

Evidence for a dissociation between causal beliefs and instrumental actions

Omar D. Perez^{1, 3} & Fabian A. Soto²

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

² Department of Psychology, Florida International University, Florida, USA

³ Nuffield College CESS-Santiago, Facultad de Administracion y Economia, Universidad de
Santiago de Chile, Santiago, Chile

Abstract

Human experiments have demonstrated that instrumental performance of an action and the causal beliefs of the effectiveness of an action in producing a reward are correlated and controlled by the probability of an action leading to a reward. The animal literature, however, shows that instrumental performance under free-operant training differs even when these reward probabilities are matched while subjects undergo training under ratio or interval schedules of reward. In two experiments, we investigated whether causal beliefs would correlate with instrumental performance under interval and ratio schedules for matched reward probabilities. In both experiments we found that performance was higher under ratio than under interval training. However, causal beliefs were similar between these two conditions despite these differences in instrumental performance. When reward probabilities were increased by experimental manipulations in Experiment 2, the causal beliefs increased but performance decreased with respect to Experiment 1. This is evidence that instrumental performance and causal action-reward attribution may not follow from the same psychological process under free-operant training.

Keywords: ratio schedule; interval schedule; causal judgments; instrumental learning; goal-directed; habits

Introduction

It has become widely-accepted that instrumental behavior comprises at least two different types of mechanisms which can be probed both behaviorally and neurally (Balleine & Dickinson, 1998; Daw & O’Doherty, 2013; Dolan & Dayan, 2013). On the one hand, instrumental actions can be controlled by a stimulus-response habit system whereby actions are elicited automatically by an environment in which actions have been previously followed by rewarding outcomes (i.e., the law of effect, Thorndike (1911)). Alternatively, they can be controlled by a goal-based system which considers the consequences of the actions and the current incentive value of the reward (Daw, 2015; Daw, Niv, & Dayan, 2005; Daw & O’Doherty, 2013; Dickinson & Balleine, 2002).

To allow organisms to fulfil their current needs and desires, goal-directed actions need to be mediated by a causal representation between their performance and the rewards they produce, a relationship established by the environment but that the agent should be capable of encoding in order to direct and energize its behavior toward valuable rewards and avoid harmful ones. Folk psychology recognizes this as the agents’ *belief* in the action-reward (action-reward) contingency (Dickinson, 1985; Dickinson & Perez, 2018).

Previous experiments in animals have demonstrated that instrumental performance can be sensitive to the causal action-reward contingency and that this contingency is related to the overall probability of an action producing a reward. Indeed, when the contingency is degraded by manipulations that equate the probability of a reward following ($P(\text{reward}|\text{action})$), and in the absence of an action ($P(\text{reward}|\text{no action})$), the rate of lever-pressing in rats is significantly reduced (Balleine & Dickinson, 1998; Dickinson & Charnock, 1985). Moreover, a wealth of data has shown that instrumental actions in humans can also be sensitive to the same causal manipulation and that, more importantly, human causal beliefs of the effectiveness of an action in producing a reward correlate with instrumental performance. Shanks and Dickinson (1991), for example, probed participants’ causal beliefs by presenting them with a contingency in which the probability of obtaining a reward after each key press was fixed and manipulated the probability of obtaining and reward in the absence of a key press and found that the causal ratings followed the difference between these two probabilities (Dickinson, Shanks, & Evenden, 1984; Hammond, 1980; Vaghi et al., 2019; Wasserman, Chatlosh, & Neunaber, 1983); instrumental performance as measured by rate of key pressing also followed the same difference.

Although the question of whether these causal beliefs are reflecting action-reward causality or simple statistical associations is still a matter of debate, most theories (see Cheng (1997); Cheng and Buehner (2012), for example) have accepted reward probability as the basis of causal learning in humans; the Shanks and Dickinson’s experiment is only one of many studies providing evidence that goal-directed action is mediated by a causal action-reward belief and that this belief correlates with the overall reward probability per action.

The idea that reward probability is a key determinant of goal-directed action is pervasive in many different areas of decision-making research, both mechanistic (Mackintosh & Dickinson, 1979) and normative (Cheng, 1997; Dolan & Dayan, 2013; Mas-Colell, Whinston,

Green, & Others, 1995; Sutton & Barto, 1998). However, the instrumental literature in animals has challenged this notion by showing that instrumental free-operant performance does not depend only upon reward probability. In free-operant experiments, animals are trained under action-reward schedules which resemble non-depleting and depleting sources that animals find in natural environments. Non-depleting resources are modelled using ratio schedules in which each action has a fixed probability q of yielding a reward, so that the ratio, or number of actions per each reward earned is equal to $1/q$. Interval schedules, by contrast, mirror depleting resources in which a period of time needs to elapse between each reward that has been earned before the agent has the opportunity to collect a new reward. Thus, what defines an interval schedule is that the number of actions performed is relatively independent of the number of rewards obtained.

The challenge to probability-based theories is illustrated by experiments in which subjects are assigned to interval training and, after observing the number of actions and rewards obtained under this type of training, the same experienced reward probabilities (or number of rewards per action) are set to train subjects under a ratio condition. Under these circumstances, and in spite of the similar reward probabilities per action, instrumental actions come to be performed at a higher rate under ratio than under interval training (Catania et al., 1977; Dickinson, Nicholas, & Adams, 1983). Although a number of probability-based theories have been proposed to explain these results, the most likely explanation is that instrumental actions are only partially controlled by reward probability, and that the difference in performance is a consequence of the different action-reward rate correlations that subjects experience under these two types of training (Dickinson, 1994; Dickinson & Perez, 2018). Because the reward rate is fixed and mostly independent of the action rate under interval training (no matter how many times an agent searches for nectar in a given patch, it still needs to wait for the depleting reward source to replete before having the opportunity to earn a new reward), the higher action-reward rate correlation experienced by subjects under ratio training should establish a stronger causal action-reward belief. From a psychological standpoint, agents should attribute a higher causal belief to their actions if they are able to control the rate at which the rewards are obtained over a situation in which there is an additional factor involved, as it is the case of time under a reward that comes from an interval source (Pérez et al., 2016). To the extent that instrumental performance is mediated by the causal action-reward belief established by the action-reward rate correlation (Dickinson & Perez, 2018), agents should come to perform the action at a higher rate under ratio than under interval training.

In terms of performance, we have recently confirmed this prediction in human subjects (2016; see also Reed, 2007). Across a series of blocks, we trained participants to key press under an interval schedule and used the reward probability experienced under the interval schedule to set the reward probability for a subsequent ratio condition. Although participants showed similar levels of performance during training, a subsequent choice test revealed a strong preference for the ratio over the interval schedule, suggesting that performance was controlled by the experienced action-reward rate correlation.

In spite of the evidence provided by this experiment, it remains unknown if participants would attribute a higher causal belief to their actions under ratio than under interval training

if the reward probabilities per action are matched between these two types of training. If the experienced action-reward rate correlation plays a role in the development of causal action-reward beliefs one would expect both action rates and causal ratings of the effectiveness of an action in producing the reward to be higher under ratio than interval training in spite of the reward probabilities being matched between these two schedules. By contrast, if reward probability is the primary variable affecting an action-reward belief -as assumed in the human instrumental learning literature- causal ratings of the effectiveness of an action should not differ between ratio- and interval-trained actions if reward probabilities are matched, and therefore should be uncorrelated with their performance. The latter result would suggest different mechanisms of instrumental performance and causal beliefs for free-operant training.

In the present experiments, we explored these predictions using a within-subject design similar to the one employed by Perez and colleagues (2016). In two experiments, we presented participants with an interval schedule followed by a ratio schedule for which the reward probability was matched to the reward probability experienced in the previous interval condition. After each condition, participants were asked to rate the causal effectiveness of their actions in producing rewards. Both experiments were run under the same conditions using key presses as the instrumental action and fictitious credits as the reward.

Exp.	Schedule	Reward probability	Rate correlation	Action rate	Rating
1	RPI	.17 [.15, .18]	.19 [.11, .27]	59.41 [53.43, 67.40]	38.8 [34.52, 43.08]
	RR	.17 [.15, .19]	.55 [.47, .62]	82.60 [73.48, 91.72]	37.84 [33.26, 42.41]
2	RI	.47 [.45, .50]	.42 [.37, .47]	33.12 [28.72, 37.53]	52.86 [49.73, 55.99]
	RR	.47 [.44, .49]	.72 [.68, .76]	56.61 [50.88, 62.34]	56.36 [52.91, 59.82]

Table 1

Mean and 95% within-subject confidence intervals for experienced reward probabilities (reward probability per action), action-reward rate correlation, action rate (in actions per min) and causal ratings in the two experiments reported in this paper.

Experiment 1

In Experiment 1, participants were required to respond by pressing the spacebar under a regulated-probability interval schedule (RPI, Pérez et al. (2016); Dawson and Dickinson (1990)), a class of interval schedule that maintains a constant reward rate (and therefore, a low action-reward rate correlation) and that, in addition, does not increase the probability of an action after a long pause between the previous and current action performed (inter-response time, or IRT). In a subsequent trial, participants were presented with a random-ratio (RR) schedule with a reward probability per action that was equivalent to the one experienced during the previous RPI trial. After each trial, they were asked to rate the causal effectiveness of their actions in producing the reward. If action rates and causal beliefs of the action-reward association are partially determined by the experienced action-reward rate correlation, we expect the ratio schedule to bring about higher action

rates and causal ratings than the RPI schedule.

Method

Subjects and apparatus. 41 participants from Florida International University participated in the study. Both experiments reported in this paper were programmed using the library Psychopy (Peirce, 2007) 1.82.4 for Python. Participants were randomly assigned to one of five interval parameters (from 8 to 15 seconds) and tested in the same room under similar conditions. The task was run in a desktop PC running Windows 10 (c). Participants were compensated by course credit and signed a written consent form before starting the experiment. The study was approved by the Institutional Review Board of the Department of Psychology at Florida International University.

Procedure. Following previous experimental procedures in humans (Dickinson et al., 1984; Reed, 2001; Shanks & Dickinson, 1991; Wasserman et al., 1983), in each trial participants were presented with a distinguishable background color that signalled which reward schedule, interval or ratio, was in effect. Each condition was presented for 3 minutes. Participants were told that they could press the spacebar any time and as many times as they wanted and that sometimes their presses would result in a reward signalled by a yellow triangle that would appear at the center of the screen (see instructions in the Appendix). Their goal in the task was to judge the extent to which their key presses produced the reward in each condition.

On top of the screen, participants were shown an amount of fictitious credits which started at 20 at the beginning of training. Given previous evidence suggesting that free-operant performance necessitates of a cost of responding to be imposed in order to show reward schedule sensitivity (Reed, 1999), every time participants performed a key press a feedback sound was played and 1 point was deducted from their credits; 20 points were added to their credits every time they were rewarded. The yellow triangle flashed at the center of the screen for 30 frames (around 0.5 seconds).

Each block of two trials started with an interval condition in which participants' key presses were rewarded according to an RPI schedule. After each key press, the reward probability for the following action was set at $p = 1/Tb_m$, where T is the programmed inter-reward interval (or inverse of the reward rate) and b_m is the local action rate considering the last m actions performed. Effectively, the RPI schedule keeps a constant reward rate based on the local action rate performed by the subject without increasing the reward probability for long pauses between actions or IRTs. The second trial in each block presented participants with an RR schedule. In these trials, each key press was rewarded with probability q , where q was the reward probability experienced under RPI training in the previous trial; the ratio requirement for the ratio schedule was therefore equal to $1/q$ actions per reward during these yoked, ratio trials.

Two features of the RPI schedule are worth noting. First, the RPI schedule maintains a constant reward rate which is relatively independent of action rate, bringing about a weak action-reward correlation. Second, the reward probability under the RPI schedule is set for the next action, not the current action. Therefore, any difference in performance between ratio and RPI conditions cannot be attributed to a higher reward probability of long IRTs

or the reward probability per action (which is controlled by the experimental design), but to the different action-reward rate correlations established by the two schedules.

Each participant was randomly assigned to an interval parameter T between 8 and 15 seconds. After responding on the interval schedule, participants were presented with a different background color indicating that a different condition was in effect. The reward probability from the previous condition was then used as the parameter for the subsequent ratio condition, thereby yielding within-subject yoking of the interval and ratio conditions between each block of two trials. Training comprised 4 blocks for a total of 8 trials per subject (4 master interval and 4 yoked ratio trials). After each trial, participants were asked to rate the causal effectiveness of their actions in producing the reward using a scale from 1-100 (Reed, 2001; Shanks & Dickinson, 1991). Since Dawson and Dickinson (1990) did not find any difference in performance when the parameter m varied between 1, 5 and 50, we assigned a value of 5 to all subjects.

Results and discussion

For all the experiments reported in this paper, participants were excluded from the final statistical analyses if they failed to respond in at least one of the interval schedules, as this resulted in a ratio parameter for the subsequent ratio trial that was undefined, effectively losing the following ratio trial. Participants were also discarded if they did not receive any rewards during a trial, as this would result in a reward probability equal to zero for the following yoked ratio schedule. Data, manuscript and analyses for this paper can be found in <https://osf.io/b7re6/>.

Figure 1 presents the results of Experiment 1. Panel a shows the action rates in each block of training, separated by schedule type. Panel b shows the mean action rates across all blocks of training. As expected, participants' action rates were higher for ratio than interval training ($F(1, 40) = 16.60, p = .0002, D = -0.45, 95\%CI = [-0.23, -0.67]$). In contrast with this observed difference in instrumental performance, we found no detectable differences between the ratio and interval conditions in participants' causal beliefs of action-reward strength ($F(1, 40) = 0.12, p = .73, D = -0.03, 95\%CI = [-0.25, 0.18], BF_{01} = 10.95$). As expected by the yoking procedure, the experienced reward probabilities did not differ between the two schedules ($F(1, 40) = 0.01, p = .91, D = 0.02, 95\%CI = [-0.19, 0.24], BF_{01} = 10.95$).

These results suggest a dissociation between instrumental performance and causal beliefs under free-operant training and are consistent with the widely-held view that causality is attributed mainly on the basis of the probability of reward per action (Chatlosh, Neunaber, & Wasserman, 1985; Cheng, 1997; Cheng & Buehner, 2012; Dickinson et al., 1984; Shanks & Dickinson, 1991; Wasserman et al., 1983). What is not clear, however, is the role that the action-reward rate correlation process would play in increasing action rates on the ratio over the interval schedule in spite of similar experienced reward probabilities.

To test whether participants experienced action-reward rate correlations were higher for the ratio condition over the interval condition we followed a computational approach recently proposed by Dickinson and Perez (Dickinson & Perez, 2018). Each block of training was divided in 10-second time samples. The number of actions and rewards in each of the samples was counted and used as a data point from which an experienced linear

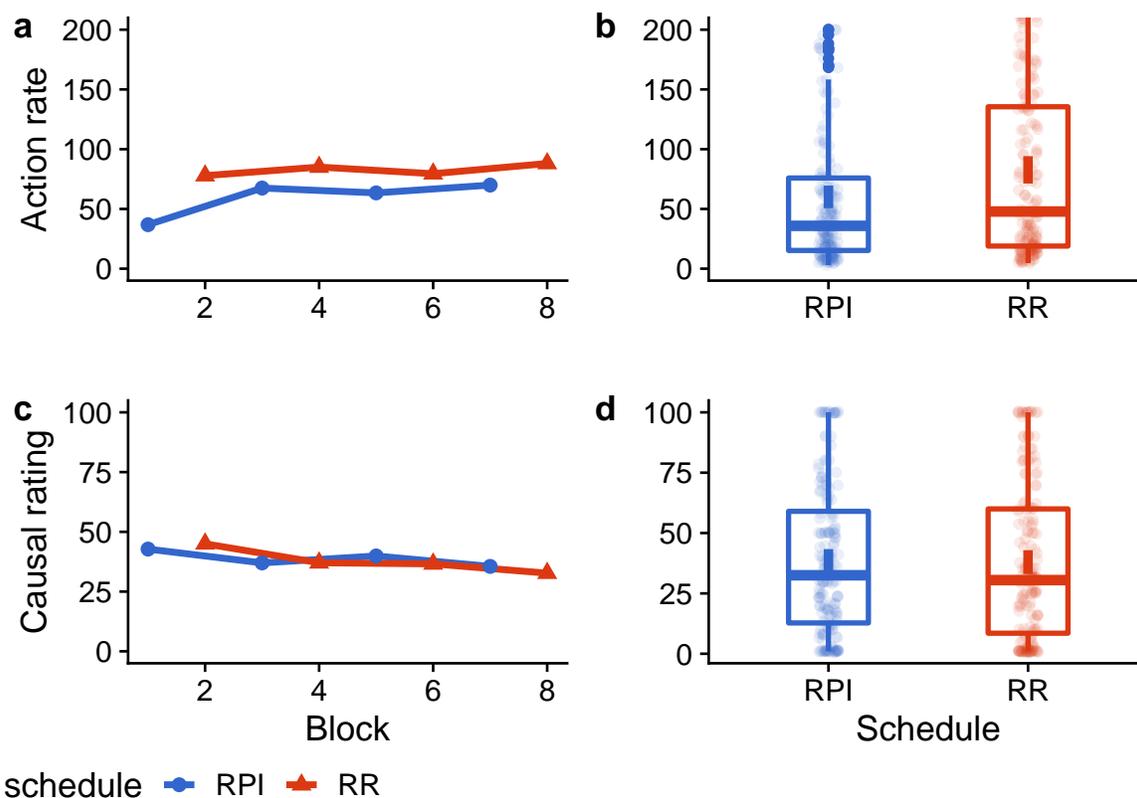


Figure 1. Action rates and causal ratings of the effectiveness of the action in Experiment 1. Panel (a): Action rates in each block of training for each schedule. Panel (b): Mean action rates across all blocks of training for each schedule. Panel (c): Causal ratings after each block of training. Panel (d): Mean causal ratings across all blocks of training. Error bars are 95% bootstrapped confidence intervals

correlation between the action and reward rates was computed. As expected, participants experienced action-reward rate correlation was higher under ratio than under interval training $F(1, 40) = 61.01, p < .0001, D = 0.69, 95\%CI = [0.46, 0.91]$, consistent with the view that this variable increases action rates when participants undergo ratio training. Since the experienced action-reward rate correlations were higher while the reward probabilities matched by design, this result lends support to the hypothesis that instrumental actions and causal beliefs might be dissociated when subjects are trained under different reward schedules.

Experiment 2

The results of Experiment 1 suggest that causal beliefs are mediated by reward probability, but that the differences in action rates are a consequence of the different experienced action-reward rate correlations. However, the conclusion that reward probability is the cardinal variable in the development of causal action-reward beliefs comes from a null result. One way of investigating further this idea would be to manipulate the reward probabilities experienced by subjects and see how the new causal beliefs would correlate

with these new reward probabilities. As noted before, that causal beliefs can be mediated by reward probability is a result that has been shown extensively in the human instrumental literature (Shanks & Dickinson, 1991; Vaghi et al., 2019; Wasserman et al., 1983).

The problem with such an idea is that reward probability cannot be controlled by the experimenter under an interval schedule. Under an interval schedule, the interval parameter can only control the maximum reward rate ($1/T$ per second) obtainable under the schedule; the experienced reward probability, in contrast, depends on the ratio of number of rewards to the number of actions performed. Because actions are beyond the experimenter's control, so is the experienced reward probability.

In spite of this difficulty, we can rely on previous data on free-operant performance to encourage higher or lower action rates for the interval condition. If one could decrease the action rates maintaining the same interval parameters (or, equivalently, reward rates) used in Experiment 1, the resulting experienced reward probabilities should be higher, since less actions would be performed for the same number of rewards. For example, Perez and colleagues (2018) trained rats to lever press under ratio, RPI and RI schedules and observed that, for the same reward rate, rats in the RI condition came to perform at a lower rate than those under the RPI condition. A similar observation was previously reported by Dawson and Dickinson (1990) using chain-pulling as the instrumental action.

To encourage lower action rates, the present experiment used an RI schedule as the master schedule. We expected the interval schedule to bring about lower action rates than in Experiment 1, thereby increasing the experienced reward probabilities for the subsequent ratio condition. If our hypothesis that reward probability mediates causal beliefs is correct, the causal ratings should be higher than in the previous experiment. Since the reward probabilities are controlled by design, these should again not differ between ratio and interval conditions. By contrast, the experienced action-reward rate correlations should still be higher under ratio training. To the extent that the experienced action-reward rate correlation on ratio and interval training increases performance for the ratio schedule, we expect participants' action rates to be higher in the ratio condition. These expected results would strengthen the hypothesis that instrumental action and causal beliefs are dissociated under instrumental free-operant training.

Method

Subjects and apparatus. 98 participants from Florida International University participated in the study. They were tested under the same conditions of Experiment 1.

Procedure. The procedure was similar to that of Experiment 1. The only difference was that participants were presented in the first trial of each block with an RI condition. After each second elapsed, the computer scheduled a reward with probability $1/T$, where T is the scheduled interval between rewards. As in Experiment 1, participants were randomly assigned one interval parameter T between 8 and 15.

Results and discussion

The rate of presses maintained by ratio and interval schedules across training blocks are shown in Panel a of Figure 2. Panel b presents the average rates across all blocks of

training. As in the previous experiment, participants action rates were higher under ratio than interval training ($F(1, 73) = 27.56, p < .0001, D = -0.44, 95\%CI = [0.28, 0.60]$), even when the reward probabilities were matched ($F(1, 73) = 0.27, p = .61, D = 0.11, 95\%CI = [-0.27, 0.06], BF_{01} = 3.01$). As expected, the experienced action-reward rate correlations were higher under ratio than under interval training ($F(1, 72.93) = 112.37, p < .0001, D = 0.70, 95\%CI = [0.54, 0.87]$). To check if our manipulation of using an RI schedule as the master schedule was successful in decreasing responding with respect to Experiment 1, and therefore whether the experienced reward probabilities were higher than in Experiment 1, we compared the mean reward probabilities of both experiments. The analysis confirmed that our manipulation succeeded in increasing reward probabilities with respect to Experiment 1 ($F(1, 113) = 91.25, p < .0001, D = 1.33, 95\%CI = [1.18, 1.47]$).

In contrast with the higher performance observed under ratio training, we did not find a detectable difference in participant's causal action-reward beliefs between ratio and interval conditions ($F(1, 73) = 2.86, p = .10, D = -0.12, 95\%CI = [-0.28, 0.04]$). However, in further support of the idea that causal beliefs are driven by reward probability, the analysis revealed that the causal beliefs in this experiment were higher than those of Experiment 1 ($F(1, 113) = 17.55, p < .0001, D = 0.53, 95\%CI = [0.39, 0.67]$).

This experiment has confirmed that the main determinant of causal action-reward beliefs is the reward probability and is consistent with the previous literature on the topic (Shanks & Dickinson, 1991; Vaghi et al., 2019; Wasserman et al., 1983). The second observation that arises from this experiment is that instrumental actions are not only dependent upon reward probability. Indeed, these results suggest that the difference in experienced action-reward rate correlation between schedules increases action rates for ratio training (Dickinson, 1994; Dickinson & Perez, 2018; Pérez et al., 2016).

Discussion

The question of primary interest in these two studies was investigating whether the widely-accepted notion that an instrumental action is mediated by a causal belief would hold under circumstances in which participants are trained under different reward schedules. These data suggest otherwise. Experiment 1 showed that a ratio schedule maintains higher levels of performance than an interval schedule even when reward probabilities are matched, a result that is consistent with previous studies in both humans (Pérez et al., 2016) and animals (Matthews, Shimoff, Catania, & Sagvolden, 1977). At the same time, there was no detectable difference in causal action-reward beliefs between the two types of training, a result that confirms the cardinal role of reward probability in the attribution of causality to instrumental actions (Dickinson et al., 1984; Shanks & Dickinson, 1991; Wasserman et al., 1983). The difference in performance was confirmed in Experiment 2. More importantly, the results showed that experiencing higher reward probabilities renders participants more likely to attribute a higher causal belief to their actions. However, as in Experiment 1, causal beliefs did not differ between the two schedules when reward probabilities were matched. Because the experienced action-reward rate correlations were higher under ratio training whereas reward probabilities between the two training schedules equated by design, one may conclude that the action-reward rate correlation plays a key role in modulating action rates,

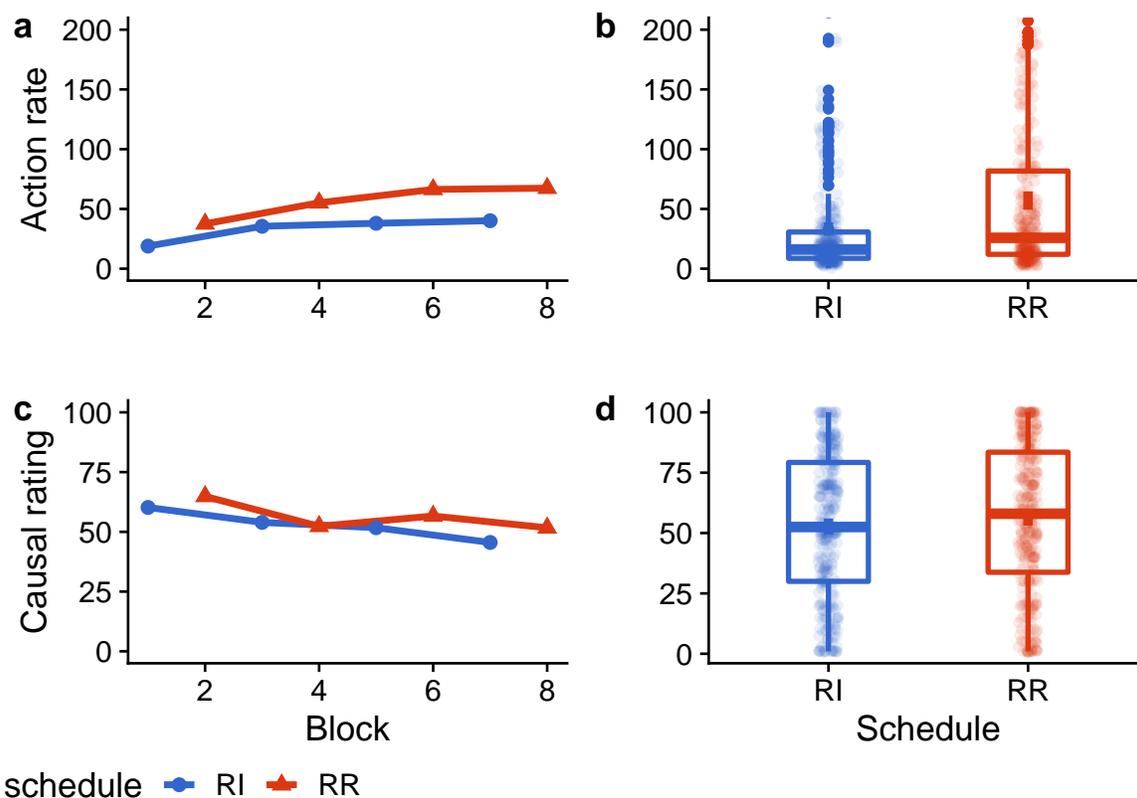


Figure 2. Action rates and causal ratings of the effectiveness of the action in Experiment 2. Panel (a): Action rates in each block of training for each schedule. Panel (b): Mean action rates across all blocks of training for each schedule. Panel (c): Causal ratings after each block of training. Panel (d): Mean causal ratings across all blocks of training. Error bars are 95% bootstrapped confidence intervals

and that reward probability is a more important factor in determining the attribution of causality to actions under free-operant training.

In addition to demonstrating that reward probability is the main determinant of a causal belief and that instrumental performance across ratio and interval schedules seem to be dissociated from these causal beliefs, the comparison between the two experiments lends further support to this dissociation. Indeed, when higher reward probabilities were encouraged by using an RI schedule in Experiment 2, instrumental performance decreased with respect to Experiment 1, but causal beliefs moved in the opposite direction, tracking these new, higher reward probabilities. Had causal beliefs been lower, or instrumental performance higher, an argument could still be made for causal beliefs explaining most of performance and for the ratio schedule contingency only explaining the residual difference between the two schedules. However, this was not case: the dissociation between causal beliefs and performance held not only within, but also between experiments.

The conclusion that action-reward rate correlation modulates instrumental performance had been previously demonstrated by both Perez and colleagues (2016) in humans and

Dawson and Dickinson (1990) in rats using RPI schedules. As explained above, the RPI schedule is programmed to maintain a fixed reward rate, as do regular interval schedules, but it does not explicitly increase reward probability with the passage of time since the last obtained reward.

A post-hoc analysis of rewarded and non-rewarded IRTs of Experiment 1, however, showed that the RPI schedule was rewarding long IRTs in a higher proportion than other emitted IRT sizes. To investigate this, we calculated the ratio of the mean rewarded IRT over the average IRT emitted by each subject and found that the IRT ratio for the RPI was higher than 1 ($mean = 1.69$, $95\%CI [1.42, 1.97]$). A possible explanation for this puzzling result is that the choice of a low memory size m allowed for changes in action rates (and consequently, changes on average IRT size) to incidentally affect the probability of reward for the next action performed. Although we have demonstrated in previous studies that the RPI schedule did not allow for this possibility using higher values of m , it might well be that a low value allows for this possibility. Whatever the explanation for this observation, it is clear that the explicit rewarding of long IRTs of the RI schedule is sufficient to slow down responding with respect to the RPI schedule in this ($t(275.31) = -4.64, p < .001$) and previous studies reported in the topic (Dawson & Dickinson, 1990; Pérez et al., 2016; Pérez, Aitken, Milton, & Dickinson, 2018).

It is also worth noting that, by itself, the experienced action-reward rate correlation cannot be the only variable explaining overall performance. As can be seen in Table 1, the action-reward rate correlation was higher for the RI schedule in Experiment 2 than it was for the RPI schedule in Experiment 1, but performance was higher in the former. The overall picture seems to suggest a full model of instrumental action that includes two variables that increase performance—the action-reward rate correlation and the reward probability—and one that moves in the opposite direction independently of these two other factors—the explicit differential reinforcement of long IRTs of the RI schedule. The particular influence of each factor, however, remains to be determined. One possibility, recently explored by Dickinson and Perez (2018), is that a goal-directed system based on the experienced action-reward rate correlation cooperate with an habitual system based on reward probability, so that total performance is directly related to a linear sum of these two components. Under this view, this model can explain the higher action rates maintained under ratio training and also the fact that ratio-trained actions are more sensitive to reward devaluation. However, if a causal action-reward belief is mediated by the higher experienced action-reward rate correlation brought about by a ratio schedule, their model should also anticipate higher causal ratings to the ratio schedule than to the interval schedule. This prediction was not confirmed in this study.

The data on causal beliefs observed in the present experiments are consistent with previous results of instrumental learning in humans, in that we found causal ratings to be directly related to the experienced reward probability. Given that the experienced action-reward rate correlations were, as expected, higher under ratio training, the question arises as to why this variable might increase performance independently of causal beliefs. Of course, one might argue that some implicit process was in effect in increasing action rates for the ratio schedule in the present experiments, or in other words, that the attribution

of causality to actions might be an explicit process whereas the transfer from experienced action-reward rate correlation to actions might not necessitate of such a process to operate. This is especially plausible given that in our experiments participants were told that the aim of the task was to judge the extent to which their actions caused the reward in each condition rather than paying attention to their performance during the task. This hypothesis is unlikely, however, because a recent study by Bradshaw and Reed demonstrated the opposite: contingency awareness is necessary to obtain differences in performance between ratio and interval schedules (2012). More importantly, such an idea would be at variance with the evidence from animal experiments in which ratio schedules are more prone to maintain goal-directed behavior, a type of control that is assumed to be mediated by a process of practical inference that includes the causal belief of the action-reward association as one critical component (Dickinson, 2012; Heyes & Dickinson, 1990).

Other explanations are procedural. For example, it is not clear what type of representation might the causal rating be tapping into. One possibility is that participants simply track reward probability to rate each condition, ignoring the higher experienced action-reward rate correlation. Alternatively, the causal rating might not be as sensitive as instrumental performance with respect to the action-reward rate correlation; in other words, the underlying theoretical function mapping the action-reward rate correlation to the performance of an action might differ from the one that maps the action-reward rate correlation to a casual belief; perhaps a bigger difference in action-reward rate correlation is necessary to detect a difference in causal ratings given to ratio- and interval-trained actions.

Overall, these data prompt two main conclusions. The first one is that higher experienced action-reward rate correlations increase performance for ratio- compared to interval-trained actions. The second, and perhaps most important one, is that causal beliefs of the effectiveness of an action are directly related to reward probability and seem to be independent of the difference in experienced action-reward rate correlations established by the schedules. It is this latter observation that lends support to the idea that instrumental performance of an action can be uncorrelated with its underlying causal attribution under free-operant training. Whether this dissociation between causal beliefs and instrumental performance is attributable to a distinction between explicit, or verbalizable learning, and implicit or automatic learning, or otherwise a consequence of different sensitivities to verbal and performance measures, remains to be tested.

Acknowledgments

We thank an anonymous reviewer for insightful comments and suggestions on this manuscript and S.D.G for the constant support to finish this paper.

Appendix

Instructions

Your task in this experiment is to judge the extent to which you can make something happen on the computer screen. In particular, your task is to find out whether pressing the space bar has any effect on whether or not the triangle appears on the screen. At any time you may choose or not to press the space bar. You can press it as often or as little as you

like. However, because of the nature of the task, it is to your advantage to press it some of the time and not to press it some of the time. (Press “c” to continue)

Sometimes the triangle will appear on the screen when you press the space bar. You must judge the extent to which pressing the space bar is the cause of the triangle lighting up.

You will be given several exposures to two different problems, each lasting 3 minutes. The relationship between pressing the space bar and whether or not the triangle lights up will be constant within each problem, but may well differ from one problem to the next. (Press “c” to continue)

Your task in this experiment is to earn as many points as you can. Every time that the triangle flashes, 10 points will be added to your score, but each press of the space bar will cost you 1 point, which will be deducted from your score.

At the end of each problem you will be told how many points you scored in that problem. You must try to get as many points as possible in each problem. . (Press “c” to continue)

At the end of each problem you will be asked to give an estimate on a rating scale of the extent to which you think that pressing the space bar caused the triangle to light up during that problem. (Press “c” to start the experiment).

Click on the scale below to indicate your judgment of the extent to which pressing the space bar caused the triangle to light up. Use a scale from 0 to 100. Zero indicates that the space bar had no effect on whether the triangle appeared on the screen, and 100 means that the space bar always caused the triangle to appear. Press either ENTER or the button below to confirm your answer. Your judgment?

References

- Balleine, B., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4-5), 407–419. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9704982>
- Bradshaw, C. M., & Reed, P. (2012). Relationship between contingency awareness and human performance on random ratio and random interval schedules. *Learning and Motivation*, *43*(1-2), 55–65. <https://doi.org/10.1016/j.lmot.2011.11.002>
- Catania, A. C., Matrrhews, T. J., Silverman, P. J., Yohalem, R., 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*(2), 155–161.
- Chatlosh, D. L., Neunaber, D. J., & Wasserman, E. A. (1985). Response-outcome contingency: Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation*, *16*(1), 1–34. [https://doi.org/10.1016/0023-9690\(85\)90002-5](https://doi.org/10.1016/0023-9690(85)90002-5)
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405. <https://doi.org/10.1037//0033-295X.104.2.367>

Cheng, P. W., & Buehner, M. J. (2012). Causal learning. *The Oxford Handbook of Thinking and Reasoning*, 210–233.

Daw, N. (2015). Of goals and habits. *Proceedings of the National Academy of Sciences*, 201518488. <https://doi.org/10.1073/pnas.1518488112>

Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711. <https://doi.org/10.1038/nn1560>

Daw, N., & O’Doherty, J. P. (2013). Multiple Systems for Value Learning. In *Neuroeconomics: Decision making and the brain: Second edition* (pp. 393–410). <https://doi.org/10.1016/B978-0-12-416008-8.00021-8>

Dawson, G. R., & Dickinson, A. (1990). Performance on ratio and interval schedules with matched reinforcement rates. *The Quarterly Journal of Experimental Psychology*, 42(789759670), 37–41. <https://doi.org/10.1080/14640749008401882>

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 308(1135), 67–78. Retrieved from <http://rstb.royalsocietypublishing.org/content/308/1135/67.short>

Dickinson, A. (1994). Instrumental conditioning. In N. J. Mackintosh (Ed.), *Animal cognition and learning* (pp. 45–78). London: Academic Press. Retrieved from <http://doi.apa.org/psycinfo/1994-98574-002>

Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1603), 2733–2742. <https://doi.org/10.1098/rstb.2012.0220>

Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. *Stevens’ Handbook of Experimental Psychology*.

Dickinson, A., & Charnock, D. J. (1985). Contingency effects with maintained instrumental reinforcement. *The Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, 37(4), 397–416. <https://doi.org/10.1080/14640748508401177>

Dickinson, A., Nicholas, D. J. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology*, 35(789759670), 35–51. <https://doi.org/10.1080/14640748308400912>

Dickinson, A., & Perez, O. D. (2018). Actions and habits: Psychological issues in dual-system theory. In R. W. Morris, A. M. Bornstein, & A. Shenhav (Eds.), *Goal-directed decision making: Computations and neural circuits* (pp. 1–37). Elsevier.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology*, 36A(1), 37–41. <https://doi.org/10.1080/14640748408401502>

Dolan, R., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325. <https://doi.org/10.1016/j.neuron.2013.09.007>

Hammond, L. J. (1980). The effect of contingency upon the appetitive conditioning of free-operant behavior. *Journal of the Experimental Analysis of Behavior*, *34*(3), 297–304.

Heyes, C., & Dickinson, A. (1990). The intentionality of animal action. *Mind {ℰ} Language*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0017.1990.tb00154.x/abstract>

Mackintosh, N. J., & Dickinson, A. (1979). Instrumental (Type II) Conditioning. In *Mechanisms of learning and motivation* (pp. 143–167).

Mas-Colell, A., Whinston, M. D., Green, J. R., & Others. (1995). No Title. In *Microeconomic theory* (Vol. 1). Oxford university press New York.

Matthews, B. A., Shimoff, E., Catania, A. C., & Sagvolden, T. (1977). Uninstructed human responding: sensitivity to ratio and interval contingencies. *Journal of the Experimental Analysis of Behavior*, *27*(3), 453–467. <https://doi.org/10.1901/jeab.1977.27-453>

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

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*, 1–33. <https://doi.org/10.1080/17470218.2016.1265996>

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.

Reed, P. (1999). Effect of perceived cost on judgments regarding the efficacy of investment. *Journal of Economic Psychology*, *20*, 657–676. Retrieved from <http://www.sciencedirect.com/science/article/pii/S016748709900029X>

Reed, P. (2001). Schedules of reinforcement as determinants of human causality judgments and response rates. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(3), 187–195. <https://doi.org/10.1037//0097-7403.27.3.187>

Reed, P. (2007). Human sensitivity to reinforcement feedback functions. *Psychonomic Bulletin {ℰ} Review*, *14*(4), 653–657. <https://doi.org/10.3758/BF03196816>

Shanks, D., & Dickinson, A. (1991). Instrumental judgment and performance under variations in action-outcome contingency and contiguity. *Memory {ℰ} Cognition*, *19*(4), 353–360. <https://doi.org/10.3758/BF03197139>

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1). MIT press Cambridge.

Thorndike, E. L. (1911). Edward Lee Thorndike. *Animal Intelligence*, *1874*, 1949.

Vaghi, M. M., Cardinal, R. N., Apergis-Schoute, A. M., Fineberg, N. A., Sule, A., & Robbins, T. W. (2019). Action-outcome knowledge dissociates from behavior in obsessive-compulsive disorder following contingency degradation. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(2), 200–209.

Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learning and Motivation*, *14*(4), 406–432.