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

Omar D. Perez[1, 2]

[1] Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California, USA.

[2] Nuffield College CESS-Santiago, Facultad de Administracion y Economia, Universidad de Santiago de Chile, Santiago, Chile.

Author Note

Correspondence concerning this article should be addressed to Omar D. Perez, 1200 East California Boulevard, Pasadena, California, CA91107. E-mail: odperez@caltech.edu.

Abstract

Recent theories of instrumental conditioning postulate that the correlation between responses and outcome rates is a critical factor in instrumental free-operant performance and goal-directed control. However, it is still not clear whether human performance can be sensitive to such variable. Using a novel within-subject design, participants were trained under ratio and interval contingencies of reinforcement matching both outcome probability and outcome rates. 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 similar correlational properties but different reward probabilities. The results are discussed in terms of a recent dual-system model of instrumental behavior.

Keywords: goal-directed,habits, instrumental conditioning,free-operant, reinforcement learning

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

Introduction

A long-standing question in the learning literature concerns the factors that determine the rate at which actions or responses are performed during instrumental free-operant conditioning. Although the answer is still not clear, the evidence suggests that the reward rate (Catania and Reynolds, 1968; de Villiers and Herrnstein, 1976) and the reward probability per response (Mazur, 1983) 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, only recently it has been proposed as part of a series of potential factors that can control different behavioral instrumental systems in parallel, affecting both performance and the degree to which behavior is sensitive to outcome devaluation—that is, the extent to which behavior is goal-directed (Perez and Dickinson, in press). 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 and Dickinson, in press). By contrast, the reward rate sets an average interval between each available reward and so variations response rate do not generally produce variations in outcome rate—the reward rate is fixed at the scheduled interval between rewards once the response rate is sufficiently high to collect all rewards. Consistent with this idea, a wealth of data have demonstrated that RR schedules generate higher response rates than RI schedules even when the reward probability

(Pérez and Soto, 2020) or reward rate (Reed, 2001) is matched between the two conditions.

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 (Niv *et al.*, 2005; Peele *et al.*, 1984; Tanno and Silberberg, 2012; Wearden and Clark, 1988). Their argument lies on 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.

To pit these two approaches against each other, Perez and Soto (2020) trained human participants under a regulated-probability interval schedule (RPI), 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). Consistent with a rate correlation view, they found that RPI training uniformly yielded lower response rates compared to an RR schedule for which the reward probability was matched with the RPI 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 respond at a higher rate in the RR schedule, an argument could be made that it is the higher reward rate obtained under RR training, but not necessarily the response-outcome rate correlation, the main factor driving the effect (Herrnstein, 1969). Whether human responding can be sensitive to the rate correlation when reward rates are controlled between conditions remains unclear. The little evidence so far comes from 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 (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, but the merits of a rate correlation approach contrasting human performance under RR and RPI schedules for matched reward rates have not yet been tested.

An additional prediction of a rate correlation approach can be made for 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 and Dickinson, in press). 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 predictions, in this study a novel within-subject design was employed where human participants experienced both types of interval schedules (RPI, RI) with the same reward rate and performance was compared to that of RR training matching both reward probabilities and rates with respect to the interval schedules. Apart from being more sensitive to the different properties of the schedules than the previous between-subject designs, this design allows to probe in a single experiment the effect of the positive rate correlation of the RR schedule over and above other variables thought to control instrumental performance. The experiment used mouse-clicks as the instrumental response and fictitious credits as the rewarding outcome.

Method

Subjects and apparatus

71 participants were tested on Windows 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 an equivalent of 10 dollars for their participation.

Procedure

Participants were asked to insert fictitious coins in different candy dispensers by clicking on them with the mouse. In each 2.5-minute 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 Figure 1A, left panel). Each mouse-click response on the active machine changed the contrast of the dispenser for around 0.2 seconds. 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 Figure 1A, 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.

In each of three blocks of training consisting of 4 trials each, participants were first presented with a master RR schedule (see Figure 1B). 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 $\frac{1}{Tb_m}$, 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 participant 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. A memory size of $m = 5$ was set for all participants for the RPI trials.

By setting the same interval parameter T for both RIy and RPIy schedules, the present design ensures that the reward rate is equated between the two interval schedules and the master ratio schedule (see Figure 1B). At the same, the design allows 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 similar equivalent correlational properties.

After participants responded in the two interval trials, the ratio requirement for the following RR schedule was set to $k = 1/p_{interval}$, 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, rate, and the differential reinforcement of long IRTs, higher performance is anticipated in the two RR and RRy schedules than each of the RIy and RPIy schedules.

Results

Statistical analyses were performed using the R programming language (Version 3.4.3; R Core Team 2017) under RStudio (RStudio Team 2015). For all the pre-planned comparisons a Welsh 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 Figure 1C. A 4(schedule) x 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:

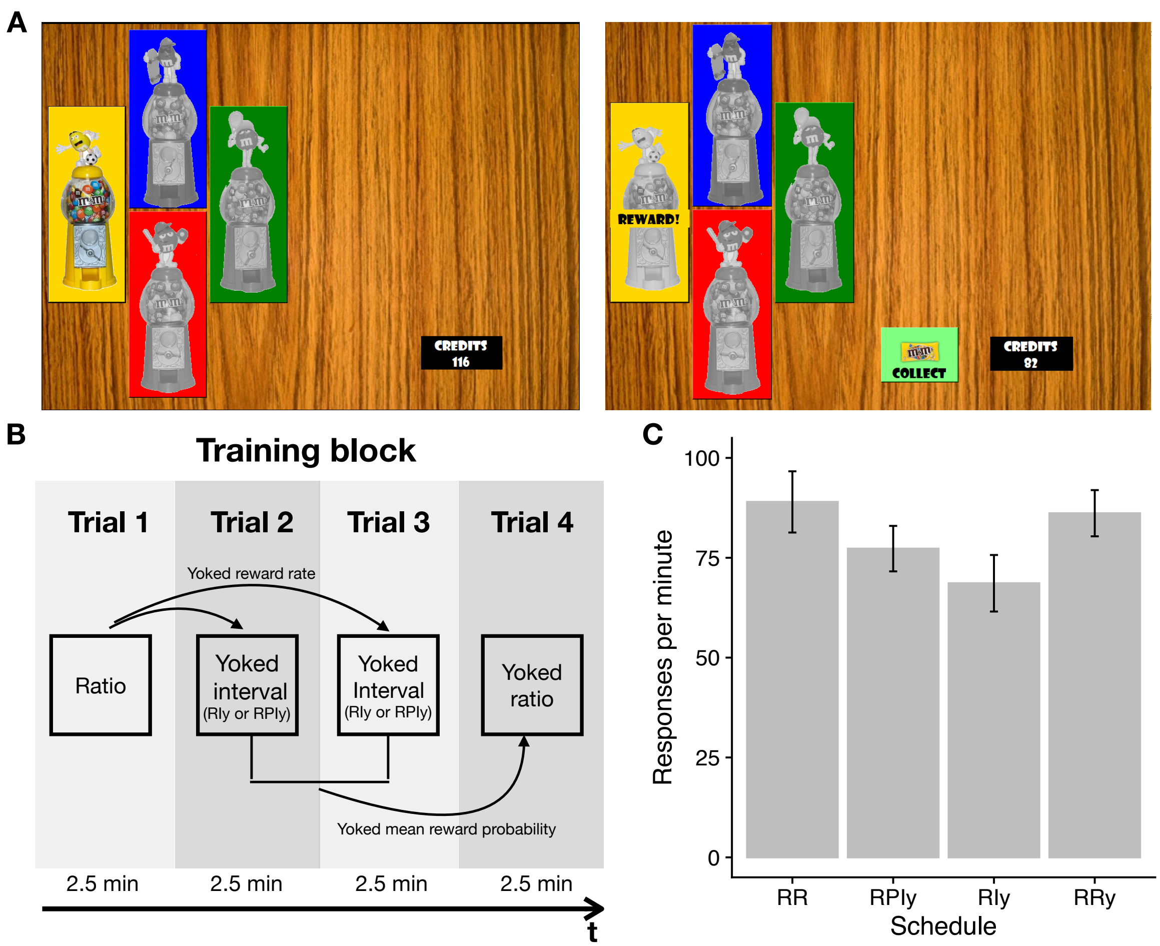

Figure 1. **Design and results of the experiment**. (A) Design of each trial of training. (B) Design of each block of training. Each participant was first trained under a master ratio schedule, followed by two types of interval schedules (RI or RPI) yoked with respect to the master ratio schedule in terms of reward rate. The final RRy 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.

$t(70) = 3.86, p < .001, d = 20.37\ CI_d = [9.83, 30.90], BF_{01} = 0.01$. This higher performance on RR schedules was replicated when comparing performance on the RRy schedule with each of the interval schedules that had matched reward probabilities—RRy vs RPIy: $t(70) = 2.59, p = 0.01, d = 8.87\ CI_d = [2.03, 15.7], BF_{01} = 0.35$; RRy 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 RR schedules was due to its 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. $F_{(1.04,72.51)} = 2.24, p = .14, \eta_p^2 = 0.03\ 90\%\ CI[0.00, 0.82]$. 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 RRy, RIy and RPIy schedules. Contrary to this prediction, a significant effect of schedule in reward probability was found, $F_{(1.04,72.51)} = 2.24, p = .14, \eta_p^2 = 0.03\ 90\%\ CI[0.00, 0.82]$. 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 RRy: $t(70) = 2.38, p = 0.02, d = 11.70\ CI_d = [1.87, 21.5], BF_{01} = 0.56$. By contrast, there was no detectable difference in reward probability between the RRy and the RPIy schedules—RRy 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 RRy 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. RR: random-ratio; RPIy: random-probability interval (yoked); RIy: random-interval (yoked); RRy: random-ratio (yoked)

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]
RRy	1.06 [0.98, 1.13]	.12 [.11, .14]	0.50 [0.08, 0.92]

Discussion

The present study demonstrates that the response-outcome rate correlation has a significant impact on human instrumental free-operant performance. Using a within-subject design where both reward probabilities and rates were matched between conditions, participants responded at a higher rate under RR schedules than under RPI (RPIy) schedules. More generally, it shows that instrumental learning in humans is 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. 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 learning and Pavlovian conditioning (Daw *et al.*, 2005; Lee *et al.*, 2014; Mackintosh and Dickinson, 1979; Vogel *et al.*, 2004).

These results are consistent with a recent model offered by Perez and Dickinson (in press) 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 and Pérez, 2018). 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 and Dickinson, in press). 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. One such schedule is the random-interval-plus-linear-feedback schedule (RI+; Reed, 2007). The RI+ schedule establishes a positive rate correlation at the same time as it reinforces long IRTs with higher probability. If RI+ performance is similar to that generated by RR training, that is evidence that the rate correlation is more important a factor than IRT reinforcement.

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 RRy schedules.

It is well-established that RI training renders behavior more insensitive to outcome devaluation, or habitual, than RR training (Dickinson *et al.*, 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 (in press) model, the experienced rate correlation determines the degree to which behavior is goal-directed, explaining why instrumental responses under RR training are 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.

In conclusion, when participants experience ratio and interval contingencies of reinforcement, their behavior does not only hold a direct relationship with the number of rewards obtained (Herrnstein, 1969), the reward probability per response (Mazur, 1983) or the reward probability of different IRTs (Tanno and Silberberg, 2012; Wearden and Clark, 1988). The present data demonstrates that humans can also be sensitive to the correlation between response and outcome rates.

Acknowledgments

I would like to thank Dr. Raymond Duch, Mauricio Lopez and Felipe Torres from CESS-Santiago for their help in running this experiment.

References

Baum, W. M. (1973). The Correlation-Based Law of Effect. *Journal of the Experimental Analysis of Behavior*, **20**: 137–153.

Catania, A. C. and 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, and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, **8**: 1704–1711.

Dawson, G. R. and Dickinson, A (1990). Performance on ratio and interval schedules with matched reinforcement rates. *The Quarterly Journal of Experimental Psychology*, **42**: 37–41.

de Villiers, P. and Herrnstein, R. (1976). Toward a law of response strength. *Psychological Bulletin*, **83**: 1131–1153.

Dickinson, A and Pérez, O. D. (2018). Actions and habits: psychological issues in dual-system theory. *Goal-Directed Decision Making*. Elsevier: 1–25.

Dickinson, A., Nicholas, D., and 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.

Herrnstein, R. J. (1969). Method and theory in the study of avoidance. *Psychological Review*, **76**: 49–69.

Lee, S. W., Shimojo, S., and O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, **81**: 687–699.

Mackintosh, N. J. and Dickinson, A (1979). Instrumental (Type II) Conditioning. *Mechanisms of learning and motivation*. Hillsdale, NJ: Erlbaum.: 143–167.

Mazur, J. E. (1983). Steady state performance on fixed, mixed, and random ratio schedules. *Journal of the Experimental Analysis of Behavior*, **2**: 293–307.

Niv, Y, Daw, N., and 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.

Peele, D. B., Casey, J., and 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.

Pérez, O. D., Aitken, M. R., Milton, A. L., and Dickinson, A. (2018). A re-examination of responding on ratio and regulated-probability interval schedules. *Learning and Motivation*, **64**: 1–8.

Pérez, O. D. and Soto, F. A. (2020). Evidence for a dissociation between causal beliefs and instrumental actions. *Quarterly Journal of Experimental Psychology*: 1747021819899808.

Perez, O. D. and Dickinson, A. (in press). A theory of actions and habits: The interaction of rate correlation and contiguity systems in free-operant behavior. *Psychological Review*.

Reed, P (2001). Schedules of reinforcement as determinants of human causality judgments and response rates. *Journal of Experimental Psychology: Animal Behavior Processes*, **27**: 187–195.

Reed, P (2007). Human sensitivity to reinforcement feedback functions. *Psychonomic Bulletin & Review*, **14**: 653–657.

Tanno, T and 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. and Silberberg, A. (2012). The copyist model of response emission. *Psychonomic Bulletin & Review*, **19**: 759–778.

Vogel, E. H., Castro, M. E., and Saavedra, M. A. (2004). Quantitative models of Pavlovian conditioning. *Brain research bulletin*, **63**: 173–202.

Wearden, J. H. and 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.