Instrumental behavior in humans is sensitive to the correlation between response rate and reward rate

Omar D. Perez^{1,2}

Accepted: 13 October 2020 / Published online: 30 November 2020
© The Psychonomic Society, Inc. 2020

Abstract

Recent theories of instrumental behavior postulate that the correlation between response and reward rate is a critical factor in instrumental goal-directed performance. However, it is still not clear whether human actions can be sensitive to rate correlation. Using a novel within-subject design, participants were trained under ratio and interval contingencies of reinforcement matching both reward probabilities and reward rates between conditions. The impact of rate correlation on performance was evident in the higher performance observed under ratio contingencies for both types of matching. Moreover, there was no difference in performance between two classes of interval schedules with equivalent correlational properties but different reward probabilities. These results are discussed in terms of a recent dual-system model of instrumental behavior.

Keywords Goal-directed · habits · Instrumental conditioning · Free-operant · Reinforcement learning

Introduction

A long-standing question in associative and reinforcement learning concerns the factors that determine the rate at which actions or responses are performed during instrumental learning. Although the answer is still not clear, the evidence suggests that the reward rate (Catania & Reynolds, 1968; de Villiers & Herrnstein, 1976; Niv, Daw, Joel, & Dayan, 2007) and the reward probability per response (Mazur, 1983; Sutton & Barto, 1998) can have a direct effect on responding. What is less clear is whether a third variable, the correlation between the number of responses performed and the number of rewards obtained in a given period of time, can also have an impact on responding.

The hypothesis that the response-outcome rate correlation is a critical determinant of instrumental performance is not new (see Baum, 1973). However, it is only recently

that it has regained attention in light of recent proposals that include the rate correlation as part of a series of potential factors that can control different behavioral instrumental systems in parallel, affecting both the performance and the degree to which behavior is sensitive to outcome devaluation—that is, the extent to which behavior is goal-directed (Perez & Dickinson 2020; see also Dolan & Dayan, 2013; Balleine & O'Doherty, 2010; Balleine, 2019 for reviews).

Evidence for the response-outcome rate correlation affecting instrumental performance comes from experiments comparing responding under random-ratio (RR) and random-interval (RI) contingencies. Under an RR schedule, there is a fixed probability of reward per each response performed and therefore a positive correlation between response rate and reward rate (Perez & Dickinson, 2020; Fig. 1). By contrast, the interval schedule sets an average interval between each available reward and so variations in response rate do not generally produce variations in reward rate—once the response rate is sufficiently high to collect all rewards, the reward rate is fixed at the scheduled interval between rewards. Consistent with this idea, a wealth of data has demonstrated that RR schedules generate higher response rates than RI schedules even when the reward probability (Pérez & Soto, 2020; Catania, Matthews, Silverman, & Yohalem, 1977) or reward rate (Reed, 2001b; Baum, 1993) is matched between the two conditions.

✉ Omar D. Perez
odperez@caltech.edu

¹ Division of the Humanities and Social Sciences, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91107, USA

² Nuffield College CESS-Santiago, Faculty of Business and Economics, University of Santiago, Santiago, Chile

Fig. 1 Relationship between response and outcome or reward rate in ratio and interval contingencies. The *left plot* shows the functional relationship between different response rates and the resulting reward rates under a ratio schedule. Each *line* corresponds to the function that is generated by using a different programmed reward probability per

response (the inverse of the ratio requirement). The *right plot* shows the same functional relationship under an interval schedule. Each *line* represents, in this case, the function that is generated by using a different programmed reward rate (the inverse of the interval parameter). Modified from Figure 2 in Perez and Dickinson (2020)

Although the ratio/interval difference in performance observed in previous studies suggests a critical role of the rate correlation on responding, several authors have taken an alternative approach (Peele, Casey, & Silberberg, 1984; Wearden & Clark, 1988; Niv, Daw, & Dayan, 2005; Tanno & Silberberg, 2012). Their argument lies in the theoretical observation that, unlike RR training where there is a fixed probability of reward per response which is independent of time, in the RI schedule this probability of reward increases with the time since the last obtained reward or the last response performed. Therefore, under an RI schedule long inter-response times (IRTs), or pauses between responses, are differentially reinforced. If subjects are sensitive to this higher probability of reward of long IRTs, they will respond at a lower rate under RI than RR schedules without requiring that subjects are sensitive to the different response-outcome rate correlation of the two schedules.

So far, there have been two main experimental approaches to assess the relative importance that rate correlation and IRT reinforcement may have on instrumental behavior. The first one, initially tested in humans by McDowell and Wixted (1986; see also Reed, 2007; Reed & Morgan, 2007), consists of employing an RI-plus-linear-feedback (RI+) schedule that simultaneously establishes a positive response-outcome rate correlation and a higher reward probability for long IRTs, effectively pitting these two factors against each other. If RI+ performance is similar to that generated by RR training, this is evidence that the rate correlation is a more important factor than IRT reinforcement in determining responding. Although the evidence on this contrast is mixed, one hypothesis that has found support in animal studies is that the differential sensitivity to the correlational or IRT properties of the RI+

schedule may be a function of the response rate attained during training (Reed, 2006). When measures are taken so that subjects respond at high rates, RI+ performance approaches that of RR training whereas RI+ performance is comparable to that of RI training when low response rates are encouraged and reward rates matched between the two schedules. This result suggests that IRT reinforcement may only become a critical factor on responding when subjects have the opportunity to experience the higher reward probability of long IRTs of the RI schedule. More recent studies in animals, however, have begun to reveal a more complex picture, in that programmed reward probabilities and rates also appear to modulate the extent to which rate correlation or IRT-reinforcement control behavior (Reed, 2015).

The second approach has employed a type of interval schedule that holds equivalent correlational properties as the regular RI schedule (in that there is no relationship between the rate of responding and rewards) but in which the effect of time on reward probability is neutralized (i.e., there is no differential reinforcement of long IRTs). The regulated-probability interval schedule (RPI) sets the probability of reward for the following response by considering the response rate performed during the last m IRTs and constantly adjusts it so that the reward rate remains constant. Using the RPI schedule, a recent study by Perez and Soto (2020, Experiment 1) suggests that the rate correlation may indeed have an impact on human responding. When these authors matched the reward probability of the RR schedule with that obtained during RPI training, they observed that RPI training uniformly yielded lower response rates compared to the RR schedule.

An alternative conclusion, however, can be drawn from this experiment. If the reward probability is equated

between the RR and RPI schedules and participants come to respond at a higher rate during RR training, an argument could be made that the main factor driving the effect is the higher reward rate obtained under RR training, but not necessarily the response-outcome rate correlation (Herrnstein, 1969). Whether human responding can be sensitive to the rate correlation when reward rates (rather than reward probabilities) are controlled between RR and RPI schedules remains unclear. The little evidence so far comes from a few experiments in rodents where the reward rates between RR and RPI schedules have been matched in different groups of subjects. Using that design, both Tanno and Sakagami (2008) and Perez et al. (2018) failed to detect a difference in rates of lever-pressing between the two schedules whereas previously, using chain-pulling, Dawson and Dickinson (1990) reported higher performance in RR than RPI schedules. In all of these studies, RI training yielded lower response rates than RPI training. Only a single study in humans by Perez et al. (2016) has explored this question using a similar design, and suggests that the RR/RPI difference may only be obtained if participants are given a choice test between the two schedules at the end of the experiment.

A rate correlation approach makes an additional prediction with regard to the instrumental performance that should be observed for RI and RPI schedules when the two schedules are programmed to have the same interval between rewards (or reward rate). Whereas theories based on reward probability predict higher performance under RPI than RI training due to the differential reinforcement of long IRTs of the RI schedule (which tends to slow responding), a rate correlation approach predicts no difference between the two schedules given their similar correlational properties (Perez & Dickinson, 2020). Again, the evidence so far comes from animal studies and it is largely at variance with the rate correlation hypothesis, as both Dawson and Dickinson (1990) and Perez et al. (2018) found lower performance in RI than RPI training. No experiments have investigated this prediction in humans.

To test these hypotheses, in this study a novel within-subject design was employed where human participants experienced both types of interval schedules, RPI and RI, using the same programmed reward rate, and performance was compared to that of RR training matching both reward probabilities and reward rates with respect to the interval conditions. Apart from being more sensitive to the different properties of the schedules than the previous between-subject designs in animals, this design is able to isolate in a single experiment the effect of the positive rate correlation of the RR schedule over and above the rest of the variables that have been proposed to control instrumental performance. The experiment used mouse-clicks as the

instrumental action and fictitious credits as the rewarding outcome.

Method

Subjects and apparatus

Seventy-one participants were tested on Windows (c) 7 machines at the Nuffield College Centre for Experimental and Social Sciences (CESS)-Santiago. Participants were recruited from the CESS-Santiago subject's pool via the CESS-Santiago mailing list. They gave informed consent and were paid the equivalent of 10 dollars for their participation.¹

Procedure

Participants were asked to insert fictitious coins in different candy dispensers by clicking on them with the mouse. In each 2.5-min trial of training, the active dispenser was colored while the rest of the dispensers were grey, indicating that they were inactive for that particular trial (see Fig. 2a, left panel). Each mouse-click response on the active machine changed the contrast of the dispenser for around 0.2 s. A sound was simultaneously played, indicating that the response in the active machine had been recorded. The spatial position of each dispenser and the assignation of each dispenser to each reward schedule was randomized across participants. When a response was rewarded, a sound was played and a banner on top of the active machine appeared with the message "Reward!". At the same time, a button with the image of an M&M candy was shown at the bottom of the screen (see Fig. 2a, right panel). Before the trial could continue, participants had to collect each earned reward by clicking on the M&M image. Participants were asked to try to figure out the best way of inserting coins in each machine so as to maximize the number of credits obtained at the end of the task. Every time they were rewarded, 40 points were added to their credits; every time a response was performed, 1 point was deducted from their credits. All participants were given a single set of instructions on the screen at the beginning of the experiment.

¹By the time this experiment was run, there was only a single study testing RPI and RR performance in humans (Perez et al. 2016). Since in that study there was no difference in performance between the conditions during training when matching reward rates, no sensible power analysis was possible prior to conducting the present study. However, it was deemed appropriate to aim for twice as many subjects as in that study.

Fig. 2 Design and results of the experiment. **a** Design of each trial of training. **b** Design of each block of training. Each participant was first trained under a master ratio schedule, followed by two types of interval schedules (RI or RPI) yoked with respect to the master ratio schedule in terms of reward rate. The final RPy schedule was matched with respect to the reward probability of the two previous interval schedules. **c** Response rates (in responses per minute) for each schedule during the final block of training. Error bars are 95% within-subject confidence intervals following the method proposed by Morey (2008)

In each of three blocks of training consisting of four trials each, participants were first presented with a master RR schedule (see Fig. 2b). The ratio requirement for the master RR schedule was randomly assigned for each participant to 10 or 15 responses per reward, such that the reward probability per response, p , was either $1/10 = .10$ or $1/15 = .07$. After responding in the master RR schedule, participants were presented with two consecutive interval schedules, RI (RIy) or RPI (RPIy; order counterbalanced across subjects). Each of these two interval schedules was matched to the previous ratio schedule in terms of reward rate; that is, if the mean interval between each reward obtained in the master RR schedule was T , the interval between each reward was set as T for the RIy and RPIy conditions. The RIy schedule programmed the availability of each reward by a Bernoulli process where in each second the probability of a reward becoming available was $1/T$. For the RPIy schedule, the probability of reward for the next response was set to $\frac{1}{T b_m}$, where b_m was the response rate performed by the participant during the last m responses. In practice, the RPI schedule sets a probability of reward for the next response that will yield an average reward rate that is equal to the programmed reward rate, assuming the subject will continue responding at the current rate b_m . Because the reward probability in the RPI is set for the next response and depends on the current response rate (and not the current IRT), the RPI is able to neutralize the property of

long IRTs increasing reward probability held by the regular RI schedule. Given that Dawson and Dickinson did not find any difference in performance across memory sizes ranging from 1 to 50, a memory size of $m = 10$ was set for all participants during RPI trials.

By setting the same interval parameter T for both RIy and RPIy schedules, the present design ensured that the reward rate was equated between the two interval schedules and the master ratio schedule (see Fig. 2b). At the same, the design allowed us to compare performance under RIy and RPIy schedules with an equivalent interval parameter, revealing whether there is an impact of IRT reinforcement on performance between two schedules with equivalent correlational properties and reward rates.

After participants responded in the two interval trials, the ratio requirement for the following RR schedule was set to $k = 1/p_{\text{interval}}$, where p_{interval} was the mean reward probability obtained across the two previous interval schedules. Given that a rate correlation approach predicts no difference between the two interval conditions (RIy, RPIy) due to their equivalent interval parameter T , this manipulation should succeed in equating the reward probabilities between each of the two interval schedules and the yoked RR schedule. To the extent that the rate correlation increases performance in the RR condition independently of reward probability, reward rate, and the differential reinforcement of long IRTs, higher performance

is anticipated in the two RR and RPy schedules than each of the RIy and RPIy schedules.

Results

Statistical analyses were performed using the R programming language (Version 3.4.3) under RStudio (RStudio Team, 2015). For all the pre-planned comparisons a Welch t test is reported. As a measure of effect size, the difference of means, d , is reported, along with a 95% confidence interval on d , CI_d . When reporting ANOVA, η_p^2 and a 90% confidence interval on this estimate is reported. Evidence for the null hypothesis of no difference between means is reported as a Bayes Factor (BF_{01}). The reliability of the results was contrasted against the usual criterion of $\alpha = .05$ (two-tailed).

The final mean response rates for each schedule are shown in Fig. 2c. A 4(schedule) $\times$ 2(parameter) mixed ANOVA with schedule as within-subject factor and parameter as between-subject factor revealed a significant effect of schedule $F_{(2.68, 185.21)} = 8.06$, $p < .001$, $\eta_p^2 = 0.10$ 90% $CI[0.04, 0.16]$, but no effect of parameter $F_{(1, 69)} = 0.01$, $p = .91$, $\eta_p^2 = 0.00$ 90% $CI[0.00, 0.01]$, nor a significant interaction between schedule and parameter, $F_{(2.68, 185.21)} = 1.02$, $p = .38$, $\eta_p^2 = 0.02$ 90% $CI[0.00, 0.04]$. Confirming the predictions of a rate correlation approach, participants responded at a higher rate in the master RR schedule than on each of the interval schedules that had matched reward rates—RR versus RPIy: $t(70) = 2.38$, $p = 0.02$, $d = 11.70$ $CI_d = [1.87, 21.5]$, $BF_{01} = 0.56$; RR versus RIy: $t(70) = 3.86$, $p < .001$, $d = 20.37$ $CI_d = [9.83, 30.90]$, $BF_{01} = 0.01$. This higher performance on RR schedules was replicated when comparing performance on the RPy schedule with each of the interval schedules that had matched reward probabilities—RPy vs RPIy: $t(70) = 2.59$, $p = 0.01$, $d = 8.87$ $CI_d = [2.03, 15.7]$, $BF_{01} = 0.35$; RPy versus RIy: $t(70) = 3.79$, $p < .0001$, $d = 17.54$ $CI_d = [8.31, 26.77]$, $BF_{01} = 0.01$. Furthermore, there was no

detectable difference in performance between the two interval schedules; RPIy vs RIy: $t(70) = 1.87$, $p = 0.07$, $d = 8.67$ $CI_d = [-0.56, 17.90]$, $BF_{01} = 1.47$.

To conclude that the higher responding observed under the two RR schedules was due to their positive response-outcome rate correlation, it is necessary to check that the additional variables thought to control instrumental responding were successfully matched by the yoking procedure employed in the study. The statistical analysis revealed no detectable difference in reward rate between the master RR schedule and the yoked RIy and RPIy schedules $F_{(1.04, 72.51)} = 2.24$, $p = .14$, $\eta_p^2 = 0.03$ 90% $CI[0.00, 0.08]$. As a consequence, the higher response rates of the master RR schedule can only be attributed to its correlational properties. Similarly, it was expected that the yoking procedure would equate the reward probabilities for the RPy, RIy and RPIy schedules. Contrary to this prediction, a significant effect of schedule on reward probability was found, $F_{(1.46, 102.13)} = 5.60$, $p = .01$, $\eta_p^2 = 0.07$ 90% $CI[0.01, 0.14]$. This effect was explained by a higher reward probability experienced on the RIy schedule compared to the other two schedules—RIy vs RPIy: $t(70) = 2.36$, $p = .02$, $d = 0.06$ $CI_d = [0.01, 0.10]$, $BF_{01} = 0.58$; RIy vs RPy: $t(70) = 2.38$, $p = 0.02$, $d = 11.70$ $CI_d = [1.87, 21.5]$, $BF_{01} = 0.56$. By contrast, there was no detectable difference in reward probability between the RPy and the RPIy schedules—RPy vs RPIy: $t(70) = 0.25$, $p = 0.80$, $d = 0.00$ $CI_d = [-0.03, 0.03]$, $BF_{01} = 7.44$. Although the yoking procedure was not completely successful in yoking the reward probabilities, the result reinforces the rate correlation hypothesis, as the performance under the RPy was higher than in the RIy even when the RIy schedule generated a higher reward probability per response. Finally, the analysis confirmed that the RIy schedule was differentially reinforcing long IRTs in a higher proportion than the RPIy schedule $t(70) = 5.31$, $p < .0001$, $d = 0.75$ $CI_d = [0.57, 1.03]$, $BF_{01} = 0.00$ (see Table 1). In spite of this difference, performance was similar between these two interval schedules.

Table 1 Mean values and 95% within-subject CIs for ratio of mean reinforced IRT to mean overall IRT (IRT ratio), reward probability per response and reward rate (in rewards per second) during the final block of training

Schedule	IRT ratio	Reward probability	Reward rate
RR	0.99 [0.87, 1.10]	.09 [.06, .11]	0.26 [0.16, 0.36]
RPIy	1.10 [1.00, 1.20]	.12 [.10, .14]	0.22 [0.08, .0.35]
RIy	1.80 [1.57, 2.02]	.18 [.14, .21]	0.20 [0.02, 0.39]
RPy	1.06 [0.98, 1.13]	.12 [.11, .14]	0.50 [0.08, 0.92]

RR: random-ratio; RPIy: random-probability interval (yoked); RIy: random-interval (yoked); RPy: random-ratio (yoked)

Discussion

The present study demonstrates that the response-outcome rate correlation has a significant impact on human instrumental performance. Using a novel design where for each participant both reward probabilities and reward rates were matched between conditions, it was observed that participants responded at a higher rate under RR schedules than under RPI (RPIy) schedules. More generally, this study shows that human instrumental actions are not only dependent on response-outcome contiguity—or the probability of each response or IRT leading to an immediate reward—but on a more extended relationship between responses and rewards (Perez & Dickinson, 2020). This result is important, as it challenges the widely held notion of reward probability as the main factor determining responding in contemporary theories of reinforcement and associative learning (Daw, Niv, & Dayan, 2005; Mackintosh & Dickinson, 1979; Lee, Shimojo, & O'Doherty, 2014; Vogel, Castro, & Saavedra, 2004).

These data are readily anticipated by a recent model by Perez and Dickinson (2020) which jointly deploys probability and rate correlation systems to determine instrumental performance and behavioral control under different training schedules. In their model, the reward probability or rate determines the strength of a habit system which is insensitive to outcome value whereas the response-outcome rate correlation determines the strength of a goal-directed system which is sensitive both to outcome value and the causal relationship between the response and the outcome (Dickinson & Pérez, 2018b). Total responding is assumed to be a direct function of the linear sum of these two strengths. Under this model, the higher rate correlation established by RR schedules brings about greater goal-directed strength. By contrast, the rate correlation is low and similar between RI and RPI schedules (see Supplemental Material in Perez & Dickinson, 2020). Given that the reward probabilities and rates between the ratio and the two types of interval schedules used in the present design were matched, the same amount of habit strength is accrued by the two interval conditions. By contrast, because the ratio schedule holds a positive response-outcome rate correlation, there is a positive goal-directed strength that increases responding with respect to the two interval conditions. Moreover, since this model does not include an IRT-reinforcement mechanism, it predicts that performance under interval contingencies should be driven by rate correlation and relatively independent of the differential reinforcement of IRTs established by the RI schedule. Their model thus captures why there was no detectable difference in

performance between the RI and the RPI schedule in the present study.

The present design has potential to be applied to other types of response-outcome contingencies to test the relative importance of each factor in performance. Including in the present design both RI+ and RPI schedules should bring about higher response rates in the RI+ than in the RPI schedule if the reward rates are comparable between the two schedules, and similar performance between the RI+ and both the RR and RPy schedules. Likewise, it is well established that RI training renders behavior more insensitive to outcome devaluation, or habitual, than RR training (Dickinson, Nicholas, & Adams, 1983). In this regard, an outstanding question is the degree to which the RPI schedule, which holds the same correlational properties as the RI schedule, will also promote habitual control faster than RR schedules. In Perez and Dickinson's (2020) model, the experienced rate correlation determines the degree to which behavior is goal-directed, explaining why instrumental responding under RR training is more sensitive to outcome revaluation than under RI training (Dickinson et al., 1983). The same mechanism anticipates that habitual or outcome-insensitive responding should be more likely to be observed under RPI than RR training if reward rates or probabilities are matched between the schedules. Under the same conditions, a similar analysis anticipates greater goal-directed control for RI+ compared to RI or RPI training.

One aspect of the present study that may exert an impact on responding, and which was not manipulated in this experiment, is the explicit cost for each response performed. In rats, manipulating the energy cost of lever pressing effectively imposes a constrain on the response rate attained by the animal, which in turn affects the sensitivity of their responding to different aspects of the programmed response-outcome contingencies (Reed, 2007). It is not unreasonable to expect a similar effect in humans. However, Reed (2001a) has demonstrated how the ratio/interval contrast only obtains if an explicit cost of responding is imposed. This evidence would explain why Perez et al. (2016) failed to find a difference in performance between RR and RPI schedules during training. In their study, participants were shown the number of credits currently spent and the number of rewards currently earned, but they were not told how much each of the rewards was worth in credits. It is therefore possible that the absence of a common currency representation between response cost and amount of reward made their participants insensitive to the experienced response-outcome contingencies. This insensitivity to response cost was clearly masking learning: when their participants were given a final choice test

between the same schedules experienced during training, they were able to deploy the learning acquired during training by choosing to respond more to the RR than to the RPI schedule that had matched reward rates during training and more to a high reward probability RR than to a low reward probability RR.

Another aspect that might merit reconsideration in future studies is the gambling-type scenario employed in this design. Clearly, such a scenario may prompt participants to employ more cognitive-oriented strategies, which adds another layer of complexity to schedule sensitivity in humans (Bradshaw, Freegard, & Reed, 2015; Bradshaw & Reed, 2012). Perhaps rate correlation sensitivity is only observed in this particular type of scenario. One line of evidence against this hypothesis, however, comes from a recent study by Perez and Soto (2020). Employing a similar design to that of Reed (2001b), these authors presented participants with a black screen and asked them to perform an instrumental action (key presses) to make a triangle appear for a fraction of a second at the center of the screen. Unbeknownst to the subjects, the response-triangle contingency was arranged in two different conditions, RR and RPI, signaled only by the phrase “Condition 1” or “Condition 2” at the bottom of the screen. There was no explicit response cost imposed on their participants, and participants were asked to rate the causal effectiveness of their actions under each condition, not to maximize rewards or respond in an instructed manner (see Bradshaw et al., 2015; Hayes, Brownstein, Zettle, Rosenfarb, & Korn, 1986). Yet, their participants showed sensitivity to the rate correlation by responding more to an RR than to an RPI condition that had matched reward probabilities.

In summary, when human participants experience ratio and interval response-outcome contingencies, their instrumental behavior does not only hold a direct relationship with the number of rewards obtained (Herrnstein, 1969), the reward probability per response (Mazur, 1983; Daw et al., 2005; Lee et al., 2014; Keramati, Dezfouli, & Piray, 2011; Solway & Botvinick, 2012) or the reward probability of different IRTs (Wearden & Clark, 1988; Tanno & Silberberg, 2012; Niv, Dayan, & Joel, 2006). By controlling all these factors in a single experimental design, the present data demonstrate that humans can also be sensitive to the correlation between response and reward rates (Perez & Dickinson, 2020; Baum, 1973; Dickinson & Perez, 2018a).

Acknowledgements I would like to thank Raymond Duch, Mauricio Lopez, and Felipe Torres from Nuffield College CESS-Santiago for their help in running this experiment, Mike Aitken for helping with computer programming and Anthony Dickinson for his comments on a previous version of this manuscript.

Open practices statement The data and scripts used to analyze the study are available at <https://osf.io/uhjzp/>. This study was not preregistered.

References

- Balleine, B. W. (2019). Hierarchical action control: Adaptive collaboration between actions and habits. *Frontiers in Psychology, 10*, 2735.
- Balleine, B., & O’Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 35*, 48–69.
- Baum, W. M. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior, 20*, 137–153.
- Baum, W. (1993). Performances on ratio and interval schedules of reinforcement: Data and theory. *Journal of the Experimental Analysis of Behavior, 2*, 245–264.
- Bradshaw, C. M., Freegard, G., & Reed, P. (2015). Human performance on random ratio and random interval schedules, performance awareness and verbal instructions. *Learning & Behavior, 43*, 272–288.
- Bradshaw, C. M., & Reed, P. (2012). Relationship between contingency awareness and human performance on random ratio and random interval schedules. *Learning and Motivation, 43*, 55–65.
- Catania, A. C., Matthews, T. J., Silverman, P. J., & Yohalem, R. (1977). Yoked variable-ratio and variable-interval responding in pigeons. *Journal of the Experimental Analysis of Behavior, 28*, 155–161.
- Catania, A. C., & Reynolds, G. S. (1968). A quantitative analysis of the responding maintained by interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 11*, 327–383.
- Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*, 1704–1711.
- Dawson, G. R., & Dickinson, A. (1990). Performance on ratio and interval schedules with matched reinforcement rates. *The Quarterly Journal of Experimental Psychology, 42*, 37–41.
- Dickinson, A., & Perez, O. D. (2018). Actions and habits: Psychological issues in dual-system theory. In Morris, R. W., Bornstein, A. M., & Shenhav, A. (Eds.) *Goal-Directed Decision Making: Computations and Neural Circuits*, (pp. 1–37): Elsevier.
- Dickinson, A., & Pérez, O. D. (2018). *Actions and habits: Psychological issues in dual-system theory. Goal-directed decision making*, (pp. 1–25). Amsterdam: Elsevier.
- Dickinson, A., Nicholas, D., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology, 35*, 35–51.
- Dolan, R., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80*, 312–325.
- Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior, 45*, 237–256.
- Herrnstein, R. J. (1969). Method and theory in the study of avoidance. *Psychological Review, 76*, 49–69.
- Keramati, M., Dezfouli, A., & Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology, 7*.
- Lee, S. W., Shimojo, S., & O’Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron, 81*, 687–699.
- Mackintosh, N. J., & Dickinson, A. (1979). *Instrumental (Type II) Conditioning. Mechanisms of learning and motivation*, (pp. 143–167). Hillsdale: Erlbaum.
- Mazur, J. E. (1983). Steady state performance on fixed, mixed, and random ratio schedules. *Journal of the Experimental Analysis of Behavior, 2*, 293–307.

- McDowell, J. J., & Wixted, J. T. T. (1986). Variable ratio schedules as variable interval schedules with linear feedback loops. *Journal of the Experimental Analysis of Behavior*, 3.
- Morey, R. D. et al. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4, 61–64.
- Niv, Y., Daw, N., & Dayan, P. (2005). How fast to work: Response vigor, motivation and tonic dopamine. *Advances in Neural Information Processing Systems 18 (NIPS 2005)*, 18, 1019–1026.
- Niv, Y., Daw, N., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191, 507–520.
- Niv, Y., Dayan, P., & Joel, D. (2006). The effects of motivation on extensively trained behavior. Leibniz Technical Report. Hebrew University, 1–25.
- Peele, D. B., Casey, J., & Silberberg, A. (1984). Primacy of interresponse-time reinforcement in accounting for rate differences under variable-ratio and variable-interval schedules. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 149–167.
- Perez, O. D., & Dickinson, A. (2020). A theory of actions and habits: the interaction of rate correlation and contiguity systems in free-operant behavior. *Psychological Review*, 127(6), 945–971. <https://doi.org/https://doi.org/10.1037/rev0000201>
- Pérez, O. D., Aitken, M. R. F., Zhukovsky, P., Soto, F. A., Urcelay, G. P., & Dickinson, A. (2016). Human instrumental performance in ratio and interval contingencies: A challenge for associative theory. *The Quarterly Journal of Experimental Psychology*, 1–13.
- Pérez, O. D., Aitken, M. R., Milton, A. L., & Dickinson, A. (2018). A re-examination of responding on ratio and regulated-probability interval schedules. *Learning and Motivation*, 64, 1–8.
- Pérez, O. D., & Soto, F. A. (2020). Evidence for a dissociation between causal beliefs and instrumental actions. *Quarterly Journal of Experimental Psychology*, 1747021819899808.
- Reed, P. (2001a). Human schedule performance with hypothetical monetary reinforcement. *European Journal of Behavior Analysis*, 2, 225–234.
- Reed, P. (2001b). Schedules of reinforcement as determinants of human causality judgments and response rates. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 187–195.
- Reed, P. (2006). Effect of required response force on rats' performance on a VI+ schedule of reinforcement. *Learning & Behavior*, 34, 379–386.
- Reed, P. (2007). Human sensitivity to reinforcement feedback functions. *Psychonomic Bulletin & Review*, 14, 653–657.
- Reed, P. (2015). Response-independent outcomes impact response rates and judgments of control differentially depending on rate of response-dependent outcomes. *Learning & Behavior*, 43, 301–311.
- Reed, P., & Morgan, T. A. (2007). Resurgence of behavior during extinction depends on previous rate of response. *Learning and Behavior*, 35, 106–114.
- RStudio Team (2015). RStudio: Integrated Development Environment for R. Boston, MA.
- Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychological Review*, 119, 120–154.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction. vol. 1. 1*. Cambridge: MIT Press.
- Tanno, T., & Sakagami, T. (2008). On the primacy of molecular processes in determining response rates under variable-ratio and variable-interval schedules. *Journal of the Experimental Analysis of Behavior*, 89, 5–14.
- Tanno, T., & Silberberg, A. (2012). The copyist model of response emission. *Psychonomic Bulletin and Review*, 19, 759–778.
- de Villiers, P. A., & Herrnstein, R. J. (1976). Toward a law of response strength. *Psychological Bulletin*, 83, 1131–1153.
- Vogel, E. H., Castro, M. E., & Saavedra, M. A. (2004). Quantitative models of Pavlovian conditioning. *Brain Research Bulletin*, 63, 173–202.
- Wearden, J. H., & Clark, R. B. (1988). Interresponse-time reinforcement and behavior under aperiodic reinforcement schedules: A case study using computer modeling. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 200–211.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.