equal.

The rule that reducing reward expectation increases variation predicted more variation during the signal with the lower probability of reward. It predicted even more variation during the intertrial interval than during the signal with the lower probability of reward, but only if non-associative differences between the two (e.g., unlearned effects of light and sound, time since last reward, time since last non-reward) were unimportant.

Because variation can be measured many ways, these predictions can be tested many ways. The standard deviation is the most efficient measure of spread when the underlying distribution is Gaussian (normal) but the underlying distribution was far from Gaussian (lower panel of Figure 1). Use of the standard deviation to measure the spread of such distributions would be quite wasteful. To find a better choice, Gharib et al. asked what summary was most sensitive to the early/late difference shown in the lower panel of Figure 1. They considered 10 candidates -- two statistics (mean and standard deviation), each computed after one of five transformations (none, square root, logarithmic, reciprocal square root, reciprocal) – and found that the mean of the logarithms produced the largest t value ($t(35) = 18$) for the early/late difference. It may be surprising that computing a mean was the most sensitive way to detect a change in spread, but that is what the data showed. The mean of logarithms is the summary we used here, so predictions about variation became predictions about means. (See Figure 11 for an estimate of how much power we would have lost had we used standard deviations.)

Experiment 1: Random-Interval Schedule

In this experiment, the end of a trial was "primed" (set to happen after the next bar press) with a certain probability each second during a trial. This probability was the same for both signals. The procedural difference between signals was what happened at the end of a trial. During an initial training phase (*100%/100%*), trials with both signals always ended with reward. After that, trials with one signal (*high food*) always ended with reward, while trials with the other signal (*low food*) ended with reward on 50% of trials (during the *100%/50%* phase) or 25% of trials (during the *100%/25%* phase).

Method

Subjects. The subjects were 15 Fischer 344 rats from Harlan, 16 months old at the start of bar-press training. They had served in other behavioral experiments, none of which involved bar pressing. They were housed individually with ad-lib water and given Purina Lab Meal 5001-M after each daily session. One rat died during the experiment. Its results are not included below.

Apparatus. The rats worked in 15 similar lever boxes. Two were 23 cm x 20 cm x 28 cm; five were 23 cm x 20 cm x 21 cm; and eight were 28 cm x 26 cm x 28 cm (Gerbrands Series 7400 operant test chambers). The roof and side walls of the box were transparent acrylic; the front and back walls were aluminum. A pellet dispenser delivered 45-mg pellets (BioServ mix T101) to a food cup that the rat accessed through a hole in the center of the front wall. Water was available ad-lib from a water bottle nozzle accessible on the rear wall. Seven boxes contained a single lever on the front wall that remained extended; it was to the left (facing the front wall) of the food cavity.

The other eight boxes contained two retractable levers (Gerbrands G6311 retractable rat lever), one on each side of the food cup. During the experiment, the stage-left lever remained extended into the box; the other lever was always retracted. To close the lever switch required a force of about 15 g. A small lamp (General Electric 1155X) was mounted on the roof of the operant chamber stage left of the food cavity in each box; it provided the light stimulus through the acrylic roof. The sound stimulus was white noise from 5-cm speakers mounted on the floor behind the food cavity. A bar press was recorded when the lever switch was closed (lever down) and then opened (lever up). Rewards were given contingent upon lever release–that is, after the switch re-opened. Each box was enclosed in a insulated fan-ventilated chamber. An IBM-PC 486 computer controlled the experimental events and collected the data. Individual durations were measured to an accuracy of about 20 msec.

Procedure. Subjects were given 10 g of chow per day for one week before training began. During the first 16 days of the rest of the experiment, the daily ration was 12 g of food immediately after the daily session. After the first 16 days, the daily ration was reduced to 10 g of food.

During magazine and bar-press training, the first 50 bar presses were all rewarded (with one pellet); each of the next 50 bar presses was rewarded with probability 0.5; and later bar presses were rewarded with probability 0.25. Reward was always given after the bar had been both pressed down and released. In addition, food was given at random intervals an average of once per minute independent of responses (variable time 1-min schedule) until the rat had made 50 bar presses. This continued

until all the subjects had made at least 100 bar presses in one 4-hr session, which took no more than 20 days.

Then the main procedure began. Sessions began with an intertrial interval. All intertrial intervals lasted 60 sec. We used the shortest intertrial interval that we believed would allow aftereffects of reward or nonreward to disappear. The duration was constant rather than variable so that every trial would be followed by a sufficient buffer. After the first intertrial interval, one of two signals (house light or white noise) was turned on. The signal was randomly chosen, with the two signals equally likely. During a trial, the end was "primed" with a probability of 1/120 each sec during the first seven days. Because bar press rates were low, the probability was increased to 1/60 each sec for the remaining days. Once the end of the trial was primed, the next bar press ended the trial and a new intertrial interval began. During the first phase (100%/100%), which lasted 22 days, both signals always ended with the delivery of a food pellet. During the next phase (100%/50%), which lasted 23 days, one signal always ended with food; the other ended with food with a probability of 0.5. The signal that always ended with food was light for even-numbered rats, and sound for the rest. The final phase (100%/25%), which lasted 38 days, was the same as the 100%/50% phase except that the signal that had ended with food with probability 0.5 now ended with food with probability 0.25. The daily session lasted 3 hr.

Data analysis. To summarize durations, we used the mean of log durations, for reasons explained in the introduction. Rates were computed by, first, finding for each rat the mean log interresponse time. This way of measuring rate gives much less weight

to very long interresponse times than the usual way of measuring rate (number of responses divided by total amount of time). We computed rate both ways and got similar results; we report the log interresponse time values. This way of measuring rate is probably more sensitive than the usual way because log interresponse times have a distribution much closer to Gaussian (normal) than do untransformed interresponse times. (The usual way of computing rate is equivalent to taking the mean of the untransformed interresponse times and then taking the reciprocal of the result.) Second, we averaged *across* rats using trimmed means (10% trimmed from each side) or medians. Use of trimmed means across rats means that the rats with the lowest and highest values were excluded from the computation of the mean. We used trimmed means or medians because such averages, unlike the mean, ensure that no one rat can have a large influence on the average across rats. (Computation of an average duration for each rat did not involve any trimming.) Wilcox and Keselman (2003) discuss the value of trimmed means at length; they recommend use of a 20% trimmed mean. All p values from t tests are one-tailed unless stated otherwise.

Results

Figure 2 shows bar-press rate (upper panel) and duration (lower panel) over the course of the experiment. Rate changed as expected. During the 100%/50% and 100%/25% phases, bar-press rates were greater during the high-food signal than during the low-food signal and greater during the low-food signal than the intertrial interval. Over the last 15 days of the 100%/50% phase, high-food versus low-food, $t(13) = 4.32, p < 0.001$; low-food versus intertrial interval, $t(13) = 11.12, p < 0.001$. Over the last 20

days of the 100%/25% phase, high-food versus low-food, $t(13) = 10.18, p < 0.001$; low-food versus intertrial interval, $t(13) = 7.89, p < 0.001$. In what follows, results from the 100%/50% phase are from the last 15 days; results from the 100%/25% phase are from the last 20 days.

Bar-press duration changed in the predicted ways (lower panel of Figure 2). Mean durations were greater (indicating greater variation) during the low-food signal than during the high-food signal and greater during the intertrial interval than during the low-food signal. During the 100%/50% phase, high-food versus low-food, $t(13) = 0.96, p = 0.18$, not reliable (a later analysis explains why); low-food versus intertrial interval, $t(13) = 8.58, p < 0.001$. During the 100%/25% phase, high-food versus low-food, $t(13) = 2.96, p = 0.005$; low-food versus intertrial interval, $t(13) = 4.42, p < 0.001$.

To confirm that the differences in mean duration reflected differences in variation, Figure 3 shows the distributions. The probability density functions (upper panel) and the inverse cumulative distribution functions (lower panel) were close to linear on a log-log scale. The measure shown in the upper panel, probability density, is not the same as probability, which is why some values are more than one. Whenever one sees a smooth theoretical distribution function, such as a bell-shaped curve (a normal distribution), one is looking at a graph of probability density, not probability. Probability density equals probability (or frequency) divided by bin width. In a histogram, the heights of the bars represent probability densities (even though the *y* axis says probability or frequency); it is the *areas* of the bars that represent probabilities

or frequencies. When the bins have unequal widths, as in the upper panel of Figure 3, the fiction that the *y* axis is probability or frequency can no longer be maintained if one wants the histogram to look like the underlying (probability density) distribution. Inverse cumulative distributions (the lower panel of Figure 3) are the form in which distributions of this type (linear on a log-log scale) are usually shown (e.g., Buchanan, 2001). Both panels show that the mean differences seen in the lower panel of Figure 2 reflect changes in distribution width, that is, changes in variability. At long durations, the differences were large. Bar presses longer than 8 sec were about 50 times more likely during intertrial intervals than during the high-food signal.

To our surprise, time between bar presses had a big effect on duration (top panel of Figure 4). As interresponse time increased within a signal, duration decreased (in contrast to the between-signal differences shown in Figure 2, where longer interresponse times – that is, lower rates – were associated with greater durations). To assess the reliability of this effect, we fit parallel lines to the two functions (high-food and low-food) separately for each rat. The slopes were reliably negative, $t(13) = 10.03$, $p < 0.001$ (100%/50%) and $t(13) = 9.63$, $p < 0.001$ (100%/25%). The low-food line was above the high-food line, $t(13) = 4.97$, $p < 0.001$ (100%/50%), and $t(13) = 5.45$, $p < 0.001$ (100%/25%), meaning that the signal differences persisted when interresponse times were equated. In the case of the 100%/50% phase, the signal difference became much clearer ($t(13) = 4.97$ versus $t(13) = 0.96$) when interresponse times were equated.

Why was interresponse time important? During a trial, food was primed (made

available for the next bar press) with a fixed probability each second. The longer the wait between bar presses, the more likely that the bar press at the end of the wait would produce food. The middle panel of Figure 4 shows how the probability of reward increased with interresponse time. To see if this could explain the effects of interresponse time, the bottom panel of Figure 4 shows duration versus probability of reward for different interresponse times–the y variable of the top panel versus the y variable of the middle panel. The results from six conditions–three (high-food signal, low-food signal, and intertrial interval) from the 100%/50% phase, and the same three from the 100%/25% phase–overlap closely. To judge the reliability of the correlation, we made a graph like the bottom panel of Figure 4 for each rat and fit straight lines (one per rat). The 14 slopes were negative, $t(13) = 11.14, p < 0.001$. Duration differences between the six conditions are almost entirely explained by differences in the probability of reward.

To ensure that the mean differences shown in Figure 4 reflected differences in variation, we computed 10% and 90% quantiles of the distributions. Figure 5 shows these values as a function of probability of reward. The spreads of the distributions are given by the vertical distance between the functions. As probability of reward increased, the spread decreased. To judge the reliability of the decrease, we fit straight lines to the upper and lower functions separately for each rat. The upper line was steeper than the lower line, $t(13) = 4.64, p < 0.001$. With raw (untransformed) durations, this would be uninteresting because most distributions spread out as they move farther from a minimum-possible value (such as zero). What is notable is that the interaction occurs

with log duration.

Discussion

The results support the initial idea–that reducing reward expectation increases variation. The prediction that a reduction in the probability of reward will increase variation was verified in two well-controlled instances: high-food signal versus low-food signal during the 100%/50% phase when interresponse times were equated (top panel of Figure 4); and high-food signal versus low-food signal during the 100%/25% phase both with (top panel of Figure 4) and without (lower panel of Figure 2) correction for interresponse time. This is the best evidence so far for the idea. The prediction was also verified in two less-well-controlled instances: low-food signal versus intertrial interval during the 100%/50% phase (lower panel of Figure 3); and the same comparison during the 100%/25% phase (lower panel of Figure 3). An analysis described in the Results section of Experiment 2 suggests that the low-food/intertrial comparison is reasonable because non-associative differences between them had little effect on duration.

The results also provide new insight into the behavior generated by variable-interval schedules. The extreme form of a variable-interval schedule is a random-interval schedule (e.g., this experiment) where reward is primed with a fixed probability each time interval (e.g., each second). The usual idea is that random-interval schedules and fixed-interval schedules are opposites: With a fixed-interval schedule, the rat or pigeon times the signal; with a random-interval schedule, it does not. Libby and Church (1975) emphasized this difference in an experiment measuring the time course of fear. When a shock was given a fixed time after the last shock, fear increased over time; when a shock

was given with a fixed probability each second after the last shock, fear was roughly constant with time. Response rates are indeed quite constant over time with a random-interval schedule (e.g., Gharib et al., 2001) but what has not been realized–at least, what this experiment suggests–is that a variable-interval schedule causes subjects to measure the time since their last response. Differential-reinforcement -of-low-rate procedures have shown that animals can do this (e.g., Doughty & Richards, 2002). The subjects of this experiment did so presumably because the time since the last bar press predicted the probability that the next response would be rewarded (middle panel of Figure 4). If the rats were not measuring this time, it would be very hard to explain why the corresponding probability of reward (middle panel of Figure 4) predicted duration so well (bottom panel of Figure 4, Figure 5). That rats are keeping track of time since the last bar press could not have been determined from rate measurements alone.

The confound of reward probability and interresponse time explains why the difference between signals did not reliably affect duration during the 100%/50% phase until interresponse times were equated. Reducing the probability of reward per trial reduced rate, but a lower rate increased the probability of reward per bar press. For example, if a 50% reduction in reward density (rewards per minute) reduces rate by 50%, the probability of reward per bar press remains the same. Because duration apparently depends on the probability of reward per bar press (bottom panel of Figure 4), the lower rate during the low-food signal tended to reduce the duration difference between the signals. Tremont (1984) found no clear increase in the variation of the force of bar-presses as the density of reward decreased on random interval schedules.

Bar-press rate, however, decreased greatly as the density of reward decreased, so the probability of reward per bar-press stayed relatively constant. This may also be why Eckerman and Lanson (1969) failed to find clear decreases in the variability of key-peck location as the interval value of a random-interval schedule decreased.

The experiment seems to have mapped an important part of the duration-versus-expectation function an unexpected bonus. The scatterplots in the bottom panel of Figure 4 and Figure 5 suggest that duration is sensitive to changes in expectation only when the probability of reward/bar press is less than 0.25 or 0.3. Of course, the scatterplots show correlations. Experiment 2 asks whether these relationships reflect causation.

The shape of the duration distributions (Figure 3) is discussed in the General Discussion.

Experiment 2: Random-Ratio Schedule

The bottom panel of Figure 4 suggested that the duration of a bar press is controlled by the probability that the bar press will be rewarded. This experiment tested this conclusion by varying the probability of reward per bar press. Like Experiment 1, it used a discrete-trials procedure with two signals that varied in the probability of reward. During a trial, each bar press ended the trial with a probability that was the same for both signals. The difference between signals was that the low-food signal ended with reward less often than the high-food signal. Thus the probability of reward per bar press was constant during a signal – unlike Experiment 1, it did not vary with the time since the last bar press -- and different between signals.

Removal of the confounding between reward probability and interresponse time allowed better measurement of the effects of both factors. Because the results from several conditions overlapped neatly when equated for reward probability (bottom panel of Figure 4), reward probability surely mattered. But interresponse time may have had some effect. Perhaps a bar press requires significant exertion. The longer the interresponse time, the more rested the rat; the more rested the rat, the shorter and less variable the bar press. Some studies have found interresponse time to have the opposite effect on variation: The longer the interresponse time, the more variable the response (Baddeley, 1996; Neuringer, 1991; Roberts & Neuringer, 1998).

Method

Subjects and apparatus. The subjects were the 14 rats that finished Experiment 1. They were fed 12 g of food after the daily session until the 27th day of the experiment, when the amount was reduced to 8-10 g per day. Their housing was the same as during Experiment 1. Three died during the experiment; their results are not included below. The apparatus was the same as the apparatus of Experiment 1.

Procedure. The procedure was the same as the procedure of the main part of Experiment 1, with one exception: The end of the trial was no longer primed with a certain probability each second. Instead, each bar press during a signal ended the trial (turned off the signal) and began the intertrial interval with a probability of 0.25. During the first phase of the experiment (100%/100%), which lasted 20 days, all signals ended with food–that is, the trial-ending bar press was always rewarded. During the second phase of the experiment (100%/25%), which lasted 31 days, one of the signals

always ended with food; the other signal ended with food with a probability of 0.25. For each rat, the signal (light or sound) with the lower probability of food in this experiment was the signal with the higher probability of food in Experiment 1.

Results

Figure 6 shows the day-by-day results. Behavior was roughly constant during the last 25 days of the experiment so those are the days used in all analyses unless stated otherwise. The high-food signal produced higher rates than the low-food signal, $t(10) = 6.07, p < 0.001$, and the low-food signal produced higher rates than the intertrial interval, $t(10) = 9.27, p < 0.001$. The duration differences were also very clear. Mean log duration was greater during the low-food signal than during the high-food signal, $t(10) = 6.22, p < 0.001$, and greater during the intertrial interval than during the low-food signal, $t(10) = 4.59, p < 0.001$. The duration distributions, shown in Figure 7, confirm that the mean differences reflect differences in variation.

In contrast to Experiment 1, interresponse time had no detectable effect on duration (upper panel of Figure 8). A signal-by-interresponse-time-by-rat analysis of variance (ANOVA) found no effect of interresponse time, $F(7,70) = 1.25, p = 0.29$, no signal-by-interresponse-time interaction, $F(7, 70) = 1.63, p = 0.14$, and a strong effect of signal, $F(1,10) = 34.86, p < 0.001$. The lower panel of Figure 8 shows probability of reward as a function of signal and interresponse time. As planned, reward probability did not depend on interresponse time. There was no effect of interresponse time, $F(7, 70) = 0.75, p = 0.63$, and no signal-by-interresponse-time interaction, $F(7, 70) = 0.38, p = 0.91$.

Because the results suggest that both rate and duration are controlled by an associative process, it is worth asking if the same associative process controls both. An indication that the answer is no comes from Figure 6. At the start of the experiment, when the two signals were treated the same, their rates became equal more quickly (upper panel) than their durations became equal (lower panel). During the second block of days (Days 6-10), there was no reliable difference in rate between the two signals, $t(10) = 0.29, p = 0.8$, but there remained (from Experiment 1) a difference in duration, $t(10) = 1.92, p = 0.04$.

Another way to measure the speed of associative change is to measure the effects of repeated nonreward. When one or more bar presses in a row are not rewarded, what happens? To find out, we calculated the number of consecutive unrewarded bar presses during the low-food signal at the beginning of every trial. This number "labels" the trial, whether low-food or high-food, and all bar presses during that trial are classified using that label. When the most recent low-food trial has ended with reward, this number is zero. It does not change during a trial and is unaffected by high-food trials. Suppose that this number is zero at the start of a session. The first trial (low-food) ends without reward after five bar presses. Now the number is five. The second trial (high-food) ends with reward after three bar presses. The number remains five. The third trial (low-food) ends without reward after ten bar presses. Now the number is 15. The fourth trial (low-food) ends with reward. Now the number is zero again. In general, the number measures how much "extinction" the low-food signal has recently experienced. We calculated this number using carryover from one session to the next, so that it did not

change systematically during a session (i.e., so that it had the same average value throughout a session).

Figure 9 shows the outcome of this analysis. Nonreward during the low-food signal reduced rate (upper panel) during both signals, but had a much bigger effect during the low-food signal. To judge reliability, we fit straight lines to the data for each rat. The slope of the low-food line was negative, $t(10) = 7.50, p < 0.001$. The slope of the high-food line was also negative, $t(10) = 3.78,\ p = 0.002$, but the low-food slope was more negative than the high-food slope, $t(10) = 6.67, p < 0.001$. The number of prior unrewarded bar presses during the low-food signal affected duration (lower panel) only during the low-food signal. The slope of the low-food line was positive, $t(10) = 2.88,\ p < 0.01$. The slope of the high-food line was not reliably different from zero, $t(10) = 0.21$, two-tailed $p = 0.8$. The low-food slope was more than the high-food slope, $t(10) = 2.96,\ p < 0.01$.

Like Figure 6, Figure 9 suggests that events that reduce associative strength (specifically, the association of the low-food signal with reward) change rate faster than duration. In each panel, compare the slope of the low-food function to the vertical distance between the high-food and low-food functions: The rate function is much steeper than the duration function. To confirm this statistically, for each measure (rate and duration), we divided the low-food slope by the overall difference between the two signals (high food and low food). The larger this fraction, the faster the measure responds to changes in probability of reward. We did the computation separately for each rat. The rate fraction was larger than the duration fraction for 9 of the 11 rats, sign

test, $p = 0.03$.

Low-food trials and intertrial intervals differed in more than expectation of reward. The difference most likely to matter is time since reward. Intertrial intervals sometimes began with food, while low-food trials never did. To assess the effect of this difference, we combined data from all days when some intertrial intervals began with food and some did not–namely, the 100%/50% and 100%/25% phases of Experiment 1 and the 100%/25% phase of Experiment 2. The intervals were divided into three groups: those that followed a high-food trial (which always ended with food); those that followed a low-food trial that ended with food; and those that followed a low-food trial that did not end with food. Figure 10 shows rate (upper panel) and duration (lower panel) for each group. The upper panel shows that food depressed rate for about a minute. The lower panel suggests that food increased duration for a shorter time. Combining results from the two types of food trials, and using data only from Seconds 11-30 during the intertrial interval, duration was reliably more after food than after no food, $t(10) = 2.62$, two-tailed $p = 0.03$. Was this effect large enough to contaminate the low-food/intertrial comparison? When all intertrial bar presses are considered, not just the small fraction between Seconds 11-30, duration was slightly *less* after food than no food but the difference was not reliable, $t(10) = 0.49$, $p = 0.6$. The duration difference between intertrial intervals that followed food and those that followed no food was tiny, 2% ± 5%, too small to explain the low-food/intertrial difference, which was 31% ± 7% (10% trimmed mean ± standard error).

To improve future analyses of similar data, we asked what summary and data

transformation were most sensitive to the effect of probability of reward. Figure 11 shows $t(10)$ values for the duration difference between the two signals as a function of summary (mean and standard deviation) and power transformation (the transformation used before the durations of each rat are summarized–e.g., averaged). The mean was far more sensitive than the standard deviation. With the mean, the most sensitive power was about -0.5 (reciprocal square root). According to a bootstrap analysis with 1000 resamples, the 95% confidence interval for the best power is -0.86 to 0.50.

Discussion

Reducing reward probability increased duration (Figures 6 and 8) even when–unlike Experiment 1–reward probability was not confounded with interresponse time (lower panel of Figure 8). This is more and even better evidence that reward expectation controls duration. Further evidence of the same thing is the effect of nonreward on duration (lower panel of Figure 9). Nonreward of bar presses during the low-food signal increased the duration of bar presses during that signal. Figure 7 shows that the changes in mean reflected changes in variability. That the effect was associative is shown by its selectivity: Nonreward during the low-food signal did not reliably change duration during the high-food signal. Its effect on rate was less selective: Nonreward during the low-food signal slightly reduced rate during the high-food signal (upper panel of Figure 9). This may have been due to extinction of a context-food association. Pearce and Hall (1979) found that a context-food association increased the bar-pressing rate of rats.

With the confound between reward probability and interresponse time removed,

the effect of interresponse time on duration could be measured. These measurements found no effect (upper panel of Figure 8) in a situation where the effect of signal was very clear. This implies that the effect of signal on duration was not due to the rate difference (i.e., the interresponse time difference) between signals.

If rate and duration are both controlled by expectation of reward (the association of reward and signal), this raises the question whether the mechanism that computes the rate-controlling expectation is the same as the mechanism that computes the duration-controlling expectation. Two findings tentatively suggest that the mechanisms are different. One is that when the two signals were treated the same, at the start of the experiment, their rates became equal much sooner (upper panel of Figure 6) than their durations (lower panel of Figure 6). The other is that rate responded to nonreward (upper panel of Figure 9) more quickly than duration (lower panel of Figure 9).

Turning to methodological issues, the effect of reward probability on duration (t (10) = 6.2)–the difference between signals–was at least as clear as its effect on rate (t (10) = 6.1). In Experiment 1 (Figure 2), by contrast, a four-fold difference in reward probability changed duration (t (13) = 3) less clearly than rate (t (13) = 10). The two experiments were almost the same in other ways, so the comparison suggests that a random-ratio procedure is better than a random-interval procedure for studying how reward probability controls duration.

Sensitivity comparisons (Figure 11) led to conclusions consistent with those of Gharib et al. (2001). Whether the best transformation is the reciprocal square root, as Figure 11 suggests, or the logarithm, as Gharib et al. concluded, is not clear but both

are much better than no transformation at all. Figure 11 also shows that we might have used the standard deviation instead of the mean -- with some transformations, the standard deviation reliably detects the difference of interest -- but that we would have lost a lot of sensitivity by doing so.

General Discussion

Summary of Results

The main results were:

1. Reducing reward probability increased mean bar-press duration. This was true between signals (lower panels of Figures 2 and 6), within signals (bottom panel of Figure 4), and between the low-food signal and the intertrial interval (lower panels of Figures 2 and 6).

2. The increases in mean duration reflected increases in the spread of the distribution (Figures 3, 5, and 7).

3. Duration differences between signals were not due to differences in interresponse time (upper panels of Figures 4 and 8). When uncorrelated with reward probability, interresponse time had little or no effect on duration (upper panel of Figure 8).

4. The correlation between probability of reward and interresponse time in Experiment 1 (middle panel of Figure 4) provided a fine-grained look at duration as a function of reward probability (bottom panel of Figure 4, Figure 5). The function suggests that duration is sensitive to reward probability only when reward probability per bar press is less than about 0.3.

5. Unrewarded bar presses during the low-food signal increased duration during the low-food signal but not during the high-food signal (Figure 9).

6. The distribution of bar-press duration resembled a Pareto distribution (Figures 3 and 7).

Control of Variation

The rule tested here–reducing expectation of reward increases variation of response form–was suggested by results from a complex procedure designed for other purposes (Gharib et al., 2001). The present work supported the rule with a much simpler and better-controlled comparison–between high- and low-food signals (Figures 2 and 6). In Experiment 1, a confounding reduced the difference between them. When the confounding was removed (Experiment 2), the difference in duration was very clear ($t(10) = 6$). The rule was also supported by the effects of unrewarded responses during the low-food signal (Figure 9). Because the aftereffects of food/nonfood had, overall, little effect (Figure 10), the rule was also supported by the low-food-signal/intertrial-interval comparisons (Figures 2 and 6).

These results support the rule because it correctly predicted them. Can the correctly-predicted outcomes can be explained in other ways? When reward expectation decreases, rate decreases, so an alternative explanation of the duration increases is that they were due to the decrease in rate. Perhaps a long interresponse time (i.e., lower rate) causes the rat to "cool off," producing longer and more variable bar presses. But this explanation predicts the opposite of what Figures 4 and 5 show, which is that within signals mean duration decreased with longer interresponse times (i.e., lower

rate). More evidence against the rate explanation is the failure to find an effect of interresponse time (Figure 8) in the context of a strong signal effect and the evidence that rate and duration are controlled by different associative processes (Figures 6 and 9). It is also contradicted by Gharib et al. (2001)'s observation of several instances where rate changed while duration remained constant.

Neuringer and his colleagues have shown that rewarding variation can cause it to increase (Neuringer, 2002). This could not explain the reward-probability effects observed here because the "variation" of a single bar press (how much it differed from recent bar presses, which is the sort of measure of variation that Neuringer has used) did not affect its probability of reward.

How general is the rule? Support for it comes from rat nose pokes (Antonitis, 1951), sequences of responses made by rats (Balsam et al., 1998; Cherot, Jones, & Neuringer, 1996), and the location of pigeon key pecks (Eckerman & Lanson, 1969; Ferraro & Branch, 1968; Millenson, Hurwitz, & Nixon, 1961).

As discussed earlier, the value to an animal of such a rule is clear. Too much deviation from what has worked can cause an action to fail -- a cost of variation. On the other hand, trying new ways of doing things may turn up better ways of doing things -- a benefit of variation. When the likelihood of reward goes down, the expected cost goes down (there is less to lose), while the expected benefit, if anything, goes up, so the optimal amount of variation increases. When the optimal amount increases, the actual amount of variation should increase. Nothing about this reasoning limits it to animal behavior, of course. Evidence of its correctness comes from the existence of similar

relationships in two other domains: genetics and business.

1. *Genetics*. When a fruit fly is exposed to warm temperature or several other environmental stressors, the functioning of what are called *heat shock proteins* changes. In the case of one such protein, Hsp90, these changes produce an increase in phenotypic variation (Queitsch, Sangster, & Lindquist, 2002; Rutherford & Lindquist,1998). Hsp90 has similar effects in mustard plants (Sangster, Lindquist, & Queitsch, in press). Under normal conditions, this protein prevents mutations from producing phenotypic changes and they build up; but when its functioning is disturbed, these mutations are expressed and can form the basis for selection. "The buffered variation is therefore released precisely under those challenging conditions when selection is most stringent [i.e., successful reproduction is less likely than usual] and novelty might be most beneficial" (Sangster, Lindquist, & Queitsch, in press). Under normal conditions, in other words, Hsp90 reduces phenotypic variation; under unusual conditions, it increases it, in such a way that the new variants can be selected.

2. *Business.* Christensen (1997) noticed that industry-leading companies regularly failed to sustain their lead when new technologies came along. The leaders in making 14-inch drives did not become the leaders in making 8-inch drives. The leaders in making 8-inch drives did not become the leaders in making 5.25-inch drives. And so on. Similar examples involved retailing, computers, mechanical excavators, and steel mills. Christensen's research suggested that the reason for this pattern was that industry-leading companies were "held captive by their customers" (p. 18)–that is, by their success and by "expected rewards" (p. 32). They did not vary enough what they

did. Less successful companies took bigger leaps into the unknown. Surowiecki (2003, p. 46) pointed out a within-company example of the same thing: "Once Boeing became the dominant player in the aviation market, in the seventies, it lost its appetite for sporty bets [involving the manufacture of new airplanes]: why risk them, when profits were rolling in?"

Why Variation Increases During Extinction

Many studies have found that variation increases during extinction (Balsam et al, 1998; Neuringer, 2002). The present work repeated this result in a new way: Variation increased during what might be called *mini-extinctions*: Periods when bar presses during the low-food signal were not rewarded (lower panel of Figure 9).

Extinction differs from training in several ways. Which increased variation? (a) *Lower response rate*. Failure to detect an effect of interresponse time on duration (Figure 11) argues that response rate makes no difference. (b) *More frustration* due to omission of expected rewards. Were this important, nonreward at the end of a trial should have increased variation. Figure 10 shows that it did not. (c) *Lower density of reward*. Because food pellets were given at the end of signals and nowhere else, time since the last food pellet was less during intertrial intervals than during signals. Yet variation was greater during intertrial intervals than during signals, opposite to what this possibility predicts. (d) *Lower expectation of reward*. The present work supports this choice, of course, because it supports the idea that reducing expectation of reward increases variation. Gharib et al. (2001) reached the same conclusion.

Distribution of Bar-Press Durations

The bar-press duration distributions (Figures 3 and 7) resembled Pareto distributions, which have a density function and an inverse cumulative distribution function that are linear on a log-log scale. Pareto distributions are often called "power-law" behavior (e.g., Buchanan, 2001, p. 45) but they are unrelated to Steven's psychophysical power law. Gharib et al. (2001) observed similar duration distributions (e.g., lower panel of Figure 1) but their measurements excluded durations of 3 sec or more.

Pareto was an economist. The first Pareto distribution was a distribution of wealth–many persons had small wealth, a few had large wealth. The linguist George Zipf discovered that word frequencies (many words used a few times, a few words used many times), city sizes, and several other things had Pareto distributions (Adamic & Huberman, 2002; Zipf, 1949). Other examples of Pareto distributions include the energy released in earthquakes, the area of forest fires, the size of experimentally-produced avalanches of rice grains, and the number of links to and from a web page (Adamic & Huberman, 2002; Bak, 1996; Buchanan, 2001). Figures 3 and 7 and the examples of Gharib et al. (2001) are the first psychological examples, the first examples from experiments with living things, and the first examples where the slope is under experimental control.

A standard conclusion from Pareto distributions (e.g., Buchanan, 2001) is that the events on right-hand side of the distribution (say, large earthquakes) were generated by the same process as events on the left-hand side (small earthquakes). Applied here, this means that the longest bar presses were generated by the same

process that generated short bar presses, in spite of the great difference between them.

All previous examples of Pareto distributions, as far as we can tell, involve, or can be plausibly assumed to involve, mechanisms with three properties. First, they consist of *many similar interacting* elements. Second, the interactions include *recruitment*–if one element is in a certain state, the elements with which it interacts become more likely to be in that state. Third, the recruitment is *lossless*–that is, the recruited element is just as potent, just as likely to recruit other elements, as the element that recruited it. If you think of the change in state as a signal that spreads through the population, then lossless recruitment is the same as lossless transmission, that is, the signal becomes neither weaker nor stronger as it spreads. Mathematical models have shown that a system with this description will produce a Pareto distribution (Reed & Hughes, 2002).

With forest fires, the interacting elements are trees. A burning tree tends to cause other trees to burn. All burn at the same temperature with roughly the same chance of causing nearby trees to burn; in other words, the thousandth tree to catch fire is no more or less "contagious" than the first. With word frequency (many words used a few times, a few words used many time), the interacting elements are persons--writers and readers. Use of a word causes others to be more likely to use that word. Each use, no matter how far from the first use, has roughly the same effect on other people's words. With city populations, the interacting elements are also persons. Residents attract others to where they live. Newcomers are as attractive as the long-term residents.

With bar-press duration, there is an obvious candidate for each of the three features. (a) *Many interacting elements*: Neurons. (b) *Recruitment*: When one neuron

fires, it makes other neurons more likely to fire. (c) *Lossless* recruitment: Neuronal recruitment is lossless because action potentials are all or nothing. This line of argument suggests that bar presses are generated by neuronal activity and last as long as the neuronal activity lasts, perhaps as long as the number of neurons firing is above a threshold. This makes sense because force is required to hold the bar down. This argument further suggests that reward expectation produces changes in distribution shape by varying the probability that one neuron will recruit another, that is, the likelihood that if one neuron fires it will cause to fire the neurons to which it sends action potentials. When reward expectation goes down, this argument suggests, certain neurons become more easily excited–a somewhat counter-intuitive conclusion.

Methodology

Neuringer, Balsam, and their colleagues revived the study of variation (Balsam et al, 1998; Neuringer, 2002). One indication they were right – the subject does deserve more attention – is the strength of the effects observed here. In Experiment 2, the effect of reward probability on duration ($t(10) = 6.2$ for the signal difference) was as strong as its infinitely-better-studied effect on rate ($t(10) = 6.1$).

One aspect of the data analysis – the use of means to measure variation – may puzzle some readers. This choice derived from asking which of several possibilities worked best (Gharib et al., 2001) – that is, most clearly showed a difference that we knew existed. A lot was gained from this choice (Figure 11). Perhaps Figure 11 will inspire some readers to ask what measure and transformation work best with their own data.

In contrast to the data analysis, the data collection brought the study of variation closer to familiar methods. To test the idea that variation is under associative control, we used an experimental design similar to the within-subject designs popularized by Rescorla (e.g., Nairne & Rescorla, 1981). While most previous work on variation has used unusual responses (e.g., Neuringer, 2002) or unusual equipment (e.g., Antonitis, 1951; Herrnstein, 1961), this work used the rat's bar press measured with ordinary equipment.

References

Adamic, L. A., & Huberman, B. A. (2002). Zipf's Law and the Internet. *Glottometrics, 3,* 143-150.

Amsel, A. (1992). *Frustration theory*. Cambridge, England: Cambridge University Press.

Antonitis, J. (1951). Response variability in the white rat during conditioning, extinction, and reconditioning. *Journal of Experimental Psychology, 42*, 273-281.

Baddeley, A. D. (1966) The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, *18*, 119-129.

Bak, P. (1996). *How nature works*. Oxford, England: Oxford University Press.

Balsam, P. D., Deich, J. D., Ohyama, T., & Stokes, P. D. (1998). Origins of new behavior. In W. O'Donohue (Ed.), *Learning and behavior therapy* (pp. 403-420). Boston: Allyn & Bacon.

Buchanan, M. (2001). *Ubiquity*. New York: Three Rivers Press.

Chance, P. (1999). *Learning and behavior* (4th ed.). Pacific Grove, CA: Brooks/Cole.

Cherot, C., Jones, A., & Neuringer, A. (1996). Reinforced variability decreases with approach to reinforcers. *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 497-508.

Christensen, C. (1997). *The innovator's dilemma*. Boston, MA: Harvard Business School Press.

Domjan, M. (1998). *The principles of learning and behavior* (4th ed.).Pacific Grove, CA:

Brooks/Cole.

Doughty, A. H., & Richards, J. B. (2002). Effects of reinforcer magnitude on responding under differential-reinforcement-of-low-rate schedules of rats and pigeons. *Journal of the Experimental Analysis of Behavior*, *78*, 17-30

Eckerman, D. A., & Lanson, R. N. (1969). Variability of response location for pigeons responding under continuous reinforcement, intermittent reinforcement, and extinction. *Journal of the Experimental Analysis of Behavior, 12,* 1969, 73-80.

Ferraro, D. P., & Branch, K. H. (1968). Variability of response location during regular and partial reinforcement. *Psychological Reports, 23,* 1023-1031.

Gharib, A., Derby, S., & Roberts, S. (2001). Timing and the control of variation. *Journal of Experimental Psychology: Animal Behavior Processes, 27*, 165-178.

Herrnstein, R. J. (1961). Stereotypy and intermittent reinforcement. *Science, 133,* 2067-2069.

Hull, D. L., Langman, R. E., & Glenn, S.S., (2001) A general account of selection: Biology, immunology and behavior. *Behavioral and Brain Sciences*, *24*, 511-573.

Lacter, G. D., & Corey, J. R. (1982). Variability of the duration of an operant. *Behaviour Analysis Letters, 2,* 97-102.

Libby, M. E, & Church, R. M. (1975). Fear gradients as a function of the temporal interval between signal and aversive event in the rat. *Journal of Comparative and Physiological Psychology, **88***, 911-916.

Lieberman, D. A. (1990). *Learning: behavior and cognition.* Belmont, CA: Wadsworth.

Margulies, S. (1961). Response duration in operant level, regular reinforcement, and

extinction. *Journal of the Experimental Analysis of Behavior, 4,* 317-321.

Millenson, J. R., Hurwitz, H. M. B., & Nixon, W. L. B. (1961). Influence of reinforcement schedules on response duration. *Journal of Experimental Analysis of Behavior, 4,* 243-250.

Nairne, J. S., & Rescorla, R. A. (1981). Second-order conditioning with diffuse auditory reinforcers in the pigeon. *Learning and Motivation, 12,* 65-91.

Neuringer, A. (1991). Operant variability and repetition as functions of interresponse time. *Journal of Experimental Psychology: Animal Behavior Processes*, ***17***, 3-12.

Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review*, ***9***, 672-705.

Notterman, J. M., & Mintz, D. E. (1965). *Dynamics of response*. New York: Wiley.

Pearce, J. M. (1997). *Animal learning and cognition: An introduction*. East Sussex, UK: Psychology Press.

Pearce, J. M., & Hall, G. (1979). The influence of context-reinforcer associations on instrumental performance. *Animal Learning & Behavior*, 7, 504-508.

Platt, J. R., Kuch, D. O., & Bitgood, S. C. (1973). Rats' lever-press durations as psychophysical judgements of time. *Journal of the Experimental Analysis of Behavior*, *19*, 239-250.

Queitsch, C., Sangster, T. A., & Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature*, *417*, 618-624.

Reed, W. J., & Hughes, B. D. (2002). From gene families and genera to income and internet file sizes: Why power laws are so common in nature. *Physical Review E*,

66, 067103-1-4.

Roberts, S., & Neuringer, A. (1998). Self-experimentation. In K. A. Lattal & M. Perrone (Eds.), *Handbook of research methods in human operant behavior* (pp. 619-655). New York: Plenum.

Rutherford, S. L., & Lindquist, S. (1998). Hsp 90 as a capacitor for morphological evolution. *Nature, 396,* 336-342.

Sangster, T., Lindquist, S., & Queitsch, C. (in press). Under cover: Causes, effects, and implications of Hsp90-mediated genetic capacitance. *BioEssays*.

Schwartz, B. (1982). Interval and ratio reinforcement of a complex sequential operant in pigeons. *Journal of the Experimental Analysis of Behavior, 37,* 349-357.

Staddon, J. E. R., & Cerutti, D. T. (2003). Operant conditioning. *Annual Review of Psychology, 54,* 115-144.

Staddon, J. E. R., & Simmelhag, V. L. (1971). The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, *78*, 3-43.

Surowiecki, J. (2003, December 15). Jet blues. *The New Yorker,* 46.

Tremont, P. J. (1984). Variability of force and interresponse time under random interval reinforcement schedules. *Behavioral Processes, 9,* 413-420

Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods, 8,* 254-274.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Cambridge, MA: Addison-Wesley.

Control of Variation by Reward Probability
29 March 2004 version

Author Note

We thank Michelle Dokey, Jonathan Herberg, and Brian Louie for their assistance and Lada Adamic and Phil Reed for helpful comments. A report of this work was given at the November 2003 meeting of the Psychonomic Society. Correspondence should be sent to Seth Roberts, Department of Psychology, University of California, Berkeley, CA 94720-1650. Email: roberts@socrates.berkeley.edu.

Figure Captions

Figure 1. Results from Gharib et al. (2001). Upper panel: Bar-press duration as a function of time into a peak-procedure trial. On food trials, food was given for the first bar press more than 40 sec after the start of the signal, at which point the trial ended. On non-food trials, no food was given and the trial lasted much longer. For each of 18 rats, the harmonic mean of the durations was computed. Then the arithmetic mean was computed across rats. Lower panel: Change in distribution of bar-press duration during a peak-procedure trial. Early in the trial = the 30 sec up to and including the first bar press more than 40 sec after the start of the signal. after the start of the signal. Late in the trial = 45 to 75 sec after the first bar press more than 40 sec after the start of the signal. Each point is a median over 36 rats. Each bin is 0.1 sec wide. Bar-press durations of 3 sec or more were not recorded.

Figure 2. Experiment 1: Bar-press rate (upper panel) and duration (lower panel) as a function of day and signal. Method of computation: 1. The days of each phase were divided into blocks close to 5 days long. 2. For each rat-signal-block combination, the mean of the logarithm of all rates (upper panel) or durations (lower panel) was computed. 3. For each signal-block combination, a 10% trimmed mean over rats was computed.

Figure 3. Experiment 1: Distribution of bar-press durations as a function of signal. Data from the last 20 days of the 100%/25% phase of Experiment 1. Upper panel: Probability distributions. Method of computation: 1. The minimum and maximum

duration over all rats and signals were used to determine an interval. 2. The interval was divided into ten (high-food function) or eight (low-food and intertrial functions) segments of equal width on a log scale. For each function, the number of segments was the maximum that would not produce any zero probabilities in the final result. A probability of zero is hard to show on a log scale. 3. These bin definitions were used to determine frequencies for each rat. 4. Frequencies were converted to probabilities by dividing by sample size. 5. Probabilities were converted to probability densities by dividing by bin width on the untransformed scale (seconds, not log seconds). 6. For each bin, a median was computed over rats. Lower panel: Inverse cumulative distribution functions. Method of computation: 1. The function was computed for each rat separately. 2. For each abscissa value, a median was computed over rats.

Figure 4. Experiment 1: Effect of interresponse time. Top panel: Duration as a function of interresponse time. Middle panel: Probability of reward as a function of interresponse time. Bottom panel: Duration as a function of probability of reward. Duration and interresponse times are on a log scale, probability of reward on a square-root scale. The 100%/50% data are from the last 15 days of that phase; the 100%/25% data are from the last 20 days of that phase. Method of computation: 1. For each rat-signal combination (14 rats, 4 signals), all interresponse times were gathered, ranked, and divided into 8 equal-sized bins by rank, a total of 448 bins. (To show as clearly as possible the correlation between interresponse time and probability of reward, the first bar press during a signal was given an interresponse time equal to the time since the signal began.) One group contained all interresponse times between the minimum and

the 12.5% quantile; the next, all interresponse times between the 12.5% quantile and the 25% quantile, and so forth. 2. For each group, the mean of the log interresponse times was computed (448 means, 32 per rat). 3. The 10% trimmed mean over rats was computed to summarize each bin (32 trimmed means). 4. For each rat-signal-bin combination, the mean log duration was computed, then averaged over rats to get 32 values. 5. For each rat-signal-bin combination, the probability of reward was computed, then averaged over rats to get 32 values. 6. Responses during intertrial intervals were not divided by interresponse time because there were too few of them. The average duration for interresponse time was computed by taking the mean within rats and a trimmed mean over rats.

Figure 5. Experiment 1: 10% and 90% quantiles of duration distributions as a function of probability of reward. These values were computed in the same way as the values in the bottom panel of Figure 4, except that each set of durations was summarized by the 10% and 90% quantiles rather than by the mean.

Figure 6. Experiment 2: Bar-press rate (upper panel) and duration (lower panel) as a function of day and signal. Same method of computation as Figure 2.

Figure 7. Experiment 2: Distribution of bar-press durations as a function of signal. Upper panel: Probability density functions. Lower panel: Inverse cumulative distribution functions. Same method of computation as Figure 3.

Figure 8. Experiment 2: Separation of the effects of signal and interresponse time. Top panel: Duration as a function of signal and interresponse time. Bottom panel: Probability of reward as a function of signal and interresponse time. Duration and

interresponse time are on a log scale, probability of reward on a square-root scale. Method of computation: The bins into which intertrial intervals were classified were chosen to be close to equal width on a log scale, with the exception of the leftmost and rightmost bins, which held all interresponse times less than and greater than a cutoff. The upper cutoff was chosen to be small enough so that all rats had some responses in the rightmost bin. The eight bins were: interresponse times of 1 sec or less, 2 sec , 3 sec, 4-5 sec, 6-8 sec, 9-15 sec, 16-24 sec, and more than 24 sec. Each point is a 10% trimmed mean. Interresponse time was the time since the previous bar press.

Figure 9. Experiment 2: Effect of consecutive unrewarded bar-presses during the low-food signal on bar-press rate (upper panel) and duration (lower panel). Every trial was classified according to the number of consecutive unrewarded bar-presses during the low-food signal that had preceded it. Values on the abscissa are plotted according to $\log(1 + x)$, where x is the number of unrewarded bar presses. For example, the value for 0 is plotted above log 1, the value for 10 above log 11, and so on. Values within a bin are assigned the geometric mean of the edges of the bin. For example, the right-most bin contains data from trials preceded by 31 to 140 unrewarded bar presses during the low-food signal. The geometric mean of 31 and 140 is 66, which is plotted above log 67.

Figure 10. Experiments 1 and 2: Behavior during intertrial intervals: Bar-press rate (upper panel) and duration (lower panel). Each point is a 10% trimmed mean. Based on all days of the 100%/50% and 100%/25% phases of Experiment 1 and the 100%/25% phase of Experiment 2. After high-food (food): after high-food trials that

ended with food. After low-food (food): after low-food trials that ended with food. After low-food (no food): after low-food trials that ended without food.

Figure 11. Experiment 2: Sensitivity as a function of power transformation and measure. Each *t* value comes from a comparison of durations during the high- and low-food signals over the last 20 days of Experiment 2.

Control of Variation by Reward Probability
29 March 2004 version

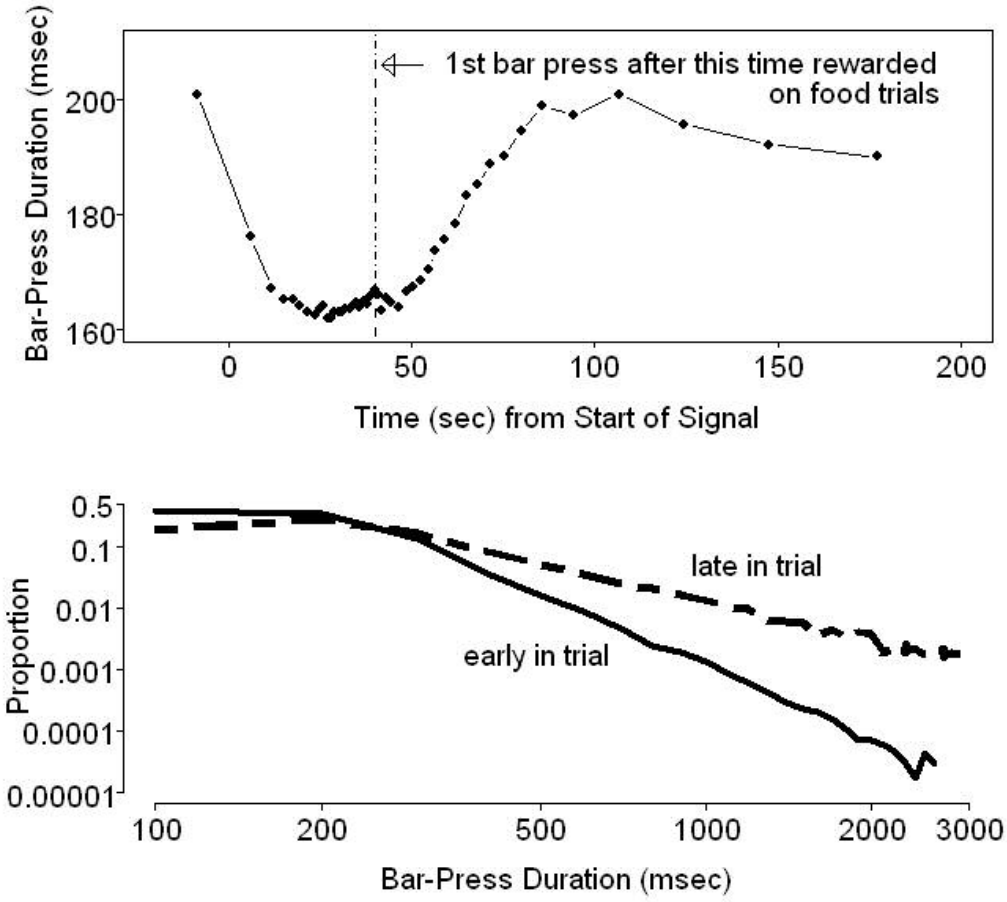


Figure 1

Control of Variation by Reward Probability
29 March 2004 version

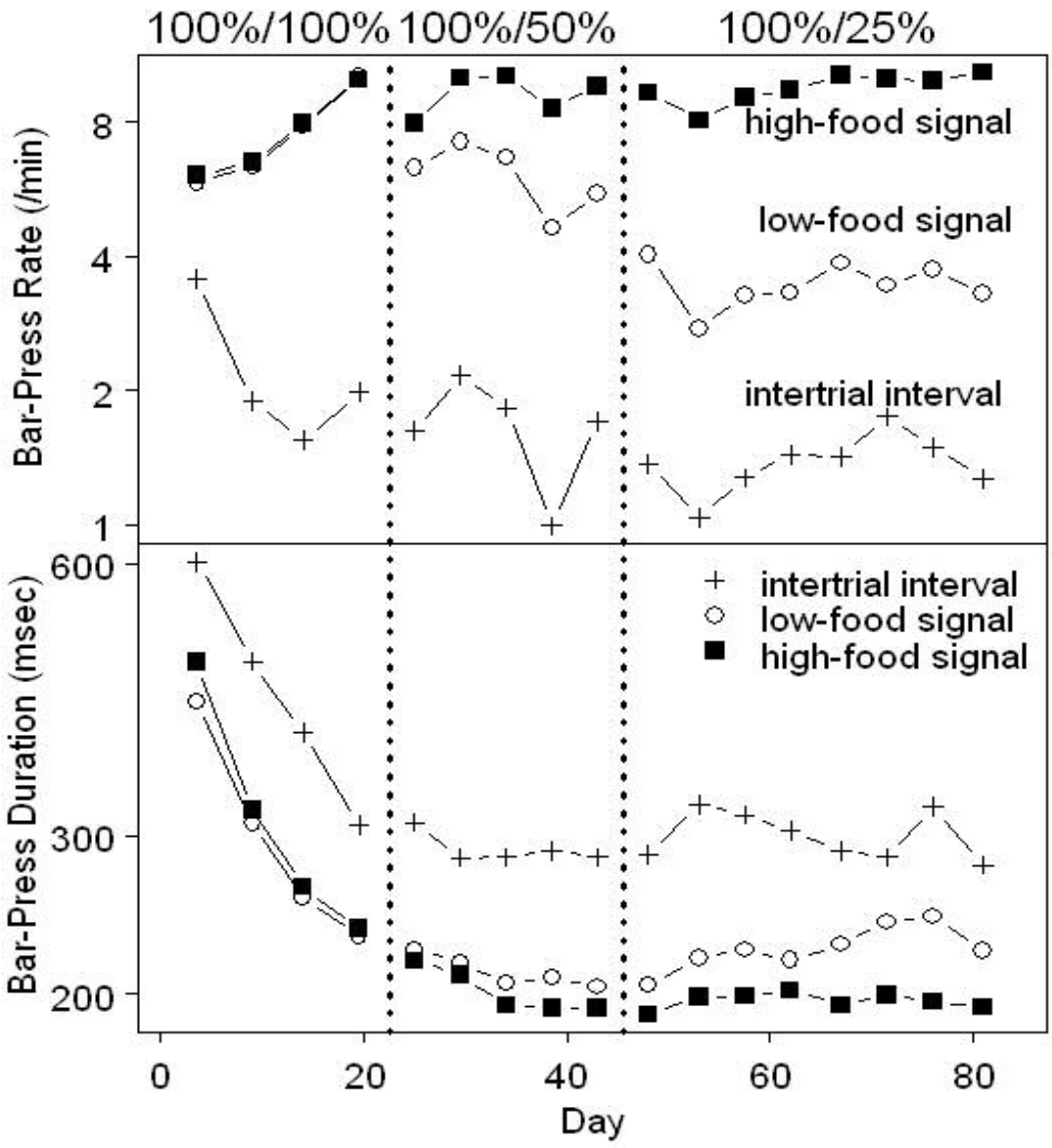
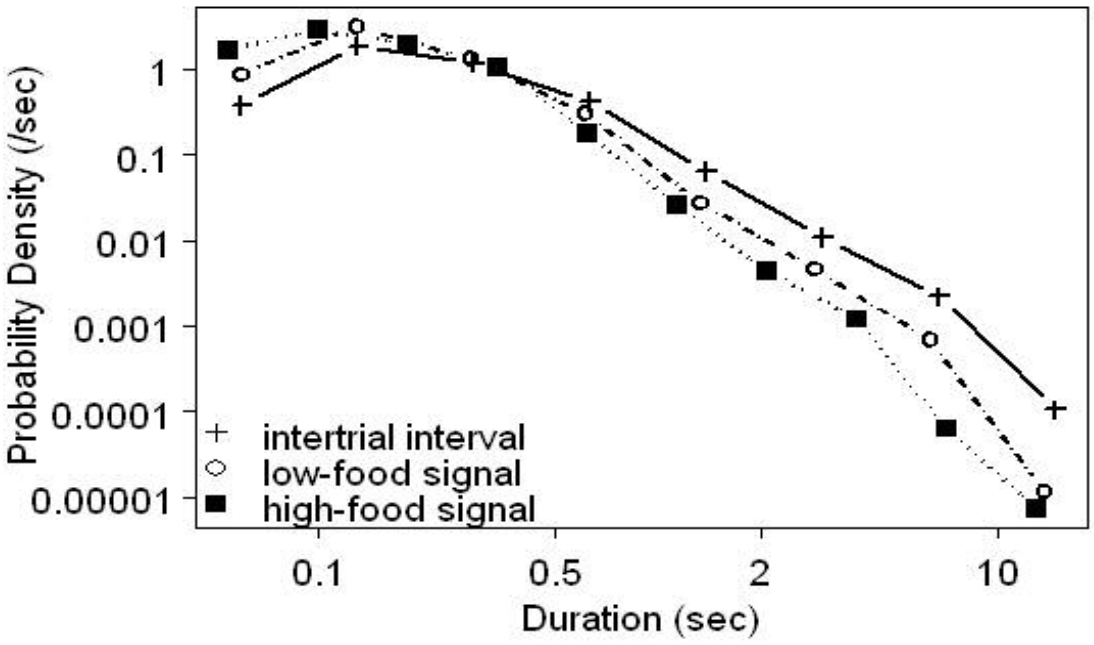


Figure 2


Probability Density (/sec)
1
0.1
0.01
0.001
0.0001
0.00001
intertrial interval
low-food signal
high-food signal
0.1
0.5
2
10
Duration (sec)

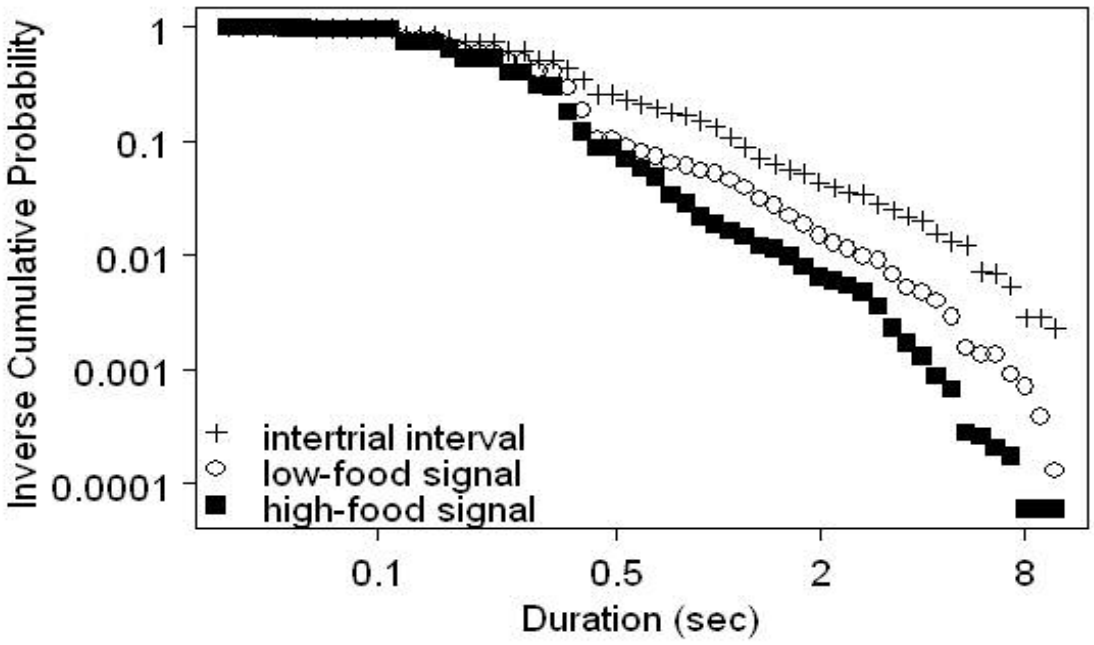

Inverse Cumulative Probability
1
0.1
0.01
0.001
0.0001
intertrial interval
low-food signal
high-food signal
0.1
0.5
2
8
Duration (sec)

Figure 3

Control of Variation by Reward Probability
29 March 2004 version

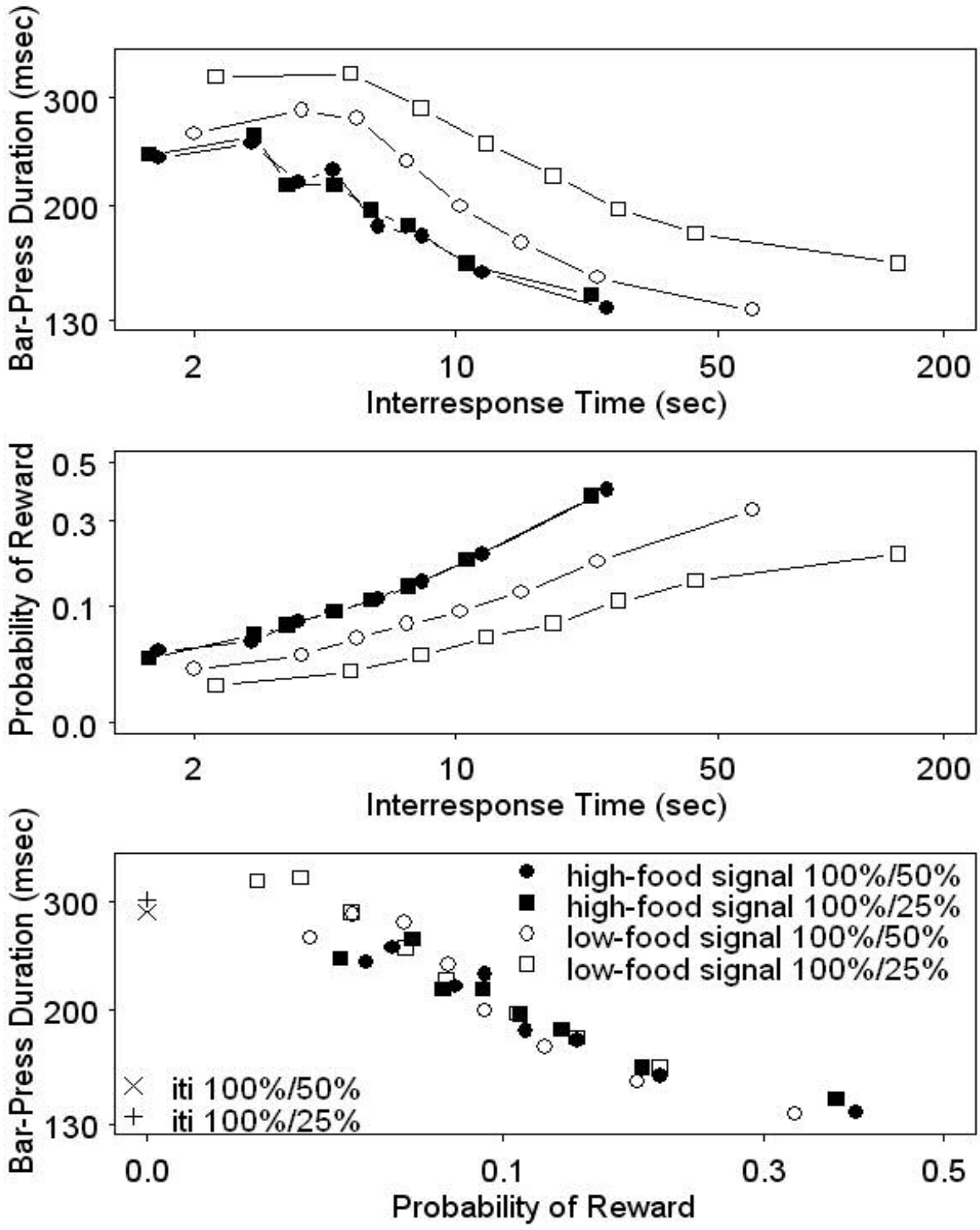


Figure 4

Control of Variation by Reward Probability
29 March 2004 version

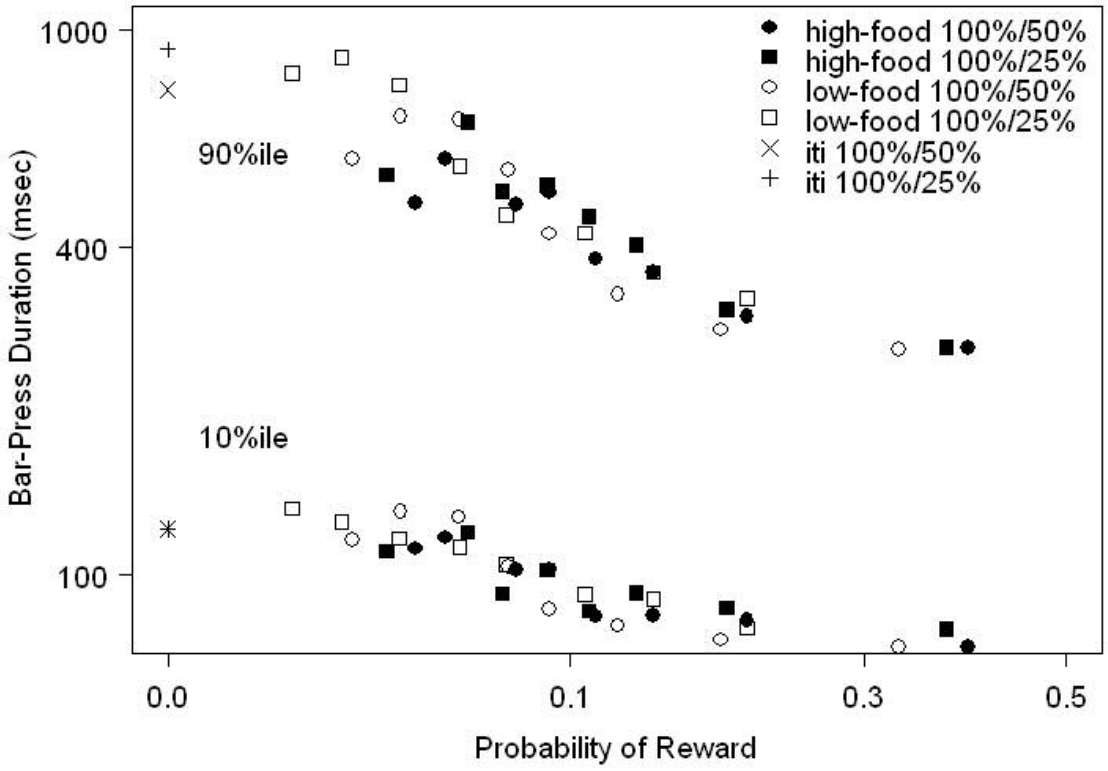


Figure 5

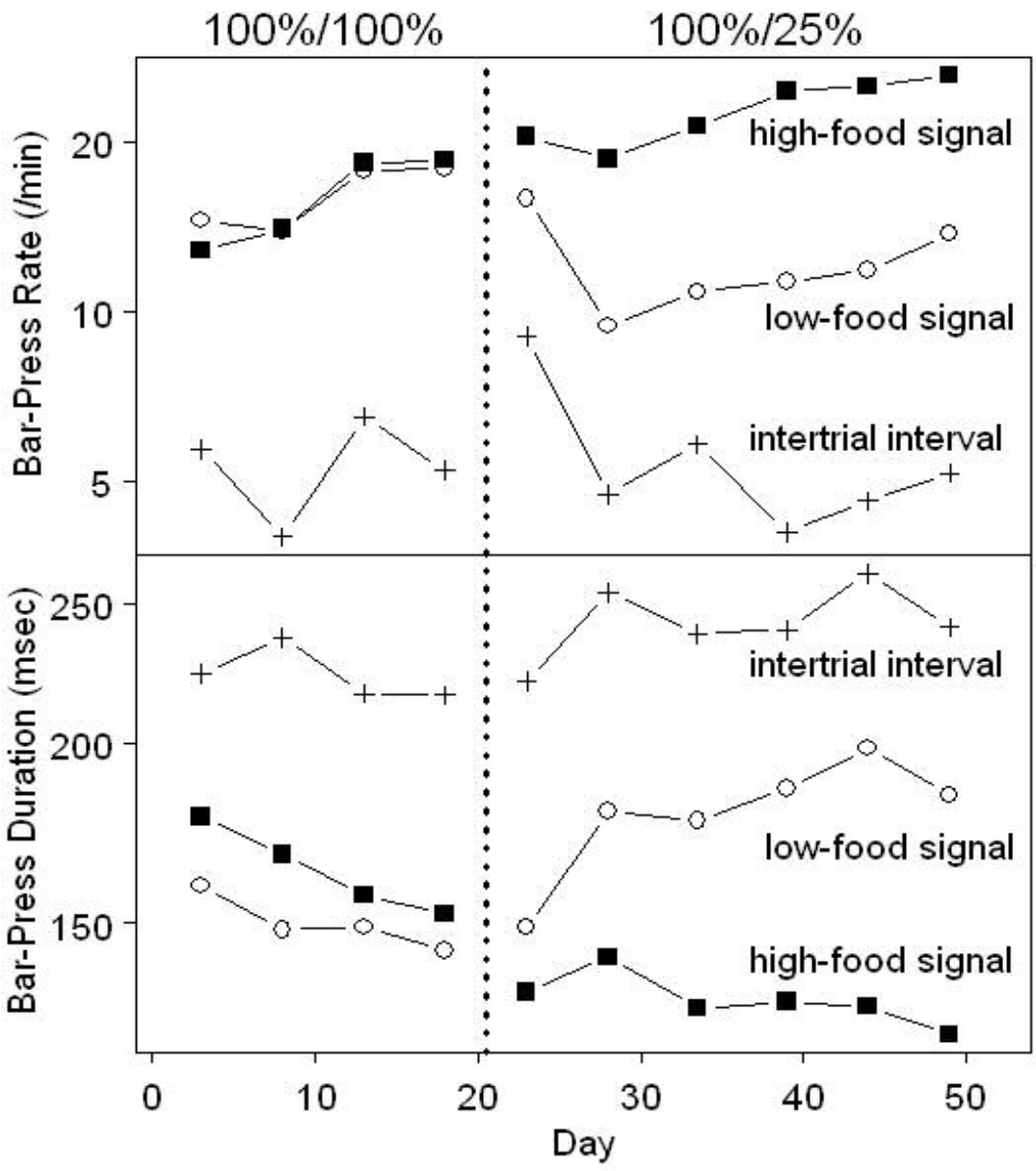

100%/100%
100%/25%
Bar-Press Rate (/min)
20
10
5
high-food signal
low-food signal
intertrial interval
Bar-Press Duration (msec)
250
200
150
intertrial interval
low-food signal
high-food signal
0
10
20
30
40
50
Day

Figure 6

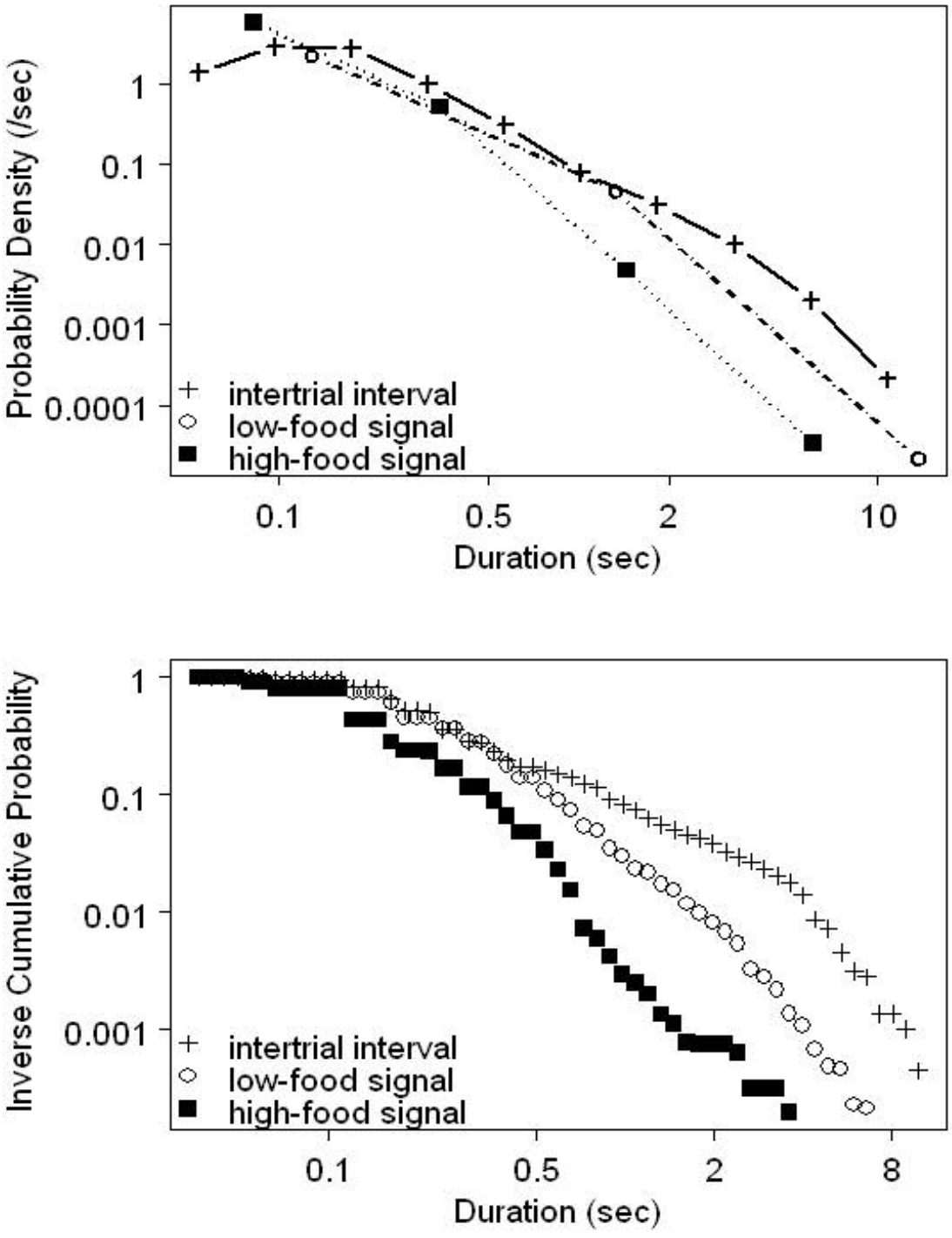

Probability Density (/sec)
1
0.1
0.01
0.001
0.0001
intertrial interval
low-food signal
high-food signal
0.1
0.5
2
10
Duration (sec)
Inverse Cumulative Probability
1
0.1
0.01
0.001
intertrial interval
low-food signal
high-food signal
0.1
0.5
2
8
Duration (sec)

Figure 7

Control of Variation by Reward Probability
29 March 2004 version

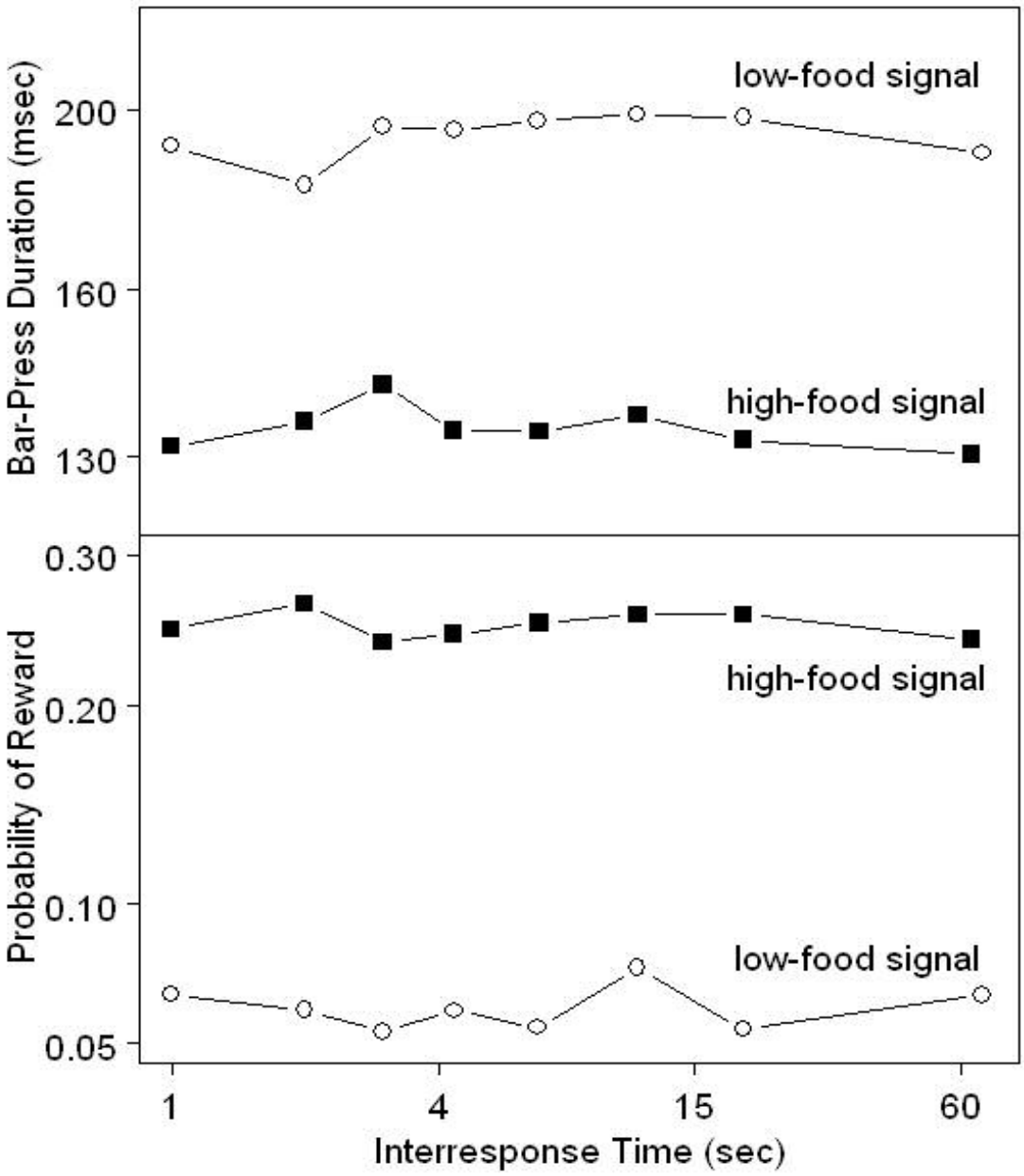


Figure 8

Control of Variation by Reward Probability
29 March 2004 version

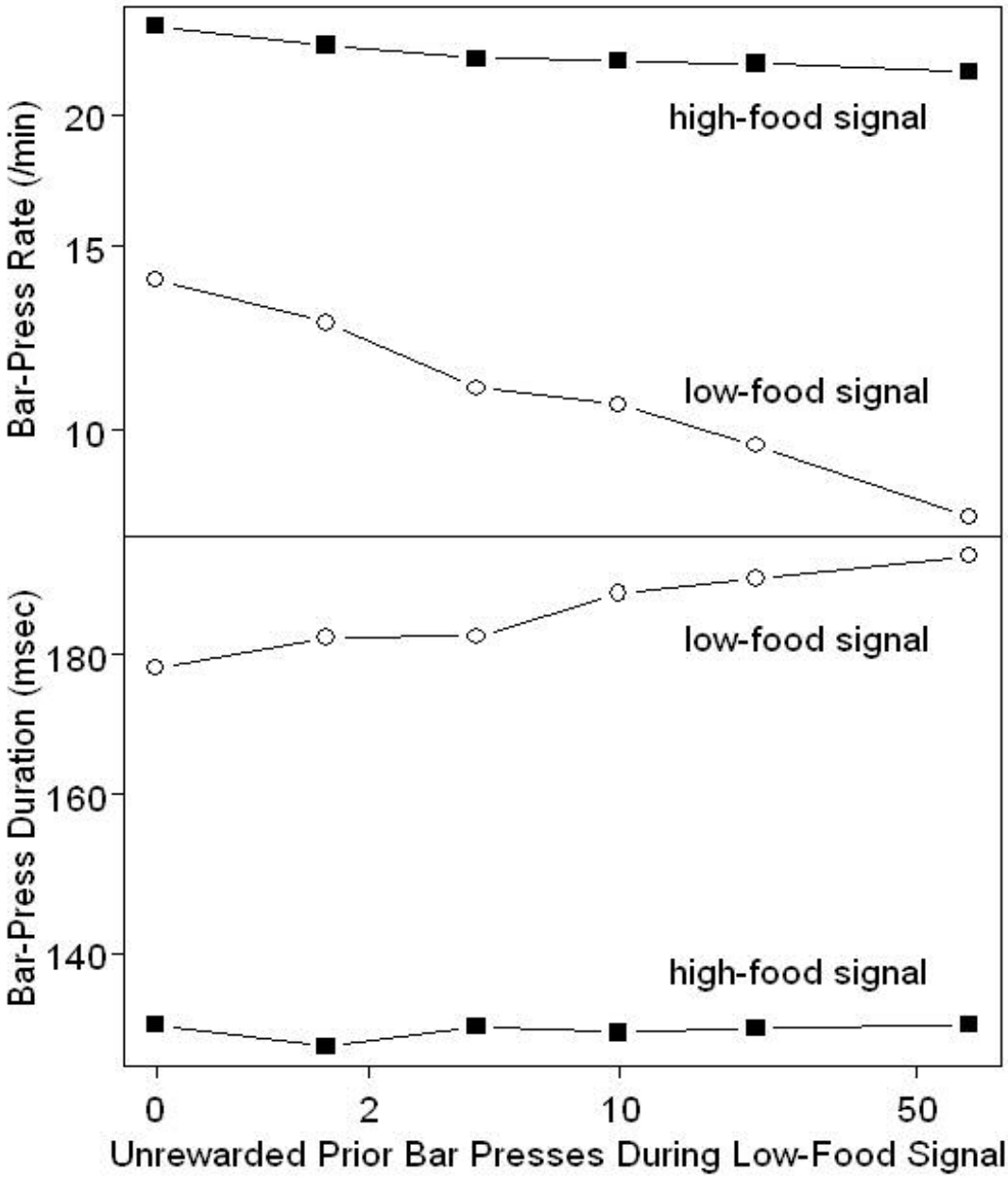


Figure 9

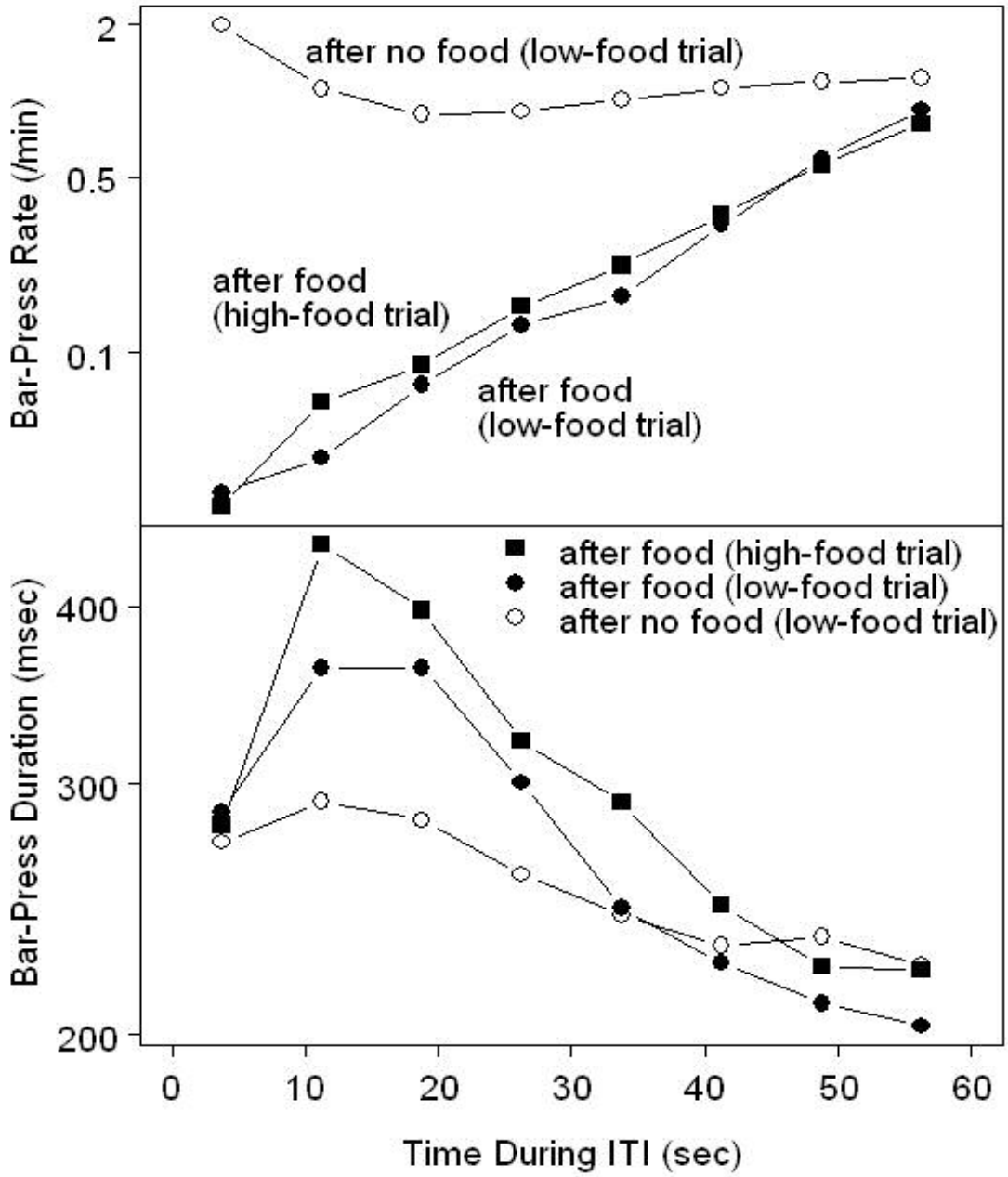

2
0.5
0.1
Bar-Press Rate (/min)
after no food (low-food trial)
after food
(high-food trial)
after food
(low-food trial)
400
300
200
Bar-Press Duration (msec)
after food (high-food trial)
after food (low-food trial)
after no food (low-food trial)
0
10
20
30
40
50
60
Time During ITI (sec)

Figure 10

Control of Variation by Reward Probability
29 March 2004 version

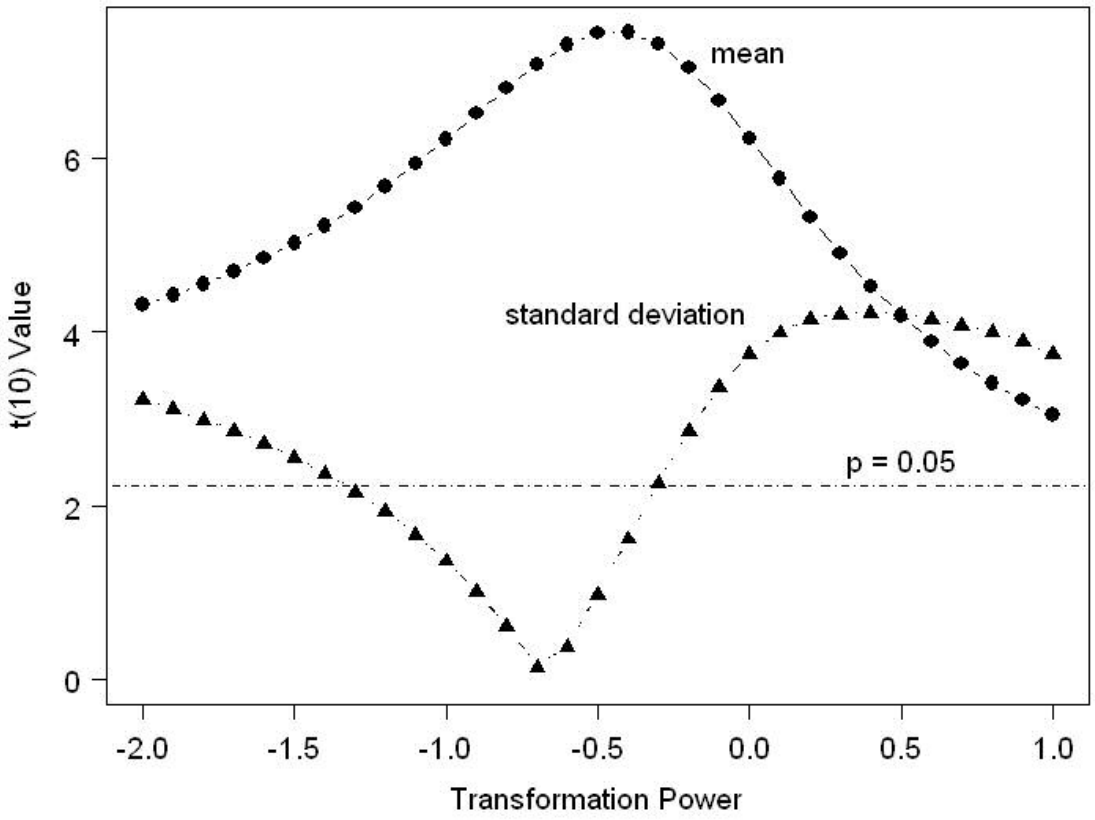


Figure 11