

Goal-Directed Control on Interval Schedules Does Not Depend on the Action–Outcome Correlation

Eric Garr, Badrunnesa Bushra, Norman Tu, and Andrew R. Delamater
City University of New York

When an organism's action is based on an anticipation of its consequences, that action is said to be goal-directed. It has long been thought that goal-directed control is made possible by experiencing a strong correlation between response rates and reward rates (Dickinson, 1985). To test this idea, we designed a set of experiments to determine whether the response rate-reward rate correlation is a reliable predictor of goal-directed control on interval schedules. In Experiment 1, rats were trained on random interval (RI) schedules in which the response rate-reward rate correlation was manipulated across groups. In tests of reward devaluation, rats behaved in a goal-directed manner regardless of the experienced correlation. In Experiment 2, rats once again experienced either a strong or weak correlation, but on RI schedules with lower overall reward densities. This time, behavior appeared habitual regardless of the experienced correlation. Experiment 3 confirmed that the density of the RI schedule influences goal-directed control, and also revealed that extensive training on these schedules resulted in goal-directed action. Finally, in Experiment 4 goal-directed responding was greater and emerged sooner on fixed than random interval schedules, but, again, was manifest after extensive training on the RI schedule. Taken together, our data suggest that goal-directed and habitual control are not determined by the correlation between response rates and reward rates. We discuss the importance of temporal uncertainty, action–outcome contiguity, and reinforcement probability in goal-directed control.

Keywords: interval schedules, actions, habits, response–reward correlation, contiguity

Supplemental materials: <http://dx.doi.org/10.1037/xan0000229.supp>

It is well known that an animal's behavior is sometimes sensitive or insensitive to changes in the value of its reinforcing outcome—that is, it can appear goal-directed or habitual (see Smith & Graybiel, 2014 for review). One important variable determining one or the other mode of behavioral control is the schedule of reinforcement. Rodents trained to respond on a random ratio (RR) schedule, for example, are more likely to show goal-directed responding than rodents trained on a random interval (RI) schedule (Dickinson, Nicholas, & Adams, 1983; Gremel et al., 2016; Gremel & Costa, 2013a, 2013b; Killcross & Coutureau, 2003; O'Hare et al., 2016; Renteria, Baltz, & Gremel, 2018). These types of schedules differ in how reward availability is controlled. Under an RR schedule,

reward is procured after n responses have been made where the value of n is randomly determined from one reward to the next. Under an RI schedule, reward is made available for procurement after t seconds have elapsed where t is randomly determined after each reward.

One long-standing hypothesis posits that instrumental responding will be goal-directed when the organism experiences a correlation between response rates and reward rates, but habitual when it experiences no correlation (Dickinson, 1985; Perez et al., 2016). Accordingly, behavior maintained on RR schedules is usually goal-directed because the animal is able to experience a positive correlation between its rate of responding and the rate of reward. Because reward availability is controlled by the number of responses on an RR schedule, higher rates of responding yield higher rates of reward. However, behavior maintained on RI schedules is usually habitual because the correlation is degraded. Because reward availability is controlled by time on an RI schedule, higher rates of responding yield negligible changes in the rate of reward above a fairly low threshold in response rate (Baum, 1973). Thus, the correlation between response rates and reward rates on an RR schedule is positive, while under an RI schedule this correlation approaches zero following a moderate amount of training. However, even under an RR schedule the correlation between response rates and reward rates will tend toward zero with extensive training. This is because as behavior becomes increasingly stereotyped, the rates of responding and reward will not vary much from session to session. Thus, Dickinson's (1985) hypothesis makes two

This article was published Online First October 17, 2019.

Eric Garr, Department of Psychology, Graduate Center and Brooklyn College, City University of New York; Badrunnesa Bushra and Norman Tu, Department of Psychology, Brooklyn College, City University of New York; Andrew R. Delamater, Department of Psychology, Graduate Center and Brooklyn College, City University of New York.

The research reported here was supported by a National Institute on Drug Abuse and National Institute of General Medical Sciences (034995) grant awarded to Andrew R. Delamater. We thank Dan Siegel for assistance with data collection.

Correspondence concerning this article should be addressed to Eric Garr, who is now at the Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218. E-mail: egarr@gradcenter.cuny.edu

key predictions: Habits will form under RI schedules after a moderate amount of training, while under RR schedules habits will form only after extensive training.

The time course of habit formation does indeed depend on whether animals are trained on RI or RR schedules, with habits forming earlier on RI schedules (Dickinson et al., 1983; Gremel et al., 2016; Gremel & Costa, 2013a, 2013b; Killcross & Coutureau, 2003; O'Hare et al., 2016; Renteria et al., 2018). While these studies provide support for the response rate-reward rate correlation hypothesis, the hypothesis itself has not been followed up on extensively. Surprisingly few studies have varied the extent of training on RR schedules, and thus it is unclear whether extensive training on RR schedules promotes a shift from goal-directed to habitual control. Adams (1982) demonstrated that extensive training on a continuous reinforcement (CRF) schedule is sufficient to induce habitual responding in rats, but CRF is quite different from RR. Garr and Delamater (2019) recently found that performance of a heterogeneous chain reinforced on a CRF schedule remained goal-directed after 60 days of training. In addition, Corbit, Nie, and Janak (2012) reported that rats trained to lever-press for a liquid sucrose reward on an RR 3 schedule remained goal-directed after 56 1-hr training sessions. In another study, O'Hare et al. (2016) trained mice on an RR schedule and concluded that responding was goal-directed after four training sessions but habitual after eight sessions. However, goal-directed and habitual control were inferred by giving mice an omission test in which each response resulted in the temporary withholding of the reward, rather than the more commonly used reinforcer devaluation test. This makes it difficult to know whether the response was truly independent of the outcome's value after eight training sessions. Also, in this study sensitivity to omission was quantified relative to a preomission baseline rather than to a noncontingent control condition (e.g., Dickinson, Squire, Varga, & Smith, 1998). As a result, ostensible differences in omission sensitivity could have been due to baseline differences, but these were not reported. Finally, the omission test is not particularly instructive at distinguishing between habitual and goal-directed responding. While it is true that habitual responding may be expected to fail an omission test, a well-learned goal-directed action could also be insensitive to such a change in contingency for the simple fact that an abrupt omission of the outcome directly conflicts with the extensive prior experience. It is thus not clear whether extensive training on an RR schedule results in a habit, as Dickinson's (1985) hypothesis would suggest.

There are additional studies that pose challenges to the response rate-reward rate correlation hypothesis even when animals are trained on interval schedules. In one study, Corbit, Chieng, and Balleine (2014) observed that rats were goal-directed following 16 training sessions in which they pressed a lever for food rewards on an RI schedule whose mean value increased across days from 15 s (2 days) to 30 s (2 days) to 60 s (12 days). Although response rate-reward rate correlation data were not reported, it is unlikely that the correlation was high after 16 training sessions. In addition, responding often remains goal-directed after extensive training on an RI schedule when multiple action-outcome contingencies are trained (Colwill & Rescorla, 1985, 1988; Kosaki & Dickinson, 2010; but see Smith & Graybiel, 2013). For example, Kosaki and Dickinson (2010) trained rats to press two concurrently available levers, one for pellets and one for sucrose solution. They found that, even after 20 sessions of RI training, responding was goal-

directed in tests of outcome devaluation. While at first blush this result seems at odds with the response rate-reward rate correlation idea, the result may, in fact, be compatible with it. In this study, rats were trained on a concurrent schedule. This means that within each session the rats could have experienced a higher local correlation between responding on one lever and a reward of one type, but not of the other. If so, this could maintain goal-directed control even after extensive training. However, this sort of explanation faces difficulties with the earlier work of Colwill and Rescorla (1985, 1988) who similarly trained rats on two different action-outcome relations on RI schedules and observed goal-directed control after extensive training. However, each action-outcome relation was trained in separate sessions in these studies, and not on a concurrent schedule. Accordingly, an appeal to local perceived correlations would not easily account for the observation of maintained goal-directed control. After extensive training, within each training session the rats should have experienced weak response rate-reward rate correlations. To account for these findings, Dickinson (1989) suggested that there is a global sense in which the animals learned that each outcome is associated with a unique action, and this type of correlation may have preserved goal-directed control. However, it is difficult to reconcile this more global idea with the notion that the experienced correlations between the *rates* of responding and rewards determines the strength of goal-directed control. Clearly, extensively trained animals can experience weak correlations between response rates and reward rates, yet can appear goal-directed.

Given the fact that the response rate-reward rate correlation idea has not been extensively studied and that training under interval schedules has resulted in seemingly divergent findings, it is important to further investigate goal-directed control on interval schedules. We sought to test the correlation idea further by directly manipulating the experienced action-outcome correlations on different interval schedules. In two experiments (Experiments 1 and 2) we trained rats on RI schedules while varying the correlation between response rates and reward rates between separate groups. If the experienced response-reward rate correlation governs action control, animals trained with a positive correlation should display goal-directed responding while animals trained with a weak correlation should appear habitual in tests of reward devaluation. As the results from Experiment 1 and Experiment 2 suggested that reward density may play a more substantial role than the response rate-reward rate correlation, in Experiment 3 we manipulated reward density when there was little opportunity for the response rate-reward rate correlation to vary. To address specific hypotheses raised by the results of Experiment 3 concerning the role of reward density and extent of training on RI schedules, Experiment 4 was designed to hold reward density constant as rats were trained on an RI or fixed interval (FI) schedule but then tested after limited, moderate, and extensive training. Collectively, the data reported here present additional challenges to the response rate-reward rate correlation hypothesis of action control, and suggest new ways of thinking about the problem.

As noted above, the notion of a response rate-reward rate correlation has been discussed in different ways in the literature. Initially, Dickinson (1985) referred to a between-day correlation, which integrates response and reward rates across days, to account for why overtraining on a CRF schedule produced goal-insensitive responding (Adams, 1982). Researchers have also noted that the

response rate-reward rate correlation can be computed based on local rates of responding and reward over short time windows (Perez et al., 2016). To our knowledge there is no consensus on exactly how animals might compute action–outcome correlations. Nonetheless, although our studies initially were designed to explore the role of between-day correlations in goal-directed control, we also present data relevant to examining a possible role for more local correlations.

Experiment 1

The aim of Experiment 1 was to examine the effect of varying the response rate-reward rate correlation across training sessions as animals underwent training on RI schedules. To this end, different groups of rats were trained on RI schedules to respond for food pellets by pressing a lever. If an animal is trained on an RI schedule for enough time to achieve a high response rate, the correlation between response rates and reward rates becomes weak (Baum, 1973). Taking advantage of this fact, we trained one group of rats on an RI schedule for 10 sessions with the express purpose of creating a weak response rate-reward rate correlation. To induce a positive correlation in a separate group, we systematically decreased the mean value of the RI schedule on every session. By decreasing the mean value of the RI schedule, the mean time between rewards decreases over sessions, which means that, by design, the mean reward rate also increases. On the assumption that rats also increase their response rates over sessions (which they are known to do, e.g., DeRusso et al., 2010; Dickinson & Charnock, 1985; LeBlanc, Maidment, & Ostlund, 2013), this would create a positive correlation between reward rates and response rates on an RI schedule. After training, devaluation tests were conducted in which rats were sated on the pellet type associated with lever pressing or on a control pellet type to gauge instrumental sensitivity to reward devaluation.

Method

Subjects. Twenty-four naïve Long-Evans rats (12 male and 12 female), bred in-house from Charles River Laboratories parentage, were housed in plastic cages ($17 \times 8.5 \times 8$ in.) in a colony room with a 14-hr light/10-hr dark cycle. Rats were housed in groups of two to four per cage with wood chip bedding and constant water access. The free-feeding body weights varied between 322 g and 365 g for males and 220 g and 247 g for females. All rats were maintained at 85% of free-feeding body weight for the duration of the experiment by supplemental feedings that occurred immediately following each daily experimental session. All procedures conformed to institutional IACUC regulations.

Apparatus. Eight operant chambers (MED Associates, ENV-008) were used for behavioral training and testing. Each chamber ($10 \times 12.5 \times 10$ in.) was located within a sound-attenuating cubicle (Med Associates, ENV-018MD, $17.75 \times 12.5 \times 23.25$ in.), and these were located in an isolated room within the laboratory (accessed only at the beginning and ends of each session). The interior of the chamber was comprised of two Plexiglas walls, two metal walls, a Plexiglas ceiling, and a grid floor with 0.25-in. diameter rods spaced 5/8 in. apart. Attached to one metal wall was a 28-v house light 8 in. above the grid floor. On the opposite metal wall was a food magazine and two retractable levers (ENV-

112CM). The food magazine (ENV-202RMA, 2.25×2.25 in.) was connected to two separate pellet dispensers via plastic tubing. The pellets used were TestDiet MLabRodent 45 mg grain pellets and Bio-Serv 45 mg purified pellets. The Bio-Serv pellets are higher in sugar content, but both pellet types are calorically very similar (3.60 and 3.30 kcal/g for Bio-serv and TestDiet, respectively). Two lever slots were located 2.5 in. above the floor and 3.5 in. to the right and left of the food magazine. For any given rat only one of these levers was used throughout the experiment. A fan inside the cubicle provided for background noise (79 dB, C weighting Realistic Sound Meter placed in the center of the chamber with the door closed). A computer in the same room controlled all chambers. Suspended wire cages ($9.75 \times 8 \times 7.25$ in.) were used for isolating rats during the 1-hr satiation periods and 20-min preference tests, and these were located in a different isolated room in the laboratory. During the satiation periods rats were given pellets in ceramic bowls that were stabilized to the cages by hooks attached to springs.

Procedure. Rats were first given magazine training with one pellet type to familiarize them with the location of pellet deliveries. Half the rats were assigned to receive the TestDiet pellet and the other half Bio-Serv (counterbalanced with sex), and the pellet type assigned to each rat remained the pellet type they would receive in the operant chamber for the duration of the experiment. During this 20-min session, pellets were delivered according to a 60-s random time schedule.

Rats were then trained to press a lever on a CRF schedule, such that each lever press yielded one pellet. At the beginning of the session, the left or right lever (counterbalanced with pellet assignment) was inserted into the chamber and remained available until 50 pellets were earned or 60 min elapsed, whichever occurred first. Following CRF training, rats were trained to press the lever to which they were assigned on an RI schedule. One group (constant group; $n = 12$, six males and six females) was trained on an RI 10-s schedule for one session per day over 10 days. The other group (changing group; $n = 12$, six males and six females) was trained on a different RI schedule each day for 10 days, such that the mean reward interval became shorter across sessions to approximate the mean value obtained in the constant group. The mean reward interval (in seconds) was set at the following values for each of the 10 sessions, in chronological order: 37, 25, 18, 14, 12, 10, 9, 8, 7, 6. For all RI schedules, pellets were delivered by consulting a list of numbers. This list was comprised of zeros and a one, such that the number of zeros was equal to $x - 1$ where x is the desired mean interval between rewards. The computer randomly drew (with replacement) from this list every second, and if the number 1 was drawn then pellet delivery was set up so that as soon as the lever was pressed a pellet was delivered. The pellet remained available until the lever was pressed, after which the random drawings continued. If a 0 was drawn, pellet delivery was not set up. This reward delivery scheme results in exponentially distributed interreward intervals (IRIs), provided that the lever pressing rate is sufficiently high. The sessions ended and the lever was withdrawn once 50 pellets were earned or 60 min elapsed, whichever occurred first. Two rats from the “constant” group failed to learn to lever press (<2 presses/min on Session 10 of RI training) and were excluded from all analyses.

Following 10 days of RI training, the devaluation cycle began. The cycle was comprised of two tests separated by a day of

retraining. Prior to the first test, rats were isolated in wire cages and given one hour of unlimited access to either the pellet type that was associated with lever pressing (“devalued” test) or the other pellet used as a control for general satiety (“valued” test). Rats were preexposed to the unfamiliar pellet type (10 pellets to consume) every day for 4 days leading up to the start of testing. Water was freely available throughout the satiation sessions. Immediately after the satiation period, rats were placed in the operant chambers and given a 5-min extinction test in which the lever was available but no rewards were delivered. After the extinction test, the remaining pellets from the satiation period were weighed and recorded, and rats were placed back in the wire cages for a 20-min preference test to gauge the effectiveness of the satiation manipulation. During the preference test, 10 g of each pellet type were available in separate bowls and rats were allowed to freely consume each pellet type. After the preference test the remaining pellets were weighed and recorded. The retraining session was run the following day. Rats pressed the lever for pellets just as they had during training on the most recent RI schedule. The second test was run the next day, with the only difference being that each rat was sated on the other pellet type. Within each group, half the rats that were assigned to the left lever were tested in the “valued” condition first and “devalued” condition second, and likewise with the rats assigned to the right lever. Similarly, half the rats assigned to earn grain pellets were tested in the “valued” condition first, and likewise with the rats assigned to the Bio-Serv pellets.

Statistical analysis. The response rate-reward rate correlation was calculated for each animal by finding the best-fitting line relating reward rates to response rates, over training sessions and within training sessions. Between-session correlations were computed for each animal based on session means, and within-session correlations were computed by measuring mean lever-press and reward counts occurring in each 60-s bin from the beginning of the session. However, the last bin in each session was excluded from analysis due to the fact that the last bin almost never included a full 60 s worth of data (session time was determined by when a rat earned the last of 50 rewards, and often this occurred well before 60 s in the last bin fully elapsed). Student’s *t* tests evaluated differences in the slopes of these best-fitting lines, and also the correlation coefficients associated with the regression lines. For the pellet preference tests that followed the 5-min extinction tests, we calculated a preference score by dividing the amount of the nonsated pellet consumed by the combined amounts of the sated and nonsated pellets consumed.

Response rate data during devaluation test sessions and pellet intake data during the satiation periods were evaluated using the recommendations of Rodger (1974). Briefly, this approach treats factorial designs by repartitioning the sum of squares from the standard factorial analysis in order to perform separate one-way ANOVAs (using pooled error terms) to explore the effect of, for example, Independent Variable A at each level of Independent Variable B (similar to simple main effects tests). In addition, the analysis also consists of a main effect test of Independent Variable B. Significant omnibus *F* scores are then further examined with a set of v_1 mutually orthogonal post hoc contrasts to determine precisely where differences exist. This approach eliminates the interaction term from the linear model together with the problems associated with interaction tests (see Rodger, 1974), but, nevertheless, examines empirical interactions with various post hoc

contrasts. Type I error rate is defined as the proportion of true null contrasts rejected in error, and this is based on Rodger’s table of critical *F* values (Rodger, 1974). We adopted an $\alpha = .05$ criterion.

One rationale for using Rodger’s method is that it is more powerful than most ANOVA techniques at detecting true effects (Rodger, 1974; Rodger & Roberts, 2013), and with the sample sizes used here the power to detect medium to large effects = 0.95. We also provide a measure of effect size based on Perlman and Rasmussen’s (1975) uniformly minimum variance unbiased estimator of the noncentrality parameter, Δ . When no differences exist in the populations from which samples are drawn, $\Delta = 0$. However, $\Delta > 0$ when true population differences exist. Here we report these estimates whenever significant omnibus *F* scores were obtained. Although the methods we employed are more powerful than the standard factorial form of analysis, we also include a *supplemental section* in which we report the results of traditional factorial analyses applied to the critical devaluation test data in all of the experiments. The findings from these analyses do not differ from those reported in the main body of the text.

Results

We first analyzed data from the 10-day training period. For each rat the number of pellet rewards per minute and the number of lever presses per minute were averaged for each training session, and the group averages are presented in Figure 1A. These data show a more positive response rate–reward rate relationship in the changing than the constant group. To verify that the training manipulation was effective in inducing a stronger positive relationship between reward rates and response rates for the changing group, we used linear regression. The mean regression coefficients, which are the slopes of the best-fitting lines relating reward rates to response rates, were significantly higher for the changing group compared with the constant group ($\bar{x} (SEM) = 0.27(0.02)$ vs. $0.07(0.01)$, respectively, $t(14) = 7.26$, $p < .05$). The mean correlation coefficients (Pearson’s *r*) were also significantly higher for the changing group ($\bar{x} (SEM) = 0.93(0.02)$ vs. $0.62(0.06)$, respectively, $t(20) = 5.41$, $p < .05$). We also analyzed the within-session correlations between response rates and rewards rates (Figure 1B). Between-groups ANOVAs were performed on each session (pooled MS error = 0.12), and between-groups differences were discovered during Sessions 5 and 10 ($F^2 s(1,164) > 6.65$, $p's < .05$). The “changing” group maintained a higher within-session correlation during Session 5 (0.40 vs. 0.02) but a lower correlation during Session 10 (−0.25 vs. 0.12).

Next we analyzed data from the devaluation tests (Figure 1C). Separate between-groups ANOVAs using a pooled error term (MS error = 45.26) compared mean lever presses per minute on valued and devalued tests, while also testing for an overall main effect of value across groups. The two groups did not differ in the rate of responding during valued $F(1, 30) = 0.25$, $p > .05$, or devalued, $F(1, 30) = 0.14$, $p > .05$, test sessions. However, responding was significantly higher on valued versus devalued tests sessions, $F(1, 20) = 5.72$, MS error = 18.82, $\Delta = 4.15$, $p < .05$. These data indicate that both groups were equally goal-directed. To quantify the relationship between goal-directed control and the action–outcome rate correlation experienced during training, a correlation was measured between a devaluation effect score during the tests (lever press rates during valued–devalued tests) and the correla-

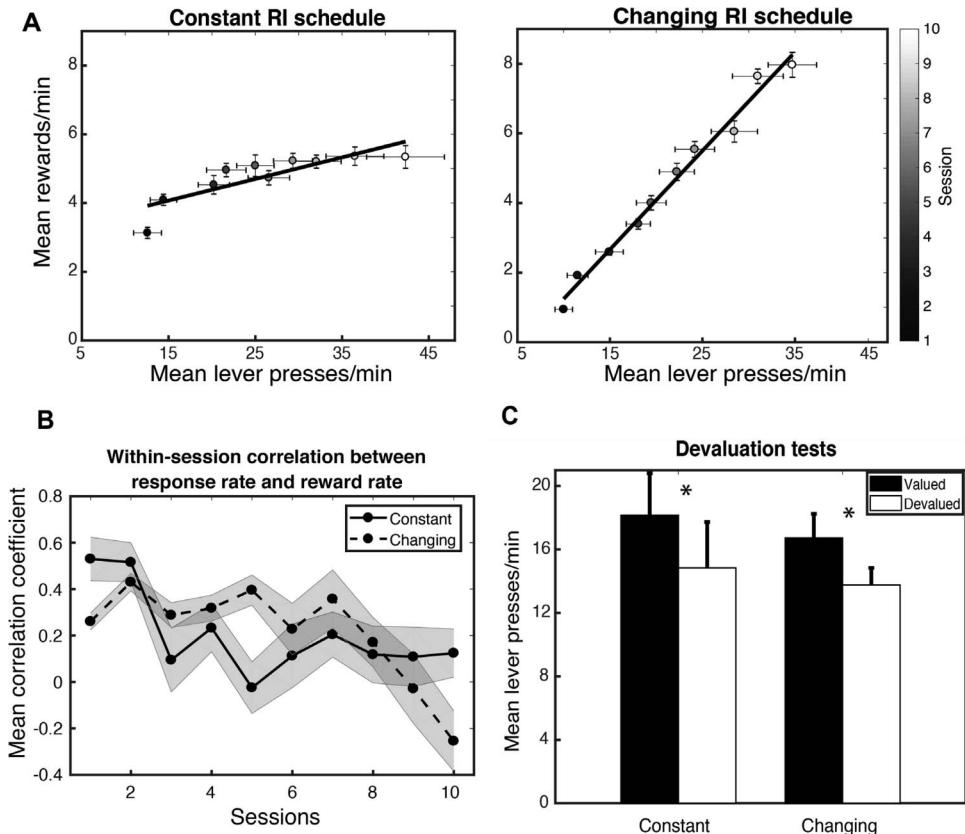


Figure 1. Data from Experiment 1. (A) Feedback functions from the 10-day training period showing the relationship between reward rates and response rates for the constant (left) and changing (right) groups. Data points represent group means. Black lines were fit to the group-averaged data using least-squares regression. (B) Within-session correlations between response rates and reward rates are plotted separately for each group. Shaded bounds are \pm SEM. (C) Data from the 5-min devaluation tests. Error bars are \pm SEM. * = statistically significant difference. RI = random interval.

tion between lever pressing rates and reward rates during the last session prior to testing (in 60-s bins). These correlations were not significantly different from zero for both training groups (see Table 1). This analysis indicates that instrumental performance during the devaluation tests cannot be reliably predicted by the action–outcome rate correlation recently experienced during instrumental training.

Data from the satiation period and pellet preference tests were also examined. The constant group consumed an average of 13.43 g (SEM: 0.86) when sated on their earned pellet type and 12.51 g (SEM = 1.19) when sated on the control pellet. The changing group consumed an average of 11.95 g (SEM = 0.52) when sated on their earned pellet type and 11.25 g (SEM = 1.01) when sated on the control pellet. An ANOVA on these data revealed that intake did not differ between earned and control pellet types, $F(1, 20) = 0.88$, MS error = 8.02, $p > .05$, and the groups did not differ in this regard (pooled $MSE = 9.15$; $F^2(1, 40) < 1.31$, $p's > .05$). Finally, during the preference tests the constant and changing group displayed a 98% and 93% preference for the nonsated pellet type, respectively. This was not a statistically significant difference, $t(20) = 1.33$, $p > .05$.

Discussion

The aim of Experiment 1 was to investigate whether the response rate–reward rate correlation experienced across training days influences the form of action control in tests of reward devaluation. Specifically, we predicted that a strong positive correlation would lead to goal-directed control while a weak correlation would lead to a habit. We found that, regardless of the experienced correlation, both groups of rats behaved in a goal-directed manner—suppressing responding when the anticipated outcome was devalued. A surprising result was that the “constant” group, having been trained on the same RI schedule for 10 sessions, did not develop a habit. It has consistently been found that even less extensive training on an RI schedule is sufficient for habit formation (e.g., Dickinson et al., 1983; Gremel et al., 2016; Gremel & Costa, 2013a, 2013b; Malvaez et al., 2018). However, every published report of habit formation in rodents trained on RI schedules has used leaner schedules than we employed with the “constant” group. Therefore, we sought to redo the experiment using leaner RI schedules.

Table 1
Correlation Coefficients Relating Goal-Directed Control During Devaluation Tests (Lever Rate Difference During Valued and Devalued Tests) and Within-Session Action-Outcome Rate Correlations During Instrumental Training

Devaluation test condition	Reinforcement schedule prior to start of testing				
	RI 6	RI 10	RI 28	RI 45	FI 45
Experiment 1, 10 day test	-.47	-.28			
Experiment 2, 10 day test			.02	-.04	
Experiment 3, 10 day test		-.23		-.44	
Experiment 3, 10 day test		.17		-.63*	
Experiment 4A, 2 day test				.25	.27
Experiment 4A, 10 day test				.24	.31
Experiment 4A, 20 day test				.49	.22
Experiment 4B, 2 day test				-.57*	
Experiment 4B, 20 day test				-.12	

Note. RI = random interval; FI = fixed interval. Goal-directed control is quantified as the difference between lever pressing rates during valued and devalued test sessions, and the action-outcome correlation is quantified as the within-session correlation between lever pressing rates and reward rates in 60-s bins during the training session just prior to the start of testing.

* $p < .05$.

Experiment 2

The aim of Experiment 2 was to examine the effect of varying the response rate-reward rate correlation across training sessions under relatively lean schedules of reinforcement. Different groups of rats were once again trained on RI schedules to respond for food pellets by pressing a lever, with one group being trained on an unchanging RI schedule to induce a weak response rate-reward rate correlation and another on an increasingly dense schedule to induce a strong positive correlation. Whereas in Experiment 1 the “constant” group was trained with an RI 10-s schedule, in the present study this group was trained on an RI 45-s schedule. On the basis of prior research, we expected this value to result in habitual responding after 10 training sessions (e.g., Lingawi & Balleine, 2012; Malvaez et al., 2018; Threlkill & Bouton, 2015). The question was whether against this background the “changing” group would display goal-directed responding. To achieve similar rates of reinforcement across days between these two groups, the changing RI group experienced mean intervals that ranged from 105 s to 28 s. Once again, we used sensory-specific satiety devaluation tests.

Method

Subjects. Eighteen naïve Long-Evans rats (12 males and six females) were housed in identical conditions as the rats in Experiment 1. The free-feeding body weights varied between 351 g and 600 g for males and 262 g and 325 g for females. All rats were maintained at 85% of free-feeding body weight for the duration of the experiment.

Apparatus. The apparatus was the same as in Experiment 1.

Procedure. Magazine training, CRF training, RI training, and devaluation testing proceeded as in Experiment 1 with the following exceptions. The constant group ($n = 9$, six males and three females) was trained on an RI 45-s schedule throughout. The

changing group ($n = 9$, six males and three females) was trained with the following set of values for each of the 10 sessions, in chronological order: 105, 85, 70, 58, 48, 40, 35, 31, 29, 28. Rats were preexposed to the novel pellet type only one time, which occurred after the final instrumental training session the day before the start of testing.

Results

As in Experiment 1, we used linear regression to verify that the training manipulation was effective in inducing a stronger positive relationship between reward rates and response rates for the “changing” group (Figure 2A). The mean slope of the regression line was significantly higher for the changing group compared with the constant group (\bar{x} (SEM) = 0.10(0.02) vs. 0.01(0.01), respectively, $t(16) = 4.33, p < .05$). The mean correlation was also significantly higher for the changing group (\bar{x} (SEM) = 0.88(0.04) vs. 0.27(0.06), respectively, $t(16) = 8.27, p < .05$). We also analyzed the within-session correlations between response rates and rewards rates in the same manner as Experiment 1 (Figure 2B). Between-groups ANOVAs were performed on each session (pooled MS error = 0.05), and one between-groups difference was discovered on Session 10, $F(1, 66) = 4.64, p < .05$, where the constant group displayed a higher within-session correlation than the changing group.

Next we analyzed data from the first set of devaluation tests (Figure 2C). Separate within-group ANOVAs (pooled MS error = 13.23) compared mean lever presses per minute on valued and devalued tests. In addition, we also tested for an overall main effect of group. There were no detectable differences in responding between valued and devalued tests for either the constant, $F(1, 16) = 0.00, p > .05$ or changing, $F(1, 16) = 1.30, p > .05$ groups. Groups did not differ in overall response rates, $F(1, 16) = 0.50$, MS error = 44.99, $p > .05$. These data indicate that neither group behaved in a goal-directed manner. We also measured the correlation between the devaluation effect score (valued–devalued lever rates) and the action–outcome correlation during training, and found these correlations do not differ significantly from zero for both groups of animals (see Table 1). This analysis once again indicates that instrumental performance during the devaluation tests cannot be reliably predicted by the action–outcome rate correlation recently experienced during instrumental training.

Data from the satiation periods and food preference tests were also examined. An ANOVA revealed that intake did not differ in those periods when rats were sated on the earned pellet type versus the control pellet (constant: \bar{x} (SEM) = 14.11(2.15) vs. 16.56(2.34); changing: \bar{x} (SEM) = 18.11(2.11) vs. 14.89(2.22); $F(1, 16) = 0.06$, MS error = 24.87, $p > .05$). During the preference test periods conducted following the extinction test sessions the constant group displayed an 83% preference for the nonsated pellet type while the changing group displayed a 98% preference. This unexpected between-groups difference was significant, $t(16) = 2.22, p < .05$.

Discussion

Like Experiment 1, the aim of Experiment 2 was to investigate whether the response rate–reward rate correlation experienced across training days influences the form of action control in tests

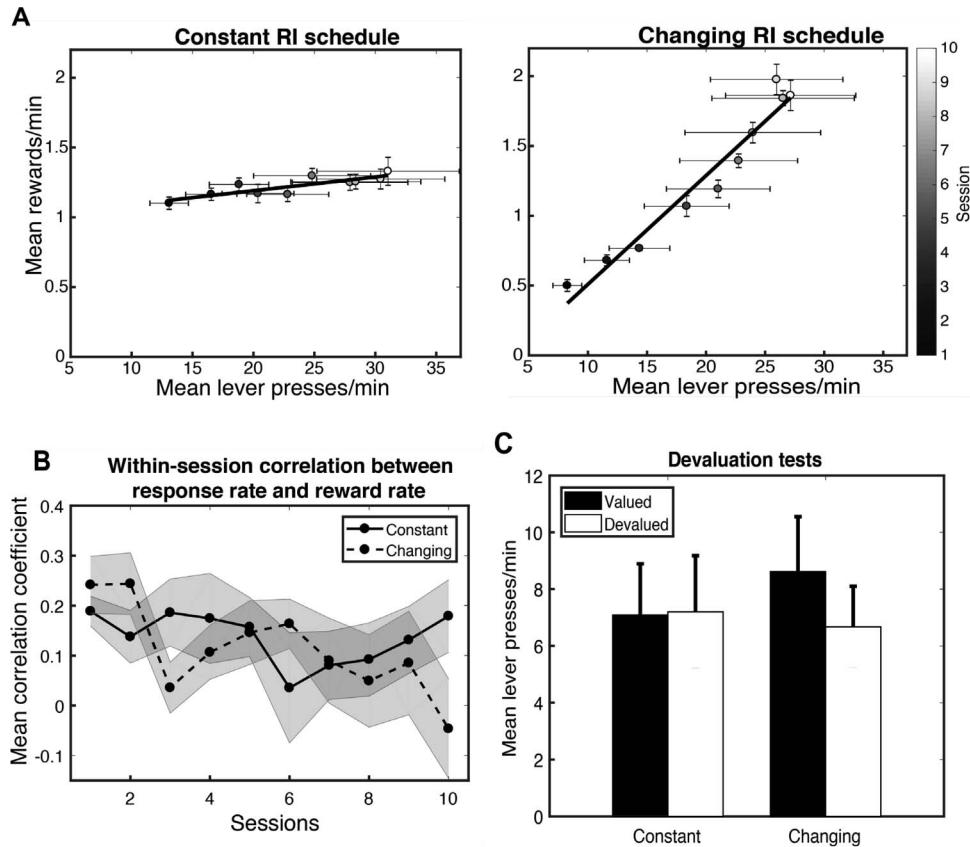


Figure 2. Data from Experiment 2. (A) Feedback functions from the 10-day training period showing the relationship between reward rates and response rates for the constant (left) and changing (right) groups. Data points represent group means. Black lines were fit to the group-averaged data using least-squares regression. (B) Within-session correlations between response rates and reward rates are plotted separately for each group. Shaded bounds are \pm SEM. (C) Data from the 5-min devaluation tests. Error bars are \pm SEM. RI = random interval.

of reward devaluation. The data collected from Experiment 1 revealed that rats responded in a goal-directed way in spite of the fact that our training procedures produced very different between-session response rate–reward rate correlations. Similarly, in Experiment 2 our training procedures produced differences in the response rate–reward rate correlations but rats responded in a seemingly habitual manner. Across both experiments the best predictor of performance in the reward devaluation tests was not the response rate–reward rate correlation, but schedule density.

The finding that dense RI schedules favor goal-directed behavior while lean RI schedules favor habits could be fundamental to understanding the psychological processes behind goal-directed control and habit formation, and needs to be further explored. However, the complex training history of our rats (i.e., experience with multiple types of RI schedules) makes it difficult to interpret those findings. Therefore, in Experiment 3 the goal was to directly assess the effect of schedule density on instrumental sensitivity to reward devaluation.

Experiment 3

The data from Experiments 1 and 2 provide preliminary evidence that schedule density is an important variable in determining

whether animals become goal-directed or habitual, independent of the correlation between response rates and reward rates. The aim of Experiment 3 was to more directly test the hypothesis that dense RI schedules promote goal-directed behavior while lean RI schedules promote habits. Two groups of rats were trained on either a relatively dense RI 10-s or a relatively lean RI 45-s schedule, and then put through tests of reward devaluation as in Experiments 1 and 2.

Another aim of the present study was to assess within-session response rate–reward rate correlations throughout training in a more fine-grained manner. In Experiments 1 and 2 we assessed within-session correlations by focusing on 60-s time bins. However, it is not known over what local interval rats might compute such correlations. In the present study, we collected data in a way that allowed us to assess these correlations over multiple time bins (10, 20, 40, 60, and 80 s).

Method

Subjects. Thirty-two naïve Long-Evans rats (16 male and 16 female) were housed in identical conditions as the rats in Experiments 1 and 2. The free-feeding body weights varied between 321 g and 498 g for males and 228 g and 294 g for females. The

experiment was run in two replications ($n = 16$ per replication, eight males and eight females).

Apparatus. The apparatus was the same as in Experiments 1 and 2.

Procedure. Magazine training, CRF training, RI training, and devaluation testing proceeded as in Experiments 1 and 2 with the following exceptions. In the first replication, one group ($n = 8$, four males and four females) was trained on an RI 45-s schedule and the other group ($n = 8$, four males and four females) was trained on an RI 10-s schedule, each for 10 daily sessions. Training sessions ended when rats earned 50 rewards. In the second replication, two groups were also trained on an RI 45-s ($n = 8$, four males and four females) or RI 10-s ($n = 8$, four males and four females) schedule, but for 20 daily sessions. Training sessions ended after either 38 or 9 min for RI 45 and RI 10 groups, respectively. These session lengths were implemented such that each group was expected to earn approximately 50 rewards per session, on average. A fixed session time was introduced in the second replication to ease the calculation of within-session correlations between response rates and reward rates. Devaluation tests were conducted in the same manner as Experiments 1 and 2 after 10 days of RI training. Rats in the second replication were also tested after 20 days of RI training. For two rats in the second replication (one in each group), the lever malfunctioned on Day 10 of training, and those rats' data were excluded from the first cycle of devaluation tests as well as analyses performed on the first 10 training sessions.

Results

By Day 10 of instrumental training, rats trained on the RI 10-s schedule achieved a higher mean rate of lever pressing compared with rats trained on the RI 45-s schedule (responses/min on Day 10 of training: \bar{x} (SEM) = 29.17(2.36) vs. 22.19(2.12), respectively, $t(30) = 2.20$, $p < .05$). Response rates did not differ after 20 days of training (responses/min: \bar{x} (SEM) = 34.82(3.42) vs. 26.77(3.01), $t(14) = 1.76$, $p > .05$). Reward rates were significantly higher for RI 10-s rats after 10 (rewards/min: \bar{x} (SEM) = 4.93(0.12) vs. 1.24(0.03), $t(28) = 30.57$, $p < .05$) and 20 (\bar{x} (SEM) = 5.56(0.31) vs. 1.34(0.07), $t(14) = 13.43$, $p < .05$) days of training.

As in Experiments 1 and 2, we examined the relationship between reward rates and response rates across training sessions by using linear regression (Figure 3A), and found that the mean regression coefficient was significantly higher for the RI 10-s rats compared with the RI 45-s rats for training Days 1 through 10 (\bar{x} (SEM) = 0.08(0.01) vs. 0.01(0.01), respectively, $t(28) = 5.95$, $p < .05$). The mean correlation between rewards rates and response rates across sessions was also significantly higher for the RI 10-s rats (\bar{x} (SEM) = 0.61(0.05) vs. 0.21(0.08), $t(28) = 4.39$, $p < .05$). For training Days 11 through 20, the mean regression coefficients did not differ between groups (\bar{x} (SEM) = 0.03(0.02) vs. 0.00(0.00), $t(14) = 1.61$, $p > .05$), but the mean correlation coefficients did (\bar{x} (SEM) = 0.30(0.11) vs. -0.10(0.12), $t(14) = 2.39$, $p < .05$).

We also computed within-session correlations for each rat in each training session by examining lever press and reward counts in successive time bins with widths ranging from 80 to 10 s, and then computing Pearson's r across all the bins in a session (Figure 3B). As in Experiments 1 and 2, the last time bin in each session

was excluded from the analysis when session length was not evenly divisible by bin width (e.g., in the case of a 9-min session, time cannot be evenly divided into 40-s bins and the residual 20 s must be left out). Due to the fact that only a subset of rats was trained beyond 10 days, the data from the first and second half of training were analyzed separately. In general, regardless of bin width, rats trained on the denser RI schedule experienced an initially higher response rate-reward rate correlation, but these correlations diminished to near zero levels and the between-groups difference disappeared by the end of the first half of training. The groups generally did not differ in the second half of training and their correlations were close to zero throughout. These data were analyzed by conducting between-groups ANOVAs on two-session blocks using pooled MS error terms. Regardless of how the data were binned, the RI 10-s group experienced a higher correlation compared with the RI 45-s group during the first block of training (F 's > 3.94, p 's < .05), but not during the last two blocks of the first half of training (F 's < 2.54, p 's > .05). There were no between-groups differences during the second half of training, except during the first block when data were organized into 10-s bins, $F(1, 28) = 6.79$, $p < .05$.

During the first set of devaluation tests (conducted after 10 training sessions, Figure 3C, left), rats trained on the RI 45-s schedule responded equally on the valued and devalued test sessions, $F(1, 28) = 0.71$, MS error = 15.59, $p > .05$, while rats trained on the RI 10-s schedule responded at a higher rate during the valued than devalued test, $F(1, 28) = 18.26$, MS error = 15.59, $\Delta = 15.96$, $p < .05$ and at an overall higher rate than the RI 45-s group, $F(1, 28) = 25.18$, MS error = 18.20, $\Delta = 22.38$, $p < .05$. Thus, a moderate amount of training on the denser RI schedule led to goal-directed lever pressing while training on the leaner RI schedule did not. Analysis of the correlations between the devaluation effect scores and the action-outcome correlations during training did not differ significantly from a correlation of zero for either training group (see Table 1). This analysis once again indicates that instrumental performance during the devaluation tests cannot be reliably predicted by the action-outcome rate correlation recently experienced during instrumental training.

During the second set of devaluation tests (conducted after 20 days of training, Figure 3C, right) both groups responded at a higher rate during the valued versus devalued condition, $F_{s}(1, 14) > 5.19$, MS error = 15.97, $\Delta = 3.45$, $p < .05$. There was no overall difference in the rate of responding between groups, $F(1, 14) = 4.25$, $p > .05$. Thus, an extensive amount of training led to goal-directed lever pressing regardless of the density of the reinforcement schedule. The correlations between the devaluation effects and the action-outcome correlations during training did not differ significantly from zero in the case of the RI 10-s group, but this correlation was significantly *negative* for the RI 45-s group (see Table 1). The significance of this correlation may be spurious, because we did not obtain significant correlations in other cases where rats were trained for 20 sessions on an RI 45-s schedule (see Experiments 4A and 4B). Moreover, the sign of the correlation coefficient is opposite to what would be expected if the experienced action-outcome correlation determines goal-directed control.

Data from the satiation periods and food preference tests were also examined. An ANOVA on the consumption data from the first set of tests revealed that animals consumed as much of their earned

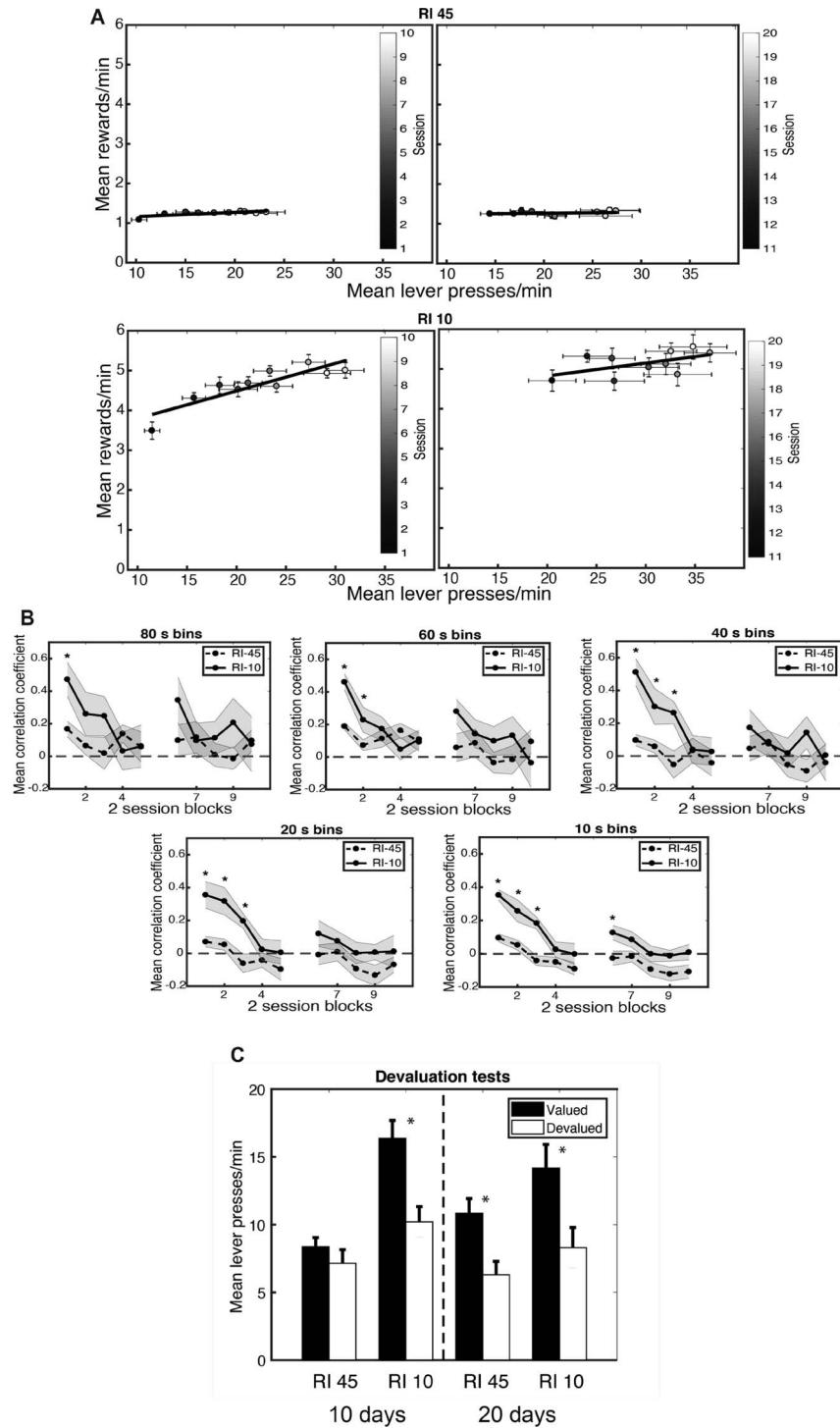


Figure 3. Data from Experiment 3. (A) Feedback functions for rats trained on an RI 45-s (top) or RI 10-s (bottom) schedule. Data points are plotted separately for the first (left) and last (right) half of training. Black lines were fit to the group-averaged data using least-squares regression. (B) Within-session correlations between response rates and reward rates are plotted separately for each group in two-session blocks. Within each graph, the gap separating Blocks 1–5 and 6–10 is meant to depict the time of the first devaluation test cycle. Graphs differ according to the bin width used to calculate the correlation coefficients (10–80 s). Shaded bounds are $\pm SEM$. (C) Data from the 5-min devaluation tests. Error bars are $\pm SEM$. * = statistically significant difference. RI = random interval.

pellet type as the control pellet type during the satiation period (RI 45: \bar{x} (SEM) = 13.63(1.09) vs. 12.60(1.09); RI 10: \bar{x} (SEM) = 14.33(1.09) vs. 14.87(1.42); $F(1, 28) = 0.06$, MS error = 15.53, $p > .05$). There was also no difference in consumption during the second set of tests (RI 45: \bar{x} (SEM) = 13.00(1.30) vs. 12.25(1.33); RI 10: \bar{x} (SEM) = 13.00(0.68) vs. 13.13(1.37); $F(1, 14) = 0.13$, MS error = 6.23, $p > .05$). During the first set of preference tests the RI 45-s group displayed a 91% preference for the nonsated pellet type while the RI 10-s group displayed a 97% preference, a difference that was not statistically reliable, $t(28) = 2.04$, $p > .05$. Groups also did not differ in preference for the nonsated pellet type in the second set of preference tests (97% vs. 98% for RI 45 and RI 10, respectively, $t(14) = 0.52$, $p > .05$).

Discussion

The main aim of Experiment 3 was to directly assess the effect of schedule density on instrumental sensitivity to reward devaluation. Consistent with our preliminary results from Experiment 2, following 10 sessions of RI training, rats trained on a dense schedule (RI 10-s) were goal-directed while rats trained on a lean schedule (RI 45-s) were not. Following 20 sessions of training, however, both groups displayed goal-directed responding. Generally, denser RI schedules can create a somewhat stronger response rate-reward rate correlation than leaner RI schedules, because early in training animals are apt to miss a substantial number of scheduled reinforcers on dense schedules due to their low level of responding (see also Baum, 1973). However, once responding becomes more regular (i.e., the interresponse times become less variable) the correlation between reward rates and response rates should diminish on an RI schedule. We observed precisely that pattern during training. Specifically, while the between-session correlation was moderately high for rats trained on the RI 10-s schedule during the first 10 sessions of training ($\bar{x} = 0.61$), it diminished during Sessions 11 through 20 ($\bar{x} = 0.30$). When the correlations were calculated within-session, a similar pattern emerged: The correlations were moderately high for the RI 10-s group but steeply diminished over training, reaching near zero levels by the time both devaluation tests commenced.

Thus, by the time the first set of devaluation tests were conducted, a similarly weak within-session relationship between response rates and reward rates in both groups was expected to give rise to habitual responding. Yet, the first set of devaluation tests revealed that only RI 45-s rats failed to develop goal-directed control. It is possible that, despite experiencing a weak within-session correlation by the end of the first half of training, the behavior of the RI 10-s group during the first devaluation tests was influenced by experiencing of a relatively high correlation earlier in training. To test this possibility, some rats continued RI training for an additional 10 sessions during which the correlations were maintained near zero. Surprisingly, we found that both groups displayed goal-directed responding, counter to the prediction that experience with a low response rate-reward rate correlation should give rise to habits. Together with the results of Experiments 1 and 2, we therefore think that the data are not easily explained by the response rate-reward rate correlation idea (Dickinson, 1985).

One notable finding from Experiment 3 is that the density of the reinforcement schedule influences how soon goal-directed control emerges over training. Why does the density of the reinforcement

schedule influence how quickly goal-directed responding emerges? One possible answer lies in the fact that animals trained on a dense RI schedule are more certain about the timing of rewards than animals trained on a lean RI schedule. This follows from the fact that on dense schedules the distribution of IIRs is less variable than on lean schedules. A second possibility is that the action-outcome contiguity is more favorable on dense than lean RI schedules.

DeRusso et al. (2010) provided evidence in support of these possibilities. Mice were trained either on equally dense FI or RI schedules and then tested in extinction following a selective satiety procedure. When tested after 2 days of training on these interval schedules, both groups were goal-directed, but when tested after 8 days of training the FI mice remained goal-directed while the RI mice were not. In addition, DeRusso et al. (2010) determined that the average time between any given response and the upcoming reward was shorter in FI than RI animals. Thus, either greater action-outcome contiguity or reduced temporal uncertainty during FI training could explain more persistent goal-directed responding.

The data collected from Experiment 3 are also in agreement with the idea that reduced temporal uncertainty and/or favorable action-outcome contiguity promote goal-directed behavior while high temporal uncertainty and/or poor action-outcome contiguity do not. However, the RI 10-s and 45-s schedules also differed in the average rate of reward. Therefore, we sought to replicate the finding from DeRusso et al. (2010) by training rats either on an FI or an RI schedule of the same density. We also sought to replicate the finding from Experiment 3 that extensive training on an RI schedule leads to goal-directed responding—something that we did not anticipate.

Experiment 4

The aim of Experiment 4 was to test whether low temporal uncertainty of outcomes and favorable contiguity between actions and outcomes promote the early emergence of goal-directed responding. The experiment was conducted in a similar manner to that of DeRusso et al. (2010) but with rats instead of mice and with additional tests conducted after more extensive instrumental training. Two groups were trained to respond on either an FI or RI schedule, both of the same reward density (45 s). In Experiment 4A, devaluation tests were then conducted after two, 10, and 20 sessions to understand how action control might change within groups over the course of training. DeRusso et al. (2010) observed that animals trained on an RI schedule developed habits within 8 days of training, while their FI counterparts remained goal-directed. It is possible that additional training on the FI schedule might also result in habitual responding. In Experiment 4B, we trained rats on an RI 45-s schedule for either two or 20 sessions before conducting devaluation tests in a between-groups replication.

Experiment 4A

Method

Subjects. Sixteen naïve Long-Evans rats (eight male and eight female) were housed in identical conditions as the rats in Experiments 1, 2, and 3. The free-feeding body weights varied between 368 g and 498 g for males and 217 g and 357 g for females.

Apparatus. The apparatus was the same as in Experiments 1, 2, and 3.

Procedure. Magazine and CRF training proceeded as in Experiments 1, 2, and 3. Following CRF training, rats were trained to press the lever on either an RI 45-s ($n = 8$, four males and four females) or FI 45-s schedule ($n = 8$, four males and four females). The FI schedule was configured so that one pellet was available every 45 s. If a pellet was made available, it remained available until the lever was pressed. If two or more 45-s cycles elapsed without a lever press, only one pellet was set up for delivery. The RI schedule was configured in the same way as Experiments 1, 2, and 3 (see Experiment 1 Method section for details). The sessions ended and the lever was withdrawn once 38 min elapsed. Rats were trained one session per day for 20 days. Three devaluation test cycles were conducted at different points during interval training: after two, 10, and 20 sessions. The devaluation cycles were conducted as in Experiments 1, 2, and 3.

Results

Overall response rates in both groups of rats increased over the 20 days of training. By the end of training RI rats responded at a higher rate than FI rats ($\bar{x} (SEM) = 31.42(1.50)$ and $17.38(1.96)$ lever presses per minute on Day 20, respectively, $t(14) = 5.68$, $p < .05$). Groups did not differ in mean reward rates by the end of training ($\bar{x} (SEM) = 1.31(0.04)$ vs. $1.32(0.00)$ rewards/min for RI and FI, respectively, $t(14) = 0.07$, $p > .05$). To estimate how temporal uncertainty of rewards affected responding across training sessions, we plotted response rates second-by-second across the IRI for both groups (Figure 4A) separately for Days 2, 10, and 20 of training. We restricted our analysis to only those IRIs that were 45 s or longer to avoid the contaminating effects of reward delivery within a 45-s period.

On Day 2 of interval training, both groups showed randomly fluctuating rates of responding that hovered around a low mean rate (Figure 4A, light gray). By Sessions 10 and 20, rats trained on the FI schedule displayed scalloping behavior (Figure 4A, right) while rats trained on the RI schedule continued to show randomly fluctuating rates of responding, albeit at an overall higher rate (Figure 4A, left). To quantify the magnitude of scalloping in the IRI on training Days 2, 10, and 20, we calculated a ratio between the mean rate of responding in Seconds 31 to 45 and Seconds 2 to 16 for each rat. We excluded responses occurring in the first two seconds because we assumed rats to be in the magazine consuming pellets during that time. One-way repeated measures ANOVAs (pooled MS error = 110.18) revealed significant differences over training in the mean ratios for FI rats, $F(2, 28) = 12.54$, $\Delta = 21.29$, $p < .05$, but not for RI rats, $F(2, 28) = 0.00$, $p > .05$. Post hoc contrasts revealed that for FI rats the mean ratio was smallest on Day 2, then Day 10, and largest on Day 20 ($\bar{x} (SEM) = 1.04(0.10)$, $8.42(1.46)$, and $26.57(9.02)$ for Days 2, 10, and 20, respectively). For RI rats the mean ratio varied nonsignificantly between 0.72 ($SEM = 0.07$) and 0.91 ($SEM = 0.04$).

We also analyzed the within-session correlations in the same manner as Experiment 3, once again binning lever press and reward counts in bins ranging from 80 to 10 s (Figure 4B). The correlations, when computed from data in 80-, 60-, and 40-s bins, were higher for the RI group compared with the FI group on training Days 10 and 20, $F_{s}(1,42) > 4.23$, MS error = 0.03,

$\Delta > 3.03$, $p < .05$. When analyzing the correlations from bins of 20 and 10 s, the groups differed only on training Day 10, $F_{s}(1,42) > 6.84$, MS error = 0.03, $\Delta > 5.51$, $p < .05$. In addition, the between-session correlation did not differ between groups (RI $\bar{x} (SEM) = 0.24(0.11)$, FI $\bar{x} (SEM) = 0.15(0.10)$; $t(14) = 0.62$, $p > .05$).

We also obtained a measure of action–outcome contiguity. For this measure, with each occurrence of a lever press we determined the time to the next reward delivery. This value was averaged across lever presses. This provides a measure of how much time elapses, on average, between each response and the next reward. The mean action–outcome contiguity score was calculated for training Days 2, 10, and 20 for each rat in each group (Figure 4C). Overall, RI rats experienced less favorable action–outcome contiguity (i.e., longer times between lever presses and rewards) than FI rats, $F(1, 14) = 794.24$, MS error = 12.60, $\Delta = 679.78$, $p < .05$. One-way repeated measures ANOVAs (pooled MS error = 42.35) revealed significant differences over training in FI rats, $F(2, 28) = 4.88$, $\Delta = 7.06$, $p < .05$, but not in RI rats, $F(2, 28) = 1.22$, $p < .05$. Post hoc contrasts for FI rats revealed that the mean action–outcome contiguity score was greater (i.e., poorer) on Day 2 than on Days 10 and 20, which did not differ.

Devaluation test cycles were conducted after 2, 10, and 20 days of instrumental training (Figure 4D). One-way repeated measures ANOVAs (pooled MS error = 17.42) were performed on each group. This analysis included valued and devalued test sessions for all test cycles. Significant differences across these conditions were observed in FI rats, $F(5, 70) = 12.40$, $\Delta = 55.23$, $p < .05$ and also in RI rats, $F(5, 70) = 2.67$, $\Delta = 7.97$, $p < .05$, but there was no overall difference in response rates between the two groups, $F(1, 14) = 0.61$, MS error = 37.02, $p > .05$. Post hoc contrasts performed on each group revealed a significant devaluation effect after 10 and 20 days of training in FI rats. RI rats displayed a significant devaluation effect after 20 days, though this was somewhat smaller than in RI rats. The correlations between the devaluation effects and the action–outcome correlations during training, computed separately for each set of tests, did not significantly differ from a correlation of zero (see Table 1). This analysis once again indicates that instrumental performance during the devaluation tests cannot be reliably predicted by the action–outcome rate correlation recently experienced during instrumental training.

The amount of pellets consumed during the 1-hr satiation periods was also analyzed. The means and SEM's for the FI group during each test cycle were as follows, for consumption of earned versus control pellet types: $11.88(1.09)$ versus $10.75(0.62)$, $14.88(1.29)$ versus $14.63(2.24)$, $16.75(1.36)$ versus $15.88(2.22)$. The means and SEM's for the RI group during each test cycle were as follows: $13.88(1.51)$ versus $11.75(0.96)$, $16.38(1.19)$ versus $15.75(1.01)$, $18.13(1.55)$ versus $15.13(2.04)$. One-way repeated measures ANOVAs were performed on each group's consumption data that included the amount of each pellet type consumed for each test cycle (pooled MS error = 16.18). This analysis revealed significant differences within each group (RI: $F(5, 70) = 2.36$, $\Delta = 6.46$, $p < .05$; FI: $F(5, 70) = 2.69$, $\Delta = 8.07$, $p < .05$). Post hoc contrasts revealed that, while both groups showed an increase in overall consumption across test cycles, neither group consumed more of the earned versus control pellet type within a given test cycle.

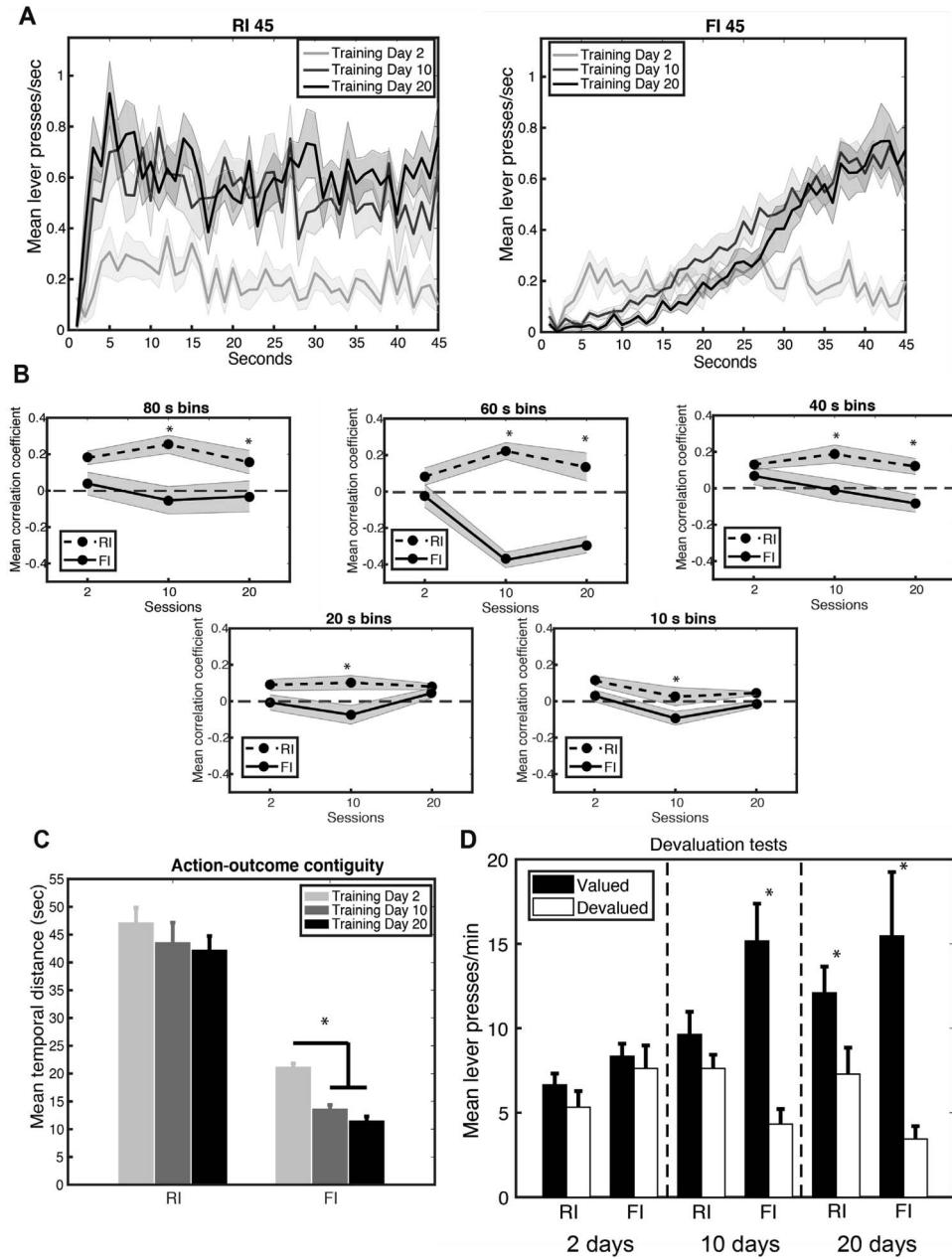


Figure 4. Data from Experiment 4A. (A) Average response rates during interreward intervals that were greater than or equal to 45 s, plotted separately for Sessions 2, 10, and 20. Data are plotted separately for groups trained on either an RI 45-s (left) or FI 45-s (right) schedule. The zero-point on the x-axis represents time of the most recent reward. Shadings around the lines represent \pm SEM. (B) Within-session correlations between response rates and reward rates, calculated over bins ranging from 10 to 80 s. Shaded bounds are \pm SEM. (C) Action–outcome contiguity for training Days 2, 10, and 20. Action–outcome contiguity is defined as the mean time between lever presses and rewards. (D) Data from devaluation tests after 2, 10, and 20 days of training. Error bars are \pm SEM. * = statistically significant difference. RI = random interval; FI = fixed interval.

During the food preference tests from each of the three devaluation test cycles, FI rats displayed 92%, 94%, and 89% preference for the nonsated pellet type while RI rats displayed 98%, 98%, and 93% preference, respectively. One-way repeated measures ANOVAs (pooled MS error = 0.01) performed

on each group's preference scores did not reveal any significant differences; RI: $F(2, 28) = 0.01, p > .05$; FI: $F(2, 28) = 0.01, p > .05$, and, further, the two groups displayed equally high preferences for the nonsated pellet type, $F(1, 14) = 3.37$, MS error = 0.01, $p > .05$.

Experiment 4B

The unexpected finding we observed in both Experiments 3 and 4A was the emergence of goal-directed control with extensive training on an RI schedule. Rats in Experiments 3 and 4A underwent multiple cycles of extinction tests, and it is possible that extensive experience with these repeated devaluation tests, not extensive training per se, may account for this finding. The aim of Experiment 4B was to replicate these results but using a between-subjects design to avoid multiple devaluation test cycles. Two groups of rats were trained on the same RI 45-s schedule as that used in Experiment 4A. One group was tested after two sessions of RI training, while the other group was tested after 20 sessions.

Method

Subjects. Thirty-two naïve Long-Evans rats (16 male and 16 female) were housed in identical conditions as the rats in the previous experiments. The free-feeding body weights varied between 536 g and 769 g for males and 255 g and 338 g for females. The experiment was run in two replications.

Apparatus. The apparatus was the same as in the previous experiments.

Procedure. Magazine training, CRF training, RI training, and devaluation test sessions proceeded as in the previous experiments with one exception. Rats were trained to press the lever on an RI 45-s schedule for either 2 ($n = 16$, eight males and eight females) or 20 ($n = 16$, seven males and eight females) days and then received one cycle of devaluation tests thereafter. The limited training group began CRF on the extensively trained group's 17th session so that testing was conducted on the same days for both groups. One rat failed to learn during CRF training and was dropped from the remainders of the experiment. The experiment was run in two replications, and data were combined across replications as there were no differences across replications.

Results

Response rates during devaluation tests were analyzed with separate one-way repeated measures ANOVAs on each group (pooled MS error = 10.26). During the devaluation tests (see Figure 5) rats trained for two sessions responded equally on valued and devalued tests, $F(1, 29) = 1.53, p > .05$, while rats trained for 20 sessions, once again, responded at a higher rate during the valued compared with the devalued tests, $F(1, 29) = 16.47, \Delta = 14.33, p < .05$, and at an overall higher rate than the two-session group, $F(1, 29) = 13.61$, MS error = 25.66, $\Delta = 11.67, p < .05$. The correlations between the devaluation effect scores and the within-session action–outcome correlations during the final day of training were not significantly different from zero for the extensively trained group, but were significantly negative for the limited training group (see Table 1). Note that the significantly negative correlation did not appear during Experiment 4A, and the direction of the correlation is opposite to what would be expected if the strength of the action–outcome rate correlation determines goal-directed control. The within-session correlations (based on a 60-s time bin) on the final day of training were weak and did not differ between the groups (r 's = 0.16).

The amount of pellets consumed during the 1-hr satiation period was analyzed with separate repeated measures one-way ANOVAs

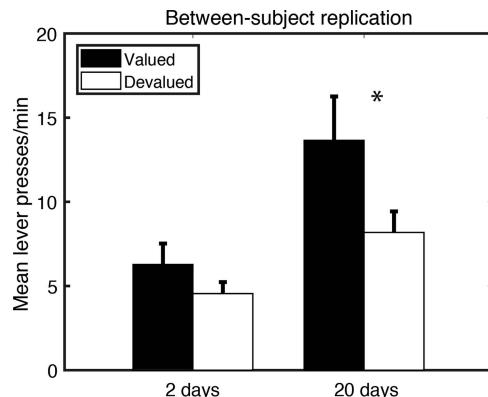


Figure 5. Data from Experiment 4B, in which devaluation tests were conducted after 2 or 20 days of training on an RI 45-s schedule. Error bars are \pm SEM. * = statistically significant difference.

(pooled MS error = 8.51). Consumption did not differ between the earned and control pellet types for either the limited training group (\bar{x} (SEM) = 10.94(0.96) vs. 10.94(0.84); $F(1, 29) = 0.00, p > .05$) or the extensively trained group (\bar{x} (SEM) = 11.80(1.08) vs. 10.33(0.67); $F(1, 29) = 1.90, p > .05$), and overall consumption did not differ between groups, $F(1, 29) = 0.02$, MS error = 16.78, $p > .05$. During the preference tests, the limited training group displayed a 90% preference for the nonsated pellet type while the extensively trained group displayed a 95% preference, a difference that was not reliable, $t(29) = 1.17, p > .05$.

Discussion

In Experiment 4A, we found that rats trained on an FI schedule did not show evidence of goal-directed control after 2 days of training, but then became goal-directed after 10 days of training and remained goal-directed after 20 days of training. Rats trained on an RI schedule also did not appear to be goal-directed after 2 or 10 days of training, but did appear so after 20 days of training. Because of the importance of this finding, seen also in Experiment 3, we sought to replicate the effect in Experiment 4B but using a between-groups comparison. The results of that study confirmed the observation, and showed that it was not due to multiple devaluation test cycles. Further discussion of these findings will be deferred until the general discussion section, but here we point out that we demonstrate with rats an effect originally reported by DeRusso et al. (2010) with mice that goal-directed responding is more likely to develop under FI than RI schedules. In other words, interval schedules do not inevitably lead to habitual control, but do seem to affect responding in more nuanced ways than have typically been considered in the literature.

General Discussion

We conducted four experiments to assess the hypothesis that goal-directed control on interval schedules depends on the experienced correlation between response rates and reward rates (Dickinson, 1985). In Experiment 1, we successfully induced a strong positive response rate-reward rate correlation in a group of rats trained on RI schedules by systematically increasing the reward

density of the schedules across sessions. Another group was maintained on an RI schedule with a constant density and thus experienced a weak correlation across sessions. In tests of selective satiety, we found that both groups displayed goal-directed control. In Experiment 2, we once again induced strong and weak response rate–reward rate correlations in two groups of rats, but using RI schedules with lower reward densities. In contrast to Experiment 1, neither group in Experiment 2 displayed goal-directed responding. In Experiment 3, we trained two groups of rats on RI schedules of different reward densities to more directly assess the impact of this variable, and found that only the group trained on the denser schedule showed goal-directed control after 10 days of training, while goal-directed responding was evident in both groups after 20 days of training. Finally, in Experiment 4 we held constant the schedule density but manipulated the randomness of the interval schedule by training rats either on an RI 45-s or an FI 45-s schedule. We found that, while responding became goal-directed after 10 and 20 days of training on the FI schedule, rats trained on the RI schedule once again became goal-directed, albeit to a lesser degree than on the FI, but only after 20 days of training. This result was obtained using both within- and between-groups experimental designs.

The results from the present experiments cast doubt on the response rate–reward rate correlation hypothesis of action control. A key prediction is that a high correlation between response rates and reward rates should promote goal-directed responding, while a weak correlation should promote habits. The results of Experiments 1 and 2 do not support this idea because in both experiments one group was trained with a strong and one with a weak between-session response–reward correlation, yet goal-directed control was only observed in Experiment 1 but not Experiment 2. It could be argued that animals had difficulty experiencing the between-session response–reward correlation in the changing RI schedules in these studies because the schedules differed rather little from day to day toward the end of training. Nonetheless, our manipulation produced clear differences in the obtained between-session response–reward correlations in both studies, and the results demonstrate that this variable was not critical. However, it may be that manipulating the within-session correlation would produce effects more consistent with the correlation idea, but we have other reasons for questioning that notion.

The data from all of our experiments provided no evidence to support the idea that within-session response–reward correlations can predict goal-directed control. Somewhat ironically, in Experiment 3 the group trained on an RI 10-s schedule showed a higher between-session correlation and also displayed superior goal-directed responding compared with the group trained on an RI 45-s schedule. However, the RI 45-s rats eventually also displayed goal-directed responding despite the between-session correlation remaining very low (see *Figure 3A*). More problematic is that in Experiment 3, despite the fact that the within-session correlations diminished to near zero levels for both groups by the end of training, both groups displayed clear goal-directed responding. This pattern of results cannot be explained by appeal to either the between- or within-session correlation idea.

The data from Experiment 4A are also problematic in that the response rate–reward rate correlation for the FI group was either equal to or lower than that of the RI group, whether computed between or within sessions. Yet, the FI group was quicker to

develop goal-directed control. It is nevertheless conceivable that the animals might compute the correlation only between consecutive rewards—that is, during the IRI. Animals typically increase their rate of responding as the time of reward nears (Ferster & Skinner, 1957; *Figure 4A*), and, therefore, the within-session correlation could become positive for animals trained on an FI schedule due to the reward period being associated with high rates of responding. While this way of computing the correlation could explain superior goal-directed control in FI but not RI rats following 10 days of training, it cannot explain the emergence of goal-directed control after 20 days of RI training. It also raises the problem of determining when different rules should be applied for computing the correlation.

The above between- and within-session response rate–reward rate correlation analyses were all based on group-averaged devaluation data. We also examined whether the within-session correlations on the final session of training could be used to predict variations across individual subjects in the magnitude of the devaluation effect (see *Table 1*). We never observed this correlation to be significantly positive across any of the experiments (for a total of 16 test conditions), and in only two cases was the correlation shown to be significantly different from zero, in the negative direction. It may be argued that this measure was insensitive because it was based on a 60-s time base used to compute the within-session response rate–reward rate correlation on the final day of training. However, for Experiments 3 and 4 we were able to examine this using a 10-s time base, and when we did we found no correlation to be significantly different from zero in any of the tests conducted.

Collectively, the data we report cannot easily be explained by Dickinson's (1985) response rate–reward rate correlation hypothesis. If the correlation hypothesis does not explain the data from the current set of experiments, what does? We next consider two other candidate mediators of action control: temporal uncertainty of outcomes and action–outcome contiguity.

DeRusso et al. (2010) suggested that temporal uncertainty of the instrumental outcome and/or action–outcome contiguity may affect action–outcome learning. Specifically, they suggested that smaller uncertainty and/or better contiguity promotes goal-directed control and greater uncertainty and/or poorer contiguity promotes habits. They observed that mice trained on an FI, but not RI, schedule appeared goal-directed after a moderate amount of training, and we replicated this result with rats. The idea can also help explain why the overall reward density of an RI schedule influences goal-directed control. By definition, on RI schedules temporal uncertainty increases and action–outcome contiguity becomes less favorable as the schedule becomes less dense. The mechanism that would allow for temporal uncertainty of outcomes to affect action–outcome learning is not immediately clear. One clue, perhaps, is that animals may experience more stress with greater temporal uncertainty (T. Robbins, personal communication, November 13, 2017). Other research has shown that stress facilitates habit formation (Schwabe, Dickinson, & Wolf, 2011), but future research would need to determine if varying the temporal uncertainty of instrumental outcomes induces variable levels of stress.

The implication that reduced temporal uncertainty might promote the early emergence of goal-directed control, however, is seemingly opposed to a recent suggestion by Thralkill, Trask,

Vidal, Alcalá, and Bouton (2018). These authors argued that reduced reward uncertainty actually promotes habits, whereas increased reward uncertainty promotes goal-directed control. Thralkill et al. (2018) observed greater goal-directed control of discriminated lever pressing in rats when there was uncertainty as to whether an outcome could be earned on any given trial. In contrast, rats responded habitually when an outcome could be earned on every trial. It is notable that in both of these conditions, rats were trained on a random interval schedule, so there still remained a high degree of temporal uncertainty within each trial. Thus, *temporal* uncertainty and *trial* uncertainty may affect behavior in different ways. In addition, while action–outcome contiguity differed in our FI and RI schedules (and, thus, was confounded with uncertainty), it is unlikely that such contiguity differences would occur in the procedures used by Thralkill et al. (2018). In addition, the current study DeRusso et al. (2010) and used free operant as opposed to Thralkill's et al. (2018) discriminated operant procedures. There is precedent for thinking that the mechanisms of learning differ between hierarchical tasks, such as discriminated operant situations, and simpler tasks such as free operant learning (Colwill & Rescorla, 1990; Holland, 1985; Rescorla, 1985).

One aspect of our results in Experiments 3 and 4, however, was not consistent with temporal uncertainty as a predictor of goal-directed or habitual responding. Despite being trained on an RI 45-s schedule, rats eventually became goal-directed after 20 sessions of training. We observed this three times in the present series of experiments. There is no reason to think that temporal uncertainty of the reward changed over training on this schedule, and,

thus, these findings challenge the temporal uncertainty idea because it is not clear why the amount of training should matter. The data, however, are not inconsistent with a role for action–outcome contiguity.

DeRusso et al. (2010) also observed that goal-directed mice trained on an FI schedule displayed more favorable action–outcome contiguity compared with habitual mice trained on an RI schedule, a result that we also confirmed with rats (Figures 4C and 4D). Moreover, the mean action–outcome contiguity is expected to be better on dense versus lean RI schedules. Other research has confirmed that when explicitly manipulating action–outcome contiguity by either delivering reward immediately or delaying reward by 20 s following a response on a CRF schedule, goal-directed control was more evident in the former case (Urcelay & Jonkman, 2019). Others have also pointed to potential differences in goal-directed learning as a function of action–outcome contiguity (Balleine, 1992; Balleine, Garner, Gonzalez, & Dickinson, 1995; Balleine, Paredes-Olay, & Dickinson, 2005; Corbit & Balleine, 2003; Killcross & Coutureau, 2003).

The one result needing further explanation was the observation that goal-directed control emerges with extensive training on a lean RI schedule. To explain this result, we hypothesize that action–outcome contiguity determines the rate of action–outcome learning, but that this learning also depends on the cumulative number of action–outcome pairings (see Figure 6). Under schedules that produce greater average temporal distances between actions and outcomes (poor contiguity), the growth in the action–outcome association will be relatively slow, while under condi-

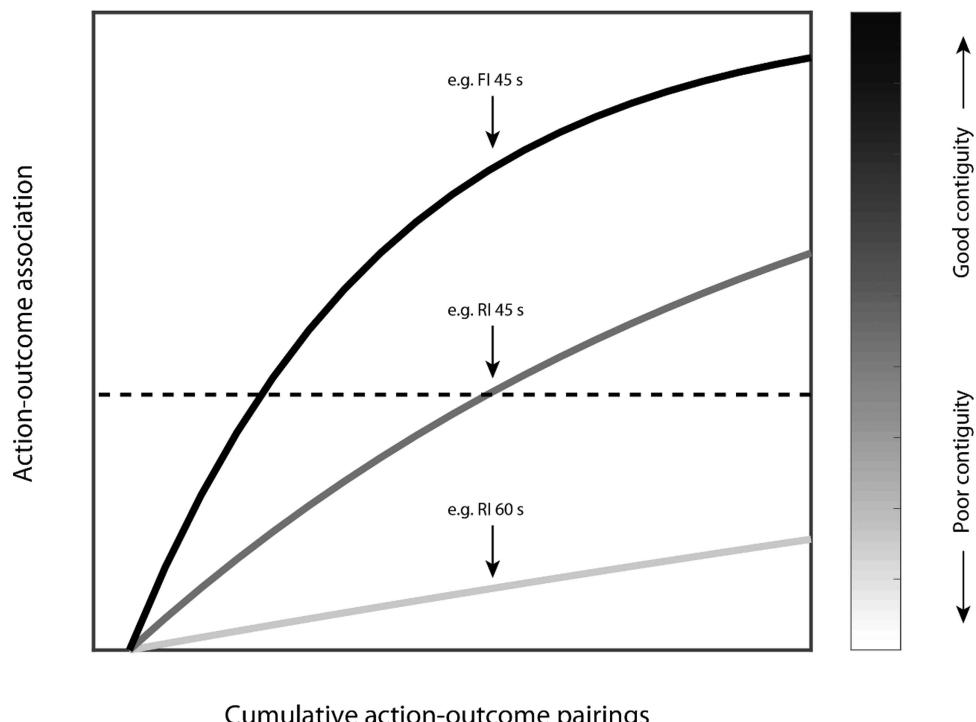


Figure 6. Theoretical relationship between the learned action–outcome association, cumulative action–outcome pairings, and average action–outcome contiguity. The average contiguity between actions and outcomes modulates the rate at which an action becomes goal-directed, as represented by the point on the x-axis at which the action–outcome association surpasses a theoretical threshold (dashed line). RI = random interval; FI = fixed interval.

tions of favorable contiguity the growth in the action–outcome association will be fast. Furthermore, if the strength of the action–outcome association must exceed a threshold before outcome devaluation effects can be empirically observed, then on relatively lean interval schedules we expect no devaluation effects to be observed after minimal training, a demonstrable devaluation effect after moderate training on an FI schedule or dense RI schedule, and the eventual emergence of a relatively weak devaluation effect after extensive training on a lean RI schedule. This framework accurately captures all of the results from our experiments. Consistent with this framework, [Shipman, Trask, Bouton, and Green \(2018\)](#) observed maintained goal-directed control after 24 sessions of training on a RI 30-s schedule—a value that is intermediate compared with the values in our RI schedules (10 s and 45 s).

It should be noted that although we emphasize the role of temporal contiguity between actions and outcomes in governing the rate of goal-sensitive learning, we do not imply that this variable is the only one that influences such learning. Earlier research has shown that the action–outcome contingency affects learning even when temporal contiguity is held constant (e.g., [Colwill & Rescorla, 1986](#); [Dickinson & Mularo, 1989](#)). These findings are not inconsistent with our perspective. They merely point out that other variables, as well as contiguity, may have important effects on the underlying learning curve.

Relatedly, other research has provided some evidence to support the view that the molar response rate–reward rate feedback function plays a role in instrumental performance, as indexed by overall response rates ([Dawson & Dickinson, 1990](#); [Perez et al., 2016](#)). Whether these observations uniquely point to the molar feedback function as a determinant of instrumental performance, as opposed to some other molecular mechanism, is unclear. For the sake of the current discussion, we think it is important to distinguish between goal-directed control and instrumental performance more generally, and to note that the variables that affect overall rates of responding may not be identical to those that control action–outcome learning.

Although this contiguity framework makes sense of the data we report here, it implies that the probability of actions becoming goal-directed increases as a function of training. While this appears to be at odds with findings from the literature that actions typically transition from goal-directed to habitual over training ([Adams, 1982](#); [DeRusso et al., 2010](#); [Killcross & Coutureau, 2003](#); [Malvaez et al., 2018](#); [O'Hare et al., 2016](#); [Smith & Graybiel, 2013](#); [Thralkill & Bouton, 2015](#); however, see [Colwill & Rescorla, 1988](#)), there are crucial methodological details to consider. First, in experiments in which subjects are trained on an RI schedule for a short period of time and then demonstrated to behave in a goal-directed manner, there is usually a training progression wherein subjects start out on CRF and are then built up to the final target RI schedule by gradually increasing the mean IRI. For example, [DeRusso et al. \(2010\)](#) gave their animals three CRF sessions, then two sessions of RI 20 s, and then an additional six sessions of RI 60 s. It seems likely that responding was goal-directed after two RI sessions because the extensive CRF and dense RI training supported favorable action–outcome contiguity and thus a relatively strong action–outcome association. We assume that the transition to a lean RI 60-s schedule resulted in a reduction in the strength of the action–outcome association that effectively brought it below the threshold for observing a devaluation effect.

A second way in which the procedures of our study differ from others is the amount of training given under an “extensive” condition. The extensive training given in Experiments 3 and 4 consisted of 20 sessions, whereas other investigators typically stop after six to 12 sessions ([Adams, 1982](#); [Dickinson, 1985](#); [Killcross & Coutureau, 2003](#); [Lingawi & Balleine, 2012](#); [Malvaez et al., 2018](#); [O'Hare et al., 2016](#); [Smith & Graybiel, 2013](#); [Thralkill & Bouton, 2015](#); [Tricomi, Balleine, & O'Doherty, 2009](#); [Wassum, Cely, Maidment, & Balleine, 2009](#)). In only one case of which we are aware, rats given 20 sessions of training on a free-operant RI schedule were demonstrated to be habitual ([Kosaki & Dickinson, 2010](#)). Notably, however, this study employed an even leaner RI schedule (RI 60 s) than ours (RI 45 s). We think this is critical as leaner RI schedules reduce the action–outcome contiguity and should thus promote slower growth rates, and possibly also a lower asymptote, in the action–outcome association (see [Figure 6](#)). But, nonetheless, it is conceivable that, given enough training sessions, responding would eventually become goal-directed even on the commonly used RI 60-s schedule ([Gremel et al., 2016](#); [Gremel & Costa, 2013a, 2013b](#); [Killcross & Coutureau, 2003](#); [Kosaki & Dickinson, 2010](#); [Lingawi & Balleine, 2012](#); [O'Hare et al., 2016](#); [Renteria et al., 2018](#); [Wassum et al., 2009](#); [Yin, Knowlton, & Balleine, 2004](#)).

There remain several additional issues worth some comment. First, while our analysis has emphasized the role of action–outcome contiguity in determining goal-directed action, whenever action–outcome contiguity varied in our studies, the probability of a response being reinforced also covaried. For instance, in Experiment 4A action–outcome contiguity was superior for the FI group ([Figure 4B](#)), however, the probability of a reward given a response was also higher for the FI group ($\bar{x} (SEM) = 0.08 (0.01)$ and $0.04(0.00)$ for FI and RI, respectively). In either case, we would expect contiguity or probability to affect the growth of the action–outcome association similarly. However, future research could attempt to disentangle the roles of these two variables. Second, we have studied goal-directed and habitual responding under interval schedules, but our analysis may also apply to training under ratio schedules. Several authors have observed that rats and mice respond habitually under RI and in a goal-directed manner under RR schedules. We anticipate that the experienced action–outcome contiguity (or reinforcement probability) would favor animals trained under a ratio schedule, but we are not aware of any attempts to verify this. Third, we and [DeRusso et al. \(2010\)](#) explored goal-directed control under fixed versus random interval schedules, but it is unknown what the time course of goal-directed control would look like under fixed versus random ratio schedules. The overall reward density and reinforcement probability can be held constant under RR and FR schedules, but we expect the action–outcome contiguities to favor the FR schedule. Finally, [Urcelay and Jonkman \(2019\)](#) recently demonstrated that simple exposure to the experimental context without the opportunity to respond renders the response goal-directed when action–outcome contiguity is poor. They interpreted this result to support the action–outcome correlation idea, but an alternative view is that their manipulation enhanced action–outcome learning (even when contiguity was poor) by diminishing competition (i.e., overshadowing) from context.

In summary, we investigated goal-directed and habitual control under a variety of training conditions on interval schedules. We found that goal-directed responding developed (a) on dense but not

lean RI schedules, (b) on an FI schedule after moderate and extensive training, and (c) on a lean RI schedule only after extensive training (albeit to a lesser degree than on an equivalently dense FI). We conclude that the response rate–reward rate correlation, construed either between- or within-sessions, fails to capture this data pattern, but that the data are explicable by noting a role for action–outcome contiguity. The emergence of goal-directed control may depend upon the strength of action–outcome learning and this can dynamically change over training.

References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, 34, 77–98. <http://dx.doi.org/10.1080/14640748208400878>

Balleine, B. (1992). Instrumental performance following a shift in primary motivation depends on incentive learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 18, 236–250. <http://dx.doi.org/10.1037/0097-7403.18.3.236>

Balleine, B. W., Garner, C., Gonzalez, F., & Dickinson, A. (1995). Motivational control of heterogeneous instrumental chains. *Journal of Experimental Psychology: Animal Behavior Processes*, 21, 203–217. <http://dx.doi.org/10.1037/0097-7403.21.3.203>

Balleine, B. W., Paredes-Olay, C., & Dickinson, A. (2005). Effects of outcome devaluation on the performance of a heterogeneous instrumental chain. *International Journal of Comparative Psychology*, 18, 257–272. Retrieved from <http://escholarship.org/uc/item/5pd9x995.pdf>

Baum, W. M. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior*, 20, 137–153. <http://dx.doi.org/10.1901/jeab.1973.20-137>

Colwill, R. M., & Rescorla, R. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 520–536. <http://dx.doi.org/10.1037/0097-7403.11.4.520>

Colwill, R. M., & Rescorla, R. A. (1986). Associative structures in instrumental learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 20, pp. 55–104). New York, NY: Academic Press.

Colwill, R. M., & Rescorla, R. A. (1988). The role of response-reinforcer associations increases throughout extended instrumental training. *Animal Learning & Behavior*, 16, 105–111. <http://dx.doi.org/10.3758/BF03209051>

Colwill, R. M., & Rescorla, R. A. (1990). Evidence for the hierarchical structure of instrumental learning. *Animal Learning & Behavior*, 18, 71–82. <http://dx.doi.org/10.3758/BF03205241>

Corbit, L. H., & Balleine, B. W. (2003). Instrumental and Pavlovian incentive processes have dissociable effects on components of a heterogeneous instrumental chain. *Journal of Experimental Psychology: Animal Behavior Processes*, 29, 99–106. <http://dx.doi.org/10.1037/0097-7403.29.2.99>

Corbit, L. H., Chieng, B. C., & Balleine, B. W. (2014). Effects of repeated cocaine exposure on habit learning and reversal by N-acetylcysteine. *Neuropsychopharmacology*, 39, 1893–1901. <http://dx.doi.org/10.1038/npp.2014.37>

Corbit, L. H., Nie, H., & Janak, P. H. (2012). Habitual alcohol seeking: Time course and the contribution of subregions of the dorsal striatum. *Biological Psychiatry*, 72, 389–395. <http://dx.doi.org/10.1016/j.biopsych.2012.02.024>

Dawson, G. R., & Dickinson, A. (1990). Performance on ratio and interval schedules with matched reinforcement rates. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, 42, 225–239. <http://dx.doi.org/10.1080/14640749008401882>

DeRusso, A. L., Fan, D., Gupta, J., Shelest, O., Costa, R. M., & Yin, H. H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Frontiers in Integrative Neuroscience*, 4, 1–8. <http://dx.doi.org/10.3389/fint.2010.00017>

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 308, 67–78. <http://dx.doi.org/10.1098/rstb.1985.0010>

Dickinson, A. (1989). Expectancy theory in animal conditioning. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and status of traditional learning theory* (pp. 279–308). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Dickinson, A., & Charnock, D. J. (1985). Contingency effects with maintained instrumental reinforcement. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, 37, 397–416. <http://dx.doi.org/10.1080/14640748508401177>

Dickinson, A., & Mulatero, C. W. (1989). Reinforcer specificity of the suppression of instrumental performance on a non-contingent schedule. *Behavioural Processes*, 19, 167–180. [http://dx.doi.org/10.1016/0376-6357\(89\)90039-9](http://dx.doi.org/10.1016/0376-6357(89)90039-9)

Dickinson, A., Nicholas, D. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, 35, 35–51. <http://dx.doi.org/10.1080/14640748308400912>

Dickinson, A., Squire, S., Varga, Z., & Smith, J. W. (1998). Omission learning after instrumental pretraining. *The Quarterly Journal of Experimental Psychology*, 51B, 271–286. <http://dx.doi.org/10.1080/713932679>

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of Reinforcement*. New York, NY: Appleton-Century-Crofts. <http://dx.doi.org/10.1037/10627-000>

Garr, E., & Delamater, A. R. (2019). Exploring the relationship between actions, habits, and automaticity in an action sequence task. *Learning & Memory*, 26, 128–132. <http://dx.doi.org/10.1101/lm.048645.118>

Gremel, C. M., Chancey, J. H., Atwood, B. K., Luo, G., Neve, R., Ramakrishnan, C., . . . Costa, R. M. (2016). Endocannabinoid modulation of orbitostriatal circuits gates habit formation. *Neuron*, 90, 1312–1324. <http://dx.doi.org/10.1016/j.neuron.2016.04.043>

Gremel, C. M., & Costa, R. M. (2013a). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, 4, 2264. <http://dx.doi.org/10.1038/ncomms3264>

Gremel, C. M., & Costa, R. M. (2013b). Premotor cortex is critical for goal-directed actions. *Frontiers in Computational Neuroscience*, 7, 110. <http://dx.doi.org/10.3389/fncom.2013.00110>

Holland, P. C. (1985). The nature of conditioned inhibition in serial and simultaneous feature negative discriminations. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 267–298). Hillsdale, NJ: Erlbaum.

Killcross, S., & Couture, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex*, 13, 400–408. <http://dx.doi.org/10.1093/cercor/13.4.400>

Kosaki, Y., & Dickinson, A. (2010). Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 36, 334–342. <http://dx.doi.org/10.1037/a0016887>

LeBlanc, K. H., Maidment, N. T., & Ostlund, S. B. (2013). Repeated cocaine exposure facilitates the expression of incentive motivation and induces habitual control in rats. *PLoS ONE*, 8, e61355. <http://dx.doi.org/10.1371/journal.pone.0061355>

Lingawi, N. W., & Balleine, B. W. (2012). Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. *The Journal of Neuroscience*, 32, 1073–1081. <http://dx.doi.org/10.1523/JNEUROSCI.4806-11.2012>

Malvaez, M., Greenfield, V. Y., Matheos, D. P., Angelillis, N. A., Murphy, M. D., Kennedy, P. J., . . . Wassum, K. M. (2018). Habits are negatively regulated by histone deacetylase 3 in the dorsal striatum. *Biological Psychiatry*, 84, 383–392. <http://dx.doi.org/10.1016/j.biopsych.2018.01.025>

O'Hare, J. K., Ade, K. K., Sukharnikova, T., Van Hooser, S. D., Palmeri, M. L., Yin, H. H., & Calakos, N. (2016). Pathway-specific striatal

substrates for habitual behavior. *Neuron*, 89, 472–479. <http://dx.doi.org/10.1016/j.neuron.2015.12.032>

Perez, O., Aitken, M., Zhukovsky, P., Soto, F. A., Urcelay, G. P., & Dickinson, A. (2016). Human instrumental performance in ratio and interval contingencies: A challenge for associative theory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 72, 311–321. <http://dx.doi.org/10.1080/17470218.2016.1265996>

Perlman, M. D., & Rasmussen, U. (1975). Some remarks on estimating a noncentrality parameter. *Communications in Statistics*, 4, 455–468. <http://dx.doi.org/10.1080/0361092750827262>

Renteria, R., Baltz, E. T., & Gremel, C. M. (2018). Chronic alcohol exposure disrupts top-down control over basal ganglia action selection to produce habits. *Nature Communications*, 9, 211. <http://dx.doi.org/10.1038/s41467-017-02615-9>

Rescorla, R. (1985). Conditioned inhibition and facilitation. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 299–326). Hillsdale, NJ: Erlbaum.

Rodger, R. S. (1974). Multiple contrasts, factors, error rate and power. *British Journal of Mathematical & Statistical Psychology*, 27, 179–198. <http://dx.doi.org/10.1111/j.2044-8317.1974.tb00539.x>

Rodger, R. S., & Roberts, M. (2013). Comparison of power for multiple comparison procedures contrasts and alternatives. *Journal of Methods and Measurement in the Social Sciences*, 4, 20–47. <http://dx.doi.org/10.2458/jmim.v4i1.17775>

Schwabe, L., Dickinson, A., & Wolf, O. T. (2011). Stress, habits, and drug addiction: A psychoneuroendocrinological perspective. *Experimental and Clinical Psychopharmacology*, 19, 53–63. <http://dx.doi.org/10.1037/a0022212>

Shipman, M. L., Trask, S., Bouton, M. E., & Green, J. T. (2018). Inactivation of prelimbic and infralimbic cortex respectively affects minimally-trained and extensively-trained goal-directed actions. *Neurobiology of Learning and Memory*, 155, 164–172. <http://dx.doi.org/10.1016/j.nlm.2018.07.010>

Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, 79, 361–374. <http://dx.doi.org/10.1016/j.neuron.2013.05.038>

Smith, K. S., & Graybiel, A. M. (2014). Investigating habits: Strategies, technologies and models. *Frontiers in Behavioral Neuroscience*, 8, 39.

Thraikill, E. A., & Bouton, M. E. (2015). Contextual control of instrumental actions and habits. *Journal of Experimental Psychology: Animal Learning and Cognition*, 41, 69–80. <http://dx.doi.org/10.1037/xan0000045>

Thraikill, E. A., Trask, S., Vidal, P., Alcalá, J. A., & Bouton, M. E. (2018). Stimulus control of actions and habits: A role for reinforcer predictability and attention in the development of habitual behavior. *Journal of Experimental Psychology: Animal Learning and Cognition*, 44, 370–384. <http://dx.doi.org/10.1037/xan0000188>

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *The European Journal of Neuroscience*, 29, 2225–2232. <http://dx.doi.org/10.1111/j.1460-9568.2009.06796.x>

Urcelay, G. P., & Jonkman, S. (2019). Delayed rewards facilitate habit formation. *Journal of Experimental Psychology: Animal Learning and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xan0000221>

Wassum, K. M., Cely, I. C., Maidment, N. T., & Balleine, B. W. (2009). Disruption of endogenous opioid activity during instrumental learning enhances habit acquisition. *Neuroscience*, 163, 770–780. <http://dx.doi.org/10.1016/j.neuroscience.2009.06.071>

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *The European Journal of Neuroscience*, 19, 181–189. <http://dx.doi.org/10.1111/j.1460-9568.2004.03095.x>

Received May 31, 2018

Revision received August 23, 2019

Accepted August 29, 2019 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <https://my.apa.org/portal/alerts/> and you will be notified by e-mail when issues of interest to you become available!